



UNIVERSIDAD NACIONAL DE COLOMBIA

Detección de puntos de cambio en la función de media para datos funcionales multivariados

Darío Alejandro Latorre Montoya

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Estadística
Bogotá, Colombia
2021

Detección de puntos de cambio en la función de media para datos funcionales multivariados

Darío Alejandro Latorre Montoya

Tesis presentado como requisito parcial para optar al título de:
Magister en Estadística

Director:
Ph.D. Rubén Darío Guevara González

Línea de Investigación:
Análisis de datos funcionales

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Estadística
Bogotá, Colombia
2021

Dedicatoria

...a Lucía, Antonella y Laura

Agradecimientos

Agradezco a mis padres que con mucho esfuerzo lograron que yo tuviera lo fundamental; a mi hermana y mis amigos por la compañía y aliento durante este proceso; a Yubieth, mi esposa por estar conmigo y ser mi soporte en los momentos en los que la realización de este trabajo parecía inalcanzable; al profesor Rubén por su orientación, dedicación y minuciosidad, pero sobre todo por las acertadas palabras para aumentar mi ánimo durante la realización de este trabajo.

Resumen

El objetivo del análisis de punto de cambio es identificar si existen cambios o no en la distribución de un proceso estocástico, determinando el tiempo del cambio cuando haya ocurrido. Para datos funcionales univariados existen metodologías de detección de cambio en la media, sin embargo, no hay propuestas explícitas para el caso multivariado. Se propone una metodología para la detección de punto de cambio en la media de datos funcionales multivariados basada en un espacio de funciones RKHS que es construido. Se define un estadístico por medio de la norma inducida por el producto interno en el RKHS. Se muestra que el estadístico usado en el caso univariado se puede generalizar para éste enfoque. El estadístico definido tiene en cuenta la estructura de covarianza multivariada en el tiempo y la funcional univariada para los procesos. Cambios en la media de procesos multivariados simulados con varias estructuras de covarianza son detectados correctamente por la propuesta. Dos aplicaciones de la metodología propuesta, la primera a una casa domótica y la otra a una plataforma hidráulica, detectan correctamente el punto de cambio. La metodología es aplicada a contaminantes en el aire de Bogotá para detectar el inicio de las medidas de cuarentena de 2020. Se desarrollan dos apps, una para realizar simulaciones y una para uso de la propuesta.

Palabras clave: Datos funcionales multivariados, punto de cambio, análisis de componentes principales, espacios de Hilbert, RKHS.

Abstract

The change point analysis aims to identify whether or not there are changes in the stochastic process distribution, providing an estimate of the change time as required. There exist several methodologies to detect changes in the mean of univariate functional data, however, there are not explicit proposals in the multivariate case. We propose a methodology to detect the change point in the mean of Multivariate Functional Data based on a RKHS functions space that is constructed. We define a statistic using the RKHS inner product. We are able to show that a statistic used in the univariate case can be generalized from the perspective of our approach. The defined statistic takes into account a multivariate covariance structure at time as well as a univariate functional for the process. The changes in the mean on simulated multivariate processes for several covariance structures are detected properly using our proposal. We are able two applications of the proposed methodology, the first to a domotic house and the other to a hydraulic platform, correctly detect the point of change. The methodology is applied to air pollutants in Bogotá to detect the start of 2020 quarantine measures. Two apps are developed, one to perform

simulations and one to use the proposal.

Keywords: Multivariate Functional Data, Change point, Principal Component Analysis, Hilbert spaces, RKHS

Contenido

Agradecimientos	vii
Resumen	ix
1 Introducción	2
2 Fundamentos	5
2.1 Espacios de Hilbert	5
2.2 Espacio $L^2[0, 1]$	8
2.3 RKHS	10
2.4 Operadores	12
2.5 Objetos Aleatorios	14
3 Datos Funcionales y Componentes Principales	19
3.1 Datos Funcionales	19
3.1.1 Datos	19
3.1.2 Suavizamiento	22
3.2 Componentes Principales	23
3.2.1 Una vista geométrica	23
3.2.2 Un problema de optimización	25
3.2.3 Componentes principales: Análisis Multivariado	27
3.2.4 Componentes principales: Datos Funcionales Univariados	29
3.2.5 Componentes principales: Datos Funcionales Multivariados	31
4 Punto de Cambio y Propuesta	36
4.1 Punto de cambio	36
4.1.1 Punto de cambio univariado funcional: Berkes	37
4.1.2 Otros enfoques	39
4.1.3 Punto de cambio y RKHS	40
4.2 Propuesta	40
4.2.1 Ideas Geométricas	41
4.2.2 Espacio de trabajo	42

4.2.3	Propuesta para la detección de punto de cambio en la media de datos funcionales multivariados	56
5	Simulaciones y Aplicaciones	63
5.1	Simulaciones	63
5.1.1	Simulación procesos 3-variados (un cambio)	65
5.1.2	Simulación procesos 3-variados (dos cambios)	76
5.1.3	Simulación procesos 4-variados	78
5.2	Aplicaciones	83
5.2.1	Temperaturas Casa Domótica	83
5.2.2	Temperaturas Plataforma Hidráulica	85
5.2.3	Calidad de aire Bogotá	87
6	Conclusiones, recomendaciones y trabajo futuro	90
6.1	Conclusiones	90
6.2	Recomendaciones	91
6.3	Trabajo futuro	92
7	Apéndice	93
	Bibliografía	96

Lista de Figuras

- 3-1. Dato funcional a la izquierda, observaciones discretizadas a la derecha. . . 20
- 3-2. Dato funcional bivariado del proceso $\mathbf{X} = (X_1, X_2)$, las componentes X_1 y X_2 , izquierda y derecha respectivamente. 20
- 3-3. Observaciones discretizadas de un dato funcional bivariado: En rojo la muestra 1 en verde la muestra 2. 21
- 3-4. Observaciones de $\mathbf{X}(t) = (X_1(t), X_2(t))$ en $t = 0, 0.2, \dots, 1$. En rojo la muestra 1 en verde la muestra 2. 21
- 3-5. Diagrama de dispersión de una muestra simulada de X de tamaño 5000 con estructura de covarianza Σ 24
- 3-6. Diagrama de dispersión de la transformación de los datos simulados. 25

- 4-1. Diagramas de dispersión de \mathbf{Y}_t para los valores $t = 0, 0.2, 0.4, 0.6, 0.8, 1$ provenientes de una simulación de 20 procesos bivariados \mathbf{Y} 42
- 4-2. Curvas ajustadas para cada uno de los valores y_1, y_2 izquierda y derecha respectivamente de la simulación de los diagramas de dispersión en la figura 4-1. 43

- 5-1. Logaritmos de los valores propios de $\mathbf{B}(t)\mathbf{B}(t)^T$ 65
- 5-2. Simulación de 50 realizaciones del proceso estocástico 3-variado de media 0, con ruidos. 66
- 5-3. Funciones medias de las tres componentes del proceso en el Escenario 1 con $v_1 = v_2 = v_3 = 0.3$. En azul claro la media antes del cambio, en azul oscuro la media después del cambio. 67
- 5-4. Las tres componentes de 50 realizaciones simuladas del proceso estocástico en el Escenario 1 con $v_1 = v_2 = v_3 = 0.3$ 68
- 5-5. Estadístico T de las simulaciones con $v_1 = v_2 = v_3 = 0.3$ en el Escenario 1. 68
- 5-6. Funciones medias de las tres componentes de los procesos en el Escenario 2 con $v_1 = v_2 = v_3 = 0.3$. En azul claro La media antes del cambio, en azul oscuro la media después del cambio. 69
- 5-7. Las tres componentes de 50 realizaciones simuladas del proceso estocástico en el Escenario 2 con $v_1 = v_2 = v_3 = 0.3$ 70
- 5-8. Estadístico T de las simulaciones con $v_1 = v_2 = v_3 = 0.3$ en el Escenario 2. 70

5-9.	50 simulaciones de ruidos 3-variados y transformados por $\mathbf{B}(t)$	72
5-10.	Funciones medias de las tres componentes de los procesos con $v_1 = v_2 = v_3 = 0.3$. En azul claro La media antes del cambio, en azul oscuro la media después del cambio.	74
5-11.	Las tres componentes de 50 realizaciones simuladas del proceso estocástico con $v_1 = v_2 = v_3 = 0.3$	75
5-12.	Logaritmos de los valores propios estimados de $\mathbf{B}(t)$. Los puntos son las estimaciones puntuales que son ajustadas a las respectivas curvas.	75
5-13.	Estadístico T de las simulaciones con $v_1 = v_2 = v_3 = 0.3$ en el Escenario 2.	75
5-14.	Funciones medias de las tres componentes del proceso con $v_1 = v_2 = v_3 = 0.05$. En negro la media antes del primer cambio, en gris la media entre el primer y el segundo cambio, y en azul la media después del segundo cambio.	77
5-15.	Las tres componentes de 50 realizaciones simuladas del proceso estocástico con $v_1 = v_2 = v_3 = 0.05$	78
5-16.	Logaritmos de los valores propios de $\mathbf{B}(t)\mathbf{B}(t)^T$	79
5-17.	50 simulaciones de ruidos de distribución Laplace y transformados por $\mathbf{B}(t)$	80
5-18.	Funciones medias de las cuatro componentes de los procesos con $v = 0.15$. En azul claro La media antes del cambio, en azul oscuro la media después del cambio.	81
5-19.	Las cuatro componentes de los 50 procesos simulados con $v = 0.15$	82
5-20.	Temperaturas casa domótica: a la izquierda datos de temperatura de la habitación a la derecha datos de temperatura externa.	84
5-21.	Raíz de las estimaciones de los valores propios de $\mathbf{B}(t)\mathbf{B}(t)^T$	85
5-22.	Salida gráfica de <i>app análisis</i> para análisis de Temperaturas Plataforma Hidráulica con $d_1 = 3, d_2 = 7, d_3 = 8, d_4 = 7$	88
5-23.	Salida gráfica de <i>app análisis</i> para análisis de Óxidos de Nitrógeno en la Estación Puente Aranda con $d_1 = d_2 = 6$	89
7-1.	Función $\sin(10\pi t) + 1$ en negro y su derivada en rojo	93
7-2.	Función media en negro antes del cambio y en rojo después	94

Lista de Tablas

4-1. Valores simulados de la distribución K_d tomado de Berkes et al. (2009) . . .	62
5-1. Detección de punto cambio usando simulaciones de Procesos 3-variados con errores usando Happ-Kurz (2020).	71
5-2. Detección de dos puntos de cambio usando simulaciones de Procesos 3-variados con errores usando Happ-Kurz (2020)	78
5-3. Detección de punto cambio y comparaciones usando simulaciones de Procesos 4-variados con errores como en Happ-Kurz (2020) y ruidos Laplace. .	83
5-4. Aplicación metodología univariada: Temperaturas sistema hidráulico . . .	86
5-5. Aplicación metodología propuesta: Temperaturas sistema hidráulico	86

1 Introducción

La detección de punto de cambio busca identificar si existen cambios o no, en la distribución de un proceso y encontrar los posibles puntos de cambio. Metodologías de punto de cambio han sido aplicadas en diferentes contextos. En Lee (2010), se realiza una revisión bibliográfica del problema de punto de cambio, allí ubican como primera investigación de la detección de cambio al estudió en Page (1954). En Lee (2010) también se realiza una clasificación del problema y de la bibliografía dependiendo del tipo de cambio que se busca y de las metodologías estadísticas usadas para abordarlo. En Chen and Gupta (2011) hay una serie de métodos para el análisis de punto de cambio que contemplan diferentes distribuciones e incluso modelos discretos. En este trabajo se realiza un estudio sobre la detección de punto de cambio en la media de datos funcionales multivariados.

El problema de detectar si hay punto de cambio en las medias o no, aunque tiene similitudes con el problema de diferencia de medias, es diferente. En el primero los datos están ordenados y se debe determinar si en algún punto de ese orden hay un cambio en la media, en el segundo los datos están etiquetados y se busca determinar si hay diferencias en las medias entre los datos marcados con diferentes etiquetas pero no hay un orden establecido para los datos. Un punto en común en los dos problemas es que al definir adecuadamente una forma de medir distancias ambos problemas se pueden solucionar. En el contexto de ambos problemas aparecen distribuciones T^2 o adaptaciones de ésta. Resultados de Hotelling (1947) son frecuentemente usados como ideas para encontrar metodologías de punto de cambio en diferentes contextos basados en la distribución T^2 .

El problema de punto de cambio para variables aleatorias ha sido ampliamente estudiado, por ejemplo en Page (1955) estudian el problema de detección de punto de cambio para un parámetro de una distribución. Ellos estudian los casos en que se conoce el parámetro inicialmente y se quiere determinar si en algún momento hay un cambio, y también el caso en el que el parámetro no es conocido. En Hawkins et al. (2003) usan metodologías de detección de punto de cambio aplicado al control estadístico de procesos suponiendo que las variables siguen distribuciones normales univariadas, y contemplando tres escenarios respecto a los parámetros de la distribución normal.

Gardner (1969) realizan un estudio para la detección de punto de cambio en la media de vectores normales para una varianza conocida. En Sen and Srivastava (1973) se realiza un

estudio para vectores normales, allí se analizan los casos en que la media antes del cambio es conocida y en el que no, pero se asume de nuevo varianza conocida. En Zamba and Hawkins (2009) se considera que las matrices de covarianza son desconocidas y además se contempla para la hipótesis alternativa un posible cambio en la media y en la estructura de varianza.

Metodologías de detección de punto de cambio son aplicadas en diferentes áreas; Control Estadístico de Procesos, en series de tiempo de variables climatológicas, hidrológicas entre otras. En control estadístico de procesos, por ejemplo, Hawkins et al. (2003) hacen una revisión al problema de punto de cambio generando una discusión, sobre el conocimiento de los parámetros de los procesos en control. Plummer and Chen (2014) usan un enfoque bayesiano para localizar puntos de cambio en el número de variaciones de copias de ADN en investigaciones de cancer. Berkes et al. (2009) aplica su metodología para datos funcionales univariados en datos de temperatura de Inglaterra. También en el contexto de datos funcionales con aplicaciones al análisis de imágenes de resonancia magnética está la propuesta de detección de punto de cambio expuesta por Aston and Kirch (2011).

El uso de funciones para datos tomados en intervalos de tiempo permite extraer información adicional a la que se obtendría trabajando con los datos discretizados. Muchas de las herramientas para datos univariados y multivariados tienen sus equivalentes para datos funcionales con la utilidad que tiene el enfoque funcional. Algunas de las metodologías existentes para Punto de Cambio en datos multivariados se han adaptado para datos funcionales y nuevas propuestas se han considerado. Una revisión a algunas propuestas es dada en el capítulo 4. El estudio de punto de cambio con ayuda de RKHS ya ha sido estudiado, sin embargo, el enfoque dado en este trabajo es distinto, las diferencias son expuestas en el capítulo 4.

En el contexto de datos funcionales multivariados no hay de manera explícita una propuesta para la detección de punto de cambio. En este trabajo se propone una metodología que se basa en las mediciones de distancia en un espacio de Hilbert del tipo RKHS. La construcción del espacio en el capítulo 4 recoge ideas geométricas y propone una forma de medir que reduce el problema de multivariado funcional a univariado funcional, la forma del estadístico T propuesto en Berkes et al. (2009) resulta después de la medición de distancia. Queda abierto el estudio de la posible implementación de otras ideas aplicadas en el espacio construido.

La propuesta para la detección de punto de cambio es dada en el capítulo 4, el capítulo 2 es de fundamentación y sirve como referencia al desarrollo de la propuesta, en el capítulo 3 se hace una presentación del manejo de datos funcionales y del análisis de componentes principales en diferentes contextos.

Se realizan simulaciones de procesos multivariados a los que se les agraga un cambio en

algún punto y la metodología propuesta es capaz de detectarlo. Para los errores en las simulaciones se usa la transformación descrita en el capítulo 4 y la propuesta de simulación en Happ and Greven (2018). Se simula un proceso 3-variado con dos puntos de cambio y la metodología detecta de manera adecuada los dos puntos de cambio. La metodología propuesta se aplica a datos bivariados de temperatura recogidos en una casa domótica y a datos de temperatura de una plataforma hidráulica, en ambos casos la propuesta es capaz de detectar un cambio que el análisis univariado por separado no es capaz de detectar. Se usa la metodología para detectar un punto de cambio en las concentraciones de óxidos de nitrógeno en la ciudad de Bogotá en el inicio de la cuarentena de 2020. Los códigos para implementar la propuesta, correr las simulaciones y desarrollar dos apps, una para las simulaciones y una para el uso de la metodología propuesta se pueden encontrar en Latorre (2019). En los apéndices se expone un enfoque para detectar cambios de forma en la media de datos funcionales univariados.

2 Fundamentos

En este primer capítulo se presentan los fundamentos teóricos que serán el soporte para las ideas presentadas a lo largo de este trabajo. El desarrollo teórico de este capítulo es justificado por la necesidad de medir distancias que indiquen diferencias entre los objetos estudiados.

En el análisis funcional, donde los objetos pertenecen a un espacio de dimensión infinita, hay muchos resultados similares al análisis de espacios vectoriales de dimensión finita, sin embargo, es necesario establecer que la dimensionalidad hace que se deba tener cuidado a la hora de hacer analogías. Los teoremas en esta sección son presentados sin demostración, algunos de ellos son resultados clásicos del análisis funcional.

La primera sección es dedicada a un concepto que es transversal a varias áreas del conocimiento, los espacios de Hilbert. En física, matemáticas y estadística se introduce este concepto que provee unas características deseables para medir distancias.

2.1. Espacios de Hilbert

La necesidad de encontrar un espacio en el que sea posible medir de manera adecuada diferencias entre objetos que pueden ser vectores o funciones, equivale a buscar un espacio métrico completo que contenga los objetos de estudio y que contemple sus características. Si bien es cierto que el espacio objetivo es un espacio métrico, hay características que son deseables y que no tienen todos los espacios métricos, por ejemplo, los espacios métricos que no tienen asociada una norma pueden perder la característica de homogeneidad.

Definición 2.1.1 (Espacio de Hilbert) *Un espacio vectorial con producto interno y completo, es llamado un espacio de Hilbert y se nota por*

$$\{\mathbf{H}, \langle \cdot, \cdot \rangle\}$$

Se hace énfasis en que el espacio de Hilbert no solo es caracterizado por los elementos del espacio vectorial, sino también por la forma bilineal del producto interno.

Ejemplo 2.1.1 *El espacio vectorial \mathbb{R}^p con el producto interno dado por $\langle u, v \rangle = u^T A v$, donde $u, v \in \mathbb{R}^p$ y A es una matriz definida positiva, es un espacio de Hilbert y se nota por $\{\mathbb{R}^p, \langle \cdot, \cdot \rangle_A\}$. En el caso particular cuando $A = I_p$, se tiene que $\{\mathbb{R}^p, \langle \cdot, \cdot \rangle_A\}$ es el espacio de Hilbert \mathbb{R}^p con el producto interno usual.*

Los espacios de Hilbert son casos particulares de los espacios de Banach y estos a su vez de los espacios métricos completos, esto se debe a que un producto interno induce una norma y esta a su vez una métrica. Se puede hablar de la métrica inducida por el producto interno teniendo en cuenta la transición.

Definición 2.1.2 (Norma y métrica inducidas) *Para cada h en el espacio de Hilbert $\{\mathbf{H}, \langle \cdot, \cdot \rangle\}$, la norma de h es:*

$$\|h\| = \sqrt{\langle h, h \rangle}$$

Para $h_1, h_2 \in \mathbf{H}$ se define la distancia entre h_1 y h_2 por

$$d(h_1, h_2) = \|h_1 - h_2\|$$

Algunas de las características de una métrica inducida por una norma es que esta es invariante por traslaciones y homogénea. En un espacio con producto interno se puede establecer una noción de perpendicularidad que da a la métrica características deseables para el problema de interés y que serán exploradas a lo largo del trabajo.

Definición 2.1.3 *Sean h_1, h_2 objetos de un espacio de Hilbert \mathbf{H} , estos se dicen ortogonales si $\langle h_1, h_2 \rangle = 0$. Una colección enumerable de elementos $\{e_1, e_2, \dots\}$ es una sucesión ortonormal si cada elemento cumple $\|e_l\| = 1$ y cada par de elementos son ortogonales.*

Las sucesiones de vectores ortonormales juegan un papel fundamental en las ideas presentadas en este trabajo. En los espacios de Hilbert \mathbb{R}^p , sin importar cual sea el producto interno, la sucesión ortonormal de la definición anterior tiene a lo más p elementos. En \mathbb{R}^p un conjunto de objetos que generan el espacio y sean linealmente independientes debe tener p elementos, y el conjunto es llamado una base para el espacio \mathbb{R}^p . Si además el conjunto es una sucesión ortonormal la base se dice base ortonormal. En el caso general para espacios de Hilbert se tiene:

Definición 2.1.4 *Una sucesión ortonormal $\{e_1, e_2, \dots\}$ en un espacio de Hilbert es llamada una base ortonormal o un sistema completo ortonormal si*

$$\overline{\text{span}\{e_1, e_2, \dots\}} = \mathbf{H}$$

donde $\text{span}\{e_1, e_2, \dots\}$ indica el conjunto generado por la sucesión y la línea superior indica la clausura topológica, esto hace que efectivamente la parte izquierda sea completa.

Definición 2.1.5 *Un espacio métrico \mathbf{H} es separable, si tiene un subconjunto denso que es enumerable.*

En la definición anterior se asume que la métrica es la inducida por el producto interno del espacio de Hilbert.

\mathbb{R}^p es separable para todo p , sin importar el producto interno. En este espacio es común el uso de bases ortonormales para expresar los vectores. El siguiente teorema muestra el caso general de esta expansión para espacios de Hilbert separables, su demostración se puede consultar en Hsing and Eubank (2015).

Teorema 2.1.1 (Expansión de Fourier) *Sea \mathbf{H} un espacio de Hilbert separable y $\{e_l\}$ una base ortonormal entonces dado $h \in \mathbf{H}$, éste se puede escribir en términos de la expansión de Fourier*

$$h = \sum_{l=1}^{\infty} \langle h, e_l \rangle e_l,$$

además

$$\|h\|^2 = \sum_{l=1}^{\infty} \langle h, e_l \rangle^2.$$

Los términos $\langle h, e_l \rangle$ son llamados la proyección de h sobre e_l , o los componentes de h en la base $\{e_l\}$.

Teorema 2.1.2 *Un espacio de Hilbert es separable si y solamente si tiene una base ortonormal.*

La prueba puede ser consultada en Hsing and Eubank (2015), sin embargo, se puede ver que las definiciones de espacio separable y de sucesión ortonormal incluyen subconjuntos contables que generarán el espacio, esto sumado a un proceso de ortonormalización da indicios de cómo es la demostración.

El siguiente teorema cuya demostración se puede encontrar en Hsing and Eubank (2015) establece una simplificación en la métrica del espacio de interés que será presentado en la siguiente sección.

Teorema 2.1.3 *Todo espacio de Hilbert sobre \mathbb{R} con una base ortonormal infinita es isomorfo al espacio vectorial de las sucesiones de números reales l^2 . El isomorfismo es un isomorfismo de espacios métricos.*

Sean \mathbb{R}^{p_1} y \mathbb{R}^{p_2} dos espacios de Hilbert con los productos internos usuales y $p_1 + p_2 = p$. Al concatenar las componentes de los vectores es natural ver los vectores resultantes como objetos de \mathbb{R}^p y definir el producto interno como el usual para este espacio. La siguiente definición formaliza esta idea intuitiva.

Definición 2.1.6 Para $k = 1, 2, \dots, p$ sea \mathbf{H}_k un espacio de Hilbert con producto interno $\langle \cdot, \cdot \rangle_k$. La suma directa de Espacios de Hilbert es un espacio vectorial \mathbf{H}

$$\mathbf{H} = \mathbf{H}_1 \oplus \mathbf{H}_2 \oplus \cdots \oplus \mathbf{H}_p$$

con producto interno $\langle \cdot, \cdot \rangle$

$$\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_1 + \langle \cdot, \cdot \rangle_2 + \cdots + \langle \cdot, \cdot \rangle_p.$$

El espacio \mathbf{H} de la definición anterior es un espacio de Hilbert con el producto interior dado.

Si los espacios $\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_p$ son espacios separables cada elemento h del espacio \mathbf{H} se puede escribir de manera única como

$$h = h_1 + h_2 + \cdots + h_p,$$

donde $h_k \in \mathbf{H}_k$ para cada k .

Del teorema 2.1.3 se puede concluir que la suma directa de finitos espacios de Hilbert isomorfos a l^2 es isomorfo a l^2 , la demostración es similar a la usada para demostrar que la unión de conjuntos enumerables y disjuntos es de nuevo enumerable, ver por ejemplo Vasudeva and Shirali (2017).

La definición 2.1.6 se puede extender para realizar la suma directa sobre cualquier conjunto de subíndices, incluso no enumerables. La extensión de la suma directa puede ser consultada en Vasudeva and Shirali (2017).

2.2. Espacio $\mathbb{L}^2[0, 1]$.

En su mayoría, la literatura sobre datos funcionales enfoca la atención en funciones definidas sobre el intervalo $[0,1]$, esto se debe a que cualquier intervalo cerrado puede ser transformado en $[0,1]$ mediante una traslación y un reescalamiento.

Definición 2.2.1 (Espacio medible) Un espacio medible es una tripla $(E, \mathcal{B}, \mathbf{m})$, en la que E es un subconjunto no vacío, \mathcal{B} es una σ -álgebra de subconjuntos de E y \mathbf{m} una medida positiva sobre E .

En este trabajo se asume que $E = [0, 1]$, la medida \mathbf{m} corresponde a la medida de Lebesgue de E y \mathcal{B} corresponde a la colección de conjuntos Lebesgue-medibles.

Definición 2.2.2 Sea $(E, \mathcal{B}, \mathbf{m})$ un espacio medible, se notará $\tilde{\mathbb{L}}^2(E, \mathcal{B}, \mathbf{m})$ el conjunto de todas las funciones definidas sobre E tal que

$$\int_E |f|^2 d\mathbf{m} < \infty.$$

Dado que el espacio E , la medida \mathbf{m} y la colección de conjuntos \mathcal{B} serán las mismas a lo largo de este trabajo, se notará este espacio por $\tilde{\mathbb{L}}^2[0, 1]$.

Proposición 1 $\tilde{\mathbb{L}}^2[0, 1]$ es un espacio vectorial.

La demostración puede ser vista en Vasudeva and Shirali (2017). La integral finita en la definición 2.2.2 cumple las condiciones para ser norma salvo que existen funciones f_0 tales que la integral en 2.2.2 es 0 y $f_0 \neq 0$. En Vasudeva and Shirali (2017) es mostrado como se pueden formar clases de funciones mediante una relación de equivalencia para definir un espacio normado y aún más, un espacio de Hilbert.

Teorema 2.2.1 (Riesz-Fischer) $\{\mathbb{L}^2[0, 1], \langle \cdot, \cdot \rangle\}$ es un espacio de Hilbert con

$$\langle f, g \rangle = \int_{[0,1]} fg d\mathbf{m}.$$

Se hace claridad en que $\mathbb{L}^2[0, 1]$ no es un conjunto de funciones, sino un conjunto de clases de funciones. Así, f es una función que representa a toda la clase de funciones que difieren de f en un conjunto de medida de 0 respecto a la medida \mathbf{m} , si

$$\int_{[0,1]} (f - f_1)^2 d\mathbf{m} = 0$$

entonces f y f_1 están en la misma clase por lo que cualquiera de ellas podría ser representante de la clase, en este caso se dice que f y f_1 son iguales en $\mathbb{L}^2[0, 1]$. Por comodidad en la notación se asume que las funciones f, g son elementos de $\mathbb{L}^2[0, 1]$.

Todos los detalles acerca de la construcción de $\mathbb{L}^2[0, 1]$ y la demostración del teorema anterior pueden ser consultados en Vasudeva and Shirali (2017). El teorema de Riesz-Fischer implica que $\mathbb{L}^2[0, 1]$ es también un espacio métrico. El producto interno con el que se trabajará para el espacio $\mathbb{L}^2[0, 1]$ es el dado en el teorema anterior a menos de que se especifique lo contrario.

Proposición 2 (Base de Fourier)

$$\mathbf{B} = \{h_1(t) = 1, h_{2n}(t) = \sqrt{2} \sin 2n\pi t, h_{2n+1}(t) = \sqrt{2} \cos 2n\pi t : n \geq 1\}$$

con $t \in [0, 1]$ es una base ortonormal para el espacio de Hilbert $\mathbb{L}^2[0, 1]$.

De la proposición anterior y del teorema 2.1.2 se tiene que el espacio de Hilbert $\mathbb{L}^2[0, 1]$ es separable. Existen diferentes bases ortonormales, aquí se trabajará en gran parte con la base de Fourier presentada en la proposición anterior.

No se debe confundir la base de Fourier de la que trata la proposición anterior con la expansión de Fourier tratada en el teorema 2.1.1 y que es general para cualquier base ortonormal.

2.3. RKHS

Los espacios de Hilbert con núcleo reproductor, RKHS por sus siglas en inglés (Reproducing Kernel Hilbert Space), son un tipo especial de espacios de Hilbert que tienen un papel fundamental para la representación de algunos procesos estocásticos y la descomposición de Karhunen-Loève, expuesta más adelante. Los RKHS son usados en diferentes áreas como en teoría estadística del aprendizaje para la minimización de la función empírica de riesgo, en física para resolver osciladores no lineales, también en la solución de ecuaciones diferenciales.

Definición 2.3.1 (rk) Una función bivariada, $K : \mathbf{D} \times \mathbf{D}$, $\mathbf{D} \subset \mathbb{R}$ es núcleo reproductor, *rk* por sus siglas en inglés (reproducing kernel), para $\{\mathbf{H}, \langle \cdot, \cdot \rangle\}$ si:

1. Para cada $t \in D$, $K(\cdot, t) \in \mathbf{H}$.
2. K satisface la propiedad de reproducir toda función $f \in \mathbf{H}$, esto es que para $t \in D$

$$f(t) = \langle f, K(\cdot, t) \rangle.$$

Cuando el espacio de Hilbert $\{\mathbf{H}, \langle \cdot, \cdot \rangle\}$ posee un *rk*, se dice un RKHS.

