



**ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO**  
**FACULTAD DE CIENCIAS**  
**CARRERA MATEMÁTICA**

**DESCRIPCIÓN DEL MÉTODO MEAN SHIFT PARA ANÁLISIS DE  
DATOS ESPACIALES**

**Trabajo de Integración Curricular**

Tipo: Proyecto de Investigación

Presentado para optar al grado académico de:

**MATEMÁTICO**

**AUTOR:** KEVIN PAUL BASANTES ASQUI

**DIRECTOR:** Ing. ESTEFANIA CRISTINA RAMOS ARAUJO , Msc.

Riobamba – Ecuador

2024

**©2024, Kevin Paul Basantes Asqui**

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento, siempre y cuando se reconozca el Derecho de Autor.

Yo, Kevin Paul Basantes Asqui, declaro que el presente Trabajo de Integración Curricular es de mi autoría y los resultados del mismo son auténticos. Los textos en el documento que provienen de otras fuentes están debidamente citados y referenciados.

Como autor asumo la responsabilidad legal y académica de los contenidos de este Trabajo de Integración Curricular; El patrimonio intelectual pertenece a la Escuela Superior Politécnica de Chimborazo.

Riobamba, 28 de mayo de 2024

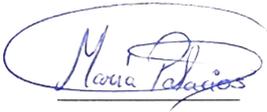
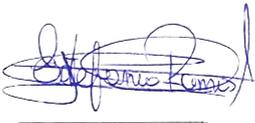


**Kevin Paul Basantes Asqui**

**060486060-1**

**ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO**  
**FACULTAD DE CIENCIAS**  
**CARRERA MATEMÁTICA**

El Tribunal del Trabajo de Integración Curricular certifica que: el Trabajo de Integración Curricular; Tipo: Proyecto de Investigación. **DESCRIPCIÓN DEL MÉTODO MEAN SHIFT PARA ANÁLISIS DE DATOS ESPACIALES**, realizado por el señor: **KEVIN PAUL BASANTES ASQUI**, ha sido minuciosamente revisado por los Miembros del Tribunal del Trabajo de Integración Curricular, el mismo que cumple con los requisitos científicos, técnicos, legales, en tal virtud el Tribunal Autoriza su presentación.

	<b>FIRMA</b>	<b>FECHA</b>
Ing. María de Lourdes Palacios Robalino <b>PRESIDENTE DEL TRIBUNAL</b>		2024-05-28
Ing. Estefanía Cristina Ramos Araujo, Mgs. <b>DIRECTOR DEL TRABAJO DE INTEGRACIÓN CURRICULAR</b>		2024-05-28
Ing. Alex Eduardo Pozo Valdiviezo, Mgs. <b>ASESOR DEL TRABAJO DE INTEGRACIÓN CURRICULAR</b>		2024-05-28

## **DEDICATORIA**

A mis padres Llaned Asqui y Eudoro Basantes que son mi principal apoyo y admiración, por su amor incondicional y sacrificios silenciosos que me enseñaron la importancia de la perseverancia y humildad. A mi hermano Rony que ha estado ahí acompañándome en todo momento y decisiones en mi etapa universitaria.

*Kevin*

## **AGRADECIMIENTO**

Agradezco a mis padres y hermano por estar conmigo siempre, enseñándome lo importante de lograr los objetivos con sacrificio y dedicación, y por su continuo apoyo en este proceso de alcanzar un sueño más en mi vida académica.

Un sincero agradecimiento a todas las personas que contribuyeron a la realización de este proyecto de investigación. A la directora, Ing. Cristina Ramos, al asesor, Ing. Alex Pozo, y a la Biof. Isabel Balvoa, por su compromiso, conocimiento y orientación a lo largo de todo el proceso de elaboración de este trabajo.

A mis profesores, cuyas palabras de sabiduría estimularon la curiosidad por el conocimiento.

Agradezo a todos con los que compartí estos años, tardes de estudios entre risas y teoremas. A mi pareja y amigos BES-K que estuvieron apoyándome en este largo viaje.

*Kevin*

## ÍNDICE DE CONTENIDOS

ÍNDICE DE TABLAS . . . . .	viii
ÍNDICE DE ILUSTRACIONES . . . . .	ix
ÍNDICE DE ANEXOS . . . . .	xi
RESUMEN . . . . .	xii
ABSTRACT . . . . .	xiii
INTRODUCCIÓN . . . . .	1
<b>CAPÍTULO I</b>	
<b>4. PROBLEMA DE INVESTIGACIÓN . . . . .</b>	<b>3</b>
1.1. Planteamiento del problema . . . . .	3
1.2. Objetivos . . . . .	3
1.2.1. <i>Objetivo general</i> . . . . .	3
1.2.2. <i>Objetivos específicos</i> . . . . .	3
1.3. Justificación . . . . .	4
<b>CAPÍTULO II</b>	
<b>5. MARCO TEÓRICO . . . . .</b>	<b>5</b>
2.1. <i>Análisis cluster</i> . . . . .	5
2.2. Mean Shift . . . . .	5
2.2.1. <i>Windows Parzen</i> . . . . .	5
2.2.2. <i>Estimador de densidad kernel univariado</i> . . . . .	6
2.2.3. <i>Funciones kernel (núcleo)</i> . . . . .	6
2.2.3.1. <i>Tipos de kernel</i> . . . . .	6

## CAPÍTULO III

6.	MARCO METODOLÓGICO . . . . .	8
3.1.	Descripción de enfoque, alcance, diseño, tipo, métodos, técnica e instrumentos de investigación. . . . .	8

## CAPÍTULO IV

7.	MARCO DE ANÁLISIS E INTERPRETACIÓN DE RESULTADOS . . . . .	9
4.1.	Resultado . . . . .	9
4.2.	Resultado de la Implementación Región Andina del Ecuador . . . . .	10
4.2.1.	<i>Base de datos</i> . . . . .	10
4.2.2.	<i>Implementación del método</i> . . . . .	10
4.2.2.1.	<i>Elección del ancho de banda (bandwidth)</i> . . . . .	11
4.2.3.	<i>Análisis por clusters</i> . . . . .	12
4.2.4.	<i>Resultados de la implementación en una estación meteorológica</i> . . . . .	15
4.2.5.	<i>Implementación del método</i> . . . . .	15
4.2.6.	<i>Análisis por cluster</i> . . . . .	17

## CAPÍTULO V

8.	CONCLUSIONES Y RECOMENDACIONES . . . . .	21
5.1.	Conclusiones . . . . .	21
5.2.	Recomendaciones . . . . .	21

## BIBLIOGRAFÍA

## ANEXOS

## ÍNDICE DE TABLAS

<b>Tabla 4-1:</b> Valores del parámetro <i>bandwidth</i> (I) . . . . .	11
<b>Tabla 4-2:</b> Grupos formados a partir del ancho de banda . . . . .	11
<b>Tabla 4-3:</b> Frecuencia de datos por cada cluster . . . . .	12
<b>Tabla 4-4:</b> Valores del parámetro <i>bandwidth</i> (II) . . . . .	15
<b>Tabla 4-5:</b> Grupos formados a partir del ancho de banda . . . . .	16
<b>Tabla 4-6:</b> Frecuencia en datos de cada cluster . . . . .	17
<b>Tabla 4-7:</b> Análisis por cada cluster <i>bandwidth</i> . . . . .	17

## ÍNDICE DE ILUSTRACIONES

<b>Ilustración 4-1:</b> Proyección climática CNRM para la región Andina del Ecuador en el período 2020-2050. . . . .	10
<b>Ilustración 4-2:</b> Relación existente del valor del <i>bandwidth</i> y el número de clusters. . . . .	11
<b>Ilustración 4-3:</b> Agrupación resultante con un <i>bandwidth</i> de 1.38. . . . .	12
<b>Ilustración 4-4:</b> Distribución cluster 0. . . . .	13
<b>Ilustración 4-5:</b> Distribución cluster 1. . . . .	13
<b>Ilustración 4-6:</b> Distribución cluster 2. . . . .	14
<b>Ilustración 4-7:</b> Distribución cluster 3. . . . .	14
<b>Ilustración 4-8:</b> Estimación mediante un <i>kernel</i> normal. . . . .	15
<b>Ilustración 4-9:</b> Relación <i>Bandwidth-Cluster</i> . . . . .	16
<b>Ilustración 4-10:</b> Temperatura mínima. . . . .	17
<b>Ilustración 4-11:</b> Serie temporal cluster 0. . . . .	18
<b>Ilustración 4-12:</b> Serie temporal cluster 1. . . . .	18
<b>Ilustración 4-13:</b> Serie temporal cluster 2. . . . .	19
<b>Ilustración 4-14:</b> Serie temporal cluster 3. . . . .	19
<b>Ilustración 4-15:</b> Serie temporal cluster 4. . . . .	20

## **ÍNDICE DE ANEXOS**

**ANEXO A: MONOGRAFÍA “DESCRIPCIÓN DEL MÉTODO MEAN SHIFT PARA ANÁLISIS DE DATOS ESPACIALES”**

## RESUMEN

El Grupo de Investigación de Energías Alternativas y Ambiente (GEAA) de la Escuela Superior Politécnica de Chimborazo (ESPOCH), actualmente trabaja en el agrupamiento de datos meteorológicos a fin de comprender los patrones climáticos de la región Andina del Ecuador. En este contexto, dentro de la ESPOCH, no se encuentra documentación accesible y en detalle acerca de un método de agrupación específico para este propósito. Por tanto, este estudio tuvo como objetivo realizar una investigación documental de carácter teórico-práctico, en el cual se describa el método de agrupación Mean Shift, sus fundamentos teóricos y su implementación en datos espaciales de temperatura en la región Sierra del Ecuador. Con esta finalidad, se utilizó un diseño metodológico de tipo cualitativo y cuantitativo con alcance descriptivo. En consecuencia, se obtuvo un documento bibliográfico el cual está constituido por 3 capítulos, en el primer capítulo se describió herramientas del Análisis Funcional y Teoría de Probabilidad como la técnica de Windows Parzen y el estimador de la función de densidad, en el segundo capítulo se estableció la estimación de la función de densidad multivariante y la localización de máximos a través del gradiente de densidad, y los criterios de convergencia mediante teoremas. Por último, en el tercer capítulo se implementó el método Mean shift a la variable temperatura mínima del modelo de Circulación atmosférica CNRM en un periodo 2020-2050 mediante la biblioteca Sklearn en Python. Los clústeres obtenidos reflejan una dependencia directa con la elección del ancho de banda. Se recomienda continuar con la teoría matemática de la elección del bandwidth óptimo, siendo este un campo que se encuentra en investigaciones activas.

**Palabras clave:** <MEAN SHIFT>, <AGRUPAMIENTO>, <DATOS TEMPERATURA MÍNIMA>, <FUNCIONES NÚCLEO>, <NO PARAMÉTRICO>

0616-DBRA-UPT-2024



## ABSTRACT

The Alternative Energies and Environment Research Group (GEAA) belonging to Escuela Superior Politécnica de Chimborazo (ESPOCH) is currently working on the clustering of meteorological data in order to understand the weather patterns of the Andean region of Ecuador. In this context, within ESPOCH, there is no accessible and detailed documentation on a specific clustering method for this purpose. Therefore, the aim of this study was to carry out a theoretical-practical documentary research, in which the Mean Shift clustering method, its theoretical foundations and its implementation in spatial temperature data in the Sierra region of Ecuador are described. Thus, a qualitative and quantitative methodological design with descriptive scope was used. Consequently, a bibliographic document containing 3 chapters was obtained, the first chapter described tools of Functional Analysis and Probability Theory such as the Windows Parzen technique and the density function estimator, the second chapter established the estimation of the multivariate density function and the localization of maxima through the gradient of a density function, and the convergence criteria by means of theorems. Finally, in the third chapter, the Mean shift method was implemented to the minimum temperature variable of the CNRM Atmospheric Circulation model during the 2020-2050 period, using the Sklearn library in Python. The obtained clusters reflect a direct dependence with the bandwidth. It is recommended to carry on with the research on mathematical theory of the optimal bandwidth choice, since it is a being a field which is under active research.

**Keywords:** <MEAN SHIFT>, <CLUSTERING>, <MINIMUM TEMPERATURE DATA>, <KERNEL FUNCTIONS>, <NON-PARAMETRIC>.



Lic. Paul Rolando Armas Pesántez. Mgs

060328987-7

## INTRODUCCIÓN

El *clustering* es el proceso no supervisado de agrupar una cantidad de observaciones de un mismo conjunto de datos con características similares. Se han diseñado y probado distintas técnicas de agrupamiento para una variedad de aplicaciones en distintas disciplinas científicas como: Biología, Medicina, Visión por computadora, Meteorología, entre otras. La técnica más reconocida es el K-means que efectúa agrupamientos con el objetivo de dividir un conjunto de datos en  $k$  grupos. Igualmente, existen otros tipos de agrupamiento como por ejemplo el método Mean Shift el cual tiene como base la localización de máximos de una función de densidad de probabilidad dado un conjunto de datos muestreados de esa función, generando así conjuntos de grupos que contienen una densidad alta, separados de otros grupos con densidad menor, (Fukunaga y Hostetler, 1995, págs. 32-40).

Esta técnica es un proceso recursivo que identifica grupos que tienen mayor frecuencia de una función de densidad mediante optimización de la misma, estimando la función de densidad a través de una función *kernel* (Cheng, 1995, págs. 790-799); (Fashing y Tomasi, 2005, págs. 471-47). El Mean Shift posee fundamentos matemáticos como el *kernel* y la técnica *Parzen Window*, que involucran su optimización y no hace suposiciones sobre la forma de las agrupaciones ni sobre un número definido de ellos.

El Grupo de Investigación de Energías Alternativas y Ambiente (GEAA) se dedica actualmente a analizar patrones climáticos mediante técnicas de agrupamiento de datos como K-means, Affinity Propagation, Gaussian Mixture y Agrupamiento por Conglomerados en la región principalmente en la provincia de Chimborazo. Sin embargo, los resultados de estos métodos no se interpretan con facilidad debido a la variación topográfica de la región.

En este contexto, se abordó la agrupación de datos espaciales de temperatura correspondientes a la región Andina del Ecuador para la comprensión de los patrones climáticos y variaciones térmicas en la región, permitiendo una identificación más precisa de las zonas con características climáticas de las zonas más similares

Este documento propone un trabajo de investigación con el objetivo de profundizar en los fundamentos teóricos del método Mean Shift y su implementación en datos meteorológicos como la temperatura. El resultado es una monografía que sirve como recurso comprensible para estudiantes de la carrera de matemáticas interesados en profundizar sus conocimientos en la rama de la matemática aplicada, proporcionando una base sólida de información sobre este tema específico.