A continuación, se presentan dos ejemplos para conjuntos \mathbf{D} finitos y dos para el intervalo $[0, 1]$

Ejemplo 2.3.1 Sean $\mathbf{D} = \{0, 1\}$, $f \in \mathbf{H}$ una función de \mathbf{D} en \mathbb{R} y considérese

$$K = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix},$$

dado que f solo tiene dos elementos en su dominio, es cómodo notar la función como se muestra a continuación,

$$f = [f(0), f(1)]^T,$$

la primera componente del vector columna corresponde al valor que toma la función f en el punto 0 y la segunda al valor tomado en el punto 1. El producto interno se define por

$$\langle f, g \rangle = f^T K^{-1} g.$$

Por el ejemplo 2.1.1 se tiene que $\{\mathbf{H}, \langle \cdot, \cdot \rangle\}$ es un espacio de Hilbert. En este caso $K(\cdot, t)$ corresponde a cada una de las columnas de la matriz K o sea

$$K(\cdot, 0) = [1, 0.9]^T \quad K(\cdot, 1) = [0.9, 1]^T$$

se puede ver que $K(\cdot, 0)$ y $K(\cdot, 1)$ están en \mathbf{H} . Ahora

$$\langle f, K(\cdot, t) \rangle = f^T K^{-1} K(\cdot, t) = f(t).$$

Con esto se verifica que K es un rk para el espacio $\{\mathbf{H}, \langle \cdot, \cdot \rangle\}$.

Ejemplo 2.3.2 Sea $\mathbf{D} = \{t_1, t_2, \dots, t_n\}$ un conjunto de n puntos distintos, sea $K(t_i, t_j) = k_{ij}$ donde la matriz $K = \{k_{ij}\}$ con $i, j = 1, 2 \dots n$ es definida positiva y simétrica. Sea $B = \{b_{ij}\} = K^{-1}$. Definiendo el producto interno

$$\langle f, g \rangle = \sum_{i=1}^n \sum_{j=1}^n f(t_i) b_{ji} g(t_j),$$

El espacio \mathbf{H} de todas las funciones de \mathbf{D} en \mathbb{R} con el producto interno $\langle f, g \rangle$ es un espacio de Hilbert con rk K . Los detalles de la verificación del ejemplo anterior y del siguiente se pueden consultar en Saitoh and Sawano (2016)

Ejemplo 2.3.3 El espacio de Sobolev $\mathbf{W}^{1,2}[0, 1]$, que corresponde al espacio de funciones sobre el intervalo $[0, 1]$, diferenciables de orden 1 y contenidas en $\mathbb{L}^2[0, 1]$, tiene por rk a

$$K(s, t) = \frac{1}{2} \exp(-|s - t|) \quad s, t \in [0, 1]$$

Un subconjunto del espacio de Sobolev en el que las funciones f definidas sobre $[0, 1]$, satisfacen

$$f(0) = f(1) = 0, \quad f'(0) = f'(1) = 0,$$

es un RKHS, así se muestra en Hsing and Eubank (2015), el rk es el indicado en el siguiente ejemplo.

Ejemplo 2.3.4 El espacio de Sobolev $\mathbf{W}_0^{1,2}[0, 1]$ tiene por rk a

$$K(s, t) = \min(s, t).$$

Del ejemplo 2.3.2 se puede concluir que una matriz de covarianza cumple las condiciones para ser un rk, por otro lado, el rk del espacio de Sobolev con la restricción del ejemplo 2.3.4 coincide con la función de covarianza del movimiento browniano.

Definición 2.3.2 (Función definida positiva) *Una función $K : \mathbf{D} \times \mathbf{D} \rightarrow \mathbb{R}$ es llamada una forma cuadrática positiva sobre \mathbf{D} o función definida positiva, cuando para toda función $f : \mathbf{D} \rightarrow \mathbb{R}$ y cualquier subconjunto finito \mathbf{F} de \mathbf{D} se tiene:*

$$\sum_{s,t \in \mathbf{F}} f(s)f(t)K(s,t) \geq 0$$

Algunos autores llaman a las funciones cumpliendo la propiedad anterior semidefinidas positivas. El siguiente teorema de existencia y unicidad conocido como el teorema de Moore–Aronszajn, como se verá más adelante, garantiza que el espacio de trabajo propuesto esté bien definido. La demostración del teorema se puede consultar en Hsing and Eubank (2015).

Teorema 2.3.1 (Teorema de Moore–Aronszajn) *Si $K(s,t)$, con $s,t \in \mathbf{D}$ es una función simétrica y definida positiva, existe un único espacio de Hilbert $\{\mathbf{H}, \langle \cdot, \cdot \rangle\}$ de funciones sobre \mathbf{D} para el cual K es su rk.*

2.4. Operadores

Los operadores lineales son funciones entre espacios de Hilbert, sus análogos para espacios vectoriales de dimensión finita son las transformaciones lineales cuya acción se puede ver a través de matrices cuando las bases de los espacios están dadas. En el caso de espacios de Hilbert de dimensión finita las transformaciones lineales tienen características que son fáciles de distinguir por el carácter finito y lineal. Sin embargo, cuando la dimensión es infinita se hace necesario establecer condiciones para los operadores que permitan el trabajo con ellos. Entre las condiciones buscadas para operadores de un espacio de Hilbert en si mismo están por ejemplo; que sean acotados, continuos, que tengan traza finita, entre otras. Para los espacios de dimensión finita cada una de las anteriores propiedades son cumplidas por todas las transformaciones lineales. Los resultados expuestos en esta sección pueden ser consultados en Hsing and Eubank (2015)

Definición 2.4.1 *Si $\mathbf{H}_1, \mathbf{H}_2$ son espacios normados con normas $\|\cdot\|_1, \|\cdot\|_2$ respectivamente, una transformación lineal \mathcal{F} de \mathbf{H}_1 en \mathbf{H}_2 es acotada si existe $M > 0$, tal que para cada elemento h en \mathbf{H}_1 se tiene*

$$\|\mathcal{F}h\|_2 \leq M \|h\|_1$$

Si el operador es acotado entonces se puede mostrar que es uniformemente continuo. La continuidad uniforme es una característica necesaria para la definición de los espacios basados en la estructura de covarianza. Los operadores acotados definidos sobre RKHS pueden ser caracterizados por sus acciones sobre el rk. Si K_1, K_2 son los kernel de los RKHS $\mathbf{H}(K_1)$ y $\mathbf{H}(K_2)$ respectivamente, y \mathcal{F} es un operador de $\mathbf{H}(K_1)$ en $\mathbf{H}(K_2)$ con \mathcal{F}^* su operador adjunto, esto es que \mathcal{F}^* es el único operador tal que

$$\langle \mathcal{F}f_1, f_2 \rangle_2 = \langle f_1, \mathcal{F}^*f_2 \rangle_1,$$

entonces se define el operador kernel para \mathcal{F} por

$$R(\cdot, s) = \mathcal{F}^*K_2(\cdot, s), \quad (2-1)$$

con esta definición se obtiene

$$\langle f, R(\cdot, s) \rangle_1 = \langle \mathcal{F}f, K_2(\cdot, s) \rangle_2 = (\mathcal{F}f)(s) \quad \text{para cada } f \in \mathbf{H}(K_1) \quad (2-2)$$

Esta expresión será usada para definir operadores sobre RKHS más adelante.

Definición 2.4.2 (Operadores clase traza y Hilbert-Schmidt) *Sea \mathcal{F} un operador lineal acotado entre los espacios de Hilbert separables \mathbf{H}_1 y \mathbf{H}_2 .*

\mathcal{F} es de clase traza si para alguna base ortonormal $\{e_j\}_{j=1,2,\dots}$ de \mathbf{H}_1 la siguiente cantidad llamada la norma traza del operador es finita

$$\|\mathcal{F}\|_{TR} := \sum_{j=1}^{\infty} \left\langle (\mathcal{F}^*\mathcal{F})^{1/2} e_j, e_j \right\rangle_1,$$

donde $\langle \cdot, \cdot \rangle_1$ es el producto interno en \mathbf{H}_1 , y $(\mathcal{F}^*\mathcal{F})^{1/2}$ es un operador semidefinido positivo, acotado y autoadjunto tal que $(\mathcal{F}^*\mathcal{F})^{1/2} (\mathcal{F}^*\mathcal{F})^{1/2} = \mathcal{F}^*\mathcal{F}$

Sea $\{e_j\}_{j=1,2,\dots}$ una base ortonormal para \mathbf{H}_1 , \mathcal{F} es un operador de Hilbert-Schmidt si

$$\sum_{j=1}^{\infty} \|\mathcal{F}e_j\|_2^2 < \infty.$$

Las expresiones anteriores corresponden respectivamente a las normas traza y de Hilbert-Schmidt y su definición no depende de la base ortonormal escogida.

Definición 2.4.3 *Un operador lineal \mathcal{F} de un espacio normado \mathbf{H}_1 a otro espacio normado \mathbf{H}_2 es compacto si para toda sucesión acotada $\{h_n\}$ de elementos de \mathbf{H}_1 , la sucesión $\{\mathcal{F}h_n\}$ tiene una subsucesión convergente en \mathbf{H}_2 .*

Si los espacios \mathbf{H}_1 y \mathbf{H}_2 son de dimensión finita entonces todos los operadores son compactos.

Teorema 2.4.1 *Todo operador \mathcal{F} de clase traza es de Hilbert-Schmidt y todo operador de Hilbert-Schmidt es compacto.*

El operador identidad en el espacio \mathbb{L}^2 no es compacto y por el teorema anterior no es de Hilbert-Schmidt. El operador identidad es claramente invertible pero no es compacto cuando el espacio es de dimensión infinita, el siguiente resultado muestra condiciones en general.

Teorema 2.4.2 *Sea \mathcal{F} un operador entre los espacios \mathbf{H}_1 y \mathbf{H}_2 de dimensión infinita, si \mathcal{F} es compacto entonces no es invertible y su imagen no es cerrada.*

Este teorema expone un contraste; la compacidad es una característica deseable para los operadores ya que esta permite hacer aproximación de los operadores cuando los espacios son de dimensión infinita, sin embargo, en muchos estudios se hace necesario el trabajo con operadores que sean invertibles, el problema está en que no es posible tener las tres características: compacidad, dimensión infinita y ser invertible. Sumado a esto está el hecho de que si el operador no tiene imagen cerrada cabe la posibilidad de que, por ejemplo, se trabaje con un proceso para el cual sus realizaciones no estén contenidas en el espacio definido por su operador de covarianza.

El producto Kronecker es un caso particular del producto tensorial definido a continuación, el resultado de este producto es una transformación entre espacios vectoriales por lo que en general no es conmutativo debido a que los espacios pueden ser de diferentes dimensiones.

Definición 2.4.4 *Sean h_1, h_2 elementos de los espacios de Hilbert $\mathbf{H}_1, \mathbf{H}_2$ respectivamente. El operador producto tensorial $h_1 \otimes h_2$ se define por*

$$\begin{aligned} (h_1 \otimes h_2) : \mathbf{H}_1 &\rightarrow \mathbf{H}_2 \\ (h_1 \otimes h_2) h &\rightarrow \langle h_1, h \rangle_1 h_2 \end{aligned}$$

El producto tensorial es un operador de tipo lineal y además es acotado.

2.5. Objetos Aleatorios

Vale aclarar que hay conceptos que se salen del alcance de este trabajo y que si bien pueden ser mencionados no serán enunciados explícitamente, en específico se hace referencia a los aspectos teóricos del operador de covarianza y a la teoría de la perturbación. Las pruebas

de los teoremas expuestos en esta sección pueden ser consultadas en Hsing and Eubank (2015).

Ya se ha especificado que de un espacio de Hilbert se induce un espacio métrico; ahora, de un espacio métrico se puede inducir una topología. Se nota por $\mathcal{B}(\mathbf{H})$ a la σ -álgebra más pequeña conteniendo los abiertos de la topología del espacio de Hilbert.

Por otro lado considere la σ -álgebra $\sigma(\mathcal{C})$ más pequeña que contiene todos los conjuntos de la forma $\{h \in \mathbf{H} : \langle h, f \rangle \in B\}$ para cada $f \in \mathbf{H}$ y B un abierto de \mathbb{R} .

Teorema 2.5.1 *Las σ -álgebras $\mathcal{B}(\mathbf{H})$ y $\sigma(\mathcal{C})$ son dos colecciones de conjuntos iguales.*

El siguiente teorema muestra los conceptos de medibilidad y de distribución.

Teorema 2.5.2 *Sea \mathcal{X} un mapeo de un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ a $(\mathbf{H}, \mathcal{B}(\mathbf{H}))$. Se tiene entonces que*

1. \mathcal{X} es medible si $\langle \mathcal{X}, f \rangle$ es medible para todo $f \in \mathbf{H}$.
2. Si \mathcal{X} es medible entonces su distribución es determinada unívocamente por $\langle \mathcal{X}, f \rangle$ para cada $f \in \mathbf{H}$.

Si el espacio de Hilbert es separable nótese que el mapeo es determinado por el comportamiento de las proyecciones sobre una base ortonormal.

Sea \mathcal{X} un elemento aleatorio de un espacio de Hilbert \mathbb{H} , en este trabajo \mathbb{H} es un subespacio de $\mathbb{L}^2[0, 1]$, definido sobre el espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$.

Se dice que \mathcal{X} es integrable si

$$E(\|\mathcal{X}\|) = E\left(\left[\int_{[0,1]} \mathcal{X}^2(t)dt\right]^{1/2}\right) < \infty$$

y es cuadrado integrable si

$$E(\|\mathcal{X}\|^2) = E\left(\int_{[0,1]} \mathcal{X}^2(t)dt\right) < \infty.$$

A continuación, se presenta la definición de media para \mathcal{X} .

Definición 2.5.1 *Si \mathcal{X} es integrable, entonces la media de \mathcal{X} es definida por la integral de Bochner*

$$\mu = E(\mathcal{X}) := \int_{\Omega} \mathcal{X} d\mathbb{P}$$

En Hsing and Eubank (2015) muestran la siguiente definición equivalente de media.

Definición 2.5.2 Si \mathcal{X} es integrable, $\mu = E(\mathcal{X})$ si para todo $f \in \mathbb{H}$

$$E(\langle \mathcal{X}, f \rangle) = \langle \mu, f \rangle$$

En esta definición se encuentra la media como extensión de la media de variables aleatorias, estas últimas resultado del producto interno en la parte izquierda de la expresión anterior.

Teorema 2.5.3 Si \mathcal{X} es cuadrado integrable entonces

$$E(\|\mathcal{X} - \mu\|^2) = E(\|\mathcal{X}\|^2) - \|\mu\|^2.$$

El operador de covarianza tiene un papel similar al que tiene la matriz de covarianza vista como transformación lineal en \mathbb{R}^p .

Definición 2.5.3 Si \mathcal{X} es cuadrado integrable, entonces el operador de covarianza para \mathcal{X} está dado por

$$\mathcal{K} = E((\mathcal{X} - \mu) \otimes (\mathcal{X} - \mu)) := \int_{\Omega} (\mathcal{X} - \mu) \otimes (\mathcal{X} - \mu) d\mathbb{P}.$$

El operador de covarianza es un operador de Hilbert-Schmidt esto implica que es acotado.

Si en \mathbb{R}^2 se consideran dos variables aleatorias X_1 y X_2 con medias $E(X_1) = \mu_1, E(X_2) = \mu_2$, y $X = (X_1, X_2)^T$, $\mu = (\mu_1, \mu_2)^T$ entonces

$$E\left((X - \mu) \otimes (X - \mu)^T\right) = \Sigma.$$

En la anterior igualdad \otimes es el producto Kronecker que es un caso particular del producto tensorial dado en la definición anterior.

Si $E(\mathcal{X}) = 0$ usando la definición 2.4.4 para el operador de covarianza se tiene:

$$E((\mathcal{X} \otimes \mathcal{X}) f) = E(\langle \mathcal{X}, f \rangle \mathcal{X}). \quad (2-3)$$

Teorema 2.5.4 El operador de covarianza \mathcal{K} admite la siguiente descomposición

$$\mathcal{K} = \sum_{j=1}^{\infty} \lambda_j e_j \otimes e_j,$$

donde los eigenvalores $\{\lambda_j\}_{j=1}^{\infty}$ son un elemento de componentes positivos de l^2 y las eigenfunciones $\{e_j\}_{j=1}^{\infty}$ forman una base ortonormal para $\overline{\text{Im}(\mathcal{K})}$ la clausura topológica de la imagen de \mathcal{K}

Es posible referirse a elementos aleatorios en el espacio de Hilbert de las funciones $\mathbb{L}^2[0, 1]$ o a los procesos estocásticos integrables definidos en el intervalo $[0, 1]$ sin lugar a confusión ya que Hsing and Eubank (2015) muestran que los enfoques son equivalentes.

Teorema 2.5.5 *Sea R un rk continuo definido sobre $[0, 1] \times [0, 1]$, y sea $\mathbf{H}(R)$ su respectivo RKHS. Un elemento aleatorio X de $\mathbf{H}(R)$ es un proceso estocástico sobre el intervalo $[0, 1]$, reciprocamente un proceso estocástico tomando valores en $\mathbf{H}(R)$ es un elemento aleatorio de $\mathbf{H}(R)$.*

Teniendo en cuenta el anterior teorema para denotar la existencia de media y covarianza de proceso se tendrá en cuenta la siguiente definición.

Definición 2.5.4 (Proceso de Segundo orden) *X es un proceso de segundo orden si la función media y covarianza están bien definidas y*

$$\mu(t) = E(X(t)), \quad K(s, t) = \text{cov}(X(s), X(t))$$

son función media y función de covarianza respectivamente.

La función de covarianza no es un operador del espacio de funciones en el mismo, sin embargo, sí hay conexión entre el operador y la función de covarianza como se muestra a continuación:

$$(\mathcal{K}f)(t) = \int_{[0,1]} K(t, s)f(s)ds. \quad (2-4)$$

El siguiente teorema relaciona las eigenfunciones y los eigenvalores del operador con la función bivariada bajo ciertas condiciones. En este trabajo el teorema es de interés cuando el kernel es la función de covarianza de un proceso X cuadrado integrable.

Teorema 2.5.6 (Teorema de Mercer) *Sea K un kernel continuo, simétrico y semidefinido positivo, con \mathcal{K} su respectivo operador integral como en la expresión 2-4. Si λ_m, e_m son los eigenvalores y eigenfunciones de \mathcal{K} luego K tiene la siguiente representación*

$$K(s, t) = \sum_{m=1}^{\infty} \lambda_m e_m(s)e_m(t)$$

y para cada s, t la suma converge absoluta y uniformemente.

El siguiente teorema expone una representación para ciertos procesos X en términos de los elementos de una base de un espacio vectorial y de variables aleatorias.

Teorema 2.5.7 (Teorema de Karhunen-Loève) *Sea X un proceso de segundo orden con función de media y covarianza continuas. Entonces:*

Existen Z_k variables aleatorias reales de media 0 tales que:

$$E(Z_i Z_j) = \delta_{ij} \lambda_i,$$

y se tiene la siguiente representación

$$X(t) - \mu(t) = \sum_{j=1}^{\infty} Z_j e_j(t),$$

donde δ_{ij} es la función delta de Kronecker, λ_j y e_j son una pareja propia del operador de covarianza mostrado en 2-4.

La expansión en la igualdad anterior recibe el nombre de expansión de Karhunen-Loève, su convergencia es fundamental para establecer el análisis de componentes principales. La hipótesis del teorema anterior también implica que el proceso es continuo en media cuadrática, esto es

$$\lim_{n \rightarrow \infty} E (X(t_n) - X(t))^2 = 0,$$

donde la sucesión $\{t_n\}$ tiende a t .

A continuación, se presenta el teorema del límite central para elementos aleatorios en un espacio de Hilbert.

Teorema 2.5.8 (Teorema del Límite Central) *Sean X_1, X_2, \dots, X_n elementos aleatorios de media 0, independientes e idénticamente distribuidos de \mathbf{H} y $E(\|X_1\|^2) < \infty$ luego*

$$\sqrt{n} \sum_{i=1}^n \mathcal{X}_i \xrightarrow{d} \xi,$$

donde ξ es un elemento gaussiano de \mathbf{H} con operador de covarianza igual a $E(\mathcal{X}_1 \otimes \mathcal{X}_1)$

La convergencia del teorema anterior es en distribución. El teorema implica que para el límite ξ en la extensión de Karhunen-Loève 2.5.7 las variables aleatorias Z_i tienen distribución normal de media 0 y varianza igual a λ_i .

3 Datos Funcionales y Componentes Principales

En el trabajo con datos funcionales es de interés el espacio \mathbb{L}^2 por las características mostradas en el capítulo 2, pero las altas dimensionalidades requieren del uso de metodologías de reducción de dimensionalidad como se explorará en este capítulo.

3.1. Datos Funcionales

Los datos funcionales son las realizaciones de uno o varios procesos estocásticos sobre un dominio, que generalmente es un intervalo de \mathbb{R} o una región de \mathbb{R}^2 , para este trabajo el dominio es el intervalo $[0, 1]$. Aunque se habla de datos funcionales los datos recolectados están discretizados sobre un dominio definido, en este caso un intervalo. Los datos discretizados requieren ser suavizados para ser considerados como funciones.

3.1.1. Datos

Un dato funcional univariado puede ser una curva, una superficie u otro elemento dependiendo del dominio, en este trabajo, dado que el dominio es un intervalo se hace referencia a una curva. En la figura **3-1** se muestra una curva ajustada a una serie de observaciones discretizadas.

El dominio de las observaciones es un subconjunto finito del dominio de las curvas, en este trabajo se considera que los datos fueron recogidos en una cantidad finita de puntos del intervalo $[0, 1]$. Por ejemplo, el intervalo $[0, 1]$ podría representar un día en el que fueron tomadas observaciones cada 2 horas, sin embargo, no es necesario que las observaciones discretizadas se hayan tomado periódicamente en el intervalo temporal.

Un dato funcional multivariado es la realización de p procesos estocásticos. En la figura **3-2** se muestra un dato funcional bivariado, si bien los dominios son iguales las imágenes de las curvas pueden variar.

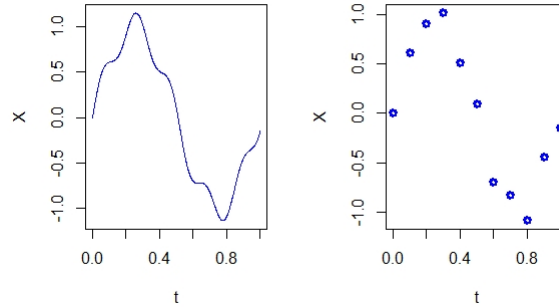


Figura 3-1: Dato funcional a la izquierda, observaciones discretizadas a la derecha.

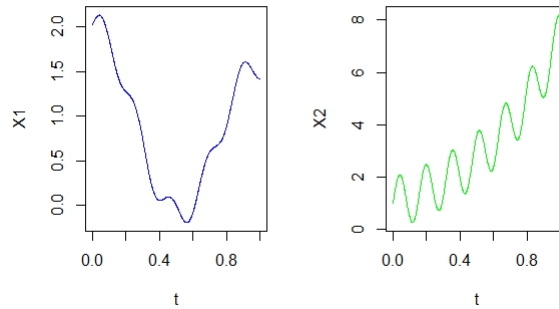


Figura 3-2: Dato funcional bivariado del proceso $\mathbf{X} = (X_1, X_2)$, las componentes X_1 y X_2 , izquierda y derecha respectivamente.

Las observaciones discretizadas de un dato funcional multivariado son un arreglo de p observaciones discretizadas de p datos funcionales univariados. Para este trabajo se asume que el dominio de las observaciones discretizadas para el dato funcional multivariado es el mismo, esto es que, en cada instante de tiempo en el que se realiza la observación se recogen la información de los p procesos. En caso de no tener en un instante de tiempo las p observaciones se podría imputar las faltantes.

Una muestra de observaciones discretizadas de datos funcionales multivariados se puede visualizar como observaciones discretizadas de los p datos funcionales o como una observación p -variada tomada a lo largo del tiempo.

Considerese una muestra de tamaño 2 de un proceso bivariado: En la figura **3-3** se muestran observaciones discretizadas correspondientes al proceso bivariado $\mathbf{X} = (X_1, X_2)$ y en la figura **3-4** se muestran como realizaciones de vectores bivariados a lo largo del tiempo.

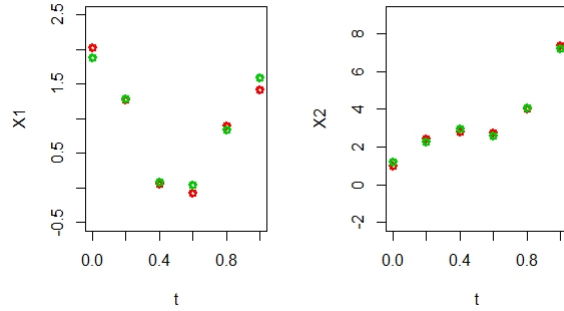


Figura 3-3: Observaciones discretizadas de un dato funcional bivariado: En rojo la muestra 1 en verde la muestra 2.

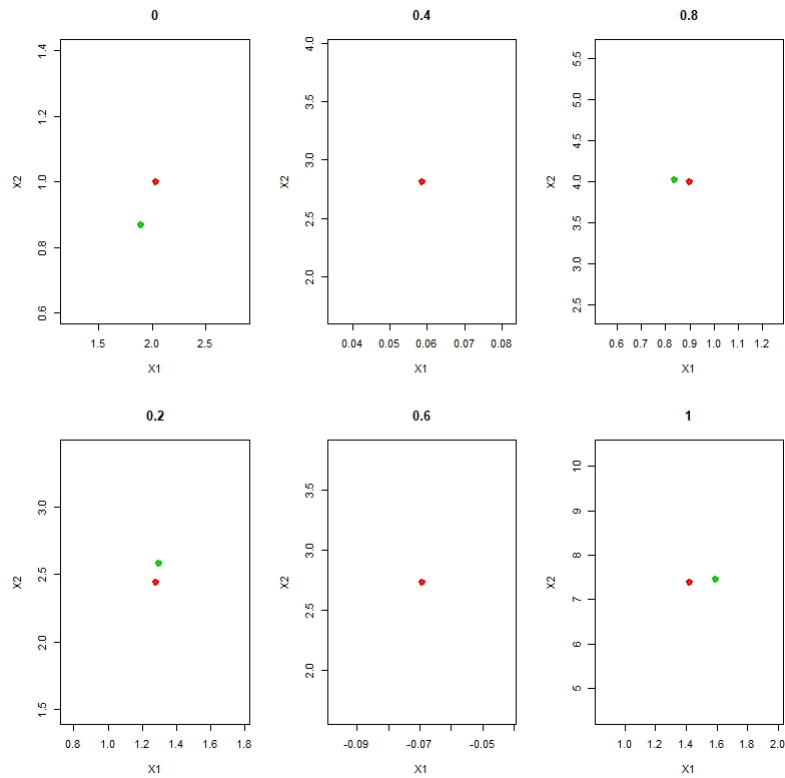


Figura 3-4: Observaciones de $\mathbf{X}(t) = (X_1(t), X_2(t))$ en $t = 0, 0.2, \dots, 1$. En rojo la muestra 1 en verde la muestra 2.

3.1.2. Suavizamiento

El suavizamiento es el proceso mediante el cual, a una observación que está discretizada se le ajusta una función. La función ajustada está en el espacio de Hilbert $\mathbb{L}^2[0, 1]$ que además es separable. Existen diferentes bases ortonormales enumerables para seleccionar, entre las más usadas se encuentran la base de Fourier, B-splines y potencias; en este trabajo se usa la base de Fourier.

Dado que el espacio $\mathbb{L}^2[0, 1]$ es de dimensión infinita lo que implica que en la base seleccionada hay infinitos elementos, una vez seleccionada la base se deben escoger el número de elementos de la base para realizar el ajuste. Si el número de elementos de la base seleccionada es pequeño es posible que se pierda información en el proceso de suavizamiento. Por otro lado, escoger un gran número de elementos de la base puede generar un sobreajuste, posiblemente generado por ruido en las observaciones.

Para solucionar los problemas de sobreajuste existen medidas que lo castigan. Por ejemplo, algunos métodos de suavizamiento castigan la rugosidad por medio de la derivada. En este trabajo para castigar el sobreajuste se utiliza el gcv por su sigla en inglés (generalized cross validation), este impone restricciones al problema de minimización del suavizamiento.

Después de escogida la base y realizado el ajuste, la curva ajustada es representada por los coeficientes que tiene en cada elemento de la base. Como la base ortonormal está fija, los procedimientos posteriores para realizar un análisis con datos funcionales se pueden realizar con los coeficientes de cada curva. Por lo tanto, si se tiene una muestra de datos funcionales univariados de tamaño N y se seleccionaron m elementos de la base para realizar el ajuste, los m coeficientes asociados a cada una de las N curvas, se pueden ver como una matriz $m \times N$, con la que se pueden realizar análisis posteriores.

Para datos funcionales p -variados sobre el intervalo $[0, 1]$ se deben realizar p suavizamientos, seleccionando para cada una de las p componentes una base ortonormal. Vale aclarar que en la propuesta desarrollada para la detección de punto de cambio en este trabajo, no se realiza directamente el suavizado a las observaciones de los p procesos, primero se transforman los datos, como se mostrará en el siguiente capítulo, y luego se realiza el suavizamiento. Una revisión de los métodos de suavizamiento y del gcv se encuentra en Ramsay (2004).

Base de Fourier

Los elementos de la base de Fourier son funciones periódicas que ayudan a describir muchos de los fenómenos en los que se usan datos funcionales, por ejemplo; temperaturas, concentraciones de contaminantes, presiones entre otras.

En la base de Fourier se incluyen las funciones senos, cosenos y una constante. Las derivadas de los elementos de la base alternan entre los elementos de la base en el caso de los senos y cosenos, mientras que la constante desaparece.

Al principio de la investigación se contempló el uso de las derivadas para detectar cambios de forma en la media. Sin embargo, dado que el operador diferenciable no es acotado un estudio formal del uso apropiado de la derivada requiere una investigación a profundidad. En el apéndice 7 se presenta una discusión sobre una variante de estudio en la detección de cambios de forma en la media funcional basada en la inspección de la derivada usando la base de Fourier, sin embargo, hay consideraciones a tener en cuenta y son presentadas en el apéndice.

3.2. Componentes Principales

En general, se usa el análisis de componentes principales con dos intenciones:

1. Realizar una reducción de dimensionalidad reteniendo las direcciones en las que más existe variabilidad, esto se traduce en trabajar en un espacio de menor dimensionalidad en el que la pérdida de información es poca.
2. Transformar los vectores o funciones de tal manera que se elimine la influencia de la estructura de covarianza entre las componentes de los vectores o funciones respectivamente.

A continuación se describen las ideas geométricas del análisis de componentes principales, su relación con un problema de optimización con restricciones y se describirán los métodos para hallar componentes principales para datos multivariados, datos funcionales y datos funcionales multivariados desde varios enfoques.

Por comodidad y dado que el análisis de componentes principales no depende de la media, porque estudia la estructura de covarianza; en este capítulo todos los objetos aleatorios a los que se hace referencia tienen por media el elemento 0 del espacio al que pertenezcan.

En el análisis de componentes principales no hay supuestos acerca de la distribución, sólo es necesario que exista la covarianza. Vale advertir que en el caso de normalidad hay características adicionales sobre las distribuciones de los valores propios, pero los enfoques aquí presentados son generales.

3.2.1. Una vista geométrica

Se presenta un ejemplo en \mathbb{R}^2 para ilustrar gráficamente. Sea X un vector aleatorio con realizaciones en el espacio \mathbb{R}^2 , sean X_1, X_2 las componentes del vector y la matriz de

covarianza dada por

$$\Sigma = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$$

En la figura **3-5** se muestra el diagrama de dispersión de una muestra simulada con dos datos externos. En el diagrama el punto azul está en la dirección en la que hay más variabilidad en los datos, mientras que el verde está en la dirección ortogonal a la de mayor variabilidad. No es correcto, desde el punto de vista estadístico, argumentar que el punto azul es atípico porque está más alejado de los demás datos y que el punto verde no es atípico.

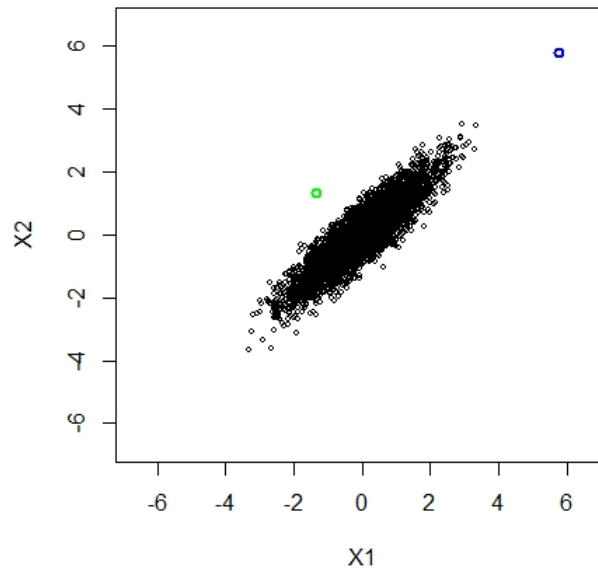


Figura 3-5: Diagrama de dispersión de una muestra simulada de X de tamaño 5000 con estructura de covarianza Σ .

Al transformar los datos, considerando la influencia de la estructura de covarianza se obtuvo el gráfico que se muestra en la figura **3-6**, la matriz de covarianza de los datos transformados es la matriz identidad. Si T es la transformación lineal entonces

$$T(X_1, X_2) = (Y_1, Y_2),$$

con $\text{cov}(Y_1, Y_2) = I_2$.

En el gráfico **3-6** se aprecia que los dos datos externos están a 6 desviaciones estándar sobre los ejes principales por lo que las decisiones acerca de si son atípicos debe ser la misma para ambos.

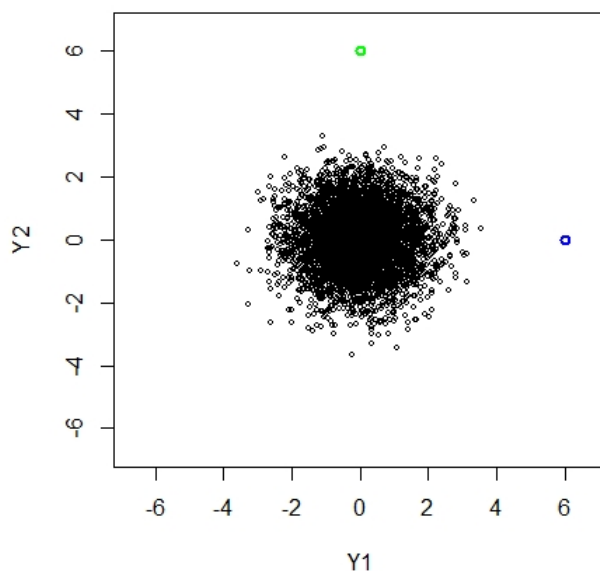


Figura 3-6: Diagrama de dispersión de la transformación de los datos simulados.

La matriz de transformación de los datos para eliminar la influencia de la estructura de covarianza es la inversa de cualquiera de las raíces de la matriz de covarianza. Más adelante se realiza la elección de una raíz específica cuando los supuestos así lo exigen.

La idea de aplicar la transformación es válida siempre que la matriz de covarianza sea invertible y esto se tiene siempre que los valores propios sean positivos. En el contexto de los operadores el hecho de que la sucesión de valores propios esté conformada solo por valores positivos no implica que sea posible aplicar la transformación. El problema de la dimensionalidad debe ser tenido en cuenta debido a que la sucesión de valores propios en el operador debería estar tendiendo a 0 para que el operador sea acotado, pero esto implica que los inversos multiplicativos son no acotados.

3.2.2. Un problema de optimización

A lo largo de este trabajo siempre que se haga referencia a isomorfismos, serán isomorfismos de espacios métricos o sea isometrías que preservan la estructura de espacio vectorial. En Vasudeva and Shirali (2017) muestran que una isometría también es un isomorfismo conservando la estructura de espacio de Hilbert. Sean $\{\mathbf{H}, \langle \cdot, \cdot \rangle\}$ un espacio de Hilbert isomorfo a $\{\mathbb{L}^2[0, 1], \langle \cdot, \cdot \rangle\}$ o a \mathbb{R}^p ; y X un objeto aleatorio de \mathbf{H} con covarianza (matriz u

operador) dada por K y $E(\|X\|^2) < \infty$.

Considere el siguiente problema de optimización que busca encontrar las direcciones para las cuales la estructura de covarianza presenta mayor variabilidad.

Algoritmo de maximización (3-1)

$$\begin{aligned} & \text{máx} && \langle Ke_1, e_1 \rangle \\ & \|e_1\| = 1 \end{aligned}$$

para $l > 1$ y $j < l$

$$\begin{aligned} & \text{máx} && \langle Ke_l, e_l \rangle \\ & \|e_l\| = 1 \\ & \langle e_l, e_j \rangle = 0 \end{aligned}$$

repetir siempre que $\langle Ke_l, e_l \rangle > 0$ y se cumplan las restricciones.

Para los espacios isomorfos a \mathbb{R}^p el proceso 3-1 tiene a lo más p pasos, mientras que para \mathbb{L}^2 puede ser un proceso infinito. En cada iteración l al valor máximo se le nota como λ_l , ya que este efectivamente corresponde al multiplicador de Lagrange de la optimización de la función Lagrangiana asociada al problema con restricción expuesto en 3-1. Debido a las propiedades de K los valores λ son todos no negativos y de aquí en adelante se asumirá que la sucesión $\{\lambda_l\}$ es estrictamente decreciente, esto es:

$$\lambda_{l'} < \lambda_l \text{ si } l' > l. \tag{3-2}$$

Teorema 3.2.1 *Sea $\{\mathbf{H}, \langle \cdot, \cdot \rangle\}$ un espacio de Hilbert isomorfo a \mathbb{R}^p o a \mathbb{L}^2 , y sea K la covarianza (matriz u operador) del objeto aleatorio X con realizaciones en \mathbf{H} . Los vectores o funciones e_l son únicos salvo por un signo.*

Para el caso \mathbb{R}^p , la demostración se puede encontrar en Anderson (2003), y para \mathbb{L}^2 en Horváth and Kokoszka (2012). Los enfoques para el tratamiento con observaciones son desarrollados para los dos casos en las mismas referencias respectivamente.

Hasta el momento no se ha establecido la relación entre lo descrito con la transformación de la sección anterior y el problema de optimización descrito en ésta; en el problema de optimización de la función Lagrangiana las derivadas para optimizar se pueden simplificar, como se mostrará más adelante, y llegar a las ecuaciones propias. O sea que la búsqueda de los λ_l, e_l del problema optimización coinciden con la búsqueda de las parejas propias. El espacio generado por los vectores propios, en la mayoría de los casos, no es isomorfo

al espacio de inicial, es decir las normas no son las mismas, por lo tanto como espacios métricos son distintos. Más aún, en el caso de \mathbb{L}^2 , el espacio generado puede ser un subconjunto propio de \mathbb{L}^2 ; por ejemplo para el movimiento browniano la estructura de covarianza, como se mencionó en 2.3.4, da lugar a un subespacio vectorial de \mathbb{L}^2 que no es completo.

Existen problemas asociados a la estabilidad de las matrices inversas; en el trabajo con datos reales sean datos multivariados o funcionales, no se dispone de la estructura de covarianza real sino de un estimado. La existencia de valores propios muy cercanos a 0 dan lugar a problemas en la transformación que pueden llevar a tomar decisiones que no son correctas.

Para evitar el problema de valores propios muy cercanos a 0, y que además es el fundamento del método de componentes principales para la reducción de dimensionalidad, se puede fijar un valor ϵ y repetir el problema de optimización planteado en 3-1 hasta que el máximo de la optimización quede por debajo del valor ϵ , después realizar la proyección del espacio sobre los vectores que resultaron y trabajar sobre este espacio proyectado. El valor de ϵ conveniente no es fijo por lo que lo más común es que el algoritmo se detenga cuando la suma de los $\{\lambda_i\}$ alcance un alto porcentaje del total de la variabilidad asociada a la estructura de covarianza. Para el uso de la metodología propuesta cuando se usan componentes principales se recomienda que la variabilidad explicada esté entre el 95 y 99%, cuando sea posible.

3.2.3. Componentes principales: Análisis Multivariado

Sea X un vector aleatorio p -variado con matriz de covarianza Σ de rango p , a continuación, se presentan los pasos para hallar las parejas propias:

1. Hallar los p valores propios: Los p valores propios de Σ pueden ser hallados como las raíces del polinomio característico $\det(\Sigma - \lambda I_p)$ donde I_p es la identidad de tamaño p .
2. Ordenar los valores propios de mayor a menor.
3. Encontrar los vectores propios asociados a cada uno de los componentes principales.
4. Hallar la varianza total de X : Esto corresponde a la traza de Σ .
5. Escoger los d valores propios más altos hasta que su suma alcance un porcentaje respecto a la varianza total, se puede tomar como 90% pero depende del problema.
6. Realizar la proyección del espacio original sobre los d vectores propios correspondientes a los valores propios más grandes.

7. Las d proyecciones de cada elemento del espacio inicial forman un nuevo espacio, estas d componentes correspondientes a las proyecciones reciben el nombre de puntajes o componentes sobre la base propia.

Sea V una matriz formada por los p vectores propios correspondientes a los valores propios ordenados de mayor a menor o sea

$$V = [e_1, e_2, \dots, e_p],$$

aquí e_1 corresponde al vector propio asociado al valor propio más grande, así hasta e_p que corresponde al vector propio asociado al valor propio más pequeño. Asumiendo la condición 3-2 y dado que para cada valor propio λ_i hay dos posibilidades para el vector propio, esto deja 2^p posibilidades para la matriz V . Esto es un hecho relevante que se tendrá en cuenta en este trabajo para la implementación de la metodología propuesta. Si hay valores propios repetidos las posibilidades para V pueden ser infinitas, por ejemplo en \mathbb{R}^2 si los valores propios son ambos 1, V podría ser cualquier matriz ortogonal.

Teniendo en cuenta que V es ortogonal y sea

$$D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$$

una matriz diagonal con los valores propios ordenados se tiene

$$\Sigma = V D V^T,$$

como D está formada por los valores propios de Σ , $D^{1/2}$ es la matriz diagonal con las raíces de estos valores propios. Por lo tanto

$$\Sigma = (V D^{1/2}) (V D^{1/2})^T$$

Si se toma $B = (V D^{1/2})$ entonces $Y = B^{-1}X$ es un vector con estructura de covarianza igual a la matriz identidad de tamaño p . El espacio RKHS del ejemplo 2.3.2 es isomorfo con el espacio resultante después de aplicar la transformación B^{-1} .

Para hallar las componentes principales de un conjunto de observaciones, considerense X_1, X_2, \dots, X_N realizaciones independientes del vector aleatorio X . Sea S un estimador de la matriz de covarianza Σ , con S una matriz invertible, para que esta condición se cumpla es necesario que $N \geq p$. La matriz S se puede tomar como

$$S = \frac{1}{N-1} \mathbb{X} \mathbb{X}^T,$$

donde \mathbb{X} es de dimensiones $p \times N$ y es la matriz en la que cada X_i es una columna, la elección entre el cociente $N-1$ y N es irrelevante ya que en cualquier caso el problema de optimización solo depende de \mathbb{X} .

El procedimiento para hallar las componentes principales es seguir los pasos indicados en 3.2.3 cambiando la matriz Σ por la matriz S . Si se está aplicando componentes principales para reducir dimensionalidad entonces se tiene que $d < p$, en este caso la proyección de cada realización X_i sobre cada vector propio e_l es una transformación de los datos que ahora tienen una estructura de covarianza dada por la identidad de tamaño d .

Si el vector aleatorio X tiene distribución normal hay características relevantes de los valores propios estimados que son estudiadas en Anderson (2003) pero que en este trabajo no son necesarias.

3.2.4. Componentes principales: Datos Funcionales Univariados

La metodología de componentes principales para un objeto aleatorio en un espacio de Hilbert está asociada al problema de optimización revisado en el teorema 3.2.1. Desde el punto de vista teórico el problema se soluciona teniendo el operador de covarianza y encontrando las parejas propias que en este caso están formadas por los valores propios y las funciones propias. Si $K(s, t)$ es la función de covarianza y \mathcal{K} el operador, ambos asociados al proceso X entonces el objetivo es hallar funciones tales que:

$$(\mathcal{K}f)(t) = \lambda f(t),$$

teniendo en cuenta la forma en la que actúa el operador de acuerdo a la expresión dada en 2-4, se tiene

$$\int_0^1 K(t, s)f(s)ds = \lambda f(t).$$

Por ejemplo, Hsing and Eubank (2015) muestran que para el operador de covarianza del movimiento browniano los valores propios están dados por

$$\lambda_j = \frac{4}{((2j - 1)\pi)^2},$$

y las funciones propias por

$$e_j = \sqrt{2} \sin\left(\frac{(2j - 1)\pi}{2}t\right),$$

también muestran que esta es una base para el espacio de Sobolev.

El proceso de encontrar funciones propias de operadores de covarianza que pueden ser representados como en la expresión 2-4, conlleva generalmente a la solución de ecuaciones diferenciales, mas en este trabajo la necesidad está en el cálculo de los componentes principales para datos funcionales y no para los objetos teóricos.

Antes de abordar el enfoque usado en este trabajo para componentes principales de datos funcionales, se realiza una analogía para explicar la pertinencia del método. El sistema numérico usado en la mayoría de aplicaciones es el conjunto de los números reales, sin embargo computacionalmente es imposible representarlos a todos. La propiedad de densidad que tienen los números racionales en los reales permite poder realizar una muy buena aproximación a todo número real; esto es análogo a lo que sucede con el espacio \mathbb{L}^2 y las bases de funciones continuas. En 2.1.5 se están definiendo espacios de Hilbert que tengan una propiedad similar a la que tienen los reales con los racionales. De la proposición 2 y el teorema 2.1.2 se dedujo que \mathbb{L}^2 tiene la propiedad de la definición 2.1.5, o sea tiene un subconjunto denso que es enumerable.

El objetivo de los componentes principales univariados funcionales, en este caso, es realizar una reducción de dimensionalidad teniendo en cuenta la estructura de covarianza.

Sea X_k un proceso de segundo orden con funciones de media y covarianza continuas que también se puede ver como un elemento aleatorio de \mathbb{L}^2 , entonces se tiene la extensión de Karhunen-Loève del teorema 2.5.7. Los datos siguen un modelo dado por

$$X_k(t) = \mu_k(t) + Y_k(t),$$

donde $\mu_k(t)$ es una función determinística de t y por lo tanto la estructura de covarianza de los procesos X_k y Y_k es la misma, pero Y_k es de media 0.

Para la aplicación de las metodologías a los datos, se supondrá que las realizaciones discretizadas de X_k tienen la siguiente estructura:

$$(x_{j,k,i}), \tag{3-3}$$

con j como subíndice temporal, k variando en $1, 2, \dots, p$ e i variando de 1 al tamaño de la muestra N . k se encuentra fijo a lo largo de este análisis, indicando que se realiza PCA para una componente del proceso.

Existen diversos enfoques para aplicar componentes principales a datos funcionales univariados, a continuación se presentan una serie de caminos para realizar la estimación.

Hsing and Eubank (2015) presentan estudios para los estimadores de media y covarianza; además muestran resultados asintóticos para las parejas propias. En los siguientes casos:

- Vía suavizamiento local.
- Vía mínimos cuadrados penalizados, empleando un suavizamiento spline con penalización.
- Vía el operador de covarianza muestral.

Horváth and Kokoszka (2012) suponen que una extensión finita de Karhunen-Loève aproxima a la expansión real, para determinar un criterio de porcentaje de varianza acumulada para las componentes principales empíricas. Ellos también muestran el uso de la función `pca.fd` de `R` para computar las componentes principales. Ramsay (2004) propone como camino para hallar las parejas propias primero el suavizamiento mediante cualquier conjunto de funciones de una base ortonormal para reducir el problema al trabajo con los componentes en la base seleccionada. Allí también se presenta un ejemplo seleccionando la base de Fourier.

En este trabajo se sigue lo descrito por Ramsay (2004) y por Hsing and Eubank (2015) para las bases de Fourier. La función `pca.fd` del paquete `fda` para `R` es usada para componentes principales univariados funcionales.

En la mayoría de análisis de componentes principales para el caso de datos multivariados, datos funcionales univariados y multivariados se tiene como supuesto que los valores propios son de multiplicidad 1. Los casos de valores propios con multiplicidad mayor a 1 tienen un tratamiento distinto y en ocasiones mucho más abstracto, en este trabajo siempre se tiene como supuesto que en cualquier análisis de componentes principales los valores propios son de multiplicidad 1 de acuerdo a la condición expuesta en 3-2.

3.2.5. Componentes principales: Datos Funcionales Multivariados

Existen varios enfoques para la adaptación de la metodología de componentes principales a los datos funcionales multivariados, algunos presentan la extensión de Karhunen-Loève.

Ramsay (2004) realiza un estudio para un proceso bivariado con datos reales. Allí se considera el producto interno del espacio de Hilbert dado por la suma directa de los espacios de cada uno de los procesos componentes sin tener en cuenta, en principio, la estructura de correlación entre las variables. Para cada dato funcional bivariado ellos forman un solo vector de valores concatenando las componentes de cada observación.

A continuación, se exponen dos enfoques de componentes principales para datos funcionales multivariados ordenados en orden crónologico de la publicación.

MFPCA-Berrendero

El enfoque descrito en esta sección fue desarrollado por Berrendero et al. (2011).

Sean $\mathbf{X} = (X_1, X_2, \dots, X_p)^T$ un proceso estocástico p -variado definido sobre el intervalo $[0, 1]$, y de media $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_p)$ en la que cada función media es 0 o sea $\mu_k = 0$ para $k = 1, 2, \dots, p$; y $\boldsymbol{\Sigma}(t) = E(\mathbf{X}(t)\mathbf{X}(t)^T)$ es una función que en cada instante de tiempo devuelve la matriz de covarianza para cada t en el intervalo $[0, 1]$.

Considerese $e(t)$ una función de $[0, 1]$ en \mathbb{R}^p , para cada $t \in [0, 1]$ se tiene que

$$\text{var} (e(t)^T \mathbf{X}(t)) = e(t)^T \Sigma(t) e(t),$$

la idea es maximizar la varianza integrada dada por

$$\int_0^1 e(t)^T \Sigma(t) e(t) dt, \quad (3-4)$$

sujeto a la restricción de que para cada t se tiene $\|e(t)\| = 1$, esta última es la norma inducida por el producto interno usual de \mathbb{R}^p . La solución del problema es única salvo por el signo en cada instante de tiempo.

Sea

$$Z_1(t) = e_1(t)^T \mathbf{X}(t),$$

donde $e_1(t)$ es una solución del problema 3-4, a $Z_1(t)$ se le llama primer componente de \mathbf{X} . El procedimiento para encontrar las demás componentes principales

$$Z_r(t) = e_r(t)^T \mathbf{X}(t),$$

es maximizar la expresión 3-4, con la condición de que para cada t el vector $e_r(t)$ es normal con la norma usual de \mathbb{R}^p y además $e_r(t)^T e_l(t) = 0$ para cada t en $[0, 1]$ y $l = 1, 2, \dots, r-1$.

Proposición 3 Para $t \in [0, 1]$, sean $\lambda_1(t) > \lambda_2(t) > \dots > \lambda_p(t) > 0$ funciones que para cada t devuelven los valores propios de la matriz que devuelve $\Sigma(t)$. Para $r = 1, 2, \dots, p$ y $s = 1, 2, \dots, p$ se tiene:

- Para todo $t \in [0, 1]$, $e_r(t)$ es un vector unitario correspondiente a $\lambda(t)$
- $\langle Z_r, Z_s \rangle_1 = 0$ para $r \neq s$, con la medida adecuada

$$\langle Z_r, Z_s \rangle_1 := E \left(\int_0^1 Z_r(t) Z_s(t) dt \right).$$

- Con la norma inducida por el producto interno definido en el ítem anterior

$$\|Z_r\|_1^2 = \int_0^1 \lambda_r(t) dt.$$

En la hipótesis de la proposición anterior implícitamente está el hecho de que la multiplicidad de los vectores propios es 1, en la siguiente proposición también es requerido esta hipótesis.

Proposición 4 Para cada $t \in [0, 1]$ la matriz Σ_t que corresponde a la imagen de la función $\Sigma(t)$, tiene p valores propios distintos. Además, $\Sigma(t)$ es diferenciable con matriz de derivadas por cada componente notada por $\dot{\Sigma}(t)$. Entonces para $r = 1, 2, \dots, p$ se tiene:

1. La función $\lambda_r(t)$ es diferenciable con derivada dada por:

$$\dot{\lambda}_r(t) = e_r(t)^T \dot{\Sigma}(t) e_r(t)$$

2. Es posible escoger $e_r(t)$ de tal manera que sea una función diferenciable respecto a t . El vector derivada está dado por:

$$\dot{e}_r(t) = - \left[\sum_{l \neq r} [\lambda_l(t) - \lambda_r(t)]^{-1} e_l(t) e_l(t)^T \right] \dot{\Sigma}(t) e_r(t)$$

A continuación se presenta una demostración alternativa a la dada por Berrendero para la primera parte de la proposición.

D/. El polinomio característico $q(\lambda, t)$ es diferenciable respecto a t , ya que la función que devuelve matrices $\Sigma(t)$ tiene componentes diferenciables y la función determinante también es diferenciable. Dado que para cada t la matriz Σ_t tiene p valores propios distintos es posible factorizar el polinomio característico como

$$q(\lambda, t) = \prod_{r=1}^p (\lambda(t) - c_r(t)),$$

en esta expresión $c_r(t)$ es una función diferenciable respecto a t por el teorema de la función implícita. Es posible ordenar los valores de $c_r(t)$ para que coincidan con los valores $\lambda_r(t)$.

En los siguientes pasos dado que todas las funciones tienen por argumento a t se omite el argumento. En $\lambda_r = e_r^T \Sigma e_r$ es posible usar de nuevo el teorema de la función implícita para deducir que e_r también es diferenciable.

Diferenciando $\lambda_r = e_r^T \Sigma e_r$ se tiene:

$$\dot{\lambda}_r = \dot{e}_r^T \Sigma e_r + e_r^T \dot{\Sigma} e_r + e_r^T \Sigma \dot{e}_r$$

en la expresión anterior dado que cada sumando es de dimensión 1 se puede aplicar traspuesta, de esta forma se puede ver que el primer y el tercer término son iguales. Ahora

$$\dot{e}_r^T \Sigma e_r = \dot{e}_r^T \lambda_r e_r = \lambda_r \dot{e}_r^T e_r$$

pero $\dot{e}_r^T e_r = 0$ ya que e_r es unitario, entonces $\dot{\lambda}_r = e_r^T \dot{\Sigma} e_r$



En Berrendero et al. (2011) presentan una medida para el porcentaje de variabilidad explicada por los componentes propios.

MFPCA-HAPP

En Happ and Greven (2018) se muestra un análisis de componentes principales para datos funcionales multivariados sobre diferentes dominios. Bajo ese análisis los dominios no son necesariamente intervalos de \mathbb{R} , pueden ser subconjuntos de \mathbb{R}^p ; de aquí que los elementos aleatorios no son necesariamente procesos estocásticos, sino que pueden ser campos aleatorios. En este trabajo se limitará el uso del enfoque dado por Happ and Greven (2018) a solamente el intervalo $[0, 1]$

En Happ and Greven (2018) se define el producto interno para el espacio de Hilbert de las funciones sobre diferentes dominios, luego para los objetos aleatorios definen el operador de covarianza y se demuestra que el operador definido tiene las características suficientes para tener la extensión del teorema de Mercer. Se expone la extensión del teorema de Karhunen-Loève 2.5.7 para los objetos aleatorios definidos y se da una caracterización para que la expansión sea finita. Presentan la metodología para su análisis de componentes principales, estudian propiedades asintóticas, realizan simulaciones, y aplican la metodología a un conjunto de neuroimágenes.

Sea $\mathbf{X} = (X_1, X_2, \dots, X_p)$ un proceso estocástico p -variado definido sobre el intervalo $[0, 1]$, para el cual la extensión de Karhunen-Loève existe. Considerese un conjunto de N observaciones con la notación ya indicada (x_{jki}) los pasos para aplicar el análisis de componentes principales presentado en Happ and Greven (2018) son:

1. Para cada proceso univariado X_k se aplica un análisis de componentes principales univariado. Esto es para k' fijo y $k' = 1, 2, \dots, p$, los datos $(x_{jk'i})$ son una observación de un dato funcional univariado. Para $(x_{jk'i})$ se obtienen las eigenfunciones $\hat{\phi}_{k',l}$ y los puntajes $\hat{\xi}_{lk'i}$, esto significa que $\hat{\xi}_{lk'i}$ es el producto interno de la i -ésima observación con la función $\hat{\phi}_{k',l}$, donde el subíndice l varía desde 1 hasta d_k .

2. Sea

$$D = \sum_{k=1}^p d_k,$$

Se define la matriz de N filas y D columnas Ξ , donde cada fila i es la concatenación de los vectores formados por los puntajes $\hat{\xi}_{lk'i}$

$$\left(\hat{\xi}_{11i}, \hat{\xi}_{12i}, \dots, \hat{\xi}_{1d_1i}, \dots, \hat{\xi}_{p1i}, \dots, \hat{\xi}_{pd_pi} \right).$$

3. Se realiza un análisis de componentes principales para la matriz cuadrada de dimensión D dada por:

$$\frac{1}{N-1} \Xi^T \Xi,$$

los valores propios se notan por v_l y la matriz con los respectivos vectores propios en cada columna notada por C .

4. Los estimadores para las funciones propias multivariadas están dadas por

$$\psi_{l,k}(t) = \sum_{q=1}^{d_k} [c_{q,l}]_k \hat{\phi}_{k,q}(t),$$

donde $[c_{q,l}]_k$ corresponde al escalar que está en la posición q del vector de d_k componentes que corresponde al bloque k del vector propio asociado a v_l , los puntajes multivariados para cada $\hat{\phi}_{k,q}(t)$ y cada observación i son las componentes de la matriz de dimensiones $N \times D$ dada por

$$\Xi \cdot C,$$

cada fila tiene los puntajes de cada observación.

En Happ-Kurz (2020) se muestra el uso de la función MFPCA del paquete `funData` para calcular las componentes principales. Existen otros enfoques para MFPCA, por ejemplo Jacques and Preda (2014) realizan un suavizamiento para los p procesos y construyen una matriz con los puntajes sobre las bases seleccionadas. Chiou et al. (2014) presentan un enfoque en el que tienen en cuenta las correlaciones entre procesos y además consideran la sensibilidad a cambios de escala dando un enfoque normalizado.

4 Punto de Cambio y Propuesta

En este capítulo se propone una metodología para detectar puntos de cambio en la media de procesos funcionales multivariados. De manera informal, el problema de punto de cambio en la media se puede reducir al análisis de las distancias de los datos a la media teniendo en cuenta la estructura de covarianza. Los espacios RKHS construidos a partir de la función de covarianza son espacios en los que se está teniendo en cuenta la estructura de covarianza para definir el producto interno, esto ya hace implícito que en la medición de distancias se contemple la estructura de covarianza.

4.1. Punto de cambio

Si X_1, X_2, \dots, X_N son objetos aleatorios para los que existe la media, en general, el problema de detección de punto de cambio en la media consiste en el contraste de las siguientes hipótesis:

$$H_0 : E(X_1) = E(X_2) = \dots = E(X_N), \quad (4-1)$$

contra la hipótesis de que existe un k^* tal que

$$E(X_1) = E(X_2) = \dots = E(X_{k^*}) = \mu_1 \neq \mu_2 = E(X_{k^*+1}) = E(X_{k^*+2}) = \dots = E(X_N). \quad (4-2)$$

En la mayoría de estudios sobre la detección de punto de cambio los valores μ_1, μ_2, k^* son desconocidos y el objetivo es estimarlos y contrastar las hipótesis. En algunos estudios se asume que la estructura de covarianza es conocida y en otros se asume que si bien no es conocida esta no cambia.

Aunque existen diversas metodologías para la detección de punto de cambio en la media de datos funcionales univariados, no hay una propuesta explícita para datos funcionales multivariados. La metodología aquí propuesta para el caso multivariado traslada el problema a una estructura en la que haya la posibilidad de aplicar las metodologías univariadas de manera adecuada. A continuación se muestra la metodología propuesta en Berkes et al. (2009) para el caso funcional univariado y que será adaptada para el caso multivariado funcional.

4.1.1. Punto de cambio univariado funcional: Berkes

Sean X_i con $i = 1, 2, \dots, N$, observaciones funcionales univariadas definidas sobre un compacto \mathbf{D} , y para este caso en específico se considera el intervalo $[0, 1]$. Las hipótesis corresponden a las planteadas en 4-1 y 4-2. Bajo la hipótesis nula se tiene el siguiente modelo:

$$X_i(t) = \mu(t) + Y_i(t), \quad (4-3)$$

donde $\mu(t)$ es la función media de X_i y $Y_i(t)$ es un proceso de media 0.

Nota 4.1.1 (Supuestos) *Los siguientes supuestos para la media y covarianza son tenidos en cuenta.*

1. *La función media $\mu(t)$ es una función en $\mathbb{L}^2[0, 1]$, los errores $Y_i(t)$ son independientes e idénticamente distribuidos de media 0 y son objetos aleatorios de $\mathbb{L}^2[0, 1]$ con*

$$E \|Y_i\|^2 < \infty.$$

Por la definición 2.5.3 se tiene que el operador de covarianza es de Hilbert-Schmidt y se obtienen los resultados de los teoremas 2.5.3 y 2.5.4 además de las expansiones de Mercer y Karhunen-Loève respectivamente en 2.5.6 y 2.5.7.