Este proyecto de investigación consta de cinco capítulos. En el primer capítulo, se exploran

tanto el planteamiento como la justificación del problema, acompañados de los objetivos de la investigación. El segundo capítulo se dedica al marco teórico, proporcionando al lector los conocimientos esenciales para comprender el método Mean Shift. El tercer capítulo ofrece detalles sobre la investigación llevada a cabo y la metodología aplicada durante el desarrollo del trabajo. En el cuarto capítulo, se expone el resultado obtenido. Finalmente, las conclusiones y recomendaciones se presentan en el capítulo cinco como apartado final de la investigación.

# CAPÍTULO I

## 1. PROBLEMA DE INVESTIGACIÓN

### 1.1. Planteamiento del problema

El Grupo de Investigación de Energías Alternativas y Ambiente (GEAA) de la Escuela Superior Politécnica de Chimborazo (ESPOCH) trabaja actualmente en el agrupamiento de datos para la comprensión de los patrones climáticos en la provincia de Chimborazo, Ecuador. Para esto se han utilizado distintos métodos de agrupamiento como el método K-means, *Affinity propagation*, *Gaussian Mixture* y el Agrupamiento por Conglomerados (*Agglomerative clustering*). Los resultados obtenidos con estos métodos no facilitan una adecuada interpretación de los datos, debido a que no consideran la variación topográfica de la provincia de Chimborazo, afectando así a la distribución espacial de los datos; además, las limitaciones de los algoritmos ya empleados para un conjunto masivo de datos.

Por lo dicho anteriormente, se considera el método *Mean-Shift*, porque resulta ser más eficaz para datos espaciales ya que se basa en una estimación de densidad por medio de una función *kernel*. Sumado a esto, se evidencia la carencia de información detallada y accesible acerca del método *Mean Shift* dentro de la carrera de Matemática en la ESPOCH. Por lo tanto, abordar este problema de falta de documentación se vuelve esencial para aprovechar todo el potencial que el método *Mean Shift* podría ofrecer en diversas áreas de investigación y aplicaciones tecnológicas.

### 1.2. Objetivos

#### 1.2.1. *Objetivo general*

Realizar una investigación documental de carácter teórico-práctico sobre los fundamentos teóricos del método de agrupamiento Mean-Shift, mediante una revisión del material bibliográfico especializado, el cual, además será implementado en Phyton para hacer una simulación con el conjunto de datos de temperatura.

#### 1.2.2. *Objetivos específicos*

- Determinar los fundamentos teóricos que sustentan el método de agrupamiento Mean Shift.
- Establecer relación sistémica entre los elementos matemáticos y la teoría del método Mean Shift

para una descripción detallada.

- Implementar en el lenguaje de programación Python el método Mean Shift para datos espaciales de temperatura en la región Sierra del Ecuador.
- Redactar una monografía acerca de los fundamentos teóricos del método Mean Shift para los estudiantes de la carrera de matemática, según la metodología de la Organización del Bachillerato Internacional.

### **1.3. Justificación**

Se plantea la necesidad de realizar una investigación teórica-práctica del método de agrupamiento Mean shift debido a la importancia del análisis de datos espaciales en diversas disciplinas como la geografía, la planificación urbana, la meteorología, la ingeniería, etc. En el campo meteorológico, este método permite la identificación de patrones de comportamiento y la frecuencia de aparición de condiciones climáticas. En este escrito se estudiará las bases y fundamentos teóricos del método, enfatizando su utilidad en aplicaciones reales que facilitarán la implementación y la selección del parámetro óptimo en el lenguaje de programación Python. Por tanto, este trabajo de titulación podrá servir como un documento referencial para los estudiantes de matemática, investigadores, y docentes, que estén interesados y se inicien en este tema. También permitirá a que se adopten nuevas investigaciones en torno a la mejora y optimización del método, lo que puede contribuir al desarrollo de la teoría de agrupamiento y análisis espacial, potenciando así, la formación académica de los estudiantes interesados en la matemática aplicada.

## CAPÍTULO II

### 2. MARCO TEÓRICO

En esta sección, se presentan los fundamentos matemáticos y conceptuales detrás del método de agrupamiento Mean Shift y su funcionamiento como parte del análisis *cluster*

#### 2.1. Análisis cluster

El Análisis Cluster es un término con el que se conoce al agrupamiento que se realiza a un conjunto de datos a través de métodos de aprendizaje no supervisado con la idea de reorganizarlas en grupos a los que se denomina *clusters*. El propósito es organizar las observaciones identificando similitudes entre los datos y agruparlos en función de su proximidad compartiendo características entre sí. (Everitt et al, 2011, págs. 1-13) Existen distintos métodos de agrupamiento como K-means, Affinity Propagation, Gaussian Mixture y Agrupamiento por conglomerados entre los más reconocidos, siendo K-means el más utilizado. De igual forma existe otro método que no hace suposiciones ni sobre la forma de las agrupaciones ni un número definido de ellos conocido como el Mean Shift.

#### 2.2. Mean Shift

El método Mean Shift es una técnica empleada en el procesamiento de imágenes y análisis de datos espaciales que se utiliza en el agrupamiento de conjuntos de datos. Fue introducido originalmente por Fukunaga y Hostetler en 1975, y posteriormente popularizado por Comaniciu y Meer en 2002. A su vez, su funcionamiento se basa en la localización de modas de una función de distribución de probabilidad mediante funciones *kernel*, además, los fundamentos teóricos sobre los que reposa el método Mean Shift se basa en la idea de la técnica no paramétrica conocida como *Windows Parzen* (Duda y Hart, 2001, págs. 161-172).

##### 2.2.1. Windows Parzen

Es una técnica que se utiliza para estimar la densidad de probabilidad de una variable aleatoria continua, expresada como:

$$f(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{V_n} \varphi\left(\frac{x-x_i}{h_n}\right),$$

donde:  $V_n$  es el volumen de la región,  $h_n$  el parámetro  $h$  y  $\varphi$  es una función de ventana. Teniendo una noción de estimador en cuanto a un *kernel* de densidad univariado

### 2.2.2. Estimador de densidad kernel univariado

El estimador de densidad *kernel* viene dado por:

$$\hat{f}(x;h) = (nh)^{-1} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)$$

aquí,  $K$  es una función que satisface  $\int K(x)dx = 1$ , denominada *kernel*, y  $h$  es un número positivo denominado como *bandwidth or window width*.

### 2.2.3. Funciones kernel (núcleo)

Una función *kernel*, en el caso unidimensional, es una función de variable real:  $K : \mathbb{R} \rightarrow \mathbb{R}$  con las siguientes propiedades.

- $\forall x \in \mathbb{R}, K(x) \in [0, \infty)$
- $\forall x \in \mathbb{R}, \int_{-1}^1 K(x)dx = 1$
- $\forall x \in \mathbb{R}, K(x), x \in [-1, 1]$
- $\forall x \in \mathbb{R}, \int_{-1}^1 xK(x)dx = 0$
- $\forall x \in \mathbb{R}, K(x) = K(-x)$
- $\forall x \in \mathbb{R}, \int_{-1}^1 x^2 K(x)dx \in \mathbb{R}^+$

#### 2.2.3.1. Tipos de kernel

- *Kernel* Rectangular o Uniforme

$$K(x) = 0.5, \quad x \in [-1, 1].$$

- *Kernel* Triangular

$$K(x) = 1 - |x|, \quad x \in [-1, 1].$$

- *Kernel* de Epanechnikov

$$K(x) = \frac{3}{4}(1 - x^2), \quad x \in [-1, 1].$$

- *Kernel* Normal o Gaussiano

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$

- *Kernel Biweight* o Cuártico

$$K(x) = \frac{15}{16} (1 - x^2)^2, \quad x \in [-1, 1].$$

- *Kernel Arco coseno*

$$K(x) = \frac{\pi}{4} \cos\left(\frac{\pi}{2}x\right), \quad x \in [-1, 1].$$

dentro de los distintos tipos de *kernel*, se destacando el uso del *kernel* Gaussiano con la finalidad de una normalización constante de la función *kernel* para que se integre a uno. El perfil de núcleo Gaussiano es de la forma:  $e^{-\frac{1}{2}x^2\sigma}$ . (Comaniciu y Meer,2002, págs. 603-619)

## CAPÍTULO III

### 3. MARCO METODOLÓGICO

#### 3.1. Descripción de enfoque, alcance, diseño, tipo, métodos, técnica e instrumentos de investigación.

En el desarrollo de este trabajo se empleó una metodología enmarcada en un enfoque mixto, integrando aspectos cualitativos y cuantitativos para ofrecer una descripción adecuada del método Mean Shift. Tratando con un estudio teórico profundo del tema, así como práctico, mediante su implementación en Python para los resultados correspondientes a la problemática planteada.

Por un lado, aludiendo el enfoque cualitativo, el trabajo involucró una revisión selectiva y pertinente de fuentes primarias y secundarias relacionados al método Mean Shift. A su vez, se dió la descripción teórica de los fundamentos y estructura del método, lo que nos orientó a una metodología con diseño documental y alcance descriptivo, enmarcados en una monografía desarrollada mediante el editor de texto  $\text{\LaTeX}$ . Por otro lado, su aspecto cuantitativo vino dado por la implementación, comprensión y el análisis de los resultados del comportamiento de la temperatura mínima en la región Andina del Ecuador. Teniendo en cuenta la no manipulación de los datos, se otorgó un diseño no experimental con alcance explicativo, dada la comprensión de las características principales que explican el comportamiento de los patrones climáticos, a través del agrupamiento o clustering.

#### **Las actividades siguientes fueron parte del desarrollo:**

1. Selección de las fuentes documentales pertinentes y de mayor interés relacionados con el tema.
2. Revisión de la bibliografía acerca del tema en estudio, para su entendimiento y comprensión.
3. Análisis descriptivo relacionado con el método Mean-Shift.
4. Implementación del método Mean Shift para la comprensión de los patrones climáticos utilizando el lenguaje de programación Python.
5. Análisis detallado de los resultados para identificar patrones climáticos significativos.
6. Consolidación del documento de titulación y de la monografía dirigida a los estudiantes de la Carrera de Matemática.

## CAPÍTULO IV

### 4. MARCO DE ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

#### 4.1. Resultado

Este trabajo de investigación se centra en un enfoque cualitativo/cuantitativo que utiliza un diseño documental descriptivo no experimental. Su resultado es una monografía titulada "Descripción del método Mean Shift para análisis de datos espaciales ". El objetivo principal de este documento es proporcionar a los lectores una comprensión más clara de la teoría matemática respecto al método en estudio, así como de su implementación en Python con datos de temperatura mínima de la región Sierra del Ecuador, para una mejor comprensión del tema.

La monografía en mención está conformada por tres capítulos en los que se aborda tanto la descripción como la implementación del método. A continuación, se detalla cada uno de ellos:

- **Capítulo I: fundamentos teóricos**

En este capítulo se encuentran las herramientas esenciales para comprender el camino que toma el método Mean Shift para su funcionamiento.

- **Capítulo II: Mean shift**

La teoría que sustenta al método se sitúa en este capítulo, se explora en detalle los conceptos y principios matemáticos que respaldan este algoritmo de clustering como el cálculo de densidad y la búsqueda de las modas, y la relación que existe con la identificación de los grupos en los datos. Además, se abordan las expresiones fundamentales que conducen la convergencia del algoritmo.

- **Capítulo III: Implementación y resultados**

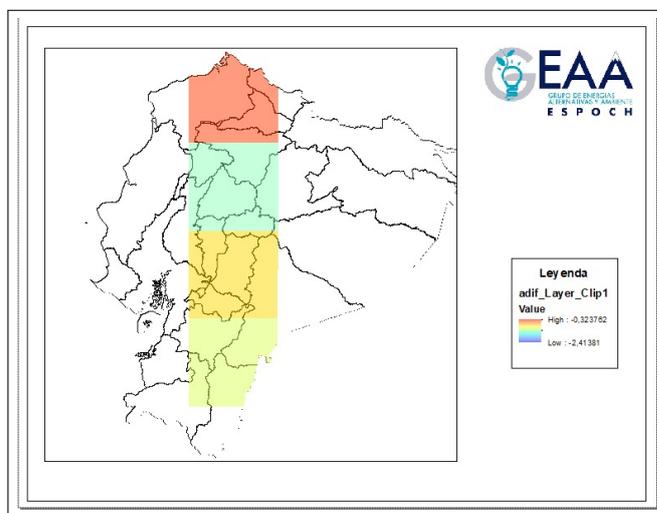
En este capítulo, se aborda la fase de implementación del método Mean Shift en Python y se analiza los resultados obtenidos. Aquí se observa cómo este algoritmo de agrupamiento se comporta en un clustering con datos meteorológicos.

De igual forma, se presentan los resultados en cuanto a la implementación realizada en dos instancias. En esta sección, se presenta la implementación del método, abarcando tanto el manejo de los datos empleados como el software utilizado.

## 4.2. Resultado de la Implementación Región Andina del Ecuador

### 4.2.1. Base de datos

Se utilizó una base de datos de los modelos de Circulación atmosférica CNRM Centre National de Recherches Meteorológicas (CNRM) y Cerfacs de la sexta fase del Proyecto de Intercomparación de Modelos Acoplados 6 (CMIP6) para la región Sierra del Ecuador. Los datos fueron proporcionados por el Grupo de Investigación de Energías Alternativas y Ambiente GEAA, para un periodo 2020-2050, con una resolución de 25km×25km.



**Ilustración 4.1:** Proyección climática CNRM para la región Andina del Ecuador en el período 2020-2050.

**Fuente:** Grupo de Investigación GEAA, 2024.

### 4.2.2. Implementación del método

La implementación del método Mean shift se realizó para el caso unidimensional de la variable temperatura mínima mediante el lenguaje de programación Python (<https://github.com/geaaepoch>).

Se importó la biblioteca sklearn, una herramienta que facilita el análisis de datos, desarrollo e implementación de algoritmos Machine Learning. En este caso se trabajó con sklearn.cluster.MeanShift, donde el parámetro de mayor relevancia es el ancho de banda (bandwidth), especificada a continuación.

#### 4.2.2.1. Elección del ancho de banda (*bandwidth*)

El ancho de banda posee un rol que influye en la estimación de la función de densidad de probabilidad determinando la forma en que los datos se acoplan a una estimación de la distribución real desconocida, se utilizó tres técnicas como Regla de Silverman, Validación cruzada y Distancia entre pares. Obteniendo los resultados mostrados en la Tabla

Tabla 4-1: Valores del parámetro *bandwidth*

Regla de silverman	Validación cruzada	Distancia entre pares
0.34	0.73	1.83

Realizado por: Basantes, K., 2024.