2. *Los eigenvalores usados en la expansión de Karhunen-Loève están ordenados y satisfacen que para algún $d > 0$*

$$\lambda_1 > \lambda_2 > \dots > \lambda_d > \lambda_{d+1}.$$

- 3.

$$E \|Y_i\|^4 < \infty.$$

4. *Las observaciones siguen el modelo:*

$$X_i(t) = \mu_\gamma(t) + Y_i(t),$$

donde γ es 1 si $i \leq c^$ y es 2 en otro caso, siendo $c^* = N\theta$ con $\theta \in (0, 1)$ y N el tamaño de la muestra.*

Se definen los siguientes estadísticos que son dos estimadores de la media.

$$\hat{\mu}_c(t) = \frac{1}{c} \sum_{i=1}^c X_i(t), \quad \check{\mu}_c(t) = \frac{1}{N-c} \sum_{i=c+1}^N X_i(t), \quad (4-4)$$

$\hat{\mu}_c$ es una media calculada con las primeras c observaciones y $\check{\mu}_c$ es una media calculada con las restantes, si no se presentan cambios en la media los dos valores son cercanos. Si c es cercano a 1 $\hat{\mu}_c$ considera pocos valores para su cálculo mientras que $\check{\mu}_c$ tiene la mayoría de la información, si c es cercano a N sucede el caso contrario, en cualquiera de los dos casos $\hat{\mu}_c$ y $\check{\mu}_c$ pueden ser muy diferentes y el test podría detectar un punto de cambio aún cuando no lo hay. Para solucionar este problema se puede multiplicar la diferencia de los valores por un peso parabólico en función de c . Se considera

$$P_c(t) = \frac{c(N-c)}{N} (\hat{\mu}_c(t) - \check{\mu}_c(t)), \quad (4-5)$$

donde $\frac{c(N-c)}{N}$ es el factor de tipo parabólico.

La idea es expresar la norma de la diferencia entre las medias, de la expresión anterior en términos de los puntajes obtenidos por las componentes principales, para esto se calculan los scores de la siguiente manera

$$\hat{\eta}_{i,l} = \left\langle X_i(t) - \bar{X}_N(t), \hat{e}_l \right\rangle, \quad (4-6)$$

donde \hat{e}_l es la estimación de la l -ésima eigenfunción, $i = 1, 2, \dots, N$ y $l = 1, 2, \dots, d$ siendo N, d el número de observaciones y número de componentes principales respectivamente.

Se considera el siguiente estadístico

$$T_N(x) = \frac{1}{N} \sum_{l=1}^d \left[\frac{1}{\hat{\lambda}_l} \left(\sum_{1 \leq i \leq Nx} \hat{\eta}_{i,l} - x \sum_{1 \leq i \leq N} \hat{\eta}_{i,l} \right)^2 \right], \quad (4-7)$$

con $x = c/N$, bajo la hipótesis nula 4-1 la anterior expresión converge en distribución a

$$T_N(x) \xrightarrow{dis} \sum_{l=1}^d B_l^2(x),$$

donde B_1, B_2, \dots, B_d denotan puentes Brownianos independientes. Berkes et al. (2009) también exponen el estadístico

$$S_{N,d} := \int_0^1 T_N(x) dx = \frac{1}{N^2} \sum_{l=1}^d \left[\frac{1}{\hat{\lambda}_l} \sum_{c=1}^N \left(\sum_{i=1}^c \hat{\eta}_{i,l} - \frac{c}{N} \sum_{i=1}^N \hat{\eta}_{i,l} \right)^2 \right], \quad (4-8)$$

éste bajo la hipótesis nula 4-1 converge en distribución a

$$S_{N,d} \xrightarrow{dis} \int_0^1 \sum_{l=1}^d B_l^2(x) dx = K_d. \quad (4-9)$$

Los valores críticos de la distribución K_d son mostrados al final del capítulo, esta distribución fue derivada por Kiefer (1959). La hipótesis nula se rechaza a un nivel de significancia α si $S_{N,d}$ es mayor que el cuantil $1 - \alpha$ de la distribución K_d .

Existen diferentes formas de usar componentes principales lo que implica diferentes formas de escoger el valor d . En este trabajo primero se realiza el suavizamiento y después se aplica el método de componentes principales para datos funcionales univariados, este procedimiento es descrito en Ramsay (2004). La selección del valor d se basa en la proporción de varianza explicada por las d primeras componentes principales de la misma forma en la que se realiza en el análisis de componentes principales multivariado.

Teniendo en cuenta que $x \in [0, 1]$, en Berkes et al. (2009) muestran que el valor x^* que maximiza $T_N(x)$, multiplicado por N corresponde a un estimador consistente del punto de cambio. Bajo el supuesto 4 de la nota 4.1.1 se tiene la unicidad del valor x^* .

Si bien es cierto que el supuesto 4 implica el estudio de sólo un punto de cambio, Berkes et al. (2009) indican que el test tiene una buena potencia incluso cuando hay múltiples puntos de cambio. Para detectar varios puntos de cambio el procedimiento es iterativo, esto es; con las N observaciones se calcula el estadístico $S_{N,d}$ como en 4-8, si se rechaza la hipótesis nula se repite el procedimiento con las observaciones antes y después del punto de cambio por separado, el procedimiento se repite hasta que no se rechace la hipótesis nula.

4.1.2. Otros enfoques

En Aston and Kirch (2011) examinan el problema de detección de punto de cambio para observaciones funcionales con dependencia y la posibilidad de ocurrencia de varios puntos de cambio. Allí también se realiza una aplicación a imágenes de resonancia magnética, para esto el dominio no es un intervalo temporal sino un dominio en el que hay una dimensión para el tiempo y dos o tres para representar la imagen. Siguiendo las ideas anteriores, en Stoehr et al. (2020) proponen una metodología para la detección de cambio en la estructura de covarianza, el procedimiento es no paramétrico y está basado en reducción de dimensionalidad, hay comentarios acerca de posibles cambios en el espacio ortogonal formado a realizar la reducción de dimensionalidad y tiene aplicación a imágenes de resonancia magnética.

Skubalska-Rafajłowicz (2013) proponen una metodología para la detección de punto de cambio en procesos de tiempo discreto de alta dimensionalidad, ellos usan proyecciones aleatorias para la reducción de la dimensionalidad y presentan cartas de control para la media. Sharipov et al. (2016) proponen el secuencial bootstrap por bloques para espacios de Hilbert y lo usan para la detección de punto de cambio.

Bardsley et al. (2017) introducen varios métodos para la detección de punto de cambio en la media de series de tiempo de curvas, los métodos son capaces de detectar cambios de forma en las curvas como la concavidad.

4.1.3. Punto de cambio y RKHS

En la revisión de literatura no se encontró ningún estudio en el que se mostrara la construcción de un RKHS como se realizará en la siguiente sección. Los trabajos encontrados corresponden a espacios RKHS para el estudio de distancias entre distribuciones, en esos espacios el kernel está definido previamente y no es necesario estimar su acción. Entre los kernel más usados en ese contexto se encuentran los kernel gaussianos y de Laplace.

Algunos estudios sobre RKHS para distribuciones son:

- En Huang et al. (2014) se realiza un estudio sobre RKHS para medir distancias entre distribuciones que revelen un punto de cambio, además proponen las cartas de control en ese contexto.
- En Garreau (2017) se realiza el estudio del problema de múltiples puntos de cambio a través de kernel, muestran la consistencia del método. También estudian diferentes tipos de penalizaciones para el problema de múltiples puntos de cambio y el orden de convergencia de los estimadores.
- Arlot et al. (2019) proponen penalizaciones para seleccionar el número de puntos de cambio. Los procedimientos tratados allí son basados en criterios de minimización de expresiones con kernel dados.

4.2. Propuesta

La noción de distancia es fundamental en la elaboración de diferentes pruebas y en la toma de decisiones, sin embargo, realizar mediciones de distancia sin contemplar las estructuras de covarianza puede llevar a decisiones erróneas. Esta propuesta toma las características que tienen las matrices y los operadores de covarianza para que a través del teorema de 2.3.1 se construya un espacio con una noción de distancia en el que las estructuras de covarianza están contempladas desde la construcción. Los procesos multivariados estudiados son objetos del espacio definido y la definición de distancia se puede llevar al estudio dado por Berkes et al. (2009) para implementar una metodología para la detección de puntos de cambio.

En la propuesta para la detección de punto de cambio en datos funcionales multivariados intervienen componentes principales del análisis multivariado tradicional con restricciones

dadas por la continuidad que se necesita para el espacio propio. En el procedimiento anterior no se realiza reducción de dimensionalidad, éste se usa para transformar y a los procesos resultantes se les aplica componentes principales funcionales univariadas para luego adaptar la propuesta de Berkes et al. (2009) para detectar cambio.

Antes de presentar la propuesta se muestran las ideas geométricas que motivaron la metodología y que sirven para entender el porqué de algunos procedimientos. Posterior a las ideas geométricas se realiza la construcción del espacio, en esta parte se presentarán las transformaciones inducidas por las estructuras de covarianza y la noción de distancia. Las ideas geométricas se expondrán con un proceso bivariado, sin embargo, para el caso p -variado las ideas son análogas.

4.2.1. Ideas Geométricas

Considerese \mathbf{Y} un proceso estocástico bivariado definido sobre el intervalo $[0, 1]$, con

$$\mathbf{Y} = (Y_1, Y_2),$$

donde Y_1, Y_2 son dos procesos de segundo orden y de media 0. Además, para cada $t \in [0, 1]$, \mathbf{Y}_t es un vector bivariado con matriz de covarianza Σ_t .

En el modelamiento de situaciones reales no es posible afirmar que la matriz de covarianza Σ_t sea constante en el tiempo o que sea una diagonal. Un supuesto más acorde a la realidad es que para valores de t cercanos a t' las matrices Σ_t y $\Sigma_{t'}$ también sean cercanas. Ahora bien, para formalizar el término cercano es necesario definir correctamente la noción de distancia entre matrices; por esto se aclara que aquí se trabaja con la norma de Frobenius y de allí se establece de manera natural la definición de distancia.

En este trabajo, siguiendo las ideas de Berrendero et al. (2011), se supone que; la función Σ , la cual en cada instante de tiempo t devuelve la matriz Σ_t , tiene componentes diferenciables, y por lo tanto continuas. La continuidad garantiza que para valores de t cercanos a t' las matrices Σ_t y $\Sigma_{t'}$ también sean cercanas con la distancia inducida por la norma de Frobenius.

En la figura 4-1 se evidencia un cambio en la estructura de covarianza de los vectores \mathbf{Y}_t para los valores $t = 0, 0.2, 0.4, 0.6, 0.8, 1$ y en la figura 4-2 se muestran cada una de las componentes del proceso \mathbf{Y} en el tiempo.

Dado que las componentes de Σ son diferenciables; si se toman t y t' cercanas, entonces la diferencia de las matrices de covarianza se puede ver como un sistema dinámico de acuerdo con la definición 1.1 presentada en Grines et al. (2016), con parámetro para el tiempo $\Delta t = t - t'$. Un estudio basado formalmente en la naturaleza de los sistemas

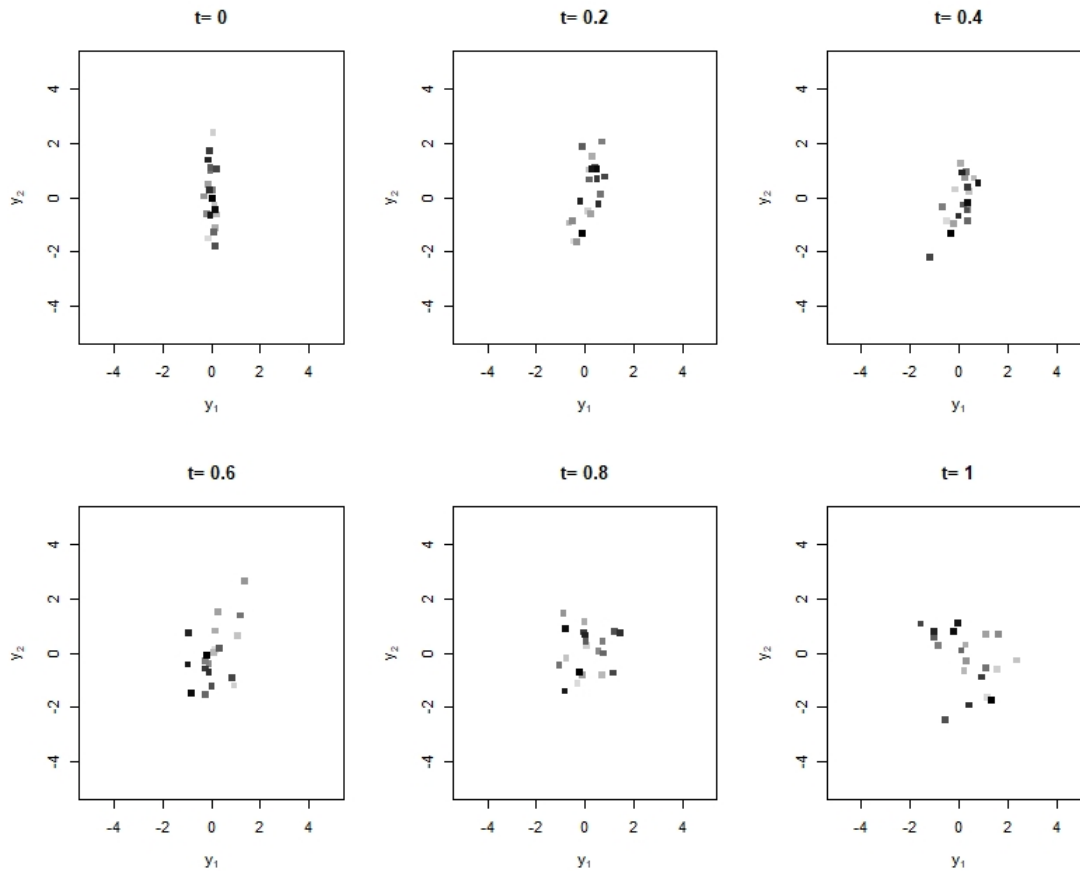


Figura 4-1: Diagramas de dispersión de \mathbf{Y}_t para los valores $t = 0, 0.2, 0.4, 0.6, 0.8, 1$ provenientes de una simulación de 20 procesos bivariados \mathbf{Y} .

dinámicos escapa del alcance de este trabajo, pero es una motivación para el análisis del cambio en los espacios de Hilbert en los que se tiene en cuenta la estructura de covarianza en cada instante de tiempo.

La idea fundamental de la propuesta es construir un espacio en el que las mediciones de distancia entre las realizaciones de los procesos tengan en cuenta los cambios de la estructura de covarianza en el tiempo ilustrados en la figura 4-1.

4.2.2. Espacio de trabajo

Se supone la existencia de dos espacios métricos, uno en el que se están obteniendo las realizaciones de los procesos y en el que hay dependencia entre las componentes de los procesos, y otro en el que por medio de una transformación se ha quitado esa dependencia.

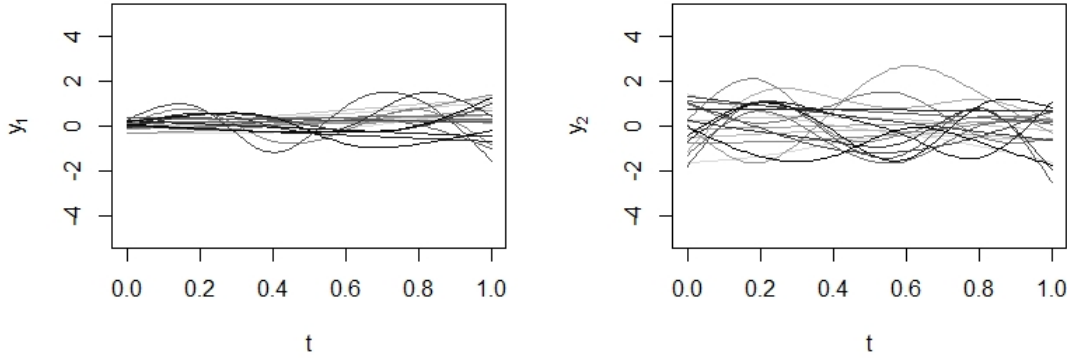


Figura 4-2: Curvas ajustadas para cada uno de los valores y_1, y_2 izquierda y derecha respectivamente de la simulación de los diagramas de dispersión en la figura 4-1.

Aprovechando que, como se mostrará más adelante, la transformación es invertible, la idea fundamental es medir las distancias de los objetos en el espacio en que hay dependencia usando las imágenes inversas que están en el espacio en el que no hay dependencia.

Definición 4.2.1 (Proceso \mathbf{Y}) Sea \mathbf{Y} un proceso funcional multivariado definido sobre el intervalo $[0, 1]$ con

$$\mathbf{Y} = (Y_1, Y_2, \dots, Y_p),$$

donde Y_1, Y_2, \dots, Y_p son procesos estocásticos con realizaciones en $\mathbb{L}^2[0, 1]$, además son procesos de segundo orden, de media 0 y función de covarianza continua.

Para cada $t \in [0, 1]$, \mathbf{Y}_t es el vector p -variado realización del proceso \mathbf{Y} en el instante t , que tiene matriz de covarianza Σ_t . La matriz Σ_t es invertible y sus valores propios cumplen la condición de formar un arreglo decreciente como se indica en 3-2. Las componentes de Σ son funciones diferenciables de t a lo largo del intervalo $[0, 1]$. Se asume que existe una transformación que al aplicarla a los procesos Y_1, Y_2, \dots, Y_p genera procesos independientes.

El teorema 3.2.1 garantiza la unicidad de los vectores propios de cada una de la matrices Σ_t salvo por un signo y la proposición 4 garantiza los siguientes resultados:

- Sean $\lambda_1(t) > \lambda_2(t) > \dots > \lambda_p(t) > 0$ las funciones tales que para cada $t \in [0, 1]$, devuelven los valores propios de Σ_t . Cada una de las funciones $\lambda_k(t)$ con $k = 1, 2, \dots, p$ son diferenciables.

- Sean $V_k(t)$ funciones con imagen en \mathbb{R}^p , tal que para cada t , $V_k(t)$ es el vector propio correspondiente a $\lambda_k(t)$. $V_k(t)$ se puede escoger de tal manera que sea diferenciable respecto a t .

Si para cada $k = 1, 2, \dots, p$ se escogen $V_k(t)$ diferenciables, lo cual es posible como se puede ver en la proposición 4; y además se imponen condiciones que fijen $V_k(0)$, siendo una de esas condiciones que la matriz teniendo por columnas a $V_k(0)$ tenga determinante 1; se tienen determinadas de manera única las funciones $V_k(t)$.

Teniendo en cuenta lo anterior, y notando por $\Sigma(t)$ al mapeo diferenciable que a cada $t \in [0, 1]$ le asigna Σ_t se tiene la siguiente expresión:

$$\Sigma(t) = \mathbf{V}(t) \mathbf{D}(t) \mathbf{V}^T(t), \quad (4-10)$$

donde $\mathbf{V}(t)$ es una función que para cada t devuelve una matriz en la que la columna k es $V_k(t)$, $\mathbf{D}(t)$ es una función en la que para cada t devuelve una matriz diagonal con los valores $\lambda_1(t), \lambda_2(t), \dots > \lambda_p(t)$.

$$\mathbf{V}(t) = [V_1(t), V_2(t), \dots, V_p(t)] \quad \mathbf{D}(t) = \text{diag}(\lambda_1(t), \lambda_2(t), \dots, \lambda_p(t)) \quad (4-11)$$

Definición 4.2.2 (Proceso $\tilde{\mathbf{Y}}_t$) Considerese para cada t la transformación lineal instantánea del proceso \mathbf{Y} definida como $\tilde{\mathbf{Y}}_t$ y dada por:

$$\tilde{\mathbf{Y}}_t := \mathbf{A}_t \mathbf{Y}_t, \quad (4-12)$$

donde \mathbf{A}_t corresponde a la evaluación en t de $\mathbf{A}(t)$ con

$$\mathbf{A}(t) = \mathbf{D}^{-1/2}(t) \mathbf{V}^T(t), \quad (4-13)$$

siendo $\mathbf{D}^{-1/2}(t)$ una función que devuelve una matriz diagonal para cada valor de t como se muestra a continuación

$$\mathbf{D}^{-1/2}(t) = \text{diag} \left(1/\sqrt{\lambda_1(t)}, 1/\sqrt{\lambda_2(t)}, \dots, 1/\sqrt{\lambda_p(t)} \right).$$

Sea $\tilde{\mathbf{Y}} = (\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_p)$, los procesos $\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_p$ son independientes.

Para $k = 1, 2, \dots, p$ se tiene:

$$\tilde{Y}_k = (A_{k,\cdot})_t \mathbf{Y}_t, \quad (4-14)$$

donde $(A_{k,\cdot})_t$ corresponde a la fila k de la matriz \mathbf{A}_t . Más adelante se evidenciará que la función $\mathbf{A}(t)$ está asociada a la existencia de la transformación que se asumió en la definición 4.2.1 y por la cual se asume que los procesos $\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_p$ son independientes.

Las entradas de las matrices $\mathbf{A}(t)$ son diferenciables y por lo tanto continuas, esto último sumado al hecho de que están definidas en el intervalo cerrado $[0, 1]$ implica que son funciones integrables, cuadrado integrables y además son acotadas. De lo anterior se tienen dos implicaciones relevantes:

- $\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_p$ son procesos integrables y cuadrado integrables.
- Para cada k las funciones de media y de covarianza del proceso \tilde{Y}_k son funciones continuas, aún más son de media 0.

Por lo anterior y teniendo en cuenta la definición 2.5.3 y el teorema 2.5.4 se tiene la siguiente definición

Definición 4.2.3 (Operador de Covarianza de \tilde{Y}_k) *Se nota el operador de covarianza de \tilde{Y}_k como $\tilde{\mathcal{K}}_k$*

$$\tilde{\mathcal{K}}_k = \sum_{m=1}^{\infty} \tilde{\lambda}_{m,k} \tilde{e}_{m,k} \otimes \tilde{e}_{m,k}, \quad (4-15)$$

donde $\tilde{\lambda}_{m,k}, \tilde{e}_{m,k}$ corresponden respectivamente a los valores y funciones propias del operador $\tilde{\mathcal{K}}_k$.

Por la continuidad de las funciones de covarianza de los procesos $\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_p$ el teorema de Mercer 2.5.6 garantiza que las funciones de covarianza se pueden escribir de la siguiente manera:

$$\tilde{K}_k(s, t) = \sum_{m=1}^{\infty} \tilde{\lambda}_{m,k} \tilde{e}_{m,k}(s) \tilde{e}_{m,k}(t) \quad (4-16)$$

las eigenfunciones $\tilde{e}_{m,k}$ son funciones en $\mathbb{L}^2[0, 1]$.

Es posible que existan $k \neq k'$ y subíndices m y m' tales que $\tilde{e}_{m,k} = \tilde{e}_{m',k'}$, pero se debe tener en cuenta que estas funciones corresponden a diferentes procesos independientes por lo que son elementos vectoriales distintos.

Dado que cada $\tilde{K}_k(s, t)$ es definida positiva y simétrica, el Teorema de Moore–Aronszajn 2.3.1 garantiza la existencia y unicidad de los espacios RKHS $\mathbf{H}(\tilde{K}_k)$.

Definición 4.2.4 (Espacio \mathbf{H}) *Teniendo en cuenta la definición 2.1.6, se considera el siguiente espacio de Hilbert:*

$$\mathbf{H} := \bigoplus_{k=1}^p \mathbf{H}(\tilde{K}_k), \quad (4-17)$$

\mathbf{H} es un espacio de Hilbert separable, ya que es suma directa de finitos espacios separables. Por lo tanto, como se mencionó en la definición 2.1.6 toda función $\tilde{\mathbf{f}} \in \mathbf{H}$ se puede escribir de manera única como

$$\tilde{\mathbf{f}} = \tilde{f}_1 \oplus \tilde{f}_2 \oplus \cdots \oplus \tilde{f}_p,$$

para cada $k = 1, 2, \dots, p$, $\tilde{f}_k \in \mathbf{H}(\tilde{K}_k)$, sin embargo, para evitar ambigüedad parece ser más adecuado escribir cada $\tilde{\mathbf{f}}$ de forma vectorial de la siguiente manera

$$\tilde{\mathbf{f}} = (\tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_p). \quad (4-18)$$

Dado que para cada k , \tilde{Y}_k es un proceso de segundo orden con funciones de media y covarianza continuas se presenta la extensión de Karhunen-Loève de acuerdo al teorema 2.5.7, de la siguiente manera:

$$\tilde{Y}_k(t) = \sum_{m=1}^{\infty} Z_{m,k} \tilde{e}_{m,k}(t), \quad (4-19)$$

teniendo en cuenta que

$$Z_{m,k} = \left\langle \tilde{Y}_k(t), \tilde{e}_{m,k}(t) \right\rangle_{\mathbf{H}}$$

donde $\tilde{e}_{m,k}(t)$ es una función propia y $Z_{m,k}$ es una variable aleatoria de media 0 y además para cada k se tiene

$$\text{cov}(Z_{m,k}, Z_{m',k}) = E(Z_{m,k} Z_{m',k}) = \begin{cases} \tilde{\lambda}_{m,k} & \text{si } m = m', \\ 0 & \text{en otro caso.} \end{cases} \quad (4-20)$$

Teniendo en cuenta la forma vectorial de las funciones en \mathbf{H} como en 4-18 se define

$$\tilde{\mathbf{e}}_{m,k}(t) = (0, 0, \dots, 0, \tilde{e}_{m,k}(t), 0, \dots, 0), \quad (4-21)$$

donde $\tilde{\mathbf{e}}_{m,k}(t)$ tiene 0 en todas las posiciones salvo en la posición k en la que está la función propia $\tilde{e}_{m,k}(t)$. De las expresiones 4-19 y dado que los procesos $\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_p$ son independientes la expansión de Karhunen-Loève para el proceso $\tilde{\mathbf{Y}}$ es

$$\tilde{\mathbf{Y}}(t) = \sum_{k=1}^p \sum_{m=1}^{\infty} Z_{m,k} \tilde{\mathbf{e}}_{m,k}(t) \quad (4-22)$$

en esta representación, las variables $Z_{m,k}$ además de lo expuesto en 4-20, cumplen:

$$\text{cov}(Z_{m,k}, Z_{m',k'}) = E(Z_{m,k} Z_{m',k'}) = \begin{cases} \tilde{\lambda}_{m,k} & \text{si } m = m' \text{ y } k = k', \\ 0 & \text{en otro caso.} \end{cases} \quad (4-23)$$

Para cada $t \in [0, 1]$ la matriz \mathbf{V}_t pertenece al grupo lineal ortogonal especial $SO(p)$; esto significa que para cada t , la matriz es ortogonal de determinante 1. $SO(p)$ es un grupo de Lie, o sea es un grupo con la operación multiplicación de matrices y además es una variedad diferenciable. La transformación que genera en el espacio Euclideo una matriz del $SO(p)$ es una isometría que preserva la orientación.

Definición 4.2.5 (Función $\mathbf{B}(t)$) *Sea $\mathbf{B}(t)$ una función que para cada t devuelve la matriz \mathbf{B}_t que cumple*

$$\mathbf{B}_t := \mathbf{A}_t^{-1}. \quad (4-24)$$

Teniendo en cuenta la expresión en 4-13 se tiene

$$\mathbf{B}(t) = \mathbf{V}(t) \mathbf{D}^{1/2}(t). \quad (4-25)$$

La función \mathbf{B} que para cada t devuelve matrices, modela los cambios en la estructura de covarianza, de la expresión 4-10 obsérvese que

$$\boldsymbol{\Sigma}(t) = \mathbf{B}(t)\mathbf{B}(t)^T,$$

además en cada instante de tiempo t por las expresiones 4-12 y 4-24 entonces

$$\mathbf{Y}_t = \mathbf{B}_t \tilde{\mathbf{Y}}_t. \quad (4-26)$$

El supuesto de cambios suaves es representado por la diferenciabilidad de las funciones $\lambda_k(t)$ y por el hecho de que las matrices en $SO(p)$ forman una variedad diferenciable. En Hall (2013) muestran diferentes resultados sobre $SO(p)$.

La intención ahora es definir el operador \mathcal{F} que modifica la estructura de covarianza en cada instante de tiempo. El operador \mathcal{F} genera los procesos que tienen dependencia, Y_1, Y_2, \dots, Y_p , como una transformación de los procesos $\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_p$, que no tienen dependencia.

Definición 4.2.6 (Operador \mathcal{F}) Se define el operador \mathcal{F} que transforma cada función $\tilde{\mathbf{f}}$ representada como en 4-18 de la siguiente manera

$$\mathbf{f}(t) = \left(\mathcal{F} \tilde{\mathbf{f}} \right) (t) = \mathbf{B}(t) \tilde{\mathbf{f}}(t). \quad (4-27)$$

El operador \mathcal{F} tiene dominio en \mathbf{H} , la función $\mathbf{f}(t)$ para cada t devuelve un vector p -variado que es la multiplicación de la matriz \mathbf{B}_t por la evaluación en t de la función $\tilde{\mathbf{f}}$, cada componente se puede ver como una función de t . Nombrando cada una de las funciones componentes se tiene:

$$\mathbf{f} = (f_1, f_2, \dots, f_p). \quad (4-28)$$

Teorema 4.2.1 (Características \mathcal{F}) El operador \mathcal{F} tiene las siguientes características:

1. Es lineal.
2. Es invertible.
3. Es acotado con las normas inducidas de \mathbb{L}^2 .
4. Tiene imagen separable.
5. No es compacto.

D/.

1. La transformación \mathcal{F} es lineal ya que para cualquier par de funciones $\tilde{f}(t), \tilde{h}(t) \in \mathbf{H}$ y $\alpha \in \mathbb{R}$ se tiene:

$$\mathbf{B}(t) \left(\alpha \tilde{\mathbf{f}}(t) + \tilde{\mathbf{h}}(t) \right) = \alpha \mathbf{B}(t) \left(\tilde{\mathbf{f}}(t) \right) + \mathbf{B}(t) \left(\tilde{\mathbf{h}}(t) \right) \quad (4-29)$$

2. Los operadores que actúan multiplicando funciones de manera similar a lo aquí planteado son operadores de tipo multiplicación. En este caso el operador es invertible con:

$$\tilde{\mathbf{f}}(t) = \left(\mathcal{F}^{-1} \mathbf{f} \right) (t) = \mathbf{A}(t) \mathbf{f}(t). \quad (4-30)$$

3. Considerese para el espacio vectorial \mathbf{H} el producto interno:

$$\langle \tilde{\mathbf{f}}, \tilde{\mathbf{g}} \rangle_{\mathbf{H}} = \sum_{j=1}^p \langle \tilde{f}_j, \tilde{g}_j \rangle_{\mathbb{L}^2} \quad (4-31)$$

donde $\langle \cdot, \cdot \rangle_{\mathbb{L}^2}$ es el producto interno usual de \mathbb{L}^2 y $\tilde{\mathbf{g}} \in \mathbf{H}$, con

$$\tilde{\mathbf{g}} = (\tilde{g}_1, \tilde{g}_2, \dots, \tilde{g}_p). \quad (4-32)$$

el producto interno definido en 4-31 induce la norma:

$$\|\tilde{\mathbf{f}}\|_{\mathbf{H}} = \sqrt{\sum_{k=1}^p \|\tilde{f}_k\|_{\mathbb{L}^2}^2}, \quad (4-33)$$

para cada $\tilde{\mathbf{f}} \in \mathbf{H}$ donde $\|\cdot\|_{\mathbb{L}^2}$ es la norma usual de $\mathbb{L}^2[0, 1]$.

Teniendo en cuenta la expresión 4-28 para \mathbf{f} y análogamente, sea

$$\mathbf{g} = (g_1, g_2, \dots, g_p),$$

con \mathbf{f}, \mathbf{g} funciones en el espacio imagen de \mathcal{F} notado por $Im(\mathcal{F})$. Se considera el siguiente producto interno:

$$\langle \mathbf{f}, \mathbf{g} \rangle_{\mathcal{F}} = \sum_{j=1}^p \langle f_j, g_j \rangle_{\mathbb{L}^2} \quad (4-34)$$

El producto interno anterior induce la norma:

$$\|\mathbf{f}\|_{\mathcal{F}} = \sqrt{\sum_{j=1}^p \|f_j\|_{\mathbb{L}^2}^2}. \quad (4-35)$$

Sean $b_{j,k}(t)$ con $j, k = 1, 2, \dots, p$ las funciones componentes de $\mathbf{B}(t)$, por la expresión 4-27 se tiene que:

$$f_j(t) = \sum_{k=1}^p b_{j,k}(t) \tilde{f}_k(t), \quad (4-36)$$

las funciones $b_{j,k}(t)$ son funciones diferenciables por la forma en la que se definió \mathbf{B} en la expresión 4-25. Lo anterior implica que $b_{j,k}(t)$ son funciones continuas y por lo tanto son acotadas, sea $M > 0$ tal que para cada j, k se tiene

$$b_{j,k}(t) \leq M.$$

El operador \mathcal{F} es un operador acotado con las normas definidas en 4-33 y 4-35 para su dominio e imagen respectivamente ya que

$$\|\mathcal{F}\tilde{\mathbf{f}}\|_{\mathcal{F}} = \|\mathbf{f}\|_{\mathcal{F}} = \sqrt{\sum_{j=1}^p \left\| \sum_{k=1}^p b_{j,k}(t) \tilde{f}_k(t) \right\|_{\mathbb{L}^2}^2} \leq \sqrt{\sum_{j=1}^p \left\| \sum_{k=1}^p M \tilde{f}_k(t) \right\|_{\mathbb{L}^2}^2} \leq \sqrt{p}M \|\tilde{\mathbf{f}}\|_{\mathbf{H}}$$

4. $Im(\mathcal{F})$ es un espacio separable, porque es imagen de un espacio con la misma característica.
5. Por el teorema 2.4.2 el operador \mathcal{F} no es compacto por que \mathbf{H} es de dimensión infinita y es invertible.

■

El operador \mathcal{F} tiene características como ser invertible, separable y acotado que son necesarias para realizar cálculos numéricos. Sin embargo, la no compacidad presenta inconvenientes de convergencia para los métodos. De acuerdo con el teorema 2.4.2 la forma más sencilla de resolver el problema es reducir el análisis a un subespacio finito adecuado.

Definición 4.2.7 (La función G_D y el espacio \mathbf{H}_D) Para $D = (d_1, d_2, \dots, d_p)$ con d_1, d_2, \dots, d_p valores enteros positivos, sea $G_D(t, s)$ una función bivariada con dominio en el producto cartesiano $[0, 1] \times [0, 1]$ y definida por:

$$G_D(t, s) = \sum_{k=1}^p \sum_{m=1}^{d_k} \tilde{\mathbf{e}}_{m,k}(t) (\mathcal{F}\tilde{\mathbf{e}}_{m,k})(s). \quad (4-37)$$

y sea

$$\mathbf{H}_D := span\{\tilde{\mathbf{e}}_{1,1}, \tilde{\mathbf{e}}_{2,1}, \dots, \tilde{\mathbf{e}}_{d_1,1}, \tilde{\mathbf{e}}_{1,2}, \tilde{\mathbf{e}}_{2,2}, \dots, \tilde{\mathbf{e}}_{d_2,2}, \dots, \tilde{\mathbf{e}}_{1,p}, \tilde{\mathbf{e}}_{2,p}, \dots, \tilde{\mathbf{e}}_{d_p,p}\}. \quad (4-38)$$

La función G_D está bien definida ya que el operador \mathcal{F} es acotado. Esta función juega un papel similar al que tienen las matrices para representar transformaciones lineales, y al igual que estas últimas no es necesariamente definida positiva.

\mathbf{H}_D es un subespacio de \mathbf{H} . Si $\tilde{\mathbf{f}} \in \mathbf{H}_D$, a continuación se presenta la expansión de Fourier como en 2.1.1

$$\tilde{\mathbf{f}}(t) = \sum_{k=1}^p \sum_{m=1}^{d_k} \phi_{m,k} \tilde{\mathbf{e}}_{m,k}(t) \quad (4-39)$$

teniendo en cuenta la expresión en 2-2 se realiza el siguiente producto interno

$$\langle \tilde{\mathbf{f}}, G_D(\cdot, s) \rangle_{\mathbf{H}} = \left\langle \sum_{k=1}^p \sum_{m=1}^{d_k} \phi_{m,k} \tilde{\mathbf{e}}_{m,k}(\cdot), \sum_{k=1}^p \sum_{m=1}^{d_k} \tilde{\mathbf{e}}_{m,k}(\cdot) (\mathcal{F} \tilde{\mathbf{e}}_{m,k})(s) \right\rangle_{\mathbf{H}}$$

dado que cada $\tilde{\mathbf{e}}_{m,k}$ son elementos de una base ortonormal se tiene

$$\langle \tilde{\mathbf{f}}, G_D(\cdot, s) \rangle_{\mathbf{H}} = \sum_{k=1}^p \sum_{m=1}^{d_k} \phi_{m,k} (\mathcal{F} \tilde{\mathbf{e}}_{m,k})(s) = \mathcal{F} \left(\sum_{k=1}^p \sum_{m=1}^{d_k} \phi_{m,k} \tilde{\mathbf{e}}_{m,k} \right) (s)$$

por lo tanto la función $G_D(t, s)$ es la representación para el operador \mathcal{F} en el espacio \mathbf{H}_D , esto es

$$\left(\mathcal{F} \tilde{\mathbf{f}} \right) (s) = \langle \tilde{\mathbf{f}}, G_D(\cdot, s) \rangle_{\mathbf{H}}. \quad (4-40)$$

Para los análisis se tendrá en cuenta las realizaciones del proceso $\tilde{Y}_k(t)$ restringidas a \mathbf{H}_D . Esto implica que se les realice un truncamiento a las respectivas extensiones de Karhunen-Loève mostradas en 4-19, los truncamientos son los siguientes:

$$\tilde{Y}_k^{TR}(t) = \sum_{m=1}^{d_k} Z_{m,k} \tilde{\mathbf{e}}_{m,k}(t), \quad (4-41)$$

a el proceso multivariado $\tilde{\mathbf{Y}}$ también se le realiza un truncamiento teniendo en cuenta las expresiones en 4-21 y 4-22

$$\tilde{\mathbf{Y}}^{TR}(t) = \sum_{k=1}^p \sum_{m=1}^{d_k} Z_{m,k} \tilde{\mathbf{e}}_{m,k}(t). \quad (4-42)$$

Las realizaciones del proceso $\tilde{\mathbf{Y}}^{TR}$ pertenecen al espacio \mathbf{H}_D por la expresión 4-38. A continuación se halla la transformada del proceso $\tilde{\mathbf{Y}}^{TR}$ usando el producto interno como en 4-40

$$\left(\mathcal{F} \tilde{\mathbf{Y}}^{TR} \right) (s) = \langle \tilde{\mathbf{Y}}^{TR}, G_D(\cdot, s) \rangle$$

reemplazando la extensión de Karhunen-Loève en 4-42 y la función bivariada G_D en 4-37 se tiene

$$\left(\mathcal{F} \tilde{\mathbf{Y}}^{TR} \right) (s) = \left\langle \sum_{k=1}^p \sum_{m=1}^{d_k} Z_{m,k} \tilde{\mathbf{e}}_{m,k}(\cdot), \sum_{k=1}^p \sum_{m=1}^{d_k} \tilde{\mathbf{e}}_{m,k}(\cdot) (\mathcal{F} \tilde{\mathbf{e}}_{m,k}) (s) \right\rangle_{\mathbf{H}}.$$

Definición 4.2.8 (El proceso \mathbf{Y}^{TR} y el espacio $\mathcal{F}\mathbf{H}_D$) Se definen $\mathcal{F}\mathbf{H}_D$ y \mathbf{Y}^{TR} , respectivamente como las imágenes, del espacio \mathbf{H}_D y del proceso $\tilde{\mathbf{Y}}^{TR}$ que tiene realizaciones en \mathbf{H}_D , del operador \mathcal{F} .

$$\mathbf{Y}^{TR} = \left(\mathcal{F} \tilde{\mathbf{Y}}^{TR} \right) = \sum_{k=1}^p \sum_{m=1}^{d_k} Z_{m,k} (\mathcal{F} \tilde{\mathbf{e}}_{m,k}) (s).$$

\mathbf{Y}^{TR} es un proceso que coincide con el proceso \mathbf{Y} en el espacio $\mathcal{F}\mathbf{H}_D$.

Teniendo en cuenta la expresión 2-3, se considera el operador autoadjunto del espacio $\mathcal{F}\mathbf{H}_D$ definido con el producto tensorial y el proceso \mathbf{Y}^{TR} .

$$E \left([\mathbf{Y}^{TR} \otimes \mathbf{Y}^{TR}] \mathbf{f} \right) = E \left(\langle \mathbf{Y}^{TR}, \mathbf{f} \rangle_{\mathcal{F}} \mathbf{Y}^{TR} \right), \quad (4-43)$$

donde $\mathbf{f} \in \mathcal{F}\mathbf{H}_D$, por lo tanto existe $\tilde{\mathbf{f}} \in \mathbf{H}_D$ tal que

$$\mathbf{f}(s) = \left(\mathcal{F} \tilde{\mathbf{f}} \right) (s) = \sum_{k=1}^p \sum_{m=1}^{d_k} \phi_{m,k} (\mathcal{F} \tilde{\mathbf{e}}_{m,k}) (s). \quad (4-44)$$

Aplicando el operador en 4-43 para $(\mathcal{F} \tilde{\mathbf{e}}_{m',k'})$

$$E \left(\langle \mathbf{Y}^{TR}, (\mathcal{F} \tilde{\mathbf{e}}_{m',k'}) \rangle_{\mathcal{F}} \mathbf{Y}^{TR} \right) = E \left(\left\langle \sum_{k=1}^p \sum_{m=1}^{d_k} Z_{m,k} (\mathcal{F} \tilde{\mathbf{e}}_{m,k}), (\mathcal{F} \tilde{\mathbf{e}}_{m',k'}) \right\rangle_{\mathcal{F}} \mathbf{Y}^{TR} \right)$$

si

$$\langle (\mathcal{F} \tilde{\mathbf{e}}_{m,k}), (\mathcal{F} \tilde{\mathbf{e}}_{m',k'}) \rangle_{\mathcal{F}} = u_{m,k,m',k'}, \quad (4-45)$$

luego

$$E \left(\langle \mathbf{Y}^{TR}, (\mathcal{F} \tilde{\mathbf{e}}_{m,k}) \rangle_{\mathcal{F}} \mathbf{Y}^{TR} \right) = E \left(\sum_{k=1}^p \sum_{m=1}^{d_k} u_{m,k,m',k'} Z_{m,k} \mathbf{Y}^{TR} \right),$$

reemplazando \mathbf{Y}^{TR}

$$E \left(\langle \mathbf{Y}^{TR}, (\mathcal{F} \tilde{\mathbf{e}}_{m,k}) \rangle_{\mathcal{F}} \mathbf{Y}^{TR} \right) = E \left(\left(\sum_{k=1}^p \sum_{m=1}^{d_k} u_{m,k,m',k'} Z_{m,k} \right) \sum_{k''=1}^p \sum_{m''=1}^{d_{k''}} Z_{m'',k''} (\mathcal{F} \tilde{\mathbf{e}}_{m'',k''}) \right),$$

y teniendo en cuenta 4-23, se tiene que los únicos términos que no se anulan son aquellos en los que $m = m''$ y $k = k''$ por lo tanto

$$E(\langle \mathbf{Y}^{TR}, (\mathcal{F} \tilde{\mathbf{e}}_{m,k}) \rangle_{\mathcal{F}} \mathbf{Y}^{TR}) = \sum_{k=1}^p \sum_{m=1}^{d_k} u_{m,k,m',k'} \tilde{\lambda}_{m,k}(\mathcal{F} \tilde{\mathbf{e}}_{m,k}). \quad (4-46)$$

El operador en 4-43 es una aproximación al operador de covarianza del proceso \mathbf{Y}

Por otro lado, se define la función bivariada $K(s, t)$.

Definición 4.2.9 (Función K)

$$K(s, t) := \sum_{k=1}^p \sum_{m=1}^{d_k} \tilde{\lambda}_{m,k}(\mathcal{F} \tilde{\mathbf{e}}_{m,k})(s) (\mathcal{F} \tilde{\mathbf{e}}_{m,k})(t), \quad (4-47)$$

donde $s, t \in [0, 1]$.

En el siguiente teorema se conecta la función K con el operador $E([\mathbf{Y}^{TR} \otimes \mathbf{Y}^{TR}] \mathbf{f})$.

Teorema 4.2.2 *Las acciones de la función K y del operador $E([\mathbf{Y}^{TR} \otimes \mathbf{Y}^{TR}] \mathbf{f})$ coinciden.*

D/. Observese que las acciones sobre los elementos de la base coinciden,

$$\langle (\mathcal{F} \tilde{\mathbf{e}}_{m',k'}), K(\cdot, s) \rangle_{\mathcal{F}} = \left\langle (\mathcal{F} \tilde{\mathbf{e}}_{m',k'}), \sum_{k=1}^p \sum_{m=1}^{d_k} \tilde{\lambda}_{m,k}(\mathcal{F} \tilde{\mathbf{e}}_{m,k})(s) (\mathcal{F} \tilde{\mathbf{e}}_{m,k}) \right\rangle_{\mathcal{F}}$$

usando la expresión 4-45 se obtiene

$$\langle (\mathcal{F} \tilde{\mathbf{e}}_{m',k'}), K(\cdot, s) \rangle_{\mathcal{F}} = \sum_{k=1}^p \sum_{m=1}^{d_k} u_{m,k,m',k'} \tilde{\lambda}_{m,k}(\mathcal{F} \tilde{\mathbf{e}}_{m,k})(s).$$

■

Dado que \mathcal{F} es un operador invertible y considerando la forma en la que es generado el espacio \mathbf{H}_D en 4-38, para $\mathbf{f} \in \mathcal{F} \mathbf{H}_D$ como en 4-44 y $\mathbf{g} \in \mathcal{F} \mathbf{H}_D$ dado por

$$\mathbf{g}(s) = (\mathcal{F} \tilde{\mathbf{g}})(s) = \sum_{k=1}^p \sum_{m=1}^{d_k} \psi_{m,k}(\mathcal{F} \tilde{\mathbf{e}}_{m,k})(s), \quad (4-48)$$

se define el siguiente producto interno.

Definición 4.2.10 (Producto Interno $\langle \cdot, \cdot \rangle_K$) Para $\mathbf{f} \in \mathcal{F}\mathbf{H}_D$ como en 4-44 y $\mathbf{g} \in \mathcal{F}\mathbf{H}_D$ como en 4-48

$$\langle \mathbf{f}, \mathbf{g} \rangle_K = \sum_{k=1}^p \sum_{m=1}^{d_k} \left\langle \frac{1}{\sqrt{\tilde{\lambda}_{m,k}}} \langle (\mathcal{F}^{-1}\mathbf{f}), \tilde{\mathbf{e}}_{m,k} \rangle_{\mathbf{H}} \tilde{\mathbf{e}}_{m,k}, \frac{1}{\sqrt{\tilde{\lambda}_{m,k}}} \langle (\mathcal{F}^{-1}\mathbf{g}), \tilde{\mathbf{e}}_{m,k} \rangle_{\mathbf{H}} \tilde{\mathbf{e}}_{m,k} \right\rangle_{\mathbf{H}} \quad (4-49)$$

Teniendo en cuenta las extensiones de las funciones \mathbf{f}, \mathbf{g} se puede simplificar la anterior expresión de la siguiente manera

$$\langle \mathbf{f}, \mathbf{g} \rangle_K = \sum_{k=1}^p \sum_{m=1}^{d_k} \frac{1}{\tilde{\lambda}_{m,k}} \phi_{m,k} \psi_{m,k}, \quad (4-50)$$

ya que

$$\langle (\mathcal{F}^{-1}\mathbf{f}), \tilde{\mathbf{e}}_{m,k} \rangle_{\mathbf{H}} = \phi_{m,k}$$

y

$$\langle (\mathcal{F}^{-1}\mathbf{g}), \tilde{\mathbf{e}}_{m,k} \rangle_{\mathbf{H}} = \psi_{m,k}$$

Teorema 4.2.3 (Existencia del Espacio de trabajo) $\{\mathcal{F}\mathbf{H}_D, \langle \cdot, \cdot \rangle_K\}$ es un RKHS.

D/. $K(s, \cdot) \in \mathcal{F}\mathbf{H}_D$ para cada valor fijo de s , y se tiene

$$\left\langle \left(\mathcal{F}^{-1} \sum_{k'=1}^p \sum_{m'=1}^{d_{k'}} \tilde{\lambda}_{m',k'} (\mathcal{F}\tilde{\mathbf{e}}_{m',k'}) (s) (\mathcal{F}\tilde{\mathbf{e}}_{m',k'}) (\cdot) \right), \tilde{\mathbf{e}}_{m,k} \right\rangle_{\mathbf{H}} = \tilde{\lambda}_{m,k} (\mathcal{F}\tilde{\mathbf{e}}_{m,k}) (s)$$

para cualquier función $\mathbf{f} \in \mathcal{F}\mathbf{H}_D$ se realiza el siguiente producto interno en el que se considera que s está fijo,

$$\langle \mathbf{f}, K(s, \cdot) \rangle_K = \sum_{k=1}^p \sum_{m=1}^{d_k} \frac{1}{\tilde{\lambda}_{m,k}} \phi_{m,k} \tilde{\lambda}_{m,k} (\mathcal{F}\tilde{\mathbf{e}}_{m,k}) (s) = (\mathcal{F}\tilde{\mathbf{f}}) (s) = \mathbf{f}(s). \quad \blacksquare$$

$\{\mathcal{F}\mathbf{H}_D, \langle \cdot, \cdot \rangle_K\}$ es el espacio de Hilbert sobre el que se trabajará, éste es un espacio de dimensión finita sobre el que se puede medir considerando las estructuras de covarianza. Teniendo en cuenta la forma en la que está definido \mathbf{H}_D si los valores d_1, d_2, \dots, d_p son grandes se presenta una mejor aproximación al espacio \mathbf{H} en el que están las realizaciones de los procesos, sin embargo, computacionalmente no se recomienda incrementar

demasiado los valores d_1, d_2, \dots, d_p debido a que se pueden generar divisiones entre valores cercanos a 0. Los valores d_1, d_2, \dots, d_p son seleccionados de tal manera que explican un alto porcentaje de variabilidad.

En el análisis multivariado la distancia de Mahalanobis contempla la estructura de varianza para medir distancias, para el estudio de datos funcionales hay propuestas para extender este concepto, Berrendero et al. (2020) tratan este tema. El enfoque dado en este trabajo fue seleccionar un subespacio de dimensión finita que permitiera desarrollar la propuesta de detección de punto de cambio.

Diferencias con otros usos de RKHS

El uso de kernels y RKHS ya ha sido usado en diversos análisis estadísticos e incluso en detección de punto de cambio, en este momento es preciso establecer las diferencias entre el espacio aquí mostrado y otras propuestas que usan RKHS.

En la definición 2.3.1, la función K tiene dominio en $\mathbf{D} \times \mathbf{D}$ con $\mathbf{D} \subset \mathbb{R}$, en el enfoque de Garreau (2017) y Arlot et al. (2019) el conjunto \mathbf{D} es cualquier espacio medible, en esta propuesta no es necesario esa generalización porque en el contexto del problema este conjunto debe ser el intervalo sobre el cual están definidos los procesos.

Los mapeos Φ en Garreau (2017) son de la forma $P \mapsto E_{X \sim P} [K(X, \cdot)]$, llamados de inmersión, en ese caso $X \sim P$ y K es una función kernel definida sobre la imagen del objeto aleatorio X . En la propuesta de este trabajo se construyó explícitamente el rk del RKHS en el que se va a trabajar.

En Hörmann et al. (2015) hay una propuesta para la detección de punto de cambio en series de tiempo de datos funcionales. Allí se muestra explícitamente el operador con el que trabajan, ellos dan indicaciones de como realizar una extensión de las ideas en Berkes et al. (2009) para el contexto de series de tiempo. Si bien es cierto que la terminología en Hörmann et al. (2015) es similar a la que hace parte de la propuesta expuesta aquí, hay grandes diferencias; allí el interés es el trabajo con las autocorrelaciones mientras que acá lo es el trabajo con la estructura de covarianza entre los procesos, esta diferencia implica diferencias en la clase de operadores. Los enfoques también son distintos, acá se toma como referencia el concepto de RKHS y se construye el espacio de Hilbert junto con su producto interno, luego se realiza el estudio de punto de cambio basado en la distancia inducida por el producto interno del RKHS.

4.2.3. Propuesta para la detección de punto de cambio en la media de datos funcionales multivariados

El tipo de procesos multivariados aquí considerados tienen una serie de supuestos para poder aplicar la metodología, pero estos son coherentes y representan adecuadamente situaciones reales.

El supuesto esencial de esta metodología es la existencia en cada tiempo t de unas direcciones principales que pueden cambiar a lo largo del tiempo pero de manera suave. Por ejemplo, la temperatura del agua a una profundidad fija y la temperatura ambiente son dos variables que están asociadas, pero es posible que el nivel de asociación cambie a lo largo del día; el supuesto está en que si bien se contemplan cambios en los niveles de asociación, esos cambios deben ser suaves, explícitamente deben ser modelados por funciones diferenciables como se expuso en la construcción del espacio de trabajo.

El producto interno en 4-49 traspone las mediciones de distancia a un espacio en el que no hay asociación.

Nota 4.2.1 (Supuestos de la propuesta) *Los siguientes supuestos hacen que los objetos encajen en el espacio definido y que sea aplicable la metodología de Berkes et al. (2009).*

1. Existe un operador \mathcal{F} invertible y un proceso funcional p -variado de media 0,

$$\tilde{\mathbf{Y}}_i = \left(\tilde{Y}_{i,1}, \tilde{Y}_{i,2}, \dots, \tilde{Y}_{i,p} \right),$$

tales que \mathcal{F} produce la transformación indicada en 4-27. Las realizaciones de los procesos $\tilde{Y}_{i,k}$ para $k = 1, 2, \dots, p$ cumplen los supuestos 1 y 3 en la nota 4.1.1, para $i = 1, 2, \dots, N$ con N el tamaño de la muestra.

2. Las observaciones son realizaciones de un proceso funcional \mathbf{X}_i con

$$\mathbf{X}_i = \boldsymbol{\mu}_\gamma + \mathbf{Y}_i$$

\mathbf{Y}_i es un proceso de media 0, $E(\mathbf{X}_i) = \boldsymbol{\mu}_\gamma$ con $\gamma = 1, 2$, y se tiene

$$\mathbf{Y}_i = \mathbf{X}_i - \boldsymbol{\mu}_\gamma$$

donde

$$\mathbf{Y}_i = \mathcal{F} \tilde{\mathbf{Y}}_i, \quad \mathbf{X}_i = \mathcal{F} \tilde{\mathbf{X}}_i, \quad \boldsymbol{\mu}_\gamma = \mathcal{F} \tilde{\boldsymbol{\mu}}_\gamma$$

Si existe un punto de cambio en c igual a parte entera de $N\theta$ con $\theta \in [0, 1]$ las observaciones provienen del proceso \mathbf{X}_i en el que desde 1 hasta el valor entero de c se tiene $\boldsymbol{\mu}_\gamma = \boldsymbol{\mu}_1$ y después $\boldsymbol{\mu}_\gamma = \boldsymbol{\mu}_2$. Si no hay punto de cambio $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$.

3. Para cada k las realizaciones $Y_{i,k}$ provienen del proceso Y_k que es una componente del proceso funcional p -variado \mathbf{Y}_i . Siendo este último la transformación de un proceso $\tilde{\mathbf{Y}}_i$ como se indicó en el anterior ítem. Para los procesos $\tilde{Y}_{i,k}$ el operador y la función de covarianza son respectivamente como en 4-15 y 4-16 y los eigenvalores cumplen la condición 2 en la nota 4.1.1.

Considérese la siguiente muestra de tamaño N

$$\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N$$

con

$$\mathbf{X}_i = (X_{i,1}, X_{i,2}, \dots, X_{i,p}),$$

para $i = 1, 2, \dots, N$.

Definición 4.2.11 Las siguientes definiciones son análogas a las expuestas en 4.1.1.

$$\hat{\mu}_{c,k}(t) = \frac{1}{c} \sum_{i=1}^c X_{i,k}(t), \quad \hat{\boldsymbol{\mu}}_c = (\hat{\mu}_{c,1}(t), \hat{\mu}_{c,2}(t), \dots, \hat{\mu}_{c,p}(t)) \quad (4-51)$$

$$\check{\mu}_{c,k}(t) = \frac{1}{N-c} \sum_{i=c+1}^N X_{i,k}(t), \quad \check{\boldsymbol{\mu}}_c = (\check{\mu}_{c,1}(t), \check{\mu}_{c,2}(t), \dots, \check{\mu}_{c,p}(t)) \quad (4-52)$$

Ahora se considera la diferencia entre los estimadores de las medias $\hat{\boldsymbol{\mu}}_c$ y $\check{\boldsymbol{\mu}}_c$ multiplicado por el factor de tipo parabólico. La siguiente expresión es igual a la dada en 4-5, la diferencia es que en este caso la expresión de la derecha corresponde a la diferencia de arreglos de medias.

$$P_c(t) = \frac{c(N-c)}{N} (\hat{\boldsymbol{\mu}}_c(t) - \check{\boldsymbol{\mu}}_c(t)) \quad (4-53)$$

el reemplazo de $\hat{\boldsymbol{\mu}}_c$ y $\check{\boldsymbol{\mu}}_c$ en la anterior expresión, conlleva a

$$P_c(t) = \sum_{1 \leq i \leq Nx} (\mathbf{X}_i(t) - \bar{\mathbf{X}}_N(t)) - \frac{c}{N} \sum_{i=1}^N (\mathbf{X}_i(t) - \bar{\mathbf{X}}_N(t)) \quad (4-54)$$

donde

$$\bar{\mathbf{X}}_N = \frac{1}{N} \sum_{i=1}^N \mathbf{X}_i$$

Definición 4.2.12 Se define $P_t(x)$ como función de $x \in (0, 1)$ mediante

$$P_t(x) = \sum_{1 \leq i \leq Nx} (\mathbf{X}_i(t) - \bar{\mathbf{X}}_N(t)) - x \sum_{i=1}^N (\mathbf{X}_i(t) - \bar{\mathbf{X}}_N(t))$$

Teniendo en cuenta los supuestos en 4.2.1 para \mathbf{X}_i , se tiene que $P_t(x) \in \mathcal{F}\mathbf{H}_D$. Por la forma del modelo en el supuesto 3, los operadores de covarianza para \tilde{Y}_i y \tilde{X}_i son iguales. El producto interno en 4-49 y su respectiva norma estimada $\|\cdot\|_{\hat{K}}$ se calculan usando una estimación para la transformada inversa de \mathcal{F} y las parejas propias estimadas para $\tilde{\lambda}_{m,k}, \tilde{e}_{m,k}$.

Definición 4.2.13 *Se define*

$$T_N(x) := \frac{1}{N} \|P_t(x)\|_{\hat{K}}^2$$

A continuación, se desarrolla la norma en la parte derecha de la definición para encontrar la forma de T_N .

Sea

$$\hat{\eta}_{i,m,k} = \left\langle \tilde{\mathbf{X}}_i - \tilde{\mathbf{X}}_N, \hat{\mathbf{e}}_{m,k} \right\rangle \quad (4-55)$$

donde

$$\tilde{\mathbf{X}}_i = \hat{\mathcal{F}}^{-1} \mathbf{X}_i \quad \tilde{\mathbf{X}}_N = \hat{\mathcal{F}}^{-1} \bar{\mathbf{X}}_N.$$

Se tiene que:

$$\mathbf{X}_i - \bar{\mathbf{X}}_N = \sum_{k=1}^p \sum_{m=1}^{m_k} \hat{\eta}_{i,m,k} \hat{\mathbf{e}}_{m,k} \quad (4-56)$$

en consecuencia

$$\tilde{\mathbf{X}}_i(t) - \tilde{\mathbf{X}}_N = \hat{\mathcal{F}} (X_i - \bar{X}_N) = \sum_{k=1}^p \sum_{m=1}^{m_k} \hat{\eta}_{i,m,k} \hat{\mathcal{F}} \hat{\mathbf{e}}_{m,k} \quad (4-57)$$

aplicando la expresión anterior a $P_t(x)$ se tiene

$$P_t(x) = \sum_{1 \leq i \leq Nx} \left(\sum_{k=1}^p \sum_{m=1}^{m_k} \hat{\eta}_{i,m,k} \hat{\mathcal{F}} \hat{\mathbf{e}}_{m,k} \right) - x \sum_{i=1}^N \left(\sum_{k=1}^p \sum_{m=1}^{m_k} \hat{\eta}_{i,m,k} \hat{\mathcal{F}} \hat{\mathbf{e}}_{m,k} \right) \quad (4-58)$$

Teniendo en cuenta la definición del producto interno en 4-49 pero usando los valores estimados:

$$\langle P_t(x), P_t(x) \rangle_{\hat{K}} = \sum_{k=1}^p \sum_{m=1}^{m_k} \langle \Theta_{m,k}, \Theta_{m,k} \rangle_{\mathbf{H}},$$

donde

$$\Theta_{m,k} := \frac{1}{\sqrt{\hat{\lambda}_{m,k}}} \hat{\mathcal{F}}^{-1} \left[\sum_{1 \leq i \leq Nx} \hat{\eta}_{i,m,k} \hat{\mathcal{F}} \hat{\mathbf{e}}_{m,k} - x \sum_{i=1}^N \hat{\eta}_{i,m,k} \hat{\mathcal{F}} \hat{\mathbf{e}}_{m,k} \right]$$

simplificando

$$\Theta_{m,k} = \frac{1}{\sqrt{\hat{\lambda}_{m,k}}} \left[\left(\sum_{1 \leq i \leq Nx} \hat{\eta}_{i,m,k} - x \sum_{i=1}^N \hat{\eta}_{i,m,k} \right) \hat{\mathbf{e}}_{m,k} \right]$$

la expresión para la norma al cuadrado se puede escribir de la siguiente manera

$$\|P_t(x)\|_{\hat{K}}^2 = \langle P_t(x), P_t(x) \rangle_{\hat{K}} = \sum_{k=1}^p \sum_{m=1}^{d_k} \left[\frac{1}{\hat{\lambda}_{m,k}} \left(\sum_{1 \leq i \leq Nx} \hat{\eta}_{i,m,k} - x \sum_{i=1}^N \hat{\eta}_{i,m,k} \right)^2 \right], \quad (4-59)$$

notese que si $p = 1$ se tiene la expresión mostrada en Berkes et al. (2009).