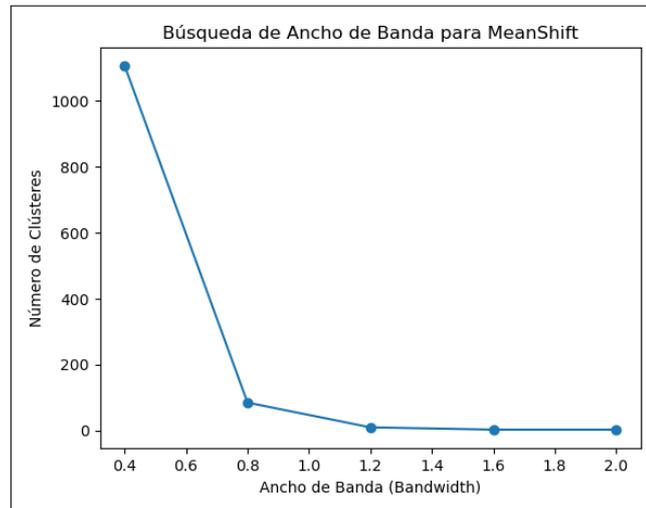


Ilustración 4.2: Relación existente del valor del *bandwidth* y el número de clusters.

Realizado por: Basantes, K., 2024.

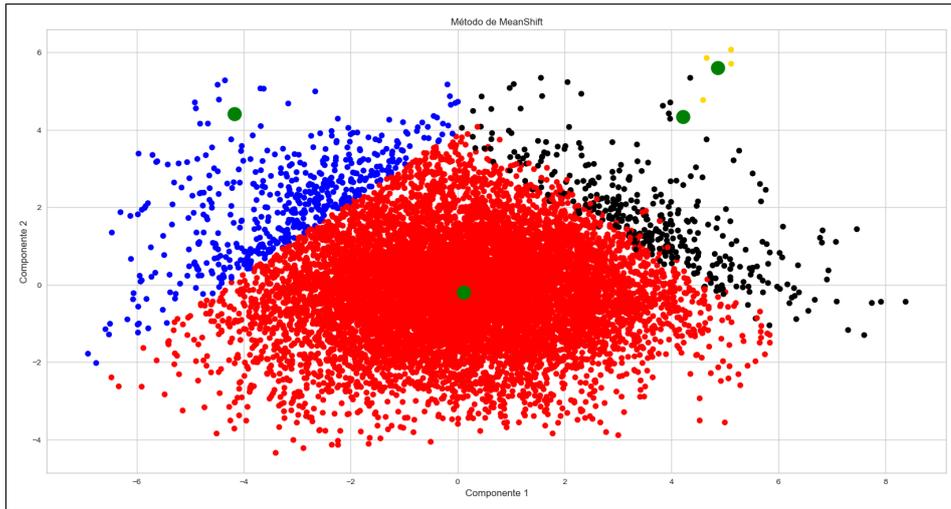
Una vez establecidos los distintos valores de *bandwidth*, se procede a ver el desempeño de cada uno de ellos en el algoritmo. Los resultados nos reflejan agrupaciones sumamente numerosas para el ancho de banda de 0.34 y 0.73, mientras que para 1.83 se obtuvo 4 grupos. Ver Tabla 4-2:

Tabla 4-2: Grupos formados a partir del ancho de banda

Bandwidth utilizado	Núm. de clusters
Regla Silverman	1441
Validación Cruzada	123
Distancia pares	4

Realizado por: Basantes, K., 2024.

Existe una relación directa entre el valor de *bandwidth* y el número de cluster como se evidencia en los resultados expuestos y en la Figura 4.2



**Ilustración 4.3:** Agrupación resultante con un *bandwidth* de 1.38.

Realizado por: Basantes, K., 2024.

### 4.2.3. Análisis por clusters

Se realizó la clasificación de las agrupaciones, es decir, se cuantificó las etiquetas obtenidas en cada cluster. En la Tabla 4-3: se observa de mayor a menor la ocurrencia de datos en cada agrupación.

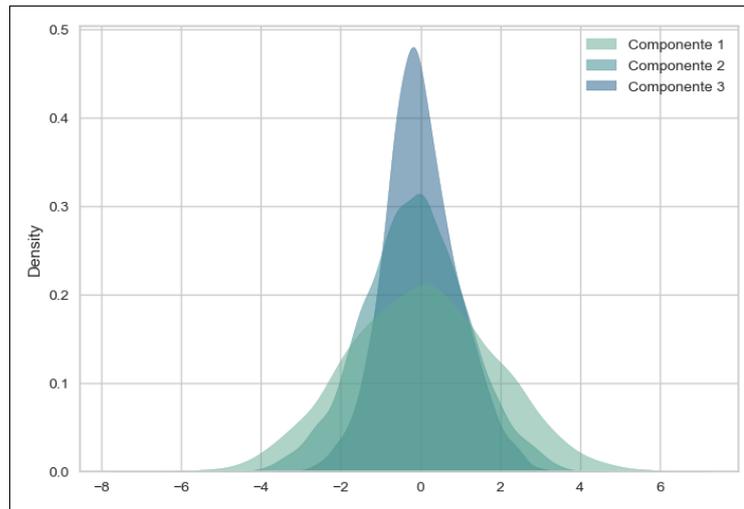
Tabla 4-3: Frecuencia de datos por cada cluster

Cluster	Frecuencia
0	10277
1	400
2	273
3	4

Realizado por: Basantes, K., 2024.

Se graficó la estimación de *kernel* de los datos por cada cluster para describir la distribución por cada componente.

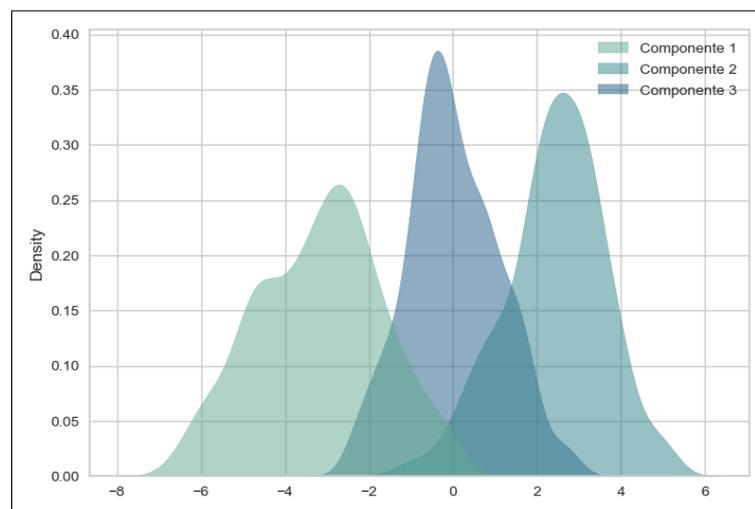
En el cluster 0 las componentes presenta una curva simétrica con respecto a la media y una desviación estándar baja. En el componente 1 se refleja una mayor dispersión de los datos con respecto a los componentes 2 y 3 como se observa en la Figura 4.4.



**Ilustración 4.4:** Distribución del clustering 0.

**Realizado por:** Basantes, K., 2024.

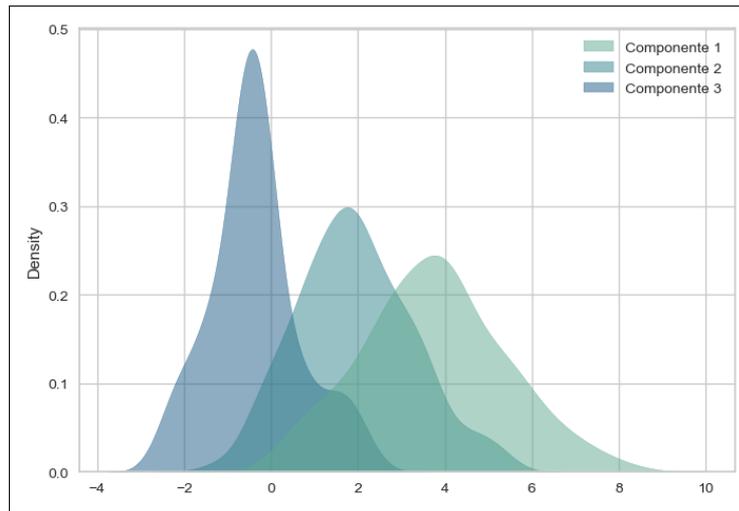
En el cluster 1 se presentó una asimetría en los componentes 1 y 2 alejados de la media como se visualiza en la Figura 4.5. El componente 3 refleja una curva cercana a la media pero no simétrica. En el cluster 2, los valores están dispersos en los 3 componentes siendo el componente 1 el que presenta una mayor dispersión respecto a los componentes 2 y 3, estas componentes se encuentran alejadas de la media y experimentan mayor desviación estándar, como se observa en la Figura 4.6.



**Ilustración 4.5:** Distribución cluster 1.

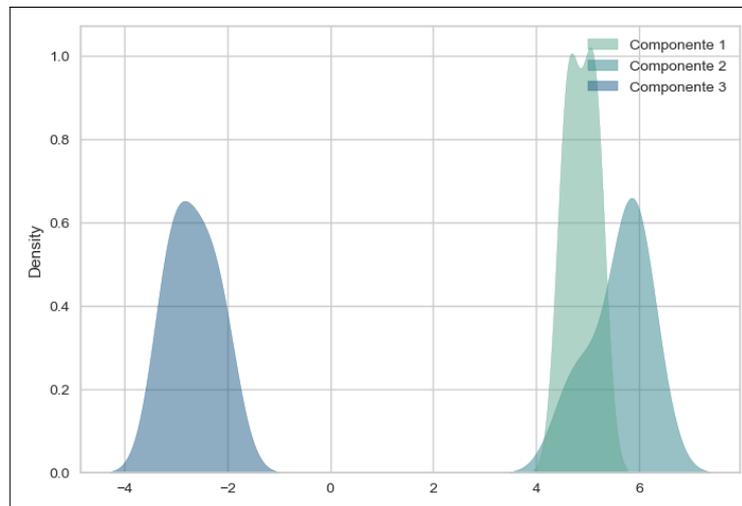
**Realizado por:** Basantes, K., 2024.

En el cluster 3, ninguno de los componentes se encuentra en la media. El componente 3 presenta valores de desviación estándar negativo mientras que las componentes 1 y 2 presentan valores positivos como lo muestra la Figura 4.7



**Ilustración 4.6:** Distribución cluster 2.

**Realizado por:** Basantes, K., 2024.



**Ilustración 4.7:** Distribución cluster 3.

**Realizado por:** Basantes, K., 2024.

Esto significa que en los próximos 50 años, (2020-2050), el modelo de cambio climático CRNM esta reproduciendo una realidad en la región andina del Ecuador, en la que el 93.81 % de los datos de temperatura mínima no experimenta variación, se encuentran con valores cercanos a la media y una desviación estándar mínima. Los eventos extremos y anomalías se presentan en bajo porcentaje y esta caracterizados en los cluster 1,2 y 3.

#### 4.2.4. Resultados de la implementación en una estación meteorológica

Se realizó una nueva implementación con la misma metodología para una estación meteorológica en la Provincia de Chimborazo, específicamente, de la Estación Meteorológica ALAO de la parroquia Pungalá en el cantón Riobamba con coordenadas -1.868806 y 78.54166432318941 en longitud e latitud respectivamente, y a una altura de 3064 metros sobre el nivel del mar.

#### 4.2.5. Implementación del método

Se calculó el ancho de banda (bandwidth) con la Regla de Silverman, la validación cruzada y la distancia entre pares, como se ve en la tabla 4-4: . Se escogió la distancia entre pares como el mejor parámetro debido a que estima un número de clusteres coherentes.

Tabla 4-4: Valores del parámetro *bandwidth*

Regla de silverman	Validación cruzada	Distancia entre pares
0.61	0.71	1.82

Realizado por: Basantes, K., 2024.

En la figura 4.8, se observa los datos correspondientes a los años 2015 y 2016 con las etiquetas ALAO15 y ALAO16 respectivamente. Estos datos se suavizan a través de una estimación de densidad por medio de un *kernel* normal, la influencia del parámetro ancho de banda de las tres técnicas se visualizan en la suavidad de la curva. El valor de 1.82 presenta un mejor suavizado de los datos en comparación con la validación cruzada y la regla de silverman.

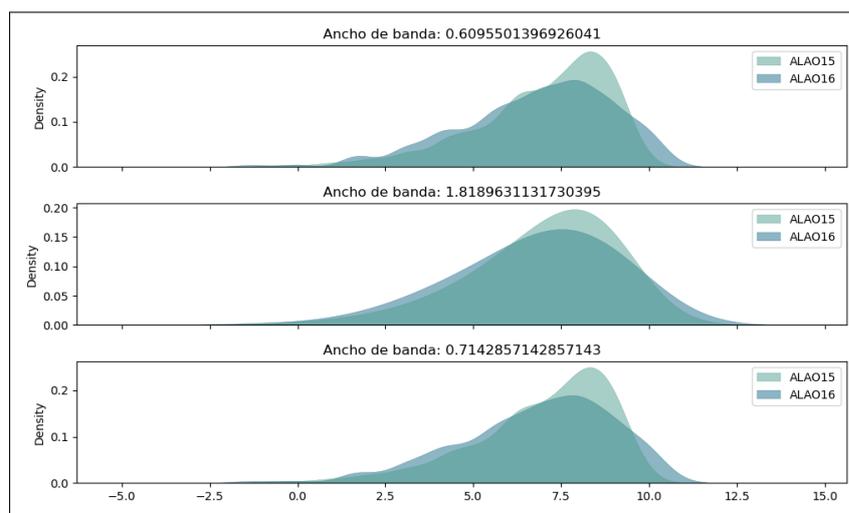
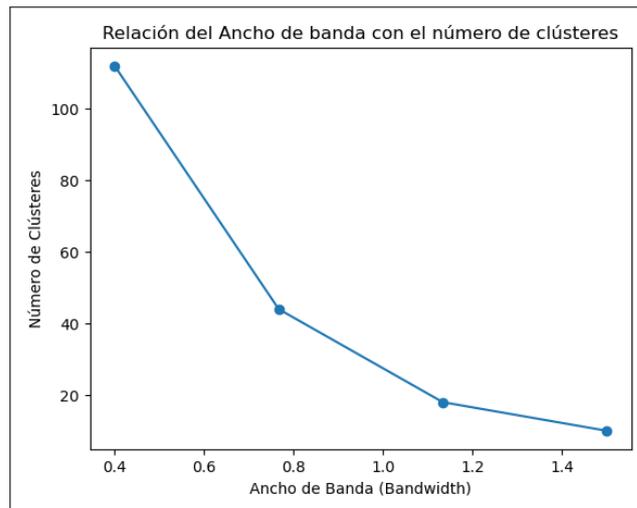


Ilustración 4.8: Estimación mediante un kernel normal.

Realizado por: Basantes, K., 2024.

La relación entre el ancho de banda y los números de cluster se presentan en la Figura 4.9,

evidenciado la relación inversamente proporcional entre el número de clústeres y el valor del *bandwidth*.



**Ilustración 4.9:** Relación *Bandwidth-Cluster*.

**Realizado por:** Basantes, K., 2024.

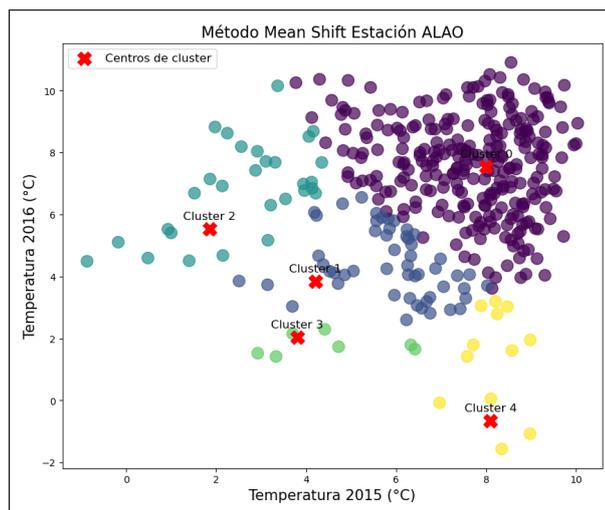
El método Mean Shift busca iterativamente mover cada punto hacia una región de mayor densidad, es decir, hacia la moda más cercana en el conjunto de datos. El proceso que realiza el algoritmo de scikitlearn (<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.MeanShift.html>) se trata de una implementación de una versión simplificada y eficiente del algoritmo Mean Shift.

**Tabla 4-5:** Grupos formados a partir del ancho de banda

Bandwidth utilizado	Núm. de clusters
Regla Silverman	54
Validación Cruzada	42
Distancia pares	5

**Realizado por:** Basantes, K., 2024.

Mediante las tres técnicas del parámetro de bandwidth se estableció el número de clusters como se observa en la Tabla 4.9. Los resultados marcan agrupaciones muy altas con la regla de silverman y la validación cruzada, con valores de 54 y 42 agrupaciones respectivamente. El método de distancia entre pares (*bandwidthestimate*) estableció un número coherente e interpretable de 5 agrupaciones utilizando todos los datos y el cuantil de 0.15. Este resultado se visualiza en la Figura 4.10.



**Ilustración 4.10:** Temperatura mínima.

Realizado por: Basantes, K., 2024.

#### 4.2.6. Análisis por cluster

Se cuantificó las etiquetas obtenidas en cada cluster. En la tabla 4-6: se observa la ocurrencia de datos en cada agrupación.

Tabla 4-6: Frecuencia en datos de cada cluster

Cluster	Frecuencia
0	265
1	52
2	29
3	7
4	12

Realizado por: Basantes, K., 2024.

Tabla 4-7: Análisis por cada cluster

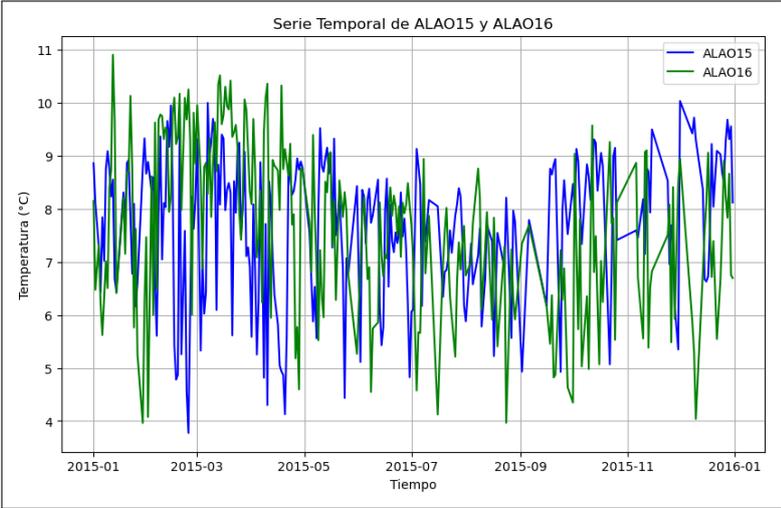
Cluster	0		1		2		3		4	
Años	2015	2016	2015	2016	2015	2016	2015	2016	2015	2016
Promedio	7.67	7.62	5.89	4.41	2.61	6.86	4.55	1.80	8.17	1.35
Desviación	1.34	1.54	1.18	1.07	1.41	1.48	1.39	0.32	0.58	1.65
Mínimo	3.77	3.62	2.51	2.60	-0.87	4.49	2.96	1.42	6.97	-1.57
Máximo	10.04	10.91	8.03	6.55	4.35	10.15	6.42	2.29	8.98	3.20

Realizado por: Basantes, K., 2024.

En la Tabla ?? se observa en detalle los valores estadísticos por cada cluster en los años 2015 y 2016. Se dispuso de esta forma para la interpretación y análisis de la temperatura mínima en la región. El análisis se lo realizó debido a su importancia como un factor de causa y riesgo en una variedad de sectores, desde la salud pública hasta la agricultura y la gestión de recursos naturales. Además del efecto en la distribución espacio temporal de la precipitación.

En el cluster 0 las temperaturas para ALAO 2015 y 2016 son similares con una diferencia de 0.05 C en la media anual, este cluster es el de mayor ocurrencia con 72 % que corresponde a 264 días. Para los dos años de estudio en el mes de marzo no experimento cambios en el valor de la temperatura,

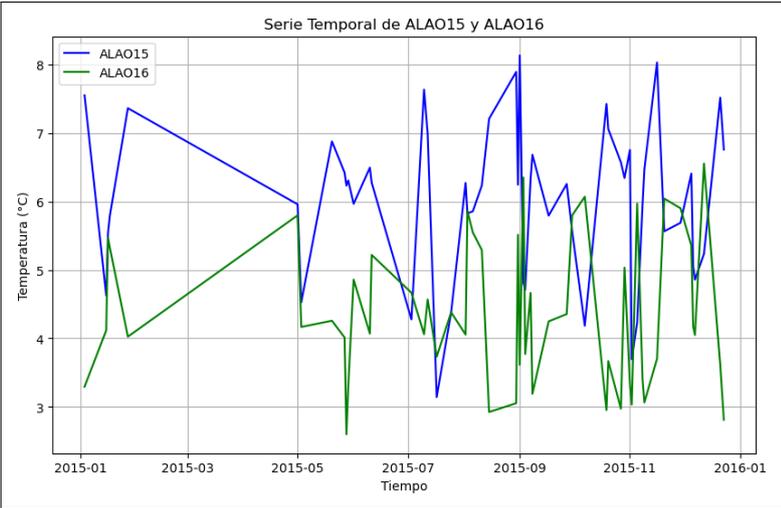
mientras que el patrón en los meses de agosto, septiembre y noviembre solo ocurrieron de 12 a 15 días, como se visualiza en la Figura 4.11



**Ilustración 4.11:** Serie temporal cluster 0.

**Realizado por:** Basantes, K., 2024.

En el cluster 1 las temperaturas promedio anual para cada año varían en dos unidades, este cluster presenta un 15 % de ocurrencia diaria que corresponde a 53 días. En los meses de agosto, septiembre y noviembre este patrón se presentó de 7 y 8 días, mientras que en el mes de abril se presenta un solo día. Existen regiones pequeñas donde los datos de temperatura son similares como se observa en la Figura 4.12 en los meses de febrero, mayo, octubre y noviembre.

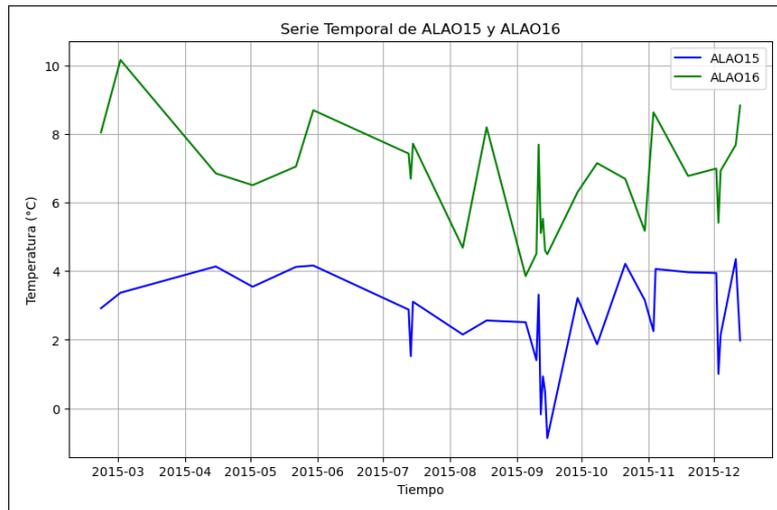


**Ilustración 4.12:** Serie temporal cluster 1.

**Realizado por:** Basantes, K., 2024.

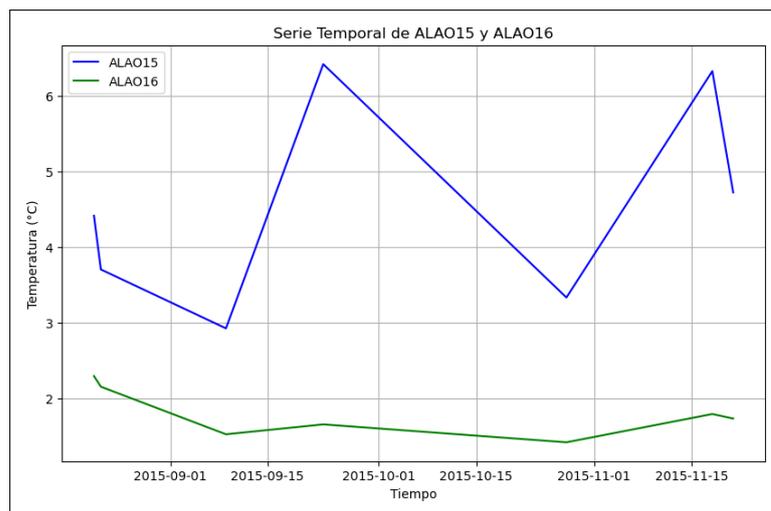
En el cluster 2, la frecuencia de ocurrencia es de 8 % que corresponde a 30 días. Se presentó una diferencia de 4C en la temperatura media anual, es decir, el año 2015 experimentó temperaturas de -0.87C hasta 4.35C, mientras que en el año 2016 experimentó superiores de 3.85 a 10.15C. Los

meses de febrero, marzo y abril poseen un solo día de ocurrencia.



**Ilustración 4.13:** Serie temporal cluster 2.

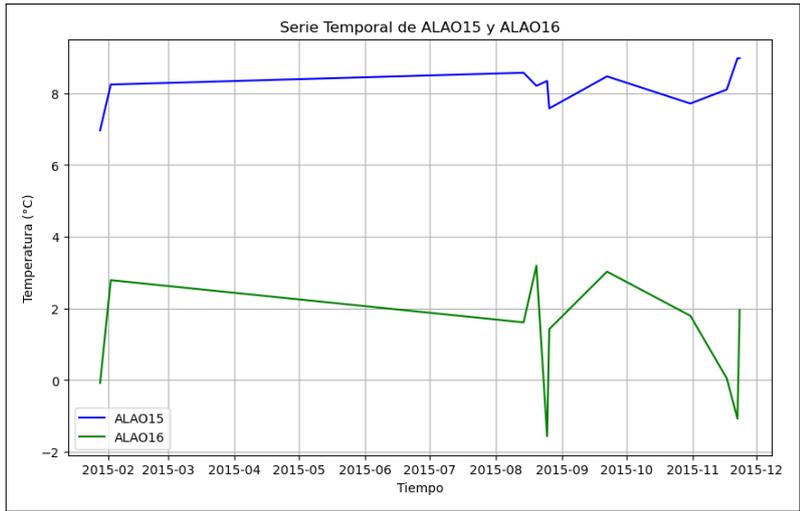
Realizado por: Basantes, K., 2024.



**Ilustración 4.14:** Serie temporal cluster 3.

Realizado por: Basantes, K., 2024.

En el cluster 3, con frecuencia de ocurrencia del 1.92 % que corresponde a 7 días. En el año 2015 existió una variación de temperatura con picos de 6.00 a 2.93C, mientras que el año 2016 se mantuvo una tendencia constante desde 2.29 hasta 1.41C como se observa en la Figura 4.14 Finalmente, se observa que en el cluster 4 con ocurrencia del 3.01 % que corresponde a 11 días, existe una variación significativa entre el año 2015 y 2016, la temperatura para el primer año entre 6.96 y 8.52C mientras que en el segundo presentó temperaturas de -1.56 a 3.20C, como se indica en la Figura 4.15



**Ilustración 4.15:** Serie temporal cluster 4.

Realizado por: Basantes, K., 2024.

## CAPÍTULO V

### 5. CONCLUSIONES Y RECOMENDACIONES

#### 5.1. Conclusiones

Con la finalización del proceso investigativo, se presentan las siguientes conclusiones basadas en los objetivos planteados al inicio de este trabajo de investigación.

- Al realizar una revisión bibliográfica del tema en estudio se logró obtener evidencia sólida y pertinente sobre la estructura y funcionamiento del Mean Shift, identificando el uso de la Teoría de la Probabilidad y el Análisis Numérico permitiendo mejorar la comprensión de este método para su respectiva descripción en detalle.
- La inmersión en la teoría existente, permitió adquirir una visión más profunda y precisa de los fundamentos teóricos que estructuran el método Mean Shift. Estableciendo que los *kernel* son importantes para el funcionamiento del método siendo las funciones que estiman la densidad de la distribución de los datos, facilitando así, el entendimiento hacia la aplicación del mismo.
- La implementación del método Mean Shift mediante el lenguaje de programación Python, permitió una comprensión inicial de los patrones climáticos en la región Sierra del Ecuador. Además de establecer una base para futuras investigaciones dentro del grupo de investigación.
- La monografía sobre los fundamentos teóricos del método Mean Shift se planteó como una herramienta educativa que no solo consolida el entendimiento del método, sino que también fomenta la implementación práctica de los conceptos aprendidos.

#### 5.2. Recomendaciones

Luego de la exposición de los resultados de este trabajo de investigación, se presentan las siguientes recomendaciones con el fin de proporcionar una orientación a los problemas encontrados.

- Se destaca la importancia de poner a disposición este documento para docentes y estudiantes con la finalidad de que tanto estudiantes como investigadores tengan acceso al material, facilitando la exploración y la expansión de conocimientos en el tema.
- Basándonos en este trabajo, se pueden explorar nuevas investigaciones enfocadas en mejorar y optimizar el método Mean Shift.