En general, sea

$$d = \sum_{k=1}^p d_k,$$

obsérvese que $P_t(x) \in \mathcal{F} \mathbf{H}_D$, por esta razón en adelante se omite el subíndice t . haciendo un ordenamiento y un cambio en los índices, la expresión 4-59, se puede escribir como:

$$\langle P(x), P(x) \rangle_{\hat{K}} = \sum_{l=1}^d \left[\frac{1}{\hat{\lambda}_l} \left(\sum_{1 \leq i \leq Nx} \hat{\eta}_{il} - x \sum_{i=1}^N \hat{\eta}_{il} \right)^2 \right],$$

esta expresión dividida en N coincide con $T_N(x)$ en Berkes et al. (2009), el problema se ha llevado al caso univariado funcional. Es decir el espacio planteado es isomorfo a un subespacio de $\mathbb{L}^2[0, 1]$. En adelante se siguen los mismos pasos que en Berkes et al. (2009).

Nota 4.2.2 (Pasos para implementar la propuesta) *El procedimiento para el caso en que los datos recogidos en el tiempo sean tomados periódicamente y no hay datos faltantes está programado en \mathbf{R} y se pueden consultar en Latorre (2019), también está disponible la app app análisis donde están implementados los pasos.*

1. Para cada instante de tiempo t , se realiza un análisis de componentes principales en \mathbb{R}^p , teniendo en cuenta las siguientes consideraciones.

- a) Para cada t en el análisis de componentes principales se ordenan los valores propios de mayor a menor, $\lambda_{t,1}, \lambda_{t,2}, \dots, \lambda_{t,p}$. Estos valores ordenados son los estimados de los elementos de la matriz diagonal que devuelve la función \mathbf{D} evaluada en t como se muestra en 4-11.
- b) Los vectores propios encontrados en cada análisis de componentes principales del ítem anterior conforman la matriz que en cada instante de tiempo t , devuelve la función \mathbf{V} siguiendo la expresión en 4-11 y además considerando las siguientes dos condiciones: Formar una matriz ortogonal de determinante 1 y estar lo más cerca posible de la matriz de vectores propios encontrada en el tiempo anterior. Para el primer instante de tiempo se toma la matriz con vectores propios más cercana a la identidad. Teniendo los vectores propios ordenados por su correspondiente valor propio de mayor a menor. Para cada t hay 2^p posibilidades para hallar la matriz que devuelve la función \mathbf{V} , esto se debe a los signos de cada uno de los p vectores. Las dos condiciones mencionadas determinan solo una matriz.
- c) Estimar \mathbf{B} . Los procedimientos anteriores están orientados a la construcción del estimado de la función \mathbf{B} teniendo en cuenta la expresión 4-25.
2. Para estimar la acción del operador \mathcal{F}^{-1} siguiendo 4-24 y 4-30 se encuentra para cada t la inversa de \mathbf{B}_t .
- a) Al obtener en cada tiempo t los valores propios para la matriz de covarianza de dimensión p y haber seleccionado una matriz formada con los vectores propios, se tiene que la matriz de covarianza estimada se puede escribir como
- $$\hat{\Sigma}_t = \hat{\mathbf{V}}_t \hat{\mathbf{D}}_t \hat{\mathbf{V}}_t^T,$$
- b) Para cada t la estimación de la matriz \mathbf{A}_t es como en 4-13 pero con los valores estimados de $\mathbf{D}_t^{-1/2}$ y \mathbf{V}_t^T .
3. Se aplica el operador inverso estimado $\hat{\mathcal{F}}^{-1}$, a las observaciones discretizadas que dan como resultado observaciones de un proceso p -variado funcional en el que no hay dependencia.
4. A las observaciones después de ser transformadas se les realiza un suavizamiento para cada una de las p componentes.
- a) Para este tipo de análisis se escogió las bases de fourier para realizar el suavizamiento y para castigar el sobre ajuste se seleccionó el gcw como se indicó en 3.1.2.

5. Después de suavizar las N curvas en cada una de las p componentes en el ítem anterior se realiza un análisis de componentes principales funcionales siguiendo lo descrito en 3.2.4
 - a) El número de armónicos para cada uno de los p análisis puede ser diferente y corresponde a los valores d_k con $k = 1, 2, \dots, p$.
6. Con los puntajes resultantes del proceso de componentes principales como en 4-55, se realiza el análisis de punto de cambio siguiendo las ideas de Berkes et al. (2009).
 - a) Se construye para cada una de las componentes el estadístico $T_{N,k}(x)$ como en 4-7 y luego se suman, para hallar $T_N(x)$ como en 4.2.13. La construcción del espacio hace que sea correcto sumar los estadísticos como se evidencia en 4-59, debido a que los espacios por construcción son ortogonales.
 - b) El estadístico $S_{N,d}$ en 4-8, donde $d = d_1 + d_2 + \dots + d_k$, es el estadístico de prueba con los criterios ya mencionados en la descripción del enfoque de Berkes et al. (2009). En la tabla 4-1 están los valores de comparación para la toma de decisión respecto a la hipótesis nula. En caso de que exista punto de cambio, o sea de que se rechace la hipótesis nula, el valor x^* en el que se alcanza el máximo del estadístico $T_N(x)$ es un estimador del punto de cambio.
7. El proceso para la detección de puntos de cambio es iterativo; en caso de que exista un punto de cambio, para los datos antes y después del punto de cambio se realiza el procedimiento desde el paso 5 hasta que no se detecten puntos de cambio.

α en %	d : Número de Componentes Principales					
	1	2	3	4	5	6
10	0.345165	0.606783	0.842567	1.065349	1.279713	1.485200
5	0.460496	0.748785	1.00139	1.239675	1.469008	1.684729
1	0.740138	1.072101	1.352099	1.626695	1.866702	2.125950
	7	8	9	10	11	12
10	1.690773	1.897365	2.096615	2.288572	2.496635	2.686238
5	1.895557	2.124153	2.322674	2.526781	2.744438	2.949004
1	2.342252	2.589244	2.809778	3.033944	3.268031	3.491102
	13	14	15	16	17	18
10	2.884214	3.066906	3.268958	3.462039	3.650724	3.837678
5	3.147604	3.336262	3.544633	3.740248	3.949054	4.136169
1	3.708033	3.903995	4.116829	4.317087	4.554650	4.734714
	19	20	21	22	23	24
10	4.024313	4.2148	4.404677	4.591972	4.778715	4.965613
5	4.327286	4.532917	4.718904	4.908332	5.101896	5.303462
1	4.974172	5.156282	5.369309	5.576596	5.759427	5.973941
	25	26	27	28	29	30
10	5.159057	5.346543	5.521107	5.714145	5.885108	6.083306
5	5.495721	5.688849	5.866095	6.068351	6.242770	6.444772
1	6.203718	6.393582	6.572949	6.771058	6.977607	7.186491

Tabla 4-1: Valores simulados de la distribución K_d tomado de Berkes et al. (2009)

5 Simulaciones y Aplicaciones

En este capítulo se simulan varios escenarios en los que se muestra la sensibilidad de la propuesta. Además, con dos conjuntos de datos reales la propuesta es capaz de detectar cambio en la media, en puntos en los que es coherente con la información suministrada por las fuentes de los datos. La metodología es aplicada a mediciones de calidad del aire en la ciudad de Bogotá.

Hay dos app disponibles, una para realizar simulaciones de procesos y detectar punto de cambio *app simulaciones*; y otra para realizar un análisis rápido de detección de punto de cambio con la propuesta desde 1 hasta 5 procesos, *app análisis*. Los códigos usados para las simulaciones, implementación de la metodología y las apps se encuentran en Latorre (2019).

En las simulaciones y en las aplicaciones para los suavizamientos se usó las funciones de la base de Fourier, el número de armónicos para los análisis de componentes principales funcionales es seleccionado tratando de explicar la mayor variabilidad posible, se recomienda un valor entre el 95 % y el 99 %.

5.1. Simulaciones

Considérese el proceso funcional multivariado

$$\mathbf{Y} = (Y_1, Y_2, \dots, Y_p), \quad (5-1)$$

como en la definición 4.2.1, el operador \mathcal{F} como en la definición 4.2.6 y la función \mathbf{B} como en 4.2.5. De acuerdo con la expresión 4-27, la forma en la que actúa el operador \mathcal{F} está dada por la función \mathbf{B} , y ésta depende de Σ como se puede apreciar en 4-10 y en 4-25. Además, para $i = 1, 2, \dots, N$ con N el tamaño de la muestra, considérese las observaciones funcionales \mathbf{X}_i con

$$\mathbf{X}_i = \boldsymbol{\mu}_\gamma + \mathbf{Y}_i, \quad (5-2)$$

como en el ítem 2 de la nota 4.2.1, y cada \mathbf{Y}_i con la misma distribución de \mathbf{Y} .

Se contrastan las siguientes hipótesis:

$$H_0 : E(\mathbf{X}_1) = E(\mathbf{X}_2) = \dots = E(\mathbf{X}_N) \quad (5-3)$$

contra la hipótesis de que existe un k^* tal que

$$E(\mathbf{X}_1) = E(\mathbf{X}_2) \dots E(\mathbf{X}_{k^*}) = \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2 = E(\mathbf{X}_{k^*+1}) = E(\mathbf{X}_{k^*+2}) \dots E(\mathbf{X}_N) \quad (5-4)$$

donde

$$\boldsymbol{\mu}_1 = (\mu_{1,1}, \mu_{2,1}, \dots, \mu_{p,1}), \quad \boldsymbol{\mu}_2 = (\mu_{1,2}, \mu_{2,2}, \dots, \mu_{p,2}),$$

denotan la media funcional antes después del punto de cambio respectivamente.

Se debe tener en cuenta que si

$$\mathbf{B}(t) = \tilde{\mathbf{B}}(t) \mathbf{O}(t),$$

donde $\mathbf{O}(t)$ es una función que para cada instante de tiempo t , devuelve una matriz ortogonal \mathbf{O}_t , entonces

$$\mathbf{B}(t) \mathbf{B}(t)^T = \tilde{\mathbf{B}}(t) \tilde{\mathbf{B}}(t)^T = \boldsymbol{\Sigma}(t).$$

Lo anterior implica que diferentes funciones como $\mathbf{B}(t)$ pueden producir la misma función $\boldsymbol{\Sigma}(t)$.

Es necesario aclarar que el objetivo no es estimar la función \mathbf{B} sino el efecto de la transformación. Teniendo en cuenta lo anterior, el procedimiento para realizar las simulaciones es mostrado a continuación.

Nota 5.1.1 (Pasos para realizar las simulaciones) *El procedimiento para realizar las simulaciones está programado en \mathbf{R} y se pueden consultar en Latorre (2019).*

1. Definir la función \mathbf{B} , teniendo en cuenta que para cada instante de tiempo t devuelve una matriz cuadrada \mathbf{B}_t . Además se debe tener el supuesto 3 en 4.2.1.
2. Generar procesos multivariados funcionales de media 0, estos procesos se asumen pertenecientes a \mathbf{H} como en la definición 4.2.4. Se aclara que los procesos decritos en este paso pueden tener ya un tipo de correlación entre ellos, sin embargo, se les adicionará la correlación dada por la transformación que genera \mathbf{B} .
3. Se definen los vectores de las funciones medias antes y después del punto de cambio.
4. En cada t se transforman los procesos del ítem 2 mediante la matriz \mathbf{B}_t , o sea los procesos del ítem 2 son transformados por el operador \mathcal{F} , estos procesos transformados corresponden a los procesos en 5-1.
5. A los procesos del ítem anterior se les agrega la media antes o después del punto de cambio dadas en el ítem 3 que generan los procesos en 5-2.

5.1.1. Simulación procesos 3-variados (un cambio)

Se presentan una serie de escenarios para procesos 3-variados en los que sólo se incluye un punto de cambio en la media funcional multivariada. En *app simulaciones* se pueden realizar las simulaciones descritas a continuación.

Se simulan 50 realizaciones de procesos estocásticos 3-variados con t en el intervalo $[0, 1]$ y tomando los valores $0, 0.01, 0.02, \dots, 1$. El punto de cambio se colocó en la realización 15; esto significa que las primeras 15 realizaciones de procesos tienen la misma media, de la realización 16 en adelante la media es diferente. La función \mathbf{B} que devuelve las matrices para cada valor de t tiene las siguientes componentes:

$$\mathbf{B}(t) = \begin{bmatrix} 2t & \exp(t) & t^3 \\ t + 1 & \sin(3t) + 2 & 1.5 \\ t^2 & \cos(4 + t) + 2 & t^2 + 1 \end{bmatrix} \quad (5-5)$$

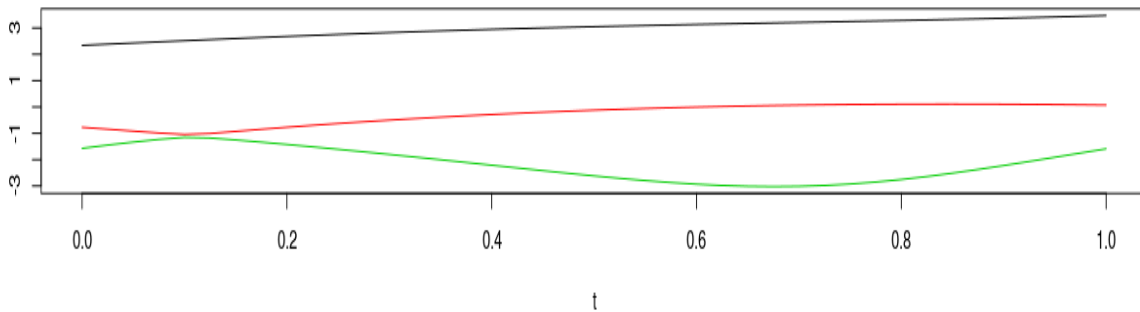


Figura 5-1: Logaritmos de los valores propios de $\mathbf{B}(t)\mathbf{B}(t)^T$

La figura 5-1 muestra que se cumplen las desigualdades indicadas en las hipótesis de la proposición 3. Se realizan dos tipos de simulaciones cambiando la forma en que se simulan los errores.

Tipo I

En este conjunto de simulaciones los errores del ítem 2 en la nota 5.1.1, se obtienen usando la librería desarrollada en Happ-Kurz (2020), además de los errores simulados se incluye un ruido Gaussiano. Los errores son reescalados antes y después del paso 4 de la nota 5.1.1 y se muestran en la figura 5-2. Es necesario aclarar que las funciones estimadas de los valores propios de $\mathbf{B}(t)\mathbf{B}(t)^T$ en este caso no se pueden comparar con las del gráfico 5-1, debido a los reescalamientos y al uso de las funciones de Happ-Kurz (2020).

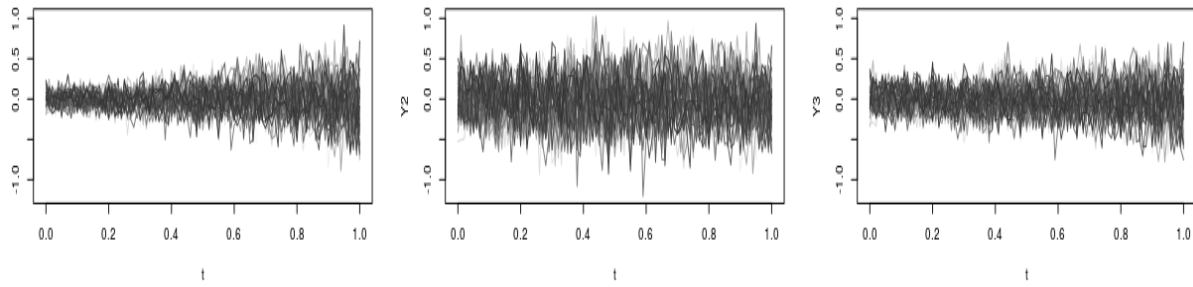


Figura 5-2: Simulación de 50 realizaciones del proceso estocástico 3-variado de media 0, con ruidos.

Las medias funcionales para cada uno de las tres componentes de los procesos antes del punto de cambio son:

$$\begin{aligned}\mu_{1,1} &= t + \exp(2t) \\ \mu_{2,1} &= 2\cos(6\pi t) + \sin(18\pi t) \\ \mu_{3,1} &= 2\sin(2\pi t) + \sin(10\pi t) - 1\end{aligned}$$

Ahora, se muestran 2 escenarios de análisis para los cambios en la media.

Escenario 1

Se consideran cambios en las tres componentes de la media; un cambio en la forma, uno en la escala y uno de desplazamiento. Las funciones medias en este caso son:

$$\begin{aligned}\mu_{1,2} &= t + \exp(2t) + v_1 \sin(12\pi t) \\ \mu_{2,2} &= (1 + v_2)(2\cos(6\pi t) + \sin(18\pi t)) \\ \mu_{3,2} &= 2\sin(2\pi t) + \sin(10\pi t) - 1 + v_3\end{aligned}$$

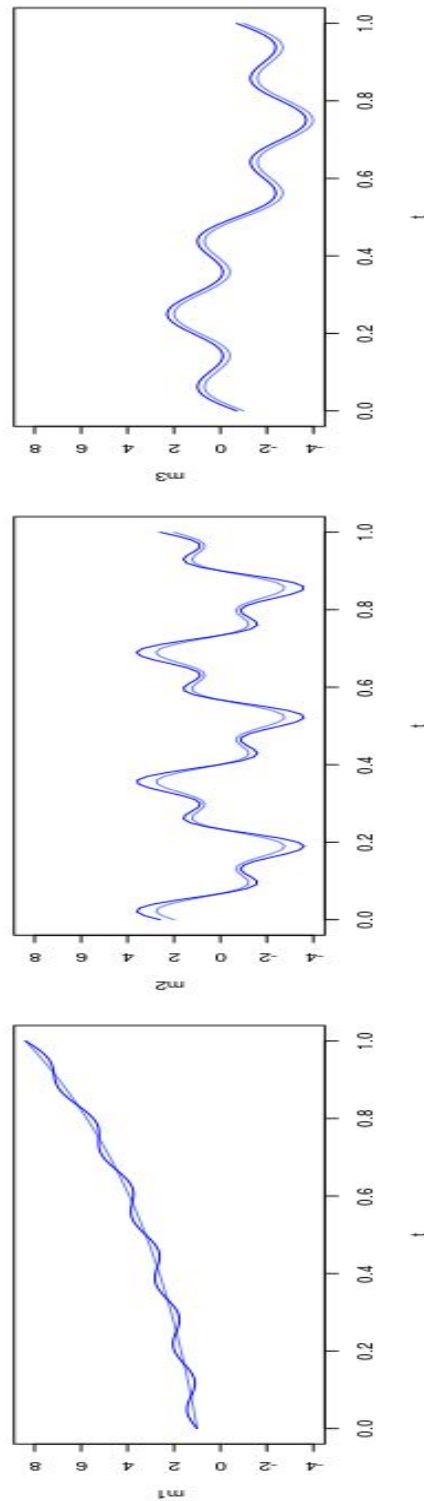


Figura 5-3: Funciones medias de las tres componentes del proceso en el Escenario 1 con $v_1 = v_2 = v_3 = 0.3$. En azul claro la media antes del cambio, en azul oscuro la media después del cambio.

Para $v_1 = v_2 = v_3 = 0.3$, como se muestra en la figura 5-3 en la media de la componente uno de los procesos se está incluyendo un cambio de forma, a la componente dos se le multiplica por un factor positivo y estrictamente mayor a 1, y a la componente tres se le realiza un desplazamiento vertical hacia arriba. En la figura 5-4, se muestran las 50 realizaciones y no es posible evidenciar que exista un cambio en la media. En el paso 4 del procedimiento 4.2.2 se seleccionan los armónicos como se muestran en la tabla 5-1. En la figura 5-5 se presenta el gráfico del estadístico T con el candidato a punto de cambio, y dado que se rechaza la hipótesis nula de no existencia de punto de cambio, el candidato coincide con el punto de cambio.

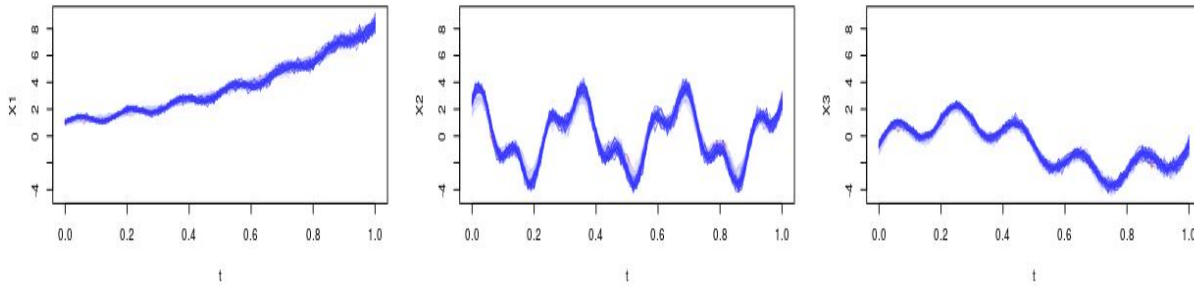


Figura 5-4: Las tres componentes de 50 realizaciones simuladas del proceso estocástico en el Escenario 1 con $v_1 = v_2 = v_3 = 0.3$.

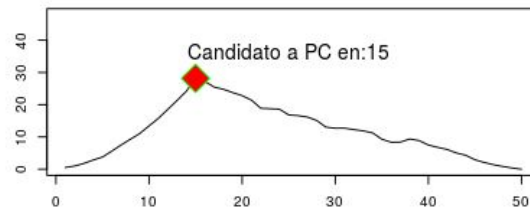


Figura 5-5: Estadístico T de las simulaciones con $v_1 = v_2 = v_3 = 0.3$ en el Escenario 1.

Escenario 2

Se consideran cambios en las tres componentes de la media; se añade una componente cuadrática, se cambia la escala de uno de los sumandos de la función y se modifica una frecuencia. Las funciones medias en este caso son:

$$\begin{aligned}\mu_{1,1} &= t + \exp(2t) + v_1 t^2 \\ \mu_{2,1} &= 2\cos(6\pi t) + (1 + v_2)\sin(18\pi t) \\ \mu_{3,1} &= 2\sin(2\pi t) + \sin(10\pi t + v_3(2\pi t)) - 1\end{aligned}$$

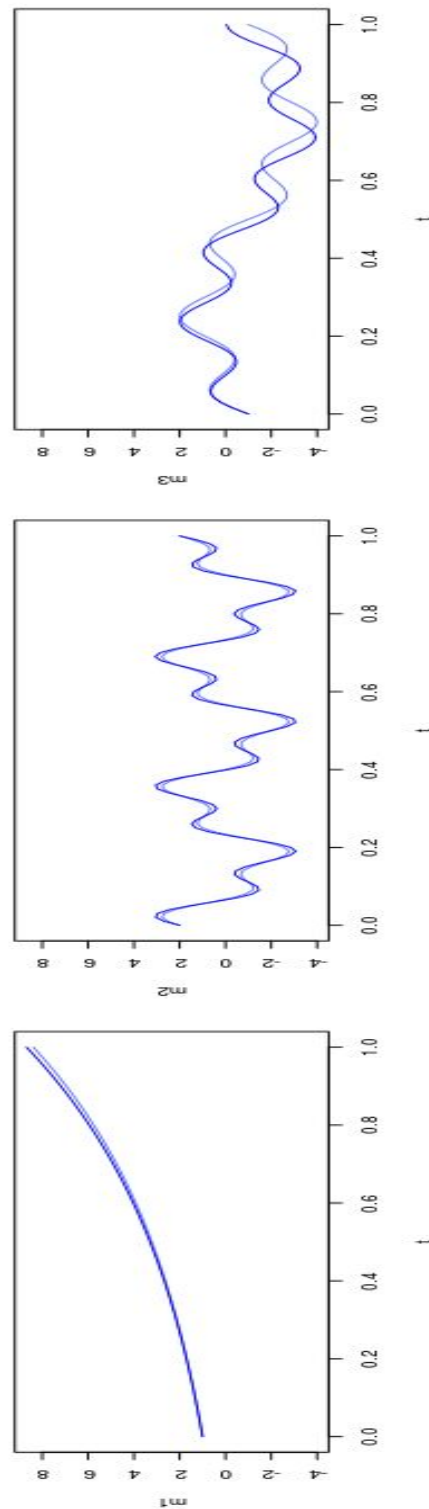


Figura 5-6: Funciones medias de las tres componentes de los procesos en el Escenario 2 con $v_1 = v_2 = v_3 = 0.3$. En azul claro La media antes del cambio, en azul oscuro la media después del cambio.

Para $v_1 = v_2 = v_3 = 0.3$, como se muestra en la figura 5-6 en la media de la componente uno de los procesos hay un desplazamiento que no es constante, a la componente dos se le multiplica por un factor positivo y estrictamente mayor a 1 al sumando de la función seno, y en la componente tres se realiza un cambio en la frecuencia de una de las funciones seno. En la figura 5-7, se muestran las 50 realizaciones, en la gráfica de la derecha es posible apreciar el cambio existente. En el paso 4 del procedimiento 4.2.2 se seleccionan los armónicos como se muestran en la tabla 5-1. En la figura 5-8 se presenta el gráfico del estadístico T con el candidato a punto de cambio, y dado que se rechaza la hipótesis nula de no existencia de punto de cambio, el candidato coincide con el punto de cambio.

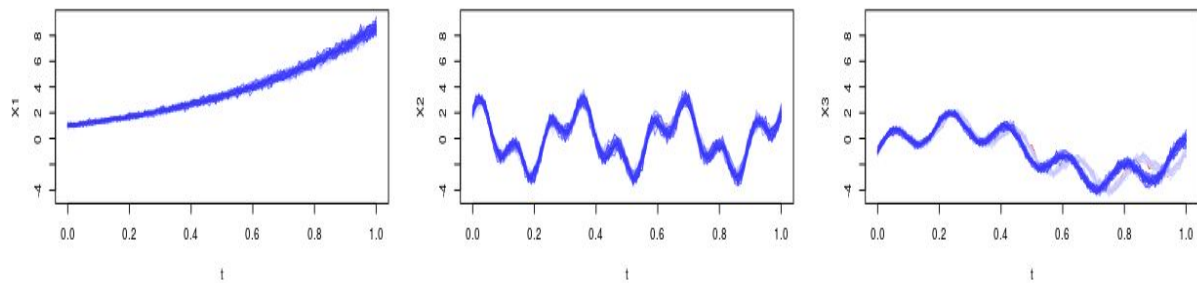


Figura 5-7: Las tres componentes de 50 realizaciones simuladas del proceso estocástico en el Escenario 2 con $v_1 = v_2 = v_3 = 0.3$.

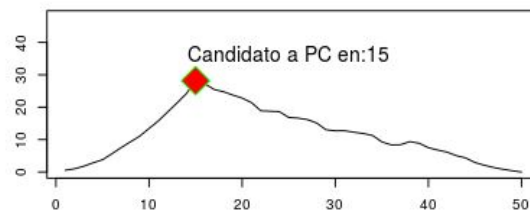


Figura 5-8: Estadístico T de las simulaciones con $v_1 = v_2 = v_3 = 0.3$ en el Escenario 2.

Magnitud Cambios			Escenario 1					Escenario 2				
v_1	v_2	v_3	c_1	c_2	c_3	P	R	d_1	d_2	d_3	P	R
0.300	0.300	0.300	8	4	11	15	S	8	6	11	15	S
0.300	0.300	0.100	8	4	11	15	S	10	4	11	15	S
0.300	0.300	0.030	8	4	11	15	S	10	6	11	15	S
0.300	0.100	0.300	10	1	11	15	S	8	8	11	15	S
0.300	0.100	0.100	10	5	10	15	S	10	5	10	15	S
0.300	0.100	0.030	10	6	10	15	S	10	8	9	15	S
0.300	0.030	0.300	10	1	11	15	S	8	8	11	15	S
0.300	0.030	0.100	11	7	11	15	S	10	6	10	15	S
0.300	0.030	0.030	11	8	10	15	S	10	8	8	15	S
0.100	0.300	0.300	8	4	11	15	S	8	8	11	15	S
0.100	0.300	0.100	9	5	11	15	S	10	5	11	15	S
0.100	0.300	0.030	9	5	11	15	S	10	8	11	15	S
0.100	0.100	0.300	10	2	10	15	S	8	8	11	15	S
0.100	0.100	0.100	10	6	9	15	S	10	7	10	15	S
0.100	0.100	0.030	10	8	8	15	S	10	10	9	15	S
0.100	0.030	0.300	10	1	9	15	S	8	8	11	15	S
0.100	0.030	0.100	10	7	8	15	S	10	7	10	15	S
0.100	0.030	0.030	11	10	9	15	S	10	10	9	16	S
0.030	0.300	0.300	8	4	11	15	S	8	8	11	15	S
0.030	0.300	0.100	9	5	10	15	S	10	6	11	15	S
0.030	0.300	0.030	9	5	11	15	S	10	8	11	15	S
0.030	0.100	0.300	10	2	10	15	S	8	8	11	15	S
0.030	0.100	0.100	10	6	9	15	S	10	8	10	15	S
0.030	0.100	0.030	10	9	8	15	S	10	10	10	15	S
0.030	0.030	0.300	10	1	9	15	S	8	7	10	15	S
0.030	0.030	0.100	10	8	8	15	S	10	7	10	15	S
0.030	0.030	0.030	10	10	9	16	S	10	10	9	15	S
0.020	0.020	0.020	10	10	9	16	S	10	10	10	16	S
0.010	0.010	0.010	10	10	10	16	S	10	10	10	16	S
0.005	0.005	0.005	10	10	10	28	N*	10	10	10	16	S
0.001	0.001	0.001	10	10	10	28	N	10	10	10	28	N

Tabla 5-1: Detección de punto cambio usando simulaciones de Procesos 3-variados con errores usando Happ-Kurz (2020).

v_1, v_2, v_3 : Magnitud de los cambios dependiendo del escenario. d_1, d_2, d_3 : Número de componentes principales para cada una de las tres variables. P: Estimación del Punto de cambio. R: ¿Se rechaza la hipótesis Nula con un nivel de significancia del 1%?, N*: No se rechaza al 1% pero sí al 5%. Las simulaciones se realizaron fijando la semilla 1234 en la aplicación *app simulaciones*. El número de elementos de la base

de Fourier fue 15.