## BIBLIOGRAFÍA

1. **CARREIRA, Miguel.** “A review of mean-shift algorithms for clustering”.[en línea],2015,(USA)[Consulta: 20 de octubre 2023]. Disponible en:  
<https://arxiv.org/abs/1503.00687>
2. **COMANICIU, Dorin & MEER; Peter.** “Mean Shift: A Robust approach Toward Feature Space Analysis”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* [en línea], 2002, (USA), vol. 24(5), págs. 603-619 [Consulta: 01 junio 2023]. Disponible en:  
<http://dx.doi.org/10.1109/34.1000236>
3. **CHACÓN, José & MONFORT, Pablo.** “A comparison of bandwidth selectors for mean shift clustering”. [en línea], 2013, (España)[Consulta: 20 noviembre 2023]. Disponible en:  
<https://arxiv.org/abs/1310.7855>
4. **CHENG, Yizong.** “Mean shift, mode seeking, and clustering”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* [en línea],1995, (USA), vol. 17(8), 790-799. [Consulta: 10 junio 2023]. Disponible en:  
<http://dx.doi.org/10.1109/34.400568>
5. **COMANICIU, D; et al.** “The variable bandwidth mean shift and data-driven scale selection”. *Proceedings Eighth IEEE International Conference on Computer Vision*[en línea],2001,(Canada), vol. 1, págs. 438-445.[Consulta: 30 octubre 2023]. Disponible en:  
<https://dx.doi.org/10.1109/ICCV.2001.937550>.
6. **DUDA, Richard; et al.** *Pattern Classificaton*. 2da ed. Jhon Wiley & Sons inc, 2001, págs. 161-172.
7. **EVERITT, Brian; et al.** *Cluster Analysis*. 5ta ed. Jhon Wiley & Sons inc, 2011, págs. 01-13.
8. **FASHING, Mark & TOMASI, Carlo.** “Mean Shift Is a Bound Optimization”. [en línea], 2005, vol. 27, págs. 471-474.[Consultado en: 18 junio 2023]. Disponible:  
<https://doi.org/10.1109/TPAMI.2005.59>
9. **FUKUNAGA, K & HOSTETLER.** “The estimation of the gradient of a density function, with applications in pattern recognition”. *IEEE Transactions on Information Theory*[en línea], 1975,,vol. 21, págs. 32-40.[Consulta:18 junio 2023]. Disponible en:  
<https://dx.doi.org/10.1109/TIT.1975.1055330>.

10. **MEER, Peter & COMANICIU, Dorin.** “Mean Shift Analysis and Applications”. *Computer Vision, IEEE International Conference on* [en línea],1999,(USA), vol. 2, 1197. [Consulta: 30 octubre 2023]. Disponible en:

<http://dx.doi.org/10.1109/ICCV.1999.790416>

11. **WAND, M.P & JONES, M.C** *Kernel Smoothing*[en línea].New York-USA:Springer-science and business media, B.V. ,1995.[Consulta: 06 noviembre 2023]. Disponible en:

<https://doi.org/10.1201/b14876>

12. **STORK, David; et al.** *Patern classification* [en línea].2da ed.USA: Wiley editorial, 2000.[Consulta: 06 noviembre 2023]. Disponible en:

[https://books.google.com/books/about/Pattern\\_Classification.html?id=Br33IRC3PkQC](https://books.google.com/books/about/Pattern_Classification.html?id=Br33IRC3PkQC)

13. **WU, K.L & YANG, M.S** “Mean shift based clustering”. *Pattern Recognition*[en línea],2007,(Taiwan),vol. 40, págs 3035-3052.[Consulta: 8 noviembre 2023]. ISSN 0031-3203. Disponible en:

<https://doi.org/10.1016/j.patcog.2007.02.006>.



## **ANEXOS**

**ANEXO A: MONOGRAFÍA “DESCRIPCIÓN DEL MÉTODO MEAN SHIFT PARA ANÁLISIS DE DATOS ESPACIALES”**

ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO



DESCRIPCIÓN DEL MÉTODO  
MEAN SHIFT PARA ANÁLISIS  
DE DATOS ESPACIALES

KEVIN BASANTES  
RIOBAMBA, 2024

# Contenidos

---

<b>Introducción</b>	<b>1</b>
<b>1 Preliminares</b>	<b>3</b>
1.1 Bases teóricas de Probabilidad . . . . .	3
1.1.1 Variables aleatorias . . . . .	3
1.1.2 Función de densidad . . . . .	5
1.1.3 Funciones <i>kernel</i> . . . . .	8
1.1.4 Funciones <i>test</i> . . . . .	11
1.2 Técnicas no paramétricas . . . . .	13
1.2.1 Estimador de densidad . . . . .	13
1.2.2 Parzen Windows . . . . .	15
1.2.3 Convergencia de la media y la covarianza . . . . .	17
1.3 Estimación de la densidad de <i>kernel</i> univariable . . . . .	19
1.3.1 Los criterios MSE y MISE . . . . .	19
1.3.2 Orden y Notación Asintótica . . . . .	22
1.3.3 Ampliación de Taylor . . . . .	22
1.3.4 Aproximaciones asintóticas de MSE y MISE . . . . .	23
1.4 El estimador de densidad de kernel multivariable . . . . .	27
1.4.1 Aproximaciones asintóticas MISE . . . . .	28
<b>2 Mean Shift</b>	<b>32</b>

## Contenidos

---

2.1	Descripción de Mean Shift . . . . .	32
2.1.1	Estimación de la función de densidad de probabilidad . . . . .	33
2.1.2	Estimación del gradiente de densidad . . . . .	35
2.2	Criterio de convergencia de Mean Shift . . . . .	37
2.2.1	Convergencia mediante el <i>Kernel Epanechnikov</i> . . . . .	37
2.2.2	Criterio suficiente de convergencia de Mean Shift . . . . .	39
2.2.3	Propiedad de una trayectoria suave . . . . .	41
2.2.4	Elección del Bandwidth . . . . .	42
<b>3</b>	<b>Implementación y Resultados</b>	<b>44</b>
3.1	Implementación Región Andina del Ecuador . . . . .	44
3.1.1	Base de datos . . . . .	44
3.1.2	Implementación del método . . . . .	46
3.1.3	Análisis por clusters . . . . .	50
3.2	Implementación en una estación meteorológica . . . . .	54
3.2.1	Base de datos . . . . .	54
3.2.2	Implementación del método . . . . .	54
3.2.3	Análisis por cluster . . . . .	57
	<b>Bibliografía</b>	<b>64</b>

# Índice de figuras

---

1.1	Tipos de Kernel . . . . .	11
2.1	Ventanas d-dimensionales utilizadas en dos iteraciones sucesivas de Mean Shift . . . . .	38
3.1	Proyección climática CNRM para la región Andina del Ecuador en el período 2020-2050. . . . .	45
3.2	Distribución de los datos de temperatura mínima . . . . .	46
3.3	Análisis de componentes principales (PCA) . . . . .	47
3.4	Suavidad de la distribución de los datos para cadaa valor de <i>bandwidth</i> . . . . .	47
3.5	Relación existente del valor del <i>bandwidth</i> y el número de clusters . . . . .	48
3.6	Agrupación resultante con un <i>bandwidth</i> de 1.38 . . . . .	49
3.7	Proyección de los cluster en un plano tridimensional . . . . .	50
3.8	Distribución Cluster 0 . . . . .	51
3.9	Distribución Cluster 1 . . . . .	52
3.10	Distribución Cluster 2 . . . . .	52
3.11	Distribución Cluster 3 . . . . .	53
3.12	Estimación mediante un kernel normal . . . . .	55
3.13	Relación Bandwidth-Clusters . . . . .	56
3.14	Temperatura mínima . . . . .	57
3.15	Serie temporal cluster 0 . . . . .	59
3.16	Serie temporal cluster 1 . . . . .	60

## Índice de figuras

---

3.17 Serie temporal cluster 2 . . . . .	61
3.18 Serie temporal cluster 3 . . . . .	61
3.19 Serie temporal cluster 4 . . . . .	62

# Índice de cuadros



3.1	Valores del parámetro <i>bandwidth</i> . . . . .	48
3.2	Grupos formados a partir del ancho de banda . . . . .	49
3.3	Frecuencia de datos en cada cluster . . . . .	50
3.4	Valores del parámetro <i>bandwidth</i> . . . . .	54
3.5	Grupos formados a partir del ancho de banda . . . . .	56
3.6	Frecuencia de datos en cada cluster . . . . .	58
3.7	Análisis por cada cluster . . . . .	58

# Introducción

---

El *clustering* es un término con el cual se conoce al agrupamiento que se realiza a un conjunto de datos a través de métodos de aprendizaje no supervisado con la idea de reorganizarlas en grupos a los que se denomina *clusters*, identificando similitudes entre los datos y agruparlos en función de su proximidad compartiendo características entre sí. La existencia de varios métodos para este propósito permite una facilidad al momento de realizar dicha tarea en un conjunto de datos. Métodos como: K-means, Agglomerative clustering, jerárquicos, entre otros son los más utilizados en distintas áreas.

Otro por destacar y en desarrollo es el método Mean Shift, presentado inicialmente en la década de 1970 por Fukunaga y Hostetler, hasta su posterior perfeccionamiento y creciente adopción en las últimas décadas, mostrándose como un método eficaz para la segmentación y agrupación de datos. Su aplicación es especialmente relevante en el ámbito de la visión por computadora, y en poca medida, también para el uso de *clustering*. Resultando característico su manejo de datos en  $n$ -dimensiones y sobre todo la no anticipación de un número de grupos previos.

En el campo de la meteorología, la agrupación de datos espaciales, en casos específicos como lo es la temperatura, son de utilidad para la comprensión del comportamiento de los patrones climáticos en una zona determinada.

Este escrito presenta un camino hacia la comprensión del procedimiento del método Mean Shift, explorando sus fundamentos teóricos desde sus bases matemáticas hasta su implementación práctica enfatizando su utilidad en aplicaciones reales como la meteorología que facilitan la implementación y la selección del parámetro óptimo del método.

La monografía consta de tres capítulos, en los que se abarca los tópicos y ejemplos de manera anítica. En el primer capítulo se describe herramientas del Análisis

## Capítulo 0. Introducción

---

Funcional y Teoría de Probabilidad como la técnica de Windows Parzen y el estimador de la función de densidad. En el segundo capítulo se estable los fundamentos en los que se basa el método como lo es la estimación de la función de densidad multivariante y la localización de máximos a través del gradiente de densidad. Además de enunciar los criterios de convergencia mediante sus teoremas y demostraciones. Por último, en el tercer capítulo se implementa el método Mean shift a la variable temperatura mínima del modelo de Circulación atmosférica CNRM en un periodo 2020-2050 mediante la biblioteca Sklearn en Python.

# 1

---

## *Preliminares*

---

### 1.1 Bases teóricas de Probabilidad

En esta sección se abordarán temas preliminares para la comprensión de la teoría del método Mean Shift. Se exponen tópicos en relación con la Teoría de Probabilidad y del Análisis Funcional mediante definiciones y ejemplos afines y por contradicción para abordarlos de mejor forma.

#### 1.1.1 Variables aleatorias

##### **Definición 1.1**

Sea  $\Omega$  un espacio muestral y  $\mathcal{F}$  una  $\sigma$ -álgebra sobre  $\Omega$ . Se denomina variable aleatoria real sobre  $(\Omega, \mathcal{F})$  a toda función

$$X : \Omega \rightarrow \mathbb{R}$$
$$\omega \mapsto X(\omega),$$

tal que para todo  $t \in \mathbb{R}$ , el conjunto  $X^{-1}(]-\infty, t]) = \{\omega \in \Omega \mid X(\omega) \leq t\} \in \mathcal{F}$

Una variable aleatoria puede tomar valores contables o no, dado que es una caracterización cuantitativa de los resultados de un espacio muestral su naturaleza puede ser de dos tipos; discreta o continua.

### Definición 1.2 Variable aleatoria discreta

Si el conjunto  $X(\Omega)$  de los valores de  $X$  es un subconjunto numerable de  $\mathbb{R}$ , se puede escribir

$$X(\Omega) = \{t_1, t_2, \dots, t_n, t_{n+1}, \dots\},$$

se dice que  $X$  es discreta.

**Ejemplo.** EL ejemplo clásico de lanzar una moneda. Se posee un espacio de posibilidades  $\Omega = \{0, 1\}$ . Existen dos resultados: 0 (cara) y 1 (cruz), o viceversa dependiendo el criterio usado.

La probabilidad asignada a los resultados en  $\Omega$  determina las probabilidades asociadas a los valores de la variable aleatoria  $X$ . La distribución de probabilidad de  $X$  establece la forma en la que se distribuye la probabilidad de 1 entre los valores del espacio  $\Omega$ .

### Definición 1.3

**La distribución de probabilidad o función masa de probabilidad** de una variable aleatoria discreta se define para cada número  $x$  como

$$p(x) = P(X = x)$$

**Ejemplo.** Uno muy empleado en el caso continuo de distribución de probabilidad es la distribución de Poisson.

### Definición 1.4

Se dice que una variable aleatoria  $X$  tiene una **distribución de Poisson** con parámetro  $\lambda (\lambda > 0)$  si la función masa de probabilidad de  $X$  es

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

El valor de  $\lambda$  es con frecuencia un valor por unidad o por unidad de área. El hecho de que

$$\sum_{x=0}^{\infty} p(x) = 1,$$

es una consecuencia de la expansión de la serie infinita de Maclaurin de  $e^\lambda$

Como se puede observar, en el caso de las variables aleatorias discretas se hace uso de aspectos matemáticos como las sumas y diferencias, mientras que en el análisis de una variable aleatoria continua se implica el uso del Cálculo, específicamente de integrales y derivadas.

### Definición 1.5 Variable aleatoria continua

Si el conjunto  $X(\Omega)$  de los valores de  $X$  es un intervalo  $I = [\alpha, \beta]$  entonces  $F_x(\alpha) = 0$  y  $F_x(\beta) = 1$  donde  $F_x$  es la función de repartición de la variable aleatoria, se dice que  $X$  es continua.

**Ejemplo.** En el estudio ecológico de un lago, se mide la profundidad en lugares determinados, luego la profundidad en ese lugar resulta ser  $X$ , y los extremos del intervalo de  $\Omega$  resultan ser sus valores máximos, es decir,  $\alpha$  es la profundidad mínima en la región muestreada y  $\beta$  es la profundidad máxima.