En la tabla **5-1** se muestran para los Escenarios 1 y 2 diferentes magnitudes en los cambios y varias posibilidades; cambios de forma, de escala, desplazamientos constantes y variables, y aumento de frecuencia. Considerando las magnitudes de los cambios y el rango de los errores observado en la figura **5-2**, se evidencia que la propuesta es sensible a detectar pequeños cambios en comparación con los errores.

Nota 5.1.2 *No es correcto realizar los pasos 4 y 5 del procedimiento 4.2.2 sin tener en cuenta los tres primeros pasos en los que se aplica la transformación de los procesos. Sin embargo, hay casos en los que ese procedimiento incorrecto, detecta de manera adecuada el punto de cambio esto debido a la magnitud de los cambios. Por ejemplo, en el Escenario 2 con $v_1, v_2, v_3 = 0.3$ aplicar los pasos 4 y 5 sin aplicar la transformación detecta adecuadamente el punto de cambio en 15, mas sin embargo, si $v_1 \leq 0.1$, $v_2 \leq 0.1$, $v_3 \leq 0.1$, ese procedimiento no detecta de manera adecuada el punto de cambio o no se rechaza la hipótesis nula; la propuesta desarrollada en este trabajo es capaz de detectar puntos de cambio por debajo de 0.1.*

Tipo II

En este caso los errores del ítem 2 en la nota 5.1.1 son ruido Gaussiano. Los errores no son reescalados por lo tanto las funciones estimadas de los valores propios de $\mathbf{B}(t)\mathbf{B}(t)^T$ en este caso son comparables con los de las figura **5-1**.

En la figura **5-9** se muestran los errores simulados sin ser reescalados.

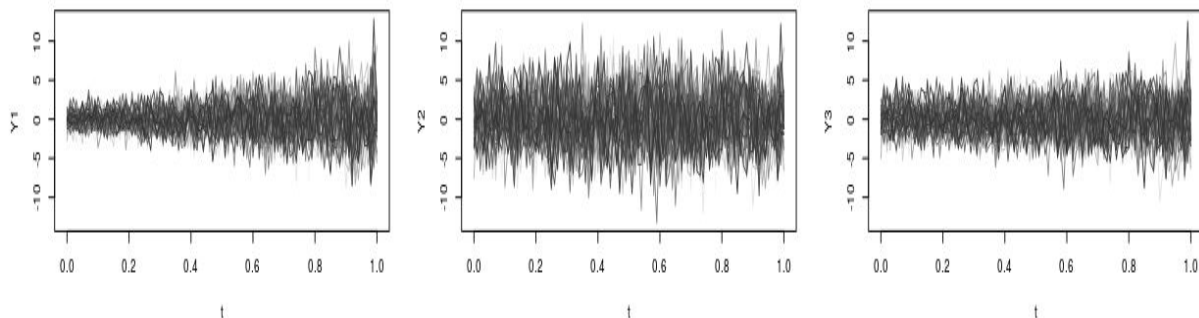


Figura 5-9: 50 simulaciones de ruidos 3-variados y transformados por $\mathbf{B}(t)$.

Las medias funcionales para cada uno de los tres procesos son:

Antes del punto de cambio

$$\begin{aligned}\mu_{1,1} &= 5(t + \exp(2t)) \\ \mu_{2,1} &= 5(2\cos(6\pi t) + \sin(18\pi t)) \\ \mu_{3,1} &= 5(2\sin(2\pi t) + \sin(10\pi t) - 1)\end{aligned}$$

y después del punto de cambio

$$\begin{aligned}\mu_{1,1} &= 5(t + \exp(2t) + v_1 \sin(12\pi t)) \\ \mu_{2,1} &= 5((1 + v_2)(2\cos(6\pi t) + \sin(18\pi t))) \\ \mu_{3,1} &= 5(2\sin(2\pi t) + \sin(10\pi t) - 1 + v_3)\end{aligned}$$

Se seleccionó para este análisis $v_1 = v_2 = v_3 = 0.3$, la forma en la que están definidas las funciones antes y después del punto de cambio son reescalamientos de las medias en el Escenario 1 de la simulación tipo I. En la figura **5-10** se muestran las medias y en **5-11** se muestran los 50 procesos.

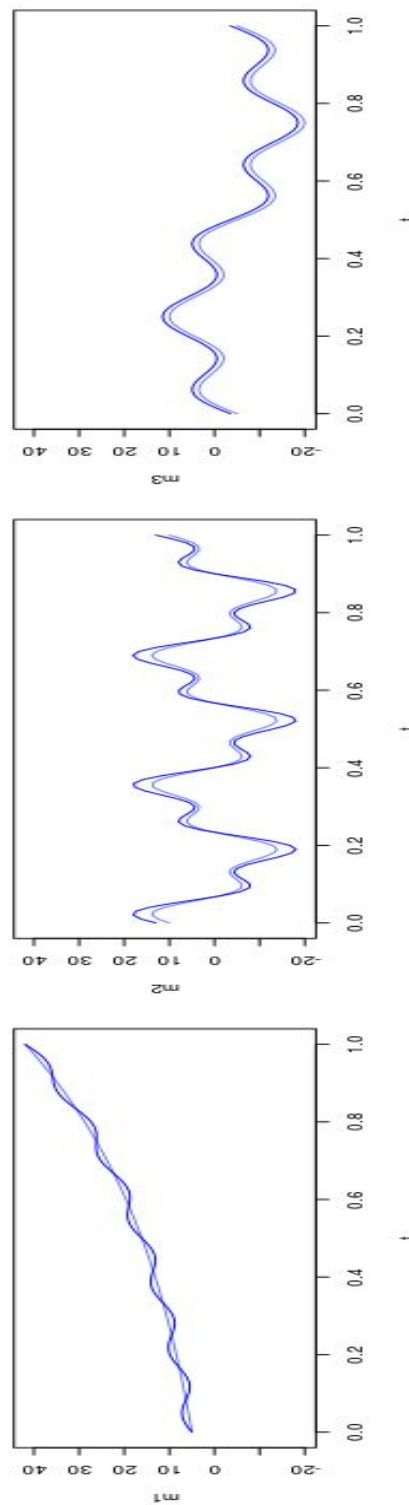


Figura 5-10: Funciones medias de las tres componentes de los procesos con $v_1 = v_2 = v_3 = 0.3$. En azul claro La media antes del cambio, en azul oscuro la media después del cambio.

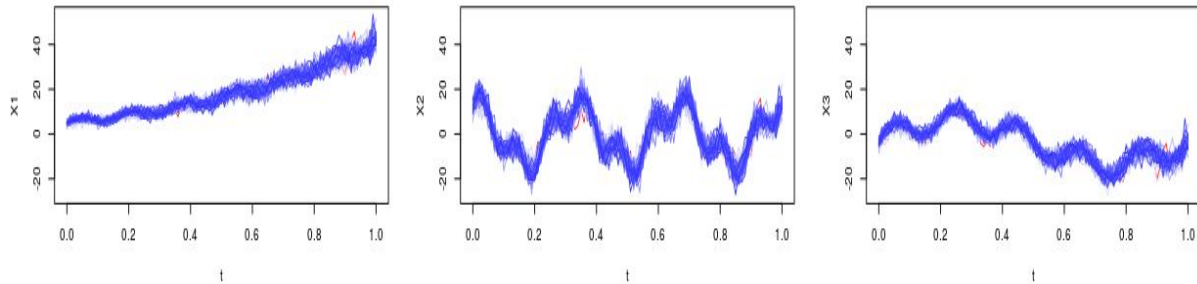


Figura 5-11: Las tres componentes de 50 realizaciones simuladas del proceso estocástico con $v_1 = v_2 = v_3 = 0.3$.

En la figura 5-12 se muestra la estimación de las funciones que en cada instante de tiempo devuelven los valores propios de la matriz \mathbf{B}_t . Se pueden comparar las figuras 5-12 y 5-1, de nuevo se hace énfasis en la imposibilidad de recuperar la función $\mathbf{B}(t)$ pero se pueden comparar las funciones que devuelven sus valores propios a lo largo del tiempo.

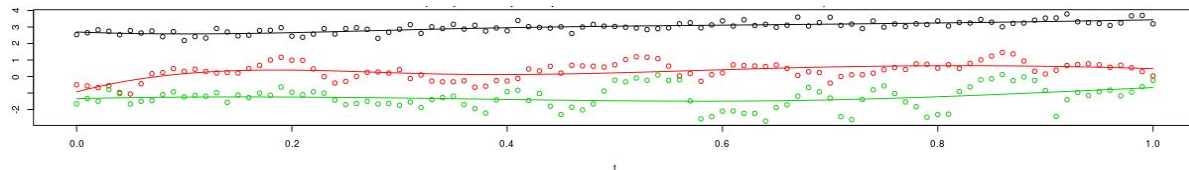


Figura 5-12: Logaritmos de los valores propios estimados de $\mathbf{B}(t)$. Los puntos son las estimaciones puntuales que son ajustadas a las respectivas curvas.

Para detectar el punto de cambio en este caso se seleccionaron 10,5,10 número de componentes principales para cada uno de los procesos después de realizar las transformaciones, el punto de cambio es detectado correctamente y la hipótesis nula que indica que no hay punto de cambio, es rechazada.

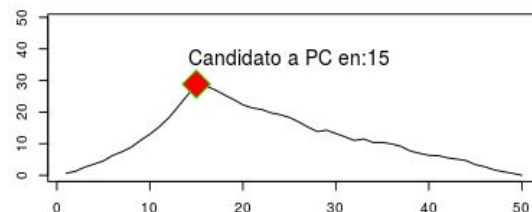


Figura 5-13: Estadístico T de las simulaciones con $v_1 = v_2 = v_3 = 0.3$ en el Escenario 2.

5.1.2. Simulación procesos 3-variados (dos cambios)

Se simulan 50 realizaciones de procesos 3-variados con $t = 0, 0.01, 0.02 \dots, 1$. Se colocan dos puntos de cambio para la media, uno en 15 y el otro en 39. La función **B** que devuelve las matrices para cada valor de t es como en 5-5. Los errores son los mismos que en la simulación Tipo I, apreciados en la figura 5-2.

Las medias funcionales son presentadas en la figura 5-14 y corresponden a:

Antes del primer cambio

$$\begin{aligned}\mu_{1,1} &= 2(t - 0.5)^2 \\ \mu_{2,1} &= 3t^3 - 0.9t \\ \mu_{3,1} &= \log(t + 1),\end{aligned}$$

entre el primer y el segundo cambio

$$\begin{aligned}\mu_{1,2} &= 2(t - 0.5)^2 + v_1 t^2 \\ \mu_{2,2} &= 3t^3 - 0.9t + v_2 \exp(t) \\ \mu_{3,2} &= \log(t + 1) + v_3,\end{aligned}$$

y después del segundo cambio

$$\begin{aligned}\mu_{1,3} &= 2(t - 0.5)^2 + v_1 \cos(8\pi t) \\ \mu_{2,3} &= 3t^3 - 0.9t + v_2 t \\ \mu_{3,3} &= (1 + v_3) \log(t + 1).\end{aligned}$$

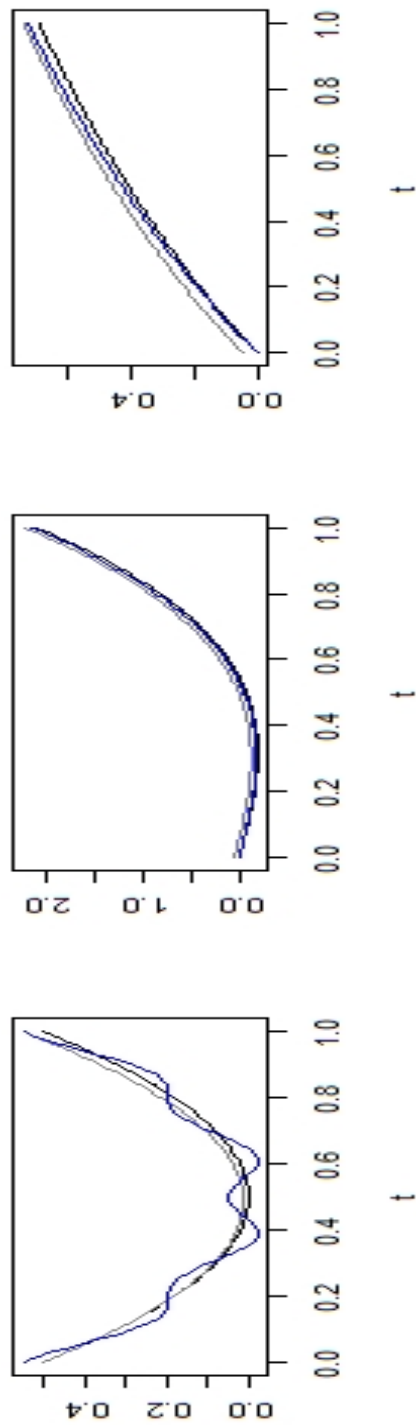


Figura 5-14: Funciones medias de las tres componentes del proceso con $v_1 = v_2 = v_3 = 0.05$. En negro la media antes del primer cambio, en gris la media entre el primer y el segundo cambio, y en azul la media después del segundo cambio.

En la figura 5-15, se muestran las 50 realizaciones y no es posible evidenciar que existan cambios en la media. La propuesta detecta primero el punto de cambio en 39 y luego en 15, en la tabla 5-2 se resumen los resultados.

Datos	d_1	d_2	d_3	P	R
1 al 50	10	10	8	39	S
1 al 39	10	10	8	15	S
1 al 15	8	9	7	9	N
16 al 39	10	10	9	28	N
40 al 50	7	6	7	45	N

Tabla 5-2: Detección de dos puntos de cambio usando simulaciones de Procesos 3-variados con errores usando Happ-Kurz (2020)

La magnitud de todos los cambios fue de 0.05. d_1, d_2, d_3 : Número de componentes principales para cada una de las tres variables. P: Estimación del Punto de cambio. R: ¿Se rechaza la hipótesis Nula con un nivel de significancia del 1%? El número de elementos de la bases de Fourier fue 15.

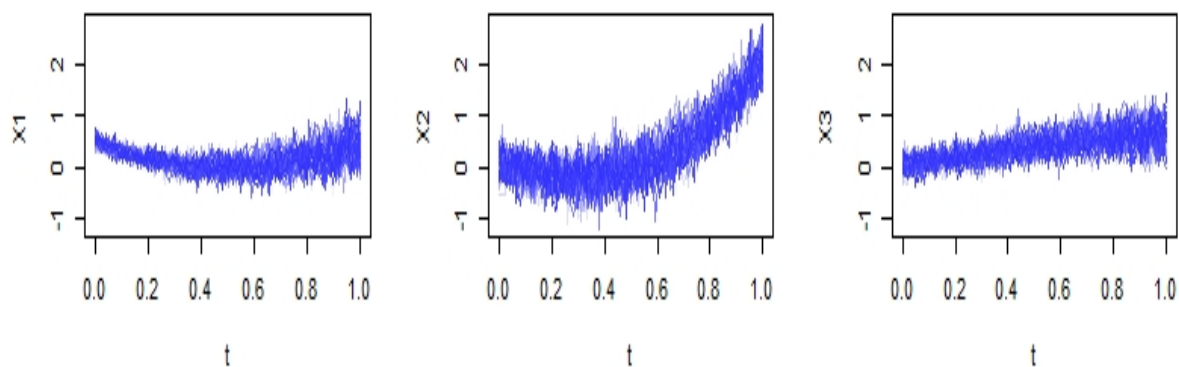


Figura 5-15: Las tres componentes de 50 realizaciones simuladas del proceso estocástico con $v_1 = v_2 = v_3 = 0.05$

5.1.3. Simulación procesos 4-variados

Se simulan 50 realizaciones de procesos 4-variados con t en el intervalo $[0, 1]$ y tomando los valores $0, 0.01, 0.02, \dots, 1$. El punto de cambio se colocó en la realización 40. La función \mathbf{B} que para cada t devuelve la matriz de transformaciones \mathbf{B}_t se muestra a continuación:

$$\mathbf{B}(t) = \begin{bmatrix} 2t & \sin(3t) + 2 & t^2 + 1 & t^2 + 1 \\ t + 1 & \cos(4 + t) + 2 & \exp(t)/3 & t^3 + 1 \\ t^2 & t^3 & \cos(2 + t) + 1 & t + 1 \\ \exp(t) & 1.5 & \cos(4t) + 1.3 & t^5 + 1 \end{bmatrix}$$

La figura 5-16 muestra que se cumplen las desigualdades indicadas en las hipótesis de la proposición 3.

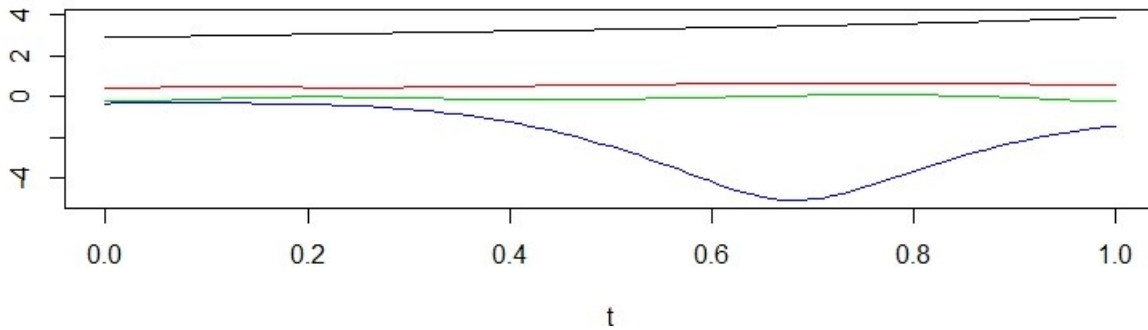


Figura 5-16: Logaritmos de los valores propios de $\mathbf{B}(t)\mathbf{B}(t)^T$

Los errores del ítem 2 en la nota 5.1.1, se obtienen usando la librería desarrollada en Happ-Kurz (2020), además de los errores simulados se incluye un ruido con distribución Laplace. Los errores son reescalados antes y después del paso 4 de la nota 5.1.1 y se muestran en la figura 5-17. Es necesario aclarar que las funciones estimadas de los valores propios de $\mathbf{B}(t)\mathbf{B}(t)^T$ en este caso no se pueden comparar con las del gráfico 5-16, debido a los reescalamientos y al uso de las funciones de Happ-Kurz (2020).

En la figura 5-18 se muestran las medias funcionales para cada uno de los cuatro procesos, antes del cambio

$$\begin{aligned} \mu_{1,1} &= 6(t - 0.5)^2 \\ \mu_{2,1} &= 24(t - 0.5)^4 \\ \mu_{3,1} &= 1.5t \\ \mu_{4,1} &= \exp(t) - 1, \end{aligned}$$

y después del cambio

$$\begin{aligned} \mu_{1,2} &= 6(t - 0.5)^2 + vt \\ \mu_{2,2} &= (1 + v)24(t - 0.5)^4 \\ \mu_{3,2} &= 1.5t + v \sin(10\pi t) \\ \mu_{4,2} &= (1 + v)(\exp(t) - 1). \end{aligned}$$

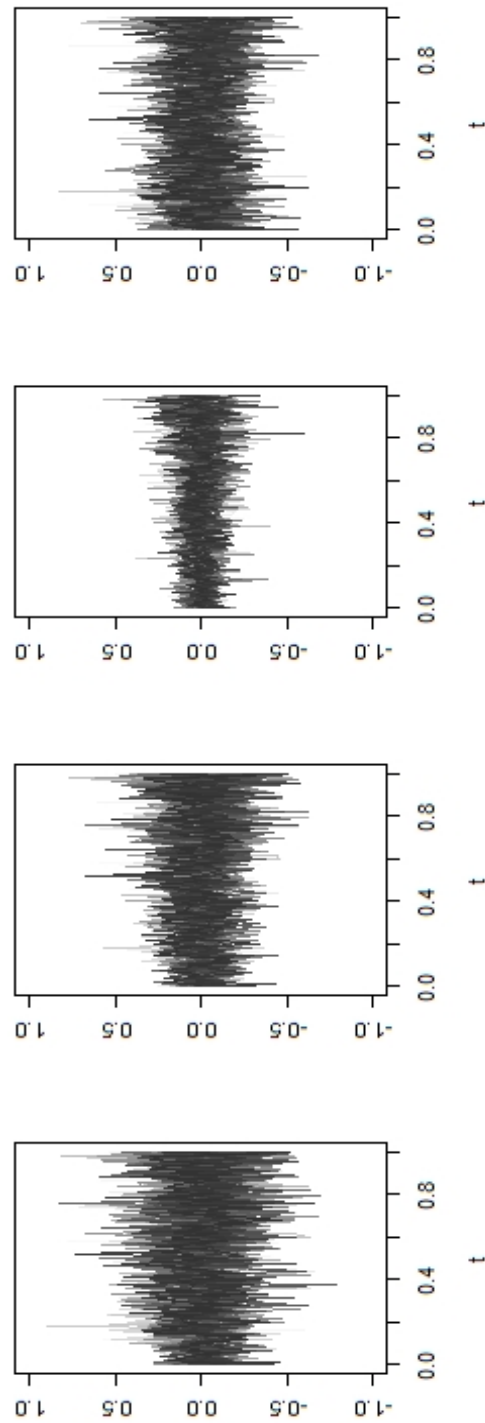


Figura 5-17: 50 simulaciones de ruidos de distribución Laplace y transformados por $\mathbf{B}(t)$.

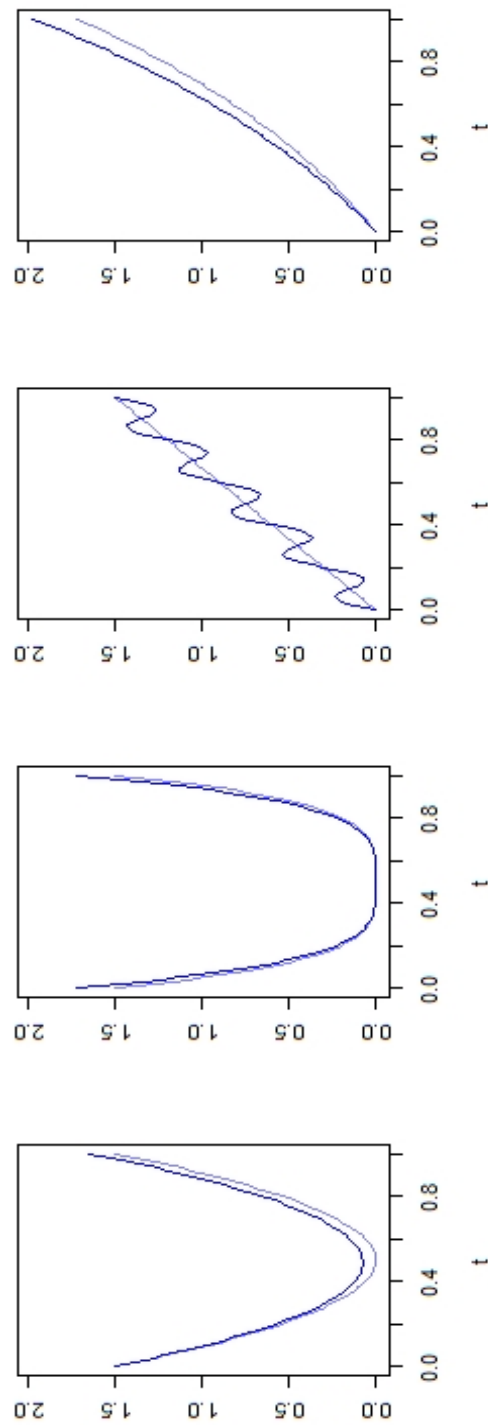


Figura 5-18: Funciones medias de las cuatro componentes de los procesos con $\nu = 0.15$. En azul claro La media antes del cambio, en azul oscuro la media después del cambio.

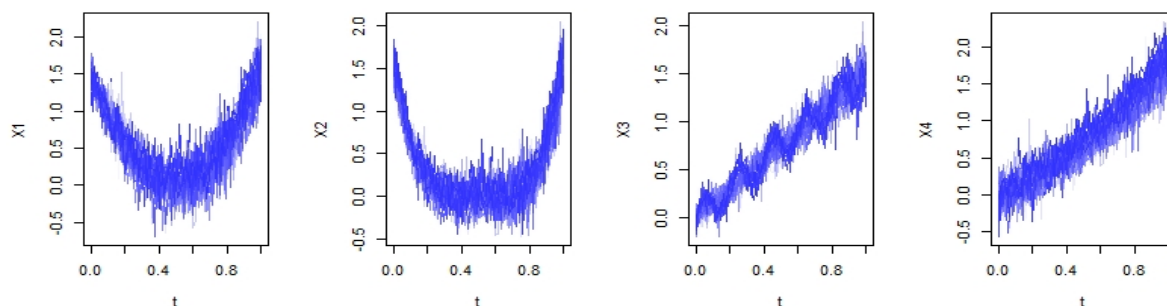


Figura 5-19: Las cuatro componentes de los 50 procesos simulados con $v = 0.15$

Nota 5.1.3 En la tabla 5-3 se muestran los resultados del análisis y se analiza el comportamiento de dos procedimientos que se pueden implementar fácilmente pero que no se deben usar, ya sea porque no tienen en cuenta toda la estructura de covarianza que existe o porque no se ha demostrado si es correcto usar algunos estadísticos.

- **Procedimiento 1** Realizar el proceso de transformación descrito en los 3 primeros pasos de la propuesta en 4.2.2 elimina la correlación en cada instante de tiempo, si los cambios son relativamente grandes en comparación con los errores no realizar esos pasos y proceder desde el paso 4 puede dar un resultado correcto.
- **Procedimiento 2** En este trabajo no se estudió la posibilidad de que la propuesta de punto de cambio en Berkes et al. (2009) se pueda extender para datos funcionales multivariados con el análisis de componentes principales de Happ and Greven (2018). Sin embargo, con los scores devueltos por el análisis de componentes principales de Happ and Greven (2018) se obtienen los análogos de los estadísticos T y se analizan esos resultados.

En la figura 5-19 se muestran las 50 realizaciones de las simulaciones. Para los análisis de componentes funcionales univariados de la propuesta dada en este trabajo y para el procedimiento 1 se utilizan 7 componentes principales para cada una de las 4 variables que suman $d = 28$, para el procedimiento 2 se seleccionan 4 componentes principales, por lo tanto $d = 4$. En todos los casos la variabilidad es mayor a 80%.

Magnitud Cambios v	Propuesta		Procedimiento 1		Procedimiento 2	
	P	R	P	R*	P	R**
0.150	40	S	40	S	40	S
0.100	40	S	39	S	40	S
0.090	40	S	39	S	40	S
0.080	40	S	39	S	40	S
0.070	40	S	39	S	40	S
0.060	40	S	39	S	40	S
0.050	40	S	32	N	39	S
0.040	40	S	32	N	30	N
0.030	40	S	17	N	30	N
0.025	40	S	17	N	30	N

Tabla 5-3: Detección de punto cambio y comparaciones usando simulaciones de Procesos 4-variados con errores como en Happ-Kurz (2020) y ruidos Laplace.

P: Estimación del Punto de cambio. R: ¿Se rechaza la hipótesis Nula con un nivel de significancia del 1%?, *: Se toma la decisión sin transformar los datos. ** Se toma la decisión asumiendo que la propuesta en Berkes et al. (2009) se puede aplicar a los scores del MFPCA de Happ and Greven (2018). 15 funciones de las bases de Fourier fueron seleccionadas para la propuesta y para el procedimiento 1.

5.2. Aplicaciones

La metodología se aplica a tres conjuntos de datos reales en los que la metodología detecta punto de cambio. Los análisis se pueden reproducir en *app análisis*. En Latorre (2019) se encuentran los códigos y los datos ya preparados. Los datos de las siguientes dos aplicaciones fueron extraídos del repositorio Dua and Graff (2017).

5.2.1. Temperaturas Casa Domótica

Los datos fueron recopilados por Zamora-Martinez et al. (2014), de un sistema de monitor montado en una casa domótica, hay más de 40 días de datos de registros, los datos fueron tomados cada 15 minutos. De la base se seleccionaron las variables temperatura exterior y temperatura de una de las habitaciones de la casa. La base estaba dividida en dos partes: La primera con registros del 13 de marzo de 2012 al 11 de abril y la segunda del 18 de abril al 1 de mayo todas del mismo año. Para el análisis realizado se trabaja con los datos del 21 de marzo al 10 de abril y del 18 de abril al 1 de mayo, debido a que los datos de la primera semana presentan varias inconsistencias. En Zamora-Martinez et al. (2014) se indica que los datos provienen de condiciones distintas por lo que el punto de cambio

debería localizarse en el 10 de abril, de acuerdo a la columna de la base de datos, este cambio corresponde al día 21. En la figura 5-20 se muestran las temperaturas externas y de la habitación para cada uno de los 35 días que hacen parte del análisis. Las decisiones son tomadas teniendo como referencia un nivel de significancia del 1%.

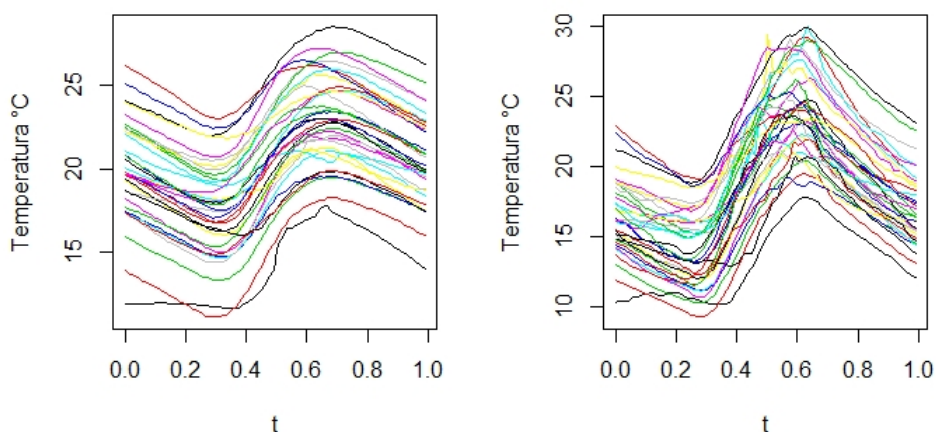


Figura 5-20: Temperaturas casa domótica: a la izquierda datos de temperatura de la habitación a la derecha datos de temperatura externa.