Tomando este ejemplo, si se discretiza la variable  $X$  midiendo la profundidad, de forma gráfica, se ilustra un histograma de probabilidad. A medida que se mide la variable  $X$  con mayor precisión los rectángulos resultantes son más angostos, siguiendo esta idea la secuencia de histogramas se aproxima a una curva. La probabilidad de que la profundidad en un punto entre  $\alpha$  y  $\beta$  es el área bajo la curva entre el intervalo.

### 1.1.2 Función de densidad

#### Definición 1.6

Si existe una función  $f : \Omega \rightarrow \mathbb{R}$  tal que:

1.  $\forall x \in \mathbb{R}, f(x) \geq 0,$
2.  $\int_{-\infty}^{\infty} f(x) dx = 1,$  y

$$3. P(a \leq X \leq b) = \int_a^b f(x)dx.$$

para cualquier  $a, b$ . Entonces  $f$  es la función de densidad de probabilidad de la variable aleatoria continua  $X$ .

La función de densidad de probabilidad (pdf) por sus siglas en inglés, es una función que describe la probabilidad que una variable aleatoria continua toma en un valor determinado.

La noción de función densidad proviene cuando se introducen las variables aleatorias en un espacio de Probabilidad, si  $X$  es una variable aleatoria, entonces  $X$  induce una medida de probabilidad en  $\mathbb{R}$  que se establece como su distribución.

La distribución de una variable aleatoria  $X$  suele describirse mediante su función de distribución acumulativa  $F(x) = P(X \leq x)$ . Cuando la función de distribución tiene la forma:

$$F(x) = \int_{-\infty}^{\infty} f(y)dy,$$

se dice que  $X$  tiene una **función de densidad**  $f$ .

### **Ejemplo. Distribución uniforme en (0,1)**

$f(x) = 1$  para cada  $x \in (0, 1)$  y 0 en otro caso. La función de distribución viene dado por:

$$F(x) = \begin{cases} 0 & x \leq 0, \\ x & 0 \leq x \leq 1, \\ 1 & x > 1. \end{cases}$$

### **Definición 1.7**

#### **Valor Esperado (Esperanza matemática)**

Si  $X \geq 0$  es una variable aleatoria en  $(\Omega, F, P)$  la esperanza matemática se define como  $\mathbb{E}[X] = \int XdP$

Por lo general,  $\mathbb{E}[X]$  se denomina media y su notación es  $\mu$ . Si  $\mathbb{E}[X]^2 < \infty$ , la varianza de  $X$  se define como  $\text{var}(X) = \mathbb{E}(X - \mu)^2$ . Para calcular la varianza

resulta útil la siguiente expresión:

$$\text{var}(X) = \mathbb{E}(X^2) - (\mathbb{E}[X])^2$$

**Ejemplo.** Si  $X$  tiene una distribución de Bernoulli con parámetro  $p$ , entonces  $P(X = 1) = p$  y  $P(X = 0) = 1 - p$ . Claramente

$$\mathbb{E}[X] = p \cdot 1 + (1 - p) \cdot 0 = p.$$

Dado que  $X^2 = X$ , se tiene  $\mathbb{E}[X^2] = \mathbb{E}[X] = p$

$$\text{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1 - p)$$

La esperanza matemática es una medida central que proporciona una idea del valor promediado esperado de una variable aleatoria, mientras que la varianza mide la dispersión que existe entre esos valores alrededor de su esperanza. Estos conceptos toman importancia al momento de realizar la estimación de una función de densidad de una variable aleatoria, habitualmente, en las estimaciones que implican el uso de funciones *kernel*.

Una función de densidad describe la distribución de probabilidad de una variable aleatoria continua a través de [1.1.2](#). Si  $f(x)$  no es una distribución que se conoce, surge la importancia de poder estimar  $f(x)$  a partir de una muestra de observaciones  $X_1, X_2, \dots, X_n$  que se suponen son resultados independientes y tienen la distribución de probabilidad igual (i.i.d). Para este propósito, las estimaciones en base a las funciones *kernel* emplean un conjunto de datos continuos cuya forma exacta de distribución es desconocida, con la finalidad de estimar la forma en la que los datos se distribuyen en un espacio.

### 1.1.3 Funciones *kernel*

#### Definición 1.8

Una función  $K : \mathbb{R} \rightarrow \mathbb{R}$  que cumple con las siguientes propiedades:

- |   |   |
|---|---|
| a) $\forall x \in [-1, 1] K(x) \in [0, \infty)$ , | d) $\int_{-1}^1 K(x)dx = 1$ ,                 |
| b) $\forall x \in [-1, 1] K(x) = 0$ ,             | e) $\int_{-1}^1 xK(x)dx = 0$ ,                |
| c) $K(x) = K(-x)$ ,                               | f) $\int_{-1}^1 x^2K(x)dx \in \mathbb{R}^+$ , |

se denomina función *kernel*

**Ejemplo. Función *kernel* Gaussiana** o también conocida como *kernel* de campana es un ejemplo común, utilizado en máquinas de vectores de soporte (SVM por sus siglas en inglés). Se define como

$$K(x) = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right),$$

donde  $x$  y  $x'$  son puntos del espacio y  $\sigma$  es un parámetro que determina el ancho de campana.

Un contraejemplo, para este caso, puede ser cualquier función que no cumpla las propiedades requeridas en [1.1.3](#)

**Ejemplo.**

$$f(x, y) = x^2 - y^2,$$

en este caso, no cumple la simetría, es decir,  $f(x, y) \neq f(y, x)$

Retomando el ejemplo de la función *kernel* Gaussiana, una parametrización permite un ajuste de la función a los datos específicos, esto dado el parámetro  $\sigma$  lo cual controla el ancho de la campana, afectando directamente la suavidad de la estimación de la densidad de probabilidad.

### Parametrización de kernels

Sea  $h \in \mathbb{R}^+$ , el kernel parametrizado en  $h$  es

$$K_h = \frac{1}{h}K\left(\frac{x}{h}\right), \quad x \in [-h, h].$$

Se mantiene las propiedades anteriores en un intervalo  $[-h, h]$ , donde  $h$  es el ancho de banda o *bandwidth* de  $K$ . Si  $h$  incrementa,  $h > 1$ , la amplitud de  $K$  aumenta, caso contrario en el factor  $\frac{1}{h}$ , que reduce el rango de  $K$  para mantener el área igual a 1. Por otro lado, si  $h$  se reduce,  $h < 1$ , y la amplitud de  $K$  se reduce, caso contrario con el factor  $\frac{1}{h}$ , el cual incrementa el rango de  $K$  para mantener el área igual a 1.

### Traslación de ejes

La función *kernel* se puede centrar en cualquier punto  $x_i \in \mathbb{R}$ ,  $i \in \{1, 2, \dots, n\}$ . El *kernel* parametrizado en  $h$  y con centro en  $x_i$  viene dado por:

$$K_h(x) = \frac{1}{h}K\left(\frac{x - x_i}{h}\right), \quad x \in [x_i - h, x_i + h].$$

La traslación que se realiza mantiene las propiedades anteriores en el intervalo  $[x_i - h, x_i + h]$ .

Algunos de los *kernels* más utilizados y conocidos se describen a continuación.

### Tipos de kernels

- *Kernel* Rectangular o Uniforme

Al hablar del *kernel* rectangular o uniforme se referencia a un rectángulo ubicado que se coloca sobre cada punto. Si este interactúa con el *kernel* de otros puntos se refleja un cambio notable en cuanto a su suma.

$$K(x) = 0.5, \quad x \in [-1, 1].$$

- *Kernel* Triangular

En este caso, se considera un triángulo que se coloca sobre cada punto. Este

al interactuar con los otros *kernels* resulta ser más liso que los rectángulos y su combinación tiene un efecto lineal.

$$K(x) = 1 - |x|, \quad x \in [-1, 1].$$

- *Kernel* de Epanechnikov

En este caso se trata de un segmento del perfil de un arco proveniente de una parábola que se coloca sobre cada punto.

$$K(x) = \frac{3}{4}(1 - x^2), \quad x \in [-1, 1].$$

- *Kernel* Normal o Gaussiano

En este caso se define como intervalo el conjunto  $\mathbb{R}$ . Por esto se considera que este *kernel* es uno de los más habituales a usar.

$$K(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$

- *Kernel Biweight* o Cuártico

Este *kernel* asigna peso a los valores dentro de una cierta distancia del punto de interés, y ese peso disminuye a medida que el valor se aleja del punto central.

$$K(x) = \frac{15}{16}(1 - x^2)^2, \quad x \in [-1, 1].$$

- *Kernel* Arco coseno

Un *kernel* poco conocido debido a su sensibilidad a la normalización de los datos y a la elección de parámetros.

$$K(x) = \frac{\pi}{4} \cos\left(\frac{\pi}{2}x\right), \quad x \in [-1, 1].$$

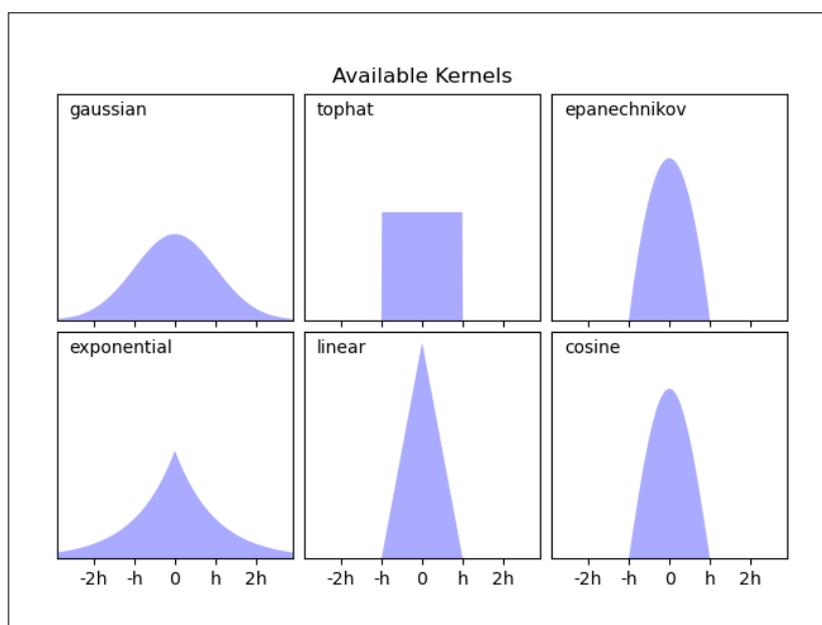


Figura 1.1: Tipos de Kernel

Fuente: Pedregosa et al., 2011.

### 1.1.4 Funciones *test*

#### Funciones soporte compacto

##### Definición 1.9

Dado  $u : \Omega \subset \mathbb{R} \rightarrow \mathbb{R}$ . El conjunto

$$\text{sop } u = \overline{\{x \in \Omega : u(x) \neq 0\}},$$

se denomina conjunto soporte de  $u$ .

De la definición se deduce de forma clara que el conjunto de  $\text{sop } u$  es un cerrado contenido en  $\overline{\Omega}$ . Además, si  $x \notin \text{sop } u$ , entonces,  $u(x) = 0$ .

Al espacio de funciones infinitamente derivables en  $\Omega$  y de soporte compacto contenido en  $\Omega$ , se le denomina espacio de funciones *tests*

La definición dada se opta en casos donde  $f$  es continua, sin embargo, en casos donde  $f$  solo es medible el soporte de la función puede ser vacío. Por ejemplo una función definida en los racionales.

**Ejemplo.** Considerando la función  $f : \mathbb{R} \rightarrow \mathbb{R}$ :

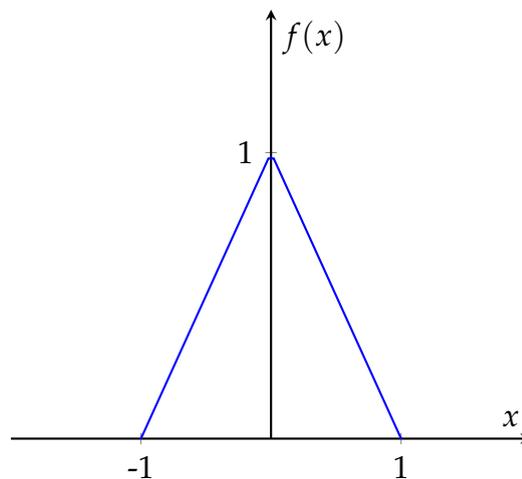
$$f(x) = \begin{cases} 1 - |x| & \text{si } |x| \leq 1 \\ 0 & \text{si } |x| > 1 \end{cases}$$

Esta función tiene soporte compacto en el intervalo  $[-1, 1]$ . Ahora, analizamos los detalles: Para  $|x| \leq 1$ , la función toma el valor  $1 - |x|$ . Esto forma una especie de triángulo con vértices  $(1, 0)$ ,  $(-1, 0)$  y  $(0, 1)$ .

El soporte de la función es el conjunto de valores de  $x$  para los cuales la función no es nula. En este caso, el soporte es  $[-1, 1]$ , ya que la función es cero fuera de este intervalo.

Para  $|x| > 1$ , la función se define como cero. Esto asegura que la función es nula fuera del intervalo  $[-1, 1]$ .

En el siguiente gráfico se puede visualizar la función dibujando el triángulo mencionado anteriormente en el plano cartesiano.



La función es 1 en el vértice superior del triángulo y disminuye linealmente hasta 0 en los extremos del intervalo  $[-1, 1]$ . En resumen, la función  $f(x) = 1 - |x|$  es un ejemplo de una función con soporte compacto, ya que solo no es nula en un conjunto cerrado  $[-1, 1]$  y es cero fuera de ese conjunto.

## 1.2 Técnicas no paramétricas

Al considerar una muestra de datos, existen diversos métodos para su análisis y procesamiento estadístico con el fin de transformar la muestra dada en información significativa para usar en un estudio específico. Para extraer las características esenciales de la muestra, se aplican dos técnicas conocidas que se dividen en dos categorías: métodos paramétricos y no paramétricos.

Los métodos no paramétricos se centran en el reconocimiento de patrones, donde no se supone nada referente a la distribución de los valores en el espacio de características. A diferencia de los métodos paramétricos, los enfoques no paramétricos no requieren información previa, resultando en métodos útiles y aplicables en diversas situaciones con estas características.

Estos métodos pueden ser empleados para la clasificación de datos de una muestra bajo un criterio de estimación de distribución. Es decir, dado un conjunto de valores  $x_1, x_2, \dots, x_n$  en un espacio de características se infiere una función de densidad de probabilidad estimada.

Para la estimación no paramétrica, existen numerosas técnicas disponibles. Dos de las más destacadas en la literatura son: Ventana de Parzen o *Windows Parzen* conocida en inglés y el método de los K-vecinos más cercanos. En este apartado, sólo se tratará de la Ventana de Parzen debido a su relación directa con la estimación de densidad *kernel*.