Realizando un análisis de punto de cambio univariado funcional para la temperatura de la habitación, tomando 2 componentes principales se rechaza la hipótesis nula o sea, se detecta punto de cambio pero se estima en el día 20. El mismo análisis para la temperatura externa, arroja con 2 componentes principales el mismo resultado que para la temperatura de la habitación.

Entre las variables temperatura externa y temperatura de la habitación existe correlación, la metodología propuesta en este trabajo la tiene en cuenta para el análisis. En la figura 5-21 se muestra la raíz de las estimaciones de los valores propios de $\mathbf{B}(t)\mathbf{B}(t)^T$ en cada t evidenciando que en ningún instante de tiempo los valores propios son iguales.

Usando la metodología aquí propuesta con 2 armónicos para la primera y 5 para la segunda componente después de la transformación, se rechaza la hipótesis nula y la estimación del punto de cambio es 21 que corresponde al 10 de abril. Para los datos antes y después del punto de cambio se aplica la metodología nuevamente. En la parte antes del 10 de abril se detecta otro punto de cambio en el 31 de marzo. Los demás análisis no revelan más puntos de cambio.

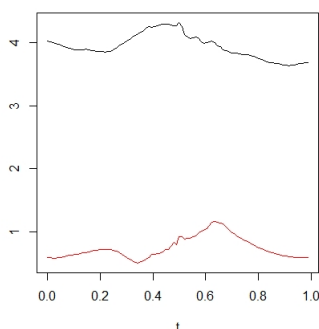


Figura 5-21: Raíz de las estimaciones de los valores propios de $\mathbf{B}(t)\mathbf{B}(t)^T$.

En este caso los dos análisis univariados rechazan la hipótesis nula y coinciden en la estimación del punto de cambio, pero lo subestiman, en contraste con la metodología multivariada que lo detecta correctamente.

5.2.2. Temperaturas Plataforma Hidráulica

En Helwig et al. (2015) desarrollan y evalúan un enfoque sistemático para el entrenamiento automatizado de monitoreo de un complejo sistema hidráulico. Ellos proveen el conjunto de datos que obtuvieron experimentalmente con una plataforma de prueba hidráulica. El sistema tiene ciclos de carga constante en los que mide valores del proceso tales como presiones, flujos de volumen y temperaturas, mientras que la condición de cuatro componentes hidráulicos varía cuantitativamente. Cada dato funcional corresponde a un minuto en el que las medidas son tomadas cada segundo.

El funcionamiento correcto se da cuando los 4 componentes hidráulicos tienen los siguientes valores:

- Enfriador: 100 correspondiente a eficiencia total.
- Válvula: 100 correspondiente a comportamiento óptimo del conmutador.
- Bomba Interna: 0 correspondiente a que no hay fugas.
- Acumulador Hidráulico: 130 correspondiente a presión óptima.

De este conjunto de datos se seleccionaron 10 datos seguidos de correcto funcionamiento y 6 también seguidos en los que hubo un cambio no tan abrupto en la válvula y en la bomba. El análisis de punto de cambio se realiza únicamente con los datos de temperatura

tomados de 4 sensores. En la figura **5-22** se muestra la salida de la aplicación para el análisis de las 4 temperaturas. Las decisiones son tomadas teniendo como referencia un nivel de significancia del 1 %.

	d	s	R	P
Temperatura S1	4	1.734647	S	8
Temperatura S1	5	1.765315	N	8
Temperatura S1	6	1.865012	N	8
Temperatura S2	4	1.35495	N	6
Temperatura S2	5	1.561917	N	6
Temperatura S2	6	1.770321	N	10
Temperatura S3	4	1.467684	N	11
Temperatura S3	5	1.616415	N	11
Temperatura S3	6	1.698793	N	11
Temperatura S4	4	0.4291891	N	9
Temperatura S4	5	0.5571559	N	9
Temperatura S4	6	0.6661212	N	10

Tabla 5-4: Aplicación metodología univariada: Temperaturas sistema hidráulico

d : Número de componentes principales. s : Estadístico de prueba para comparar con los valores de la tabla **4-1**. R : ¿Se rechaza la hipótesis Nula con un nivel de significancia del 1 %?. P : Estimacion del Punto de cambio. 15 funciones de las bases de Fourier fueron seleccionadas.

Como se puede apreciar en la tabla **5-4** con las pruebas univariadas la mejor conclusión sería que no hay punto de cambio. En la tabla **5-5** se presentan los resultados con la metodología propuesta y se puede concluir correctamente que hay punto de cambio en el dato 10.

d_1	d_2	d_3	d_4	s	R	P
3	7	8	7	6.273834	S	10
4	7	8	7	6.559753	S	10
3	8	8	7	6.354078	N*	10
3	7	9	7	6.430983	S	10
3	7	8	8	6.3464	N*	10

Tabla 5-5: Aplicación metodología propuesta: Temperaturas sistema hidráulico

s : Estadístico de prueba para comparar con los valores de la tabla **4-1**. R : ¿Se rechaza la hipótesis Nula con un nivel de significancia del 1 %?, *: el cuantil 0.99 de la distribución K_d con $d = 26$ es 6.393582. P :

Estimacion del Punto de cambio. 15 funciones de las bases de Fourier fueron seleccionadas.

5.2.3. Calidad de aire Bogotá

Se analizan las variables monóxido de nitrógeno NO y dióxido de nitrógeno NO₂ con registros cada hora desde el 3 de febrero hasta el 31 de mayo de 2020. Los datos usados provienen de la estación de Puente Aranda de la Red de Monitoreo de Calidad del Aire de Bogotá. Para imputar los datos faltantes se promedian los datos de la hora posterior y anterior al dato faltante, o se usan los datos a la misma hora y el mismo día de la siguiente y anterior semana.

En el análisis se rechaza la hipótesis Nula con un nivel de significancia del 1 %, 15 funciones de las bases de Fourier fueron seleccionadas para realizar el análisis y fueron seleccionados 6 armónicos en el análisis de componentes principales funcionales para cada una de las dos componentes.

La metodología propuesta en este trabajo detecta cambio en el día 46 que corresponde al 19 de marzo de 2020, un día antes de que iniciará el simulacro de cuarentena en Bogotá que se prolongó por varios meses. En la figura **5-23** se muestra la salida de la aplicación para el análisis de las 2 variables.

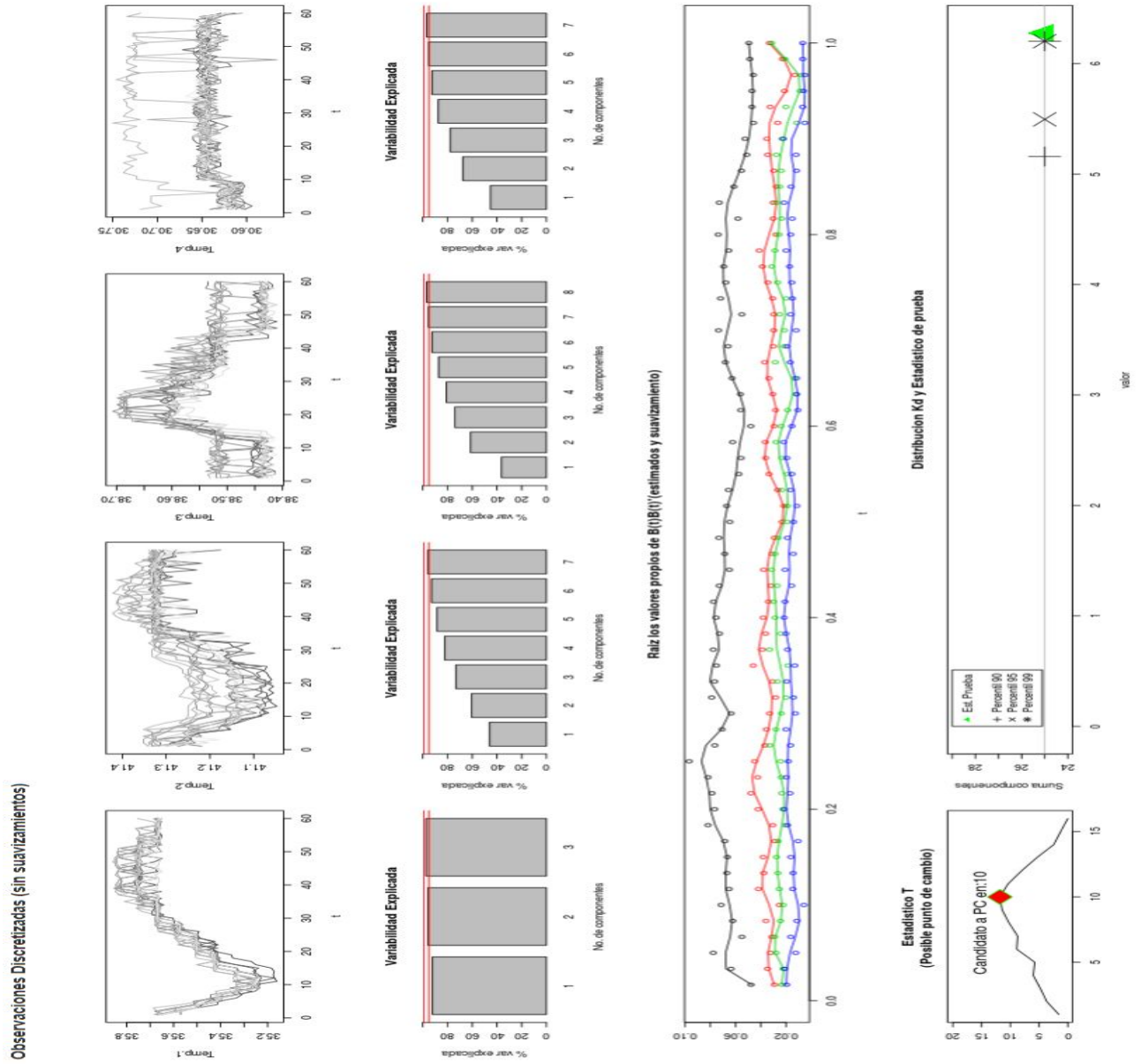


Figura 5-22: Salida gráfica de *app análisis* para análisis de Temperaturas Plataforma Hidráulica con $d_1 = 3, d_2 = 7, d_3 = 8, d_4 = 7$

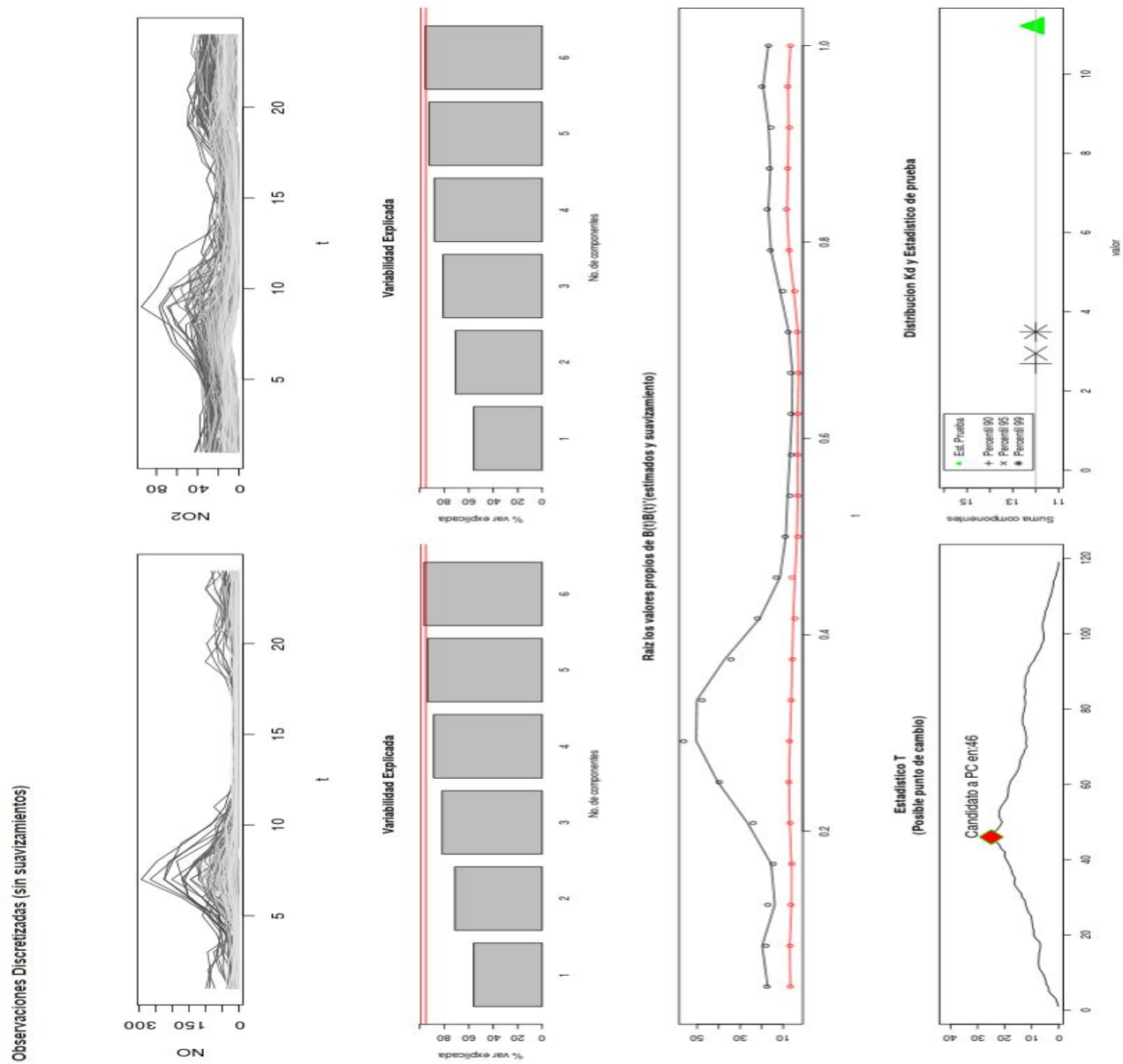


Figura 5-23: Salida gráfica de *app análisis* para análisis de Óxidos de Nitrógeno en la Estación Puente Aranda con $d_1 = d_2 = 6$

6 Conclusiones, recomendaciones y trabajo futuro

En la metodología aquí propuesta para la detección de punto de cambio en datos funcionales multivariados se construye un operador que toma datos funcionales multivariados con dependencia y la elimina, después de este procedimiento a los datos funcionales multivariados sin dependencia se les aplica una adaptación de la metodología usada por Berkes et al. (2009).

6.1. Conclusiones

La metodología propuesta tiene un buen desempeño debido a que considera dos tipos de dependencia; entre los procesos y dentro de cada proceso. Se advierte que, como se evidenció en las simulaciones y las aplicaciones, no es aconsejable aplicar procedimientos univariados funcionales para la detección de punto de cambio a cada una de las componentes de datos funcionales multivariados sin realizar ninguna transformación, debido a que al no considerar la dependencia entre los procesos se pueden generar conclusiones incorrectas, generalmente hay pérdida de sensibilidad.

Las transformaciones de los datos pueden ayudar a identificar características no observadas en los datos sin transformar, aunque en la propuesta de este trabajo las transformaciones son lineales, el hecho de que se realicen a lo largo del tiempo da lugar a que se puedan identificar de manera adecuada los puntos de cambio.

La metodología es capaz de detectar pequeños cambios de forma, traslaciones y en general cambios que afecten a alguna de las funciones que compone cada uno de los procesos multivariados.

Como se observó en el procedimiento 2 de la subsección 5.1.3, el uso de la metodología de Berkes et al. (2009) para la detección de punto de cambio usando los coeficientes resultantes del análisis de componentes principales propuesto por Happ and Greven (2018) no debe ser concluyente.

Tanto la metodología propuesta como las simulaciones quedaron programadas y están disponibles dos apps: la primera para realizar simulaciones de procesos y detectar punto

de cambio *app simulaciones*; y la segunda para realizar un análisis rápido de detección de punto de cambio con la propuesta desde 1 hasta 5 procesos, *app análisis*. Los códigos usados para las simulaciones, implementación de la metodología y las apps se encuentran en Latorre (2019).

6.2. Recomendaciones

Existen varias metodologías para el análisis de punto de cambio para datos funcionales univariados. A continuación, se realizan comentarios sobre el análisis de detección de punto de cambio en datos funcionales y específicamente en datos funcionales multivariados.

Un mal ajuste o un sobreajuste en el proceso de suavizamiento puede afectar los resultados del análisis de detección de cambio. En la subsección 3.1.2 se muestra un procedimiento recomendado para realizar el suavizamiento. En las simulaciones y aplicaciones realizadas en este trabajo se usaron entre 15 y 35 elementos de la base de Fourier y se usó el gcv. Se aconseja incluir un número considerable de elementos de la base seleccionada y el uso de gcv para controlar el sobreajuste.

En la elección del número de componentes principales univariado funcional se debe tener en cuenta que si bien el objetivo es que el porcentaje de variabilidad retenida sea alta, incluir demasiadas componentes puede hacer que en los cálculos se divida entre valores propios muy cercanos a 0 por lo que el análisis no tendría validez. Para elegir el número de componentes principales se recomienda no incluir componentes en las que el porcentaje de variabilidad explicada esté por debajo del 1 %.

Si las observaciones discretizadas tienen datos faltantes para algunos valores de t dentro del intervalo de interés, es posible realizar un suavizamiento con las observaciones que se tienen para imputar los valores de las variables que no fueron tomadas en determinados tiempos.

Hay cambios en la forma de la media que pueden no ser detectados por algunas metodologías, por ejemplo el uso del gcv para controlar el sobreajuste en el suavizamiento puede atenuar cambios de forma que luego no son detectados por metodologías de puntos de cambio como la de Berkes et al. (2009). El enfoque dado en los apéndices 7 puede resultar útil para detectar este tipo de cambios, sin embargo, hay características del operador diferenciable que dejan abierto un estudio a profundidad sobre su uso.

Cuando los datos funcionales multivariados midan diferentes tipos de variable, por ejemplo temperatura y presión como las dos componentes de un proceso, se debe primero estandarizar cada una de las dos medidas. Para el proceso de estandarización se aconseja calcular las medias y las desviaciones fijando el valor de t y hacer tantos procesos de estandarización como valores de t .

6.3. Trabajo futuro

Uno de los aportes de este trabajo es la construcción del espacio mostrado en la sección 4.2.2, si bien en este trabajo el espacio fue usado para la detección de punto de cambio quedan abiertas las posibilidades de usarlo para otros problemas, como por ejemplo, la diferencia de medias en datos funcionales multivariados.

En la propuesta dada en este trabajo para la detección de cambio se uso el espacio definido y la propuesta de Berkes et al. (2009), sin embargo, se puede estudiar la posibilidad de adaptar propuestas como la de Sharipov et al. (2016) al espacio construido.

La propuesta dada en este trabajo asume que la muestra es independiente e idénticamente distribuida, queda abierto un estudio para series de tiempo multivariadas de datos funcionales basado en el espacio aquí planteado incorporando, por ejemplo las ideas en Hörmann et al. (2015).

La detección de cambio en la forma de la media obtenida al aplicar la transformación derivada, deja abiertas dos posibles líneas de estudio, por un lado la posibilidad de que el operador diferenciable pueda actuar de manera correcta para este tipo de problemas y por otro la definición de un operador pseudodiferencial con las características necesarias para implementar de manera correcta una metodología.

A nivel teórico y teniendo en cuenta que una de las motivaciones en la construcción del espacio es el comportamiento de los sistemas dinámicos, queda abierta la posibilidad de un estudio más formal que incorpore la teoría de los sistemas dinámicos y el comportamiento de las variedades diferenciables. También queda abierto el estudio de la teoría asintótica de los estimadores.

El supuesto fundamental para la construcción de la metodología es la continuidad en las componentes principales, queda abierta una investigación que revise, si al aplicar otros tipos de metodologías en cada instante de tiempo como kmeans, los centroides tengan dependencia continua y se pueda aplicar un método de clusterización para datos funcionales multivariados teniendo en cuenta las ideas desarrolladas en este trabajo.

7 Apéndice

Como se mencionó en la sección 2.4 una de las características deseadas en el manejo con operadores es que sean acotados. El operador diferencial es un operador no acotado, por esta razón no se realiza, al menos en este trabajo, un estudio formal de este tema. En la figura 7-1 se muestra una función cuya norma es menor que la norma de su derivada. Las funciones periódicas con altas frecuencias generan problema para el análisis usando derivadas ya que sus normas no pueden acotar las de sus derivadas.

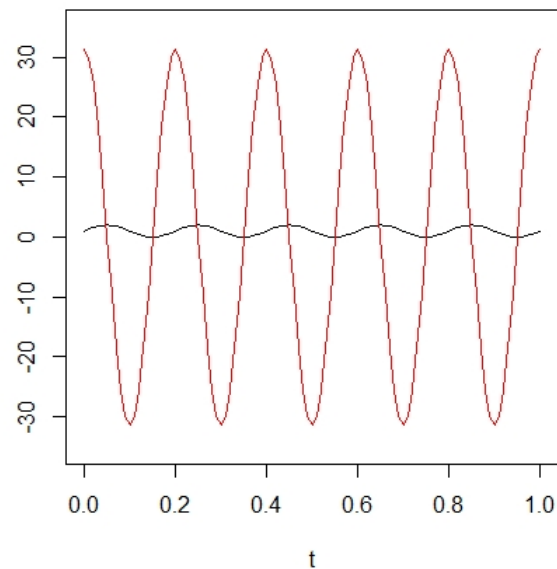


Figura 7-1: Función $\sin(10\pi t) + 1$ en negro y su derivada en rojo

Se presenta un procedimiento que puede ayudar a detectar cambios de forma en la media funcional, sin embargo, el hecho de que el operador diferenciable no sea acotado puede aumentar de manera exagerada diferencias que no necesariamente se deban a cambios de forma en la media. Se presenta un análisis para la detección de punto de cambio usando la derivada pero no se demostró que su uso sea correcto.

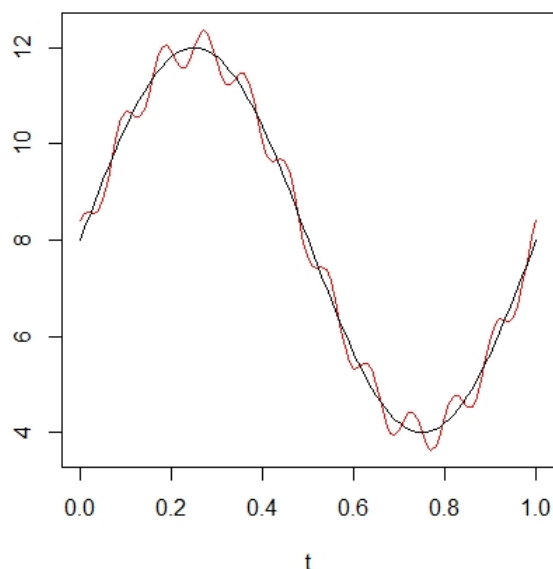


Figura 7-2: Función media en negro antes del cambio y en rojo después

Nota 7.0.1 (Procedimiento para el uso de información en las derivadas) Si para el suavizamiento se selecciona la base de Fourier es fácil encontrar la forma en la que actúa el operador derivada. En Latorre (2019) hay una función que realiza la transformación lineal del operador diferenciable así como el análisis mostrado más adelante. En la app app simulaciones hay una opción para analizar el uso de las derivadas.

1. Se realiza el suavizamiento de los datos discretizados usando la base de Fourier.
2. Teniendo en cuenta que la base de Fourier está compuesto por funciones senos y cosenos, las derivadas de los elementos de la base alterna de un elemento a otro multiplicado por una constante.
3. A los valores suavizados se les aplica la transformación lineal de la derivada.
4. Con los coeficientes después de la transformación se realiza el análisis de punto de cambio, si las funciones son diferenciables es válido el uso de la metodología en Berkes et al. (2009) ya que las funciones derivadas estarían en $\mathbb{L}^2[0, 1]$.

El procedimiento anterior puede ser usado en conjunto con la metodología propuesta.

Se simularon 159 procesos y se incluyó un cambio en la media en el dato 95, la metodología expuesta en Berkes et al. (2009) no detecta el cambio y al añadir la información de la

derivada se encuentra el punto de cambio. Las funciones medias antes y después del cambio son respectivamente

$$\mu_1(t) = 4 \sin(2\pi t) + 8 \quad \mu_2(t) = 4 \sin(2\pi t) + 8 + 0.4 \cos(22\pi t)$$

en la figura **7-2** se aprecia que el cambio está en la forma por lo que es posible que algunos métodos para punto de cambio no detecten este tipo de cambios.

Con los datos sin aplicar la transformación de la derivada, el estadístico s con $d = 12$ tiene un valor de 2.13 que implica que no se detecta punto de cambio, el máximo de T se encuentra en 46 muy lejos del punto de cambio en 95. Con la información de las derivadas, el estadístico s con $d = 15$ tiene un valor de 13.356 que implica que se rechaza la hipótesis nula y estima correctamente el punto de cambio en 95.

Bibliografía

- Anderson, T. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics.
- Arlot, S., Celisse, A., and Harchaoui, Z. (2019). A kernel multiple change-point algorithm via model selection. *Journal of Machine Learning Research*, 20(162):1–56.
- Aston, J. A. and Kirch, C. (2011). Estimation of the distribution of change-points with application to fmri data.
- Bardsley, P., Horváth, L., Kokoszka, P., and Young, G. (2017). Change point tests in functional factor models with application to yield curves. *The Econometrics Journal*, 20(1):86–117.
- Berkes, I., Gabrys, R., Horváth, L., and Kokoszka, P. (2009). Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(5):927–946.
- Berrendero, J. R., Bueno-Larraz, B., and Cuevas, A. (2020). On mahalanobis distance in functional settings. *Journal of Machine Learning Research*, 21(9):1–33.
- Berrendero, J. R., Justel, A., and Svarc, M. (2011). Principal components for multivariate functional data. *Computational Statistics & Data Analysis*, 55(9):2619–2634.
- Chen, J. and Gupta, A. K. (2011). *Parametric statistical change point analysis: with applications to genetics, medicine, and finance*. Springer Science & Business Media.
- Chiou, J.-M., Chen, Y.-T., and Yang, Y.-F. (2014). Multivariate functional principal component analysis: A normalization approach. *Statistica Sinica*, pages 1571–1596.
- Dua, D. and Graff, C. (2017). UCI machine learning repository.
- Gardner, L. (1969). On detecting changes in the mean of normal variates. *The Annals of Mathematical Statistics*, 40(1):116–126.
- Garreau, D. (2017). *Change-point detection and kernel methods*. PhD thesis.

- Grines, V. Z., Medvedev, T. V., and Pochinka, O. V. (2016). *Dynamical systems on 2-and 3-manifolds*, volume 46. Springer.
- Hall, B. C. (2013). *Lie Groups, Lie Algebras, and Representations*. Springer.
- Happ, C. and Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association*, 113(522):649–659.
- Happ-Kurz, C. (2020). Object-oriented software for functional data. *Journal of Statistical Software*, 93(5):1–38.
- Hawkins, D. M., Qiu, P., and Kang, C. W. (2003). The changepoint model for statistical process control. *Journal of quality technology*, 35(4):355–366.
- Helwig, N., Pignanelli, E., and Schütze, A. (2015). Condition monitoring of a complex hydraulic system using multivariate statistics. In *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, pages 210–215. IEEE.
- Hörmann, S., Kidziński, L., and Hallin, M. (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(2):319–348.
- Horváth, L. and Kokoszka, P. (2012). *Inference for functional data with applications*, volume 200. Springer Science & Business Media.
- Hotelling, H. (1947). Multivariate quality control. techniques of statistical analysis. *McGraw-Hill, New York*.
- Hsing, T. and Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*. John Wiley & Sons.
- Huang, S., Kong, Z., and Huang, W. (2014). High-dimensional process monitoring and change point detection using embedding distributions in reproducing kernel hilbert space. *IIE Transactions*, 46(10):999–1016.
- Jacques, J. and Preda, C. (2014). Model-based clustering for multivariate functional data. *Computational Statistics & Data Analysis*, 71:92–106.
- Kiefer, J. (1959). K-sample analogues of the kolmogorov-smirnov and cramér-v. mises tests. *The Annals of Mathematical Statistics*, pages 420–447.

- Latorre, D. (2019). Code change point for mfa thesis. https://github.com/dalatorrem/code_tesis_PCMFD.git.
- Lee, T.-S. (2010). Change-point problems: bibliography and review. *Journal of Statistical Theory and Practice*, 4(4):643–662.
- Page, E. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3/4):523–527.
- Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41(1/2):100–115.
- Plummer, P. J. and Chen, J. (2014). A bayesian approach for locating change points in a compound poisson process with application to detecting dna copy number variations. *Journal of Applied Statistics*, 41(2):423–438.
- Ramsay, J. O. (2004). Functional data analysis. *Encyclopedia of Statistical Sciences*, 4.
- Saitoh, S. and Sawano, Y. (2016). *Theory of reproducing kernels and applications*. Springer.
- Sen, A. K. and Srivastava, M. S. (1973). On multivariate tests for detecting change in mean. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 173–186.
- Sharipov, O., Tewes, J., and Wendler, M. (2016). Sequential block bootstrap in a hilbert space with application to change point analysis. *Canadian Journal of Statistics*, 44(3):300–322.
- Skubalska-Rafajłowicz, E. (2013). Random projections and hotelling’s t_2 statistics for change detection in high-dimensional data streams. *International Journal of Applied Mathematics and Computer Science*, 23(2):447–461.
- Stoehr, C., Aston, J. A., and Kirch, C. (2020). Detecting changes in the covariance structure of functional time series with application to fmri data. *Econometrics and Statistics*.
- Vasudeva, H. L. and Shirali, S. (2017). *Elements of Hilbert spaces and operator theory*. Springer.
- Zamba, K. and Hawkins, D. M. (2009). A multivariate change-point model for change in mean vector and/or covariance structure. *Journal of Quality Technology*, 41(3):285–303.
- Zamora-Martinez, F., Romeu, P., Botella-Rocamora, P., and Pardo, J. (2014). On-line learning of indoor temperature forecasting models towards energy efficiency. *Energy and Buildings*, 83:162–172.