### 1.2.1 Estimador de densidad

La idea fundamental en la que se basa un estimador de densidad  $P$  en un punto  $x$  viene dada por la probabilidad de que un vector en  $\mathbb{R}^n$  caiga dentro de una región  $\Omega$ , la cual viene dado por:

$$P = \int_{\Omega} p(x') dx'. \quad (1.1)$$

Se nota que  $P$  es un suavizado de la función de densidad  $p(x)$ , y se puede estimar el valor de  $p$  estimando la probabilidad  $P$ . Bajo la hipótesis de que las muestras

## Capítulo 1. Fundamentos teóricos

---

$x_1, \dots, x_n$  sean independientes e idénticamente distribuidos (i.i.d) y el principio de que un vector caiga dentro de una región, se sigue una ley de probabilidad dada por la ley binomial

$$P_k = \binom{n}{k} P^k (1 - P)^{n-k}, \quad (1.2)$$

con la esperanzada para  $k$  dada por  $\mathbb{E}[k] = nP$ . Además cuando  $n$  es lo suficientemente grande y  $P$  no está muy cerca de 0 o 1, la distribución binomial se asemeja a una distribución normal, de acuerdo al teorema del límite central. Ahora, cuando  $n$  es grande, la distribución es simétrica alrededor de la media, y la mayor densidad de probabilidad se concentra cerca de la esperanza, resultando cerca de  $nP$ . Por tanto, dado que esperamos que la mayoría de los valores de  $k$  están cerca de  $nP$ , la razón  $\frac{k}{n}$  tiende a ser cercana a  $P$  cuando  $n$  es lo suficientemente grande.

Asumiendo que  $P(x)$  es continua y que  $p$  no varía demasiado dentro de la región  $\Omega$  se puede aproximar

$$\int_{\Omega} p(x') dx' \approx p(x) \int_{\Omega} dx' = p(x)V, \quad (1.3)$$

donde  $V$  es el volumen de la región y  $x$  es un punto dentro de  $\Omega$ . Se puede obtener una estimación de  $p(x)$  a través de las ecuaciones 1.1, 1.3 y  $\mathbb{E}[k]$

$$\begin{aligned} \mathbb{E}[k] &= nP \\ k &= n \int_{\Omega} p(x') dx' \approx np(x)V, \end{aligned}$$

despejando  $p(x)$

$$p(x) \approx \frac{k}{nV}.$$

Ciertas consideraciones aparecen a tomar en cuenta con respecto a la elección de los parámetros a utilizar para estimar  $p(x)$ . Por ejemplo, Si se fija  $V$  y se hace crecer  $n$  hacia el infinito,  $\frac{1}{n}$  converge con resultado al área promediada de  $p(x)$  en  $\Omega$ . Sin embargo se desea obtener  $p(x)$  en lugar de una versión promediada, es decir, se busca la estimación de la densidad en un punto. Si  $V$  se aproxima a cero y  $n$  se

mantiene constante, la región no contendrá ninguna muestra puesto que resultará siendo muy pequeña y la estimación resultaría inútil.

Desde una vista teórica la solución al planteamiento con muestras ilimitadas se da a través de un conjunto de regiones  $\Omega_1, \dots, \Omega_n$  que contengan a  $x$  donde,  $V_n$  el volumen de  $\Omega_n$ ,  $k_n$  el número de muestras que caen dentro de  $\Omega_n$  y  $p_n$  es la  $n$ -ésima estimación para  $p(x)$ . Así,  $p_n$  viene dado por:

$$p_n(x) = \frac{k_n}{V_n} \quad (1.4)$$

Para establecer que  $p_n(x)$  converge a  $p(x)$ , se debe cumplir las siguientes condiciones:

1.  $\lim_{n \rightarrow \infty} V_n = 0$ ,  $V_n$  disminuirá a medida en que el tamaño de la muestra aumente
2.  $\lim_{n \rightarrow \infty} k_n = \infty$ , asegura la convergencia (en probabilidad) a  $P$
3.  $\lim_{n \rightarrow \infty} \frac{K_n}{n} = 0$ , resulta necesario para que  $p(x)$  converja. Además, indica que a pesar de que un gran número de puntos caigan dentro de  $V_n$ , formarán una fracción insignificanemente pequeña del número total de muestras.

### 1.2.2 Parzen Windows

Suponiendo que  $\Omega_n$  es una región con la forma de un hipercubo  $d$ -dimensional, definimos  $h_n$  como la longitud de uno de sus lados y  $V_n$  como su volumen dado por:

$$V_n = h_n^d. \quad (1.5)$$

Se puede obtener una expresión para  $k_n$  que es el número de muestras que caen dentro de  $\Omega_n$  mediante una función de ventana (*window function*)  $d$  – dimesional:

$$\omega(u) = \begin{cases} 1 & |u_j| \leq 1/2 \\ 0 & otherwise \end{cases} \quad j = 1, \dots, d \quad (1.6)$$

## Capítulo 1. Fundamentos teóricos

---

De esta forma,  $\omega(u)$  determina los valores que caen dentro del hipercubo centrado en el origen y de longitud uno. Así, se puede deducir que la función  $\omega\left(\frac{x-x_i}{h_n}\right)$  dará el valor de uno en el caso de que  $x_i$  se encuentre dentro de  $\Omega_n$ .

El número de muestras que caen dentro del hipercubo viene dado por:

$$k_n = \sum_{i=1}^n \omega\left(\frac{x-x_i}{h_n}\right), \quad (1.7)$$

y mediante la ecuación 1.4 se tiene

$$p_n(x) = \frac{1}{n} \frac{1}{h_n^d} \sum_{i=1}^n \omega\left(\frac{x-x_i}{h_n}\right). \quad (1.8)$$

Se evidencia que la estimación de  $p(x)$  se expresa como un promedio de funciones de  $x$  con las muestras  $x_i$ , de esta forma la función ventana cumple el papel de interpolador.

$$\int p_n(x) dx = \frac{1}{n} \frac{1}{h_n} \sum_{i=1}^n \int \omega\left(\frac{x-x_i}{h_n}\right) dx = \frac{1}{n} \frac{1}{h_n} \sum_{i=1}^n h_n = \frac{1}{n} \sum_{i=1}^n 1 = 1.$$

$V_n = h_n^d$  es el normalizador del interpolador ya que es igual a su integral.

Un ejemplo es la función ventana Gaussiana con valores centrados en  $x_i$  y con un ancho  $\sigma$ , su expresión viene dado por

$$p_n(x) = \frac{1}{n} \frac{1}{\sqrt{2\pi}\sigma} \sum_{i=1}^n e^{-\frac{(x-x_i)^2}{2\sigma^2}},$$

donde  $e^{-\frac{(x-x_i)^2}{2\sigma^2}}$  es la función ventana.

### 1.2.3 Convergencia de la media y la covarianza

#### Convergencia de la media

Consideramos  $\bar{p}_n(x)$  la media de  $p_n(x)$

$$\begin{aligned}\bar{p}_n(x) &= \mathbb{E}(p_n(x)) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \frac{1}{V_n} \omega \left( \frac{x - x_i}{h_n} \right) \right] \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{V_n} \mathbb{E} \left[ \omega \left( \frac{x - x_i}{h_n} \right) \right],\end{aligned}$$

al considerar  $x_i$  valores i.i.d, podemos reemplazarlas por una constante  $s$

$$\begin{aligned}&= \frac{1}{n} \sum_{i=1}^n \frac{1}{V_n} \mathbb{E} \left[ \omega \left( \frac{x - s}{h_n} \right) \right] \\ &= \frac{1}{n} \frac{1}{V_n} \mathbb{E} \left[ \omega \left( \frac{x - s}{h_n} \right) \right] \\ &= \int \frac{1}{V_n} \omega \left( \frac{x - s}{h_n} \right) p(s) ds \\ &= \int \delta_n(x - s) p(s) ds.\end{aligned}$$

Donde  $\delta_n(x) = \frac{1}{V_n} \omega \left( \frac{x}{h_n} \right)$ .

La expresión se puede observar como una convolución entre la densidad desconocida y la función de ventana. A medida que  $V_n$  se aproxima a cero,  $\delta_n(x - v)$  se aproxima a una función delta centrada en  $x$ . Así, si  $p$  es continua en  $x$  y el  $\lim_{n \rightarrow \infty} V_n = 0$ , se asegura que  $\bar{p}_n(x)$  se aproximará a  $p(x)$

**Convergencia de la covarianza**

$$\begin{aligned}
 \sigma_n^2(x) &= \sum_{i=1}^n \mathbb{E} \left[ \left( \frac{1}{nV_n} \omega \left( \frac{x-x_i}{h_n} \right) - \frac{1}{n} \bar{p}_n(x) \right)^2 \right] \\
 &= \sum_{i=1}^n \mathbb{E} \left[ \frac{1}{n^2 V_n^2} \omega^2 \left( \frac{x-x_i}{h_n} \right) - \frac{2}{n^2 V_n} \omega \left( \frac{x-x_i}{h_n} \right) \bar{p}_n(x) + \frac{1}{n^2} \bar{p}_n^2(x) \right] \\
 &= \sum_{i=1}^n \mathbb{E} \left[ \frac{1}{n^2 V_n^2} \omega^2 \left( \frac{x-x_i}{h_n} \right) \right] - \mathbb{E} \left[ \frac{2}{n^2 V_n} \omega \left( \frac{x-x_i}{h_n} \right) \bar{p}_n(x) \right] + \mathbb{E} \left[ \frac{1}{n^2} \bar{p}_n^2(x) \right] \\
 &= \sum_{i=1}^n \frac{1}{n^2 V_n^2} \mathbb{E} \left[ \omega^2 \left( \frac{x-x_i}{h_n} \right) \right] - \frac{2}{n^2 V_n} \bar{p}_n(x) \mathbb{E} \left[ \omega \left( \frac{x-x_i}{h_n} \right) \right] + \frac{1}{n^2} \bar{p}_n^2(x) \\
 &= \sum_{i=1}^n \frac{1}{n^2 V_n^2} \mathbb{E} \left[ \omega^2 \left( \frac{x-x_i}{h_n} \right) \right] - \frac{2}{n^2 V_n} \bar{p}_n(x) \bar{p}_n(x) V_n + \frac{1}{n^2} \bar{p}_n^2(x) \\
 &= \sum_{i=1}^n \frac{1}{n^2 V_n^2} \mathbb{E} \left[ \omega^2 \left( \frac{x-x_i}{h_n} \right) \right] - \frac{2}{n^2 V_n} \bar{p}_n^2(x) + \frac{1}{n^2} \bar{p}_n^2(x) \\
 &= \sum_{i=1}^n \frac{1}{n^2 V_n^2} \mathbb{E} \left[ \omega^2 \left( \frac{x-x_i}{h_n} \right) \right] - \frac{1}{n^2} \bar{p}_n^2(x) \\
 &= \sum_{i=1}^n \frac{1}{n^2 V_n^2} \mathbb{E} \left[ \omega^2 \left( \frac{x-s}{h_n} \right) \right] - \frac{1}{n^2} \bar{p}_n^2(x) \\
 &= n \left( \frac{1}{n^2 V_n^2} \mathbb{E} \left[ \omega^2 \left( \frac{x-s}{h_n} \right) \right] - \frac{1}{n} \bar{p}_n^2(x) \right) \\
 &= \frac{1}{n V_n^2} \mathbb{E} \left[ \omega^2 \left( \frac{x-s}{h_n} \right) \right] - \frac{1}{n} \bar{p}_n^2(x) \\
 &= \frac{1}{n V_n} \int \frac{1}{V_n} \omega^2 \left( \frac{x-s}{h_n} \right) p(s) ds - \frac{1}{n} \bar{p}_n^2(x),
 \end{aligned}$$

se puede acotar la expresión eliminando el segundo término

$$\begin{aligned}
 \sigma_n^2(x) &\leq \frac{1}{n V_n} \int \frac{1}{V_n} \omega^2 \left( \frac{x-s}{h_n} \right) p(s) ds \\
 &= \frac{1}{n V_n} \int \omega \left( \frac{x-s}{h_n} \right) \omega \left( \frac{x-s}{h_n} \right) p(s) ds \\
 &\leq \frac{1}{n V_n} \int \omega \left( \frac{x-s}{h_n} \right) \sup(\omega(\cdot)) p(s) ds \\
 &= \frac{1}{n V_n} \sup(\omega(\cdot)) \int \omega \left( \frac{x-s}{h_n} \right) p(s) ds \\
 &= \frac{1}{n V_n} \sup(\omega(\cdot)) \bar{p}_n(x) \\
 &= \frac{\sup(\omega(\cdot)) \bar{p}_n(x)}{n V_n}.
 \end{aligned}$$

## 1.3 Estimación de la densidad de *kernel* univariable

### El estimador de densidad *kernel* univariable

La función viene dada por:

$$\bar{f}(x) = (nh^{-1}) \sum_{i=1}^n K\left(\frac{(x - x_i)}{h}\right).$$

Donde  $K$  es una función *kernel* que satisface  $\int K(x)dx = 1$  y  $h$  es un número positivo denominado *bandwidth*. Utilizando la siguiente parametrización del *kernel*: ( $K_h(s) = h^{-1}K(\frac{s}{h})$ ), la estimación se puede reescribir

$$\bar{f}(x) = n^{-1} \sum_{i=1}^n K_h(x - x_i). \quad (1.9)$$

### 1.3.1 Los criterios MSE y MISE

Se tiene la estimación de la función de densidad  $p(x)$  y resulta necesario el análisis del rendimiento del estimador de densidad. Para esto se utiliza los errores asociados a los estimación.

#### Error cuadrático medio (MSE)

El Error Cuadrático Medio (MSE siglas en inglés) viene dado por

$$\text{MSE}(\bar{\phi}) = \mathbb{E}(\bar{\phi} - \phi)^2. \quad (1.10)$$

Donde  $\bar{\phi}$  es la estimación y  $\phi$  la función a estimar. Esta expresión se le puede relacionar con la varianza y el sesgo, es decir

$$\text{MSE}(\bar{\phi}) = \text{var}(\bar{\phi}) + (\mathbb{E}[\bar{\phi}] - \phi)^2. \quad (1.11)$$

Definimos  $\bar{\phi} = \bar{f}(x)$ , así

$$\begin{aligned}
 \mathbb{E}[\bar{f}(x)] &= \mathbb{E}\left[n^{-1} \sum_{i=1}^n K_h(x - x_i)\right] \\
 &= \mathbb{E}[K_h(x - s)] \\
 &= \int K_h(x - y)f(y)dy \\
 &= (K_h * f)(x).
 \end{aligned} \tag{1.12}$$

al ser variables aleatorias (i.i.d) se puede colocar una solo variable como  $s$ , además, por definición de  $\mathbb{E}$  e introduciendo la notación de convolución se obtiene la expresión final.

Esto permite escribir el sesgo de la forma

$$\begin{aligned}
 \text{sesgo}(\bar{\phi}) &= \mathbb{E}[\bar{\phi}] - \phi \\
 &= \mathbb{E}[\bar{f}(x)] - f(x) \\
 &= (K_h(x) * f(x)) - f(x).
 \end{aligned}$$

La convolución es una operación de suavizado, así, el sesgo es la diferencia entre un suavizado de  $f$  y la propia  $f$ . En cuanto a la varianza se tiene

$$\begin{aligned}
 \text{Var}(\bar{f}(x)) &= \text{Var}\left(n^{-1} \sum_{i=1}^n K_h(x - x_i)\right) \\
 &= n^{-1}\text{Var}(K_h(x - s)),
 \end{aligned}$$

con la definición de  $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

$$\begin{aligned}
 &= n^{-1}\mathbb{E}[K_h^2(x - s)] - (\mathbb{E}[K_h(x - s)])^2 \\
 &= n^{-1}\left(\int K_h^2(x - y)f(y)dy - \left(\int k_h(x - y)f(y)dy\right)^2\right) \\
 &= n^{-1}(k_h^2 * f)(x) - (K_h * f)^2(x).
 \end{aligned} \tag{1.13}$$

Reemplazando en la Ecuación 1.11

$$\text{MSE}(\bar{f}(x)) = n^{-1}((K_h^2 * f)(x) - (K_h * f)^2(x)) + ((K_h * f)(x) - f(x))^2.$$

El análisis dado se establece para un punto fijo, sin embargo, para un enfoque general se debe realizar estimando sobre toda la línea real. Así, se considera un criterio que mida de forma general la distancia entre  $\bar{f}(x)$  y  $f(x)$

$$\text{ISE}(\bar{f}(x)) = \int (\bar{f}(x) - f(x))^2 dx.$$

Este se lo conoce como el Error Cuadrático integrado y se lo define para un estimador de la densidad de  $\bar{f} \in L_2$ .

**Error cuadrático medio integrado (MISE)**

Suprime la aleatoriedad procedente de cada muestra individual promediando los resultados obtenidos

$$\begin{aligned} \text{MISE}(\bar{f}) &= \mathbb{E}[\text{ISE}(\bar{f})] \\ &= \mathbb{E} \int (\bar{f}(x) - f(x))^2 dx \\ &= \int \mathbb{E}[\bar{f}(x) - f(x)]^2 dx \\ &= \int \text{MSE}(\bar{f}(x)) dx, \end{aligned}$$

se da un intercambio de la esperanza matemática con la integral mediante el teorema de Fubini-Tonelli.

Recordando;

$$\text{MSE}(\bar{f}) = \frac{1}{n} [(k_h^2 * f)(x) - (K_h * f)^2(x)] + [(k_h * f)(x) - f(x)]^2.$$

Además:

$$\begin{aligned} i) \int ((k_h * f)(x) - f(x))^2 dx &= \int ((k_h * f)(x) - 2(k_h * f)(x) + f^2(x)) dx \\ &= \int (k_h * f)^2(x) - 2 \int (k_h * f)(x) dx + \int f^2(x) dx \end{aligned}$$

$$\begin{aligned} ii) \int \text{Var}(\bar{f}) dx &= \frac{1}{n} \int (k_h^2 * f)(x) - (k_h * f)^2(x) dx \\ &= \frac{1}{n} \int (k_h^2 * f)(x) dx - \int (k_h * f)^2(x) dx \\ &= \frac{1}{nh} \int k^2(x) dx - \frac{1}{n} \int (k_h * f)^2(x) dx. \end{aligned}$$

Se obtiene

$$\text{MISE} \bar{f}(x) = \frac{1}{n} \int [(K_h^2 * f)(x) - (K_h * f)^2(x)] dx + \int [(K_h * f)(x) - f(x)]^2 dx,$$

y con ayuda de las expresiones *i)* y *ii)* se tiene

$$\begin{aligned} \text{MISE} \bar{f}(x) &= \frac{1}{nh} \int K^2(x) dx + (1 - \frac{1}{n}) \int (K_h * f)^2(x) dx \\ &\quad - 2 \int (K_h * f(x)) f(x) dx + \int f(x)^2 dx. \end{aligned} \tag{1.14}$$

### 1.3.2 Orden y Notación Asintótica

La notación de orden  $O$  y  $o$  se define para funciones generales de valor real (Serfiing, 1980, pág. 1), y se trata de una herramienta matemática útil para un análisis simple de estimador de núcleo de densidad.

#### Definición 1.10

Sea  $a_n$  y  $b_n$  sucesiones de números reales. Decimos que  $a_n$  es de orden  $b_n$  o  $a_n$  es tan grande que  $b_n$  cuando  $n \rightarrow \infty$ , y se denota como:

$a_n = O(b_n)$  cuando  $n \rightarrow \infty$ , si y sólo si  $\limsup_{n \rightarrow \infty} |a_n/b_n| < \infty$ .

Se dice que  $a_n$  es de orden pequeño  $b_n$  y se escribe como:

$$a_n = o(b_n) \text{ cuando } n \rightarrow \infty \text{ si y solo si } \lim_{n \rightarrow \infty} |a_n/b_n| = 0$$

**Ejemplo.** Sea  $c_n = \frac{n+4}{n^3+1}$ , podemos escribir  $c_n = O(n^{-2})$

### 1.3.3 Ampliación de Taylor

Una herramienta útil para abarcar la obtención de la aproximación asintótica es la ampliación de Taylor, teniendo en cuenta el siguiente teorema:

#### Teorema 1.1

Supongamos que  $f$  es una función real definida en  $\mathbb{R}$  y sea  $x \in \mathbb{R}$ . Además, asumimos que  $f$  tiene  $p$  derivadas continuas en un intervalo  $(x - \delta, x + \delta)$  para un  $\delta > 0$ . Luego para cualquier sucesión  $\alpha_n$  convergente a cero,

$$f(x + \alpha_n) = \sum_{j=0}^p (\alpha_n^j / j!) f(x)^j + o(\alpha_n^p).$$

El teorema de Taylor nos permite aproximar los valores de una función cerca de un punto dado en términos de derivadas de orden superior en ese punto, siempre que la función sea lo suficientemente suave como para que existan dichas derivadas.

**Ejemplo.** Sea  $\alpha_n$  una sucesión que converge a cero. Mediante el teorema de Taylor tenemos:

$$(x + \alpha_n)^{-1/2} = x^{-1/2} - \frac{1}{2}\alpha_n x^{-3/2} + \frac{3}{8}\alpha_n^2 x^{-5/2} + o(\alpha_n^2),$$

para todo  $x > 0$ .

#### 1.3.4 Aproximaciones asintóticas de MSE y MISE

El desafío con las expresiones MSE y MISE radica en su dependencia complicada del ancho de banda. Esto dificulta la comprensión de cómo afecta el ancho de banda al rendimiento del estimador de densidad del núcleo. En esta sección, exploramos una forma de abordar este problema mediante la derivación de aproximaciones para muestras grandes de los términos principales de varianza y sesgo. Estas aproximaciones son mucho más simples y nos brindan una comprensión más profunda del impacto del ancho de banda. También nos permiten determinar la tasa de convergencia del estimador de densidad del núcleo y encontrar el ancho de banda óptimo para MISE.

Se establece las siguientes suposiciones:

1. Se establece la hipótesis de que la función de densidad  $f$  satisface las siguientes condiciones:
  - La segunda derivada de  $f$ , denotada como  $f''$ , es una función continua.
  - La función  $f$  es integrable al cuadrado, lo cual implica que  $\int |f(x)|^2 dx$  es finita..
  - $f$  es una función monótona, lo que implica que su valor crece o decrece de manera constante a medida que  $x$  aumenta.
2. El ancho de banda  $h = h_n$  es una sucesión no aleatoria de números positivos. Para simplificar la notación en nuestros cálculos, omitiremos la dependencia de  $h$  respecto a  $n$ . También asumiremos que  $h$  satisface

$$\lim_{n \rightarrow \infty} h = 0 \text{ y } \lim_{n \rightarrow \infty} nh = \infty,$$

cuando  $n$  tiende a infinito. Esto implica que  $h$  tiende a cero, pero a un ritmo más lento que  $n^{-1}$ .

3. Se define el núcleo  $K$  como una función de densidad de probabilidad acotada, denotada como  $K(x)$ , que satisface las siguientes propiedades:

- Existe una constante  $M > 0$  tal que  $|K(x)| \leq M$  para todo  $x$  en su dominio.
- El cuarto momento de  $K(x)$ , denotado como  $\int x^4 K(x) dx$ , es finito.
- Simetría respecto al origen, es decir,  $K(x) = K(-x)$  para todo  $x$  en su dominio.

En esencia, una función monótona se caracteriza por ser monótona tanto en el intervalo  $(-\infty, -M)$  como en  $(M, \infty)$  para algún valor  $M > 0$ . Introducimos, además una notación para simplificar los procedimientos;  $\mu_2(K) = \int z^2 K(z) dz$ . Vamos a considerar primeramente la estimación  $f(x)$  en  $x \in \mathbb{R}$ . A partir de la expresión 1.12 y mediante un cambio de variable:

$$\begin{aligned} \mathbb{E}[\bar{f}(x)] &= \mathbb{E}K_h(x - X) \\ &= \int K_h(x - X) \\ &= \int K_h(x - y)f(y)dy \\ &= \int K(z)f(x - zh)dz. \end{aligned}$$

Desarrollando la serie de Taylor de  $f(x - hz)$  en torno a  $x$ , obtenemos:

$$\begin{aligned} f(x - hz) &= \frac{hz^0}{0!}f^{(0)}(x) + \frac{hz}{1!}f'(x) + \frac{hz^2}{2!}f''(x) + o(h^2) \\ &= f(x) + hzf'(x) + \frac{1}{2}h^2z^2f''(x) + o(h^2). \end{aligned}$$

Así se tiene

$$\begin{aligned} \mathbb{E} &= \int K(z)f(x - hz)dz \\ &= \int K(z)[f(x) - hzf'(x) + \frac{1}{2}h^2z^2f''(x) + o(h^2)]dz \\ &= \int K(z)f(x)dz + \frac{1}{2} \int K(z)h^2z^2f''(x)dz + \int K(z)o(h^2)dz \\ &= f(x) + \frac{1}{2}h^2f''(x) \int K(z)z^2dz + o(h^2), \end{aligned}$$

utilizando:

$$\int K(z)dz = 1 \quad \int zK(z)dz = 0 \quad \text{y} \quad \int z^2K(z)dz < \infty.$$

Con uso de la notación  $\mu_2(k) = \int z^2K(z)dz$ , la expresión del sesgo viene dado por

$$\begin{aligned} \mathbb{E}[\bar{f}(x)] - f(x) &= \frac{1}{2}h^2 f''(x) \int K(z)z^2 dz + o(h^2) \\ &= \frac{1}{2}h^2 f''(x)\mu_2(x)^2 + o(h^2) \end{aligned} \quad (1.15)$$

Notamos que el sesgo es de orden  $h^2$  lo cual implica que  $\bar{f}(x)$  es asintóticamente insesgada.

El sesgo tiende a ser significativo cuando la magnitud de la segunda derivada de  $f$  es elevada, lo cual sucede en áreas con una curvatura densa. En muchas distribuciones, esto se manifiesta en los picos con sesgo negativo y los valles con sesgo positivo, indicando que  $\bar{f}(x)$  tiende a atenuar estas características en términos generales. (Wand., Jones,1995, págs. 20-21 ).

Para la varianza, notamos la expresión 1.13

$$\begin{aligned} \text{Var}(\bar{f}(x;h)) &= n^{-1}[(K_h^2 * f)(x) - (K_h * f)^2(x)] \\ &= n^{-1} \left[ \int K_h^2(x-y)f(y)dy - \left( \int K_h(x-y)f(y)dy \right)^2 \right] \\ &= n^{-1} \left[ h^{-1} \int K^2(z)f(x-hz)dz - \left( \int K(z)f(x-hz)dz \right)^2 \right] \\ &= nh^{-1} \int K^2(z)f(x-hz)dz - n^{-1} \left( \int K(z)f(x-hz)dz \right)^2 \\ &= nh^{-1} \int K^2(z)f(x-hz)dz - n^{-1}[\mathbb{E}[\bar{f}(x)]]^2 \\ &= nh^{-1} \int K^2(z)(f(x) + o(1))dz - n^{-1}(f(x) + o(1))^2 \\ &= nh^{-1} \int K^2(z)f(x)dz + o((nh)^{-1}) \\ &= nh^{-1} \int K^2(z)dzf(x) - o((nh)^{-1}). \end{aligned}$$

## Capítulo 1. Fundamentos teóricos

---

Una notación conveniente es  $R(g) = \int g(x)^2 dx$  para cualquier función  $g$  que sea cuadrado integrable. Esto nos facilita expresar la varianza como:

$$\text{Var}(\bar{f}(x)) = (nh)^{-1}R(K)f(x) + o((nh)^{-1}). \quad (1.16)$$

Dado que la varianza tiene un orden de  $(nh)^{-1}$ , la condición 2 garantiza que  $\text{Var}(\bar{f}(x))$  tienda a cero. Con 1.16 y elevando al cuadrado la expresión dada por 1.15 se obtiene:

$$\begin{aligned} (\mathbb{E}[\bar{f}(x)] - f(x))^2 &= \left(\frac{1}{2}h^2\mu_2(K)f''(x) + o(h^2)\right)^2 \\ &= \frac{1}{4}h^4\mu_2(K)f''(x)^2 + h^2\mu_2(K)f''(x)o(h^2) + o(h^4), \end{aligned}$$

y de acuerdo a 1.10:

$$\begin{aligned} \text{MSE}(\bar{f}) &= \text{Var}(\bar{f}) + (\mathbb{E}[\bar{f}(x)] - f(x))^2 \\ &= (nh)^{-1} \int K^2 dz f(x) + o((nh)^{-1}) + \frac{1}{4}h^4\mu_2(K)f''(x)^2 + h^2\mu_2(K)f''(x)o(h^2) + o(h^4) \\ &= (nh)^{-1}R(K)f(x) + \frac{1}{4}h^4\mu_2(K)^2 + o((nh)^{-1} + h^4). \end{aligned}$$

Si llevamos a cabo la integración de esta expresión, en virtud de nuestra suposición de que  $f$  es integrable, se obtiene:

$$\text{MISE}(\bar{f}(x)) = \text{AMISE}(\bar{f}(x)) + o((nh)^{-1} + h^4),$$

donde,

$$\text{AMISE}(\bar{f}(x)) = (nh)^{-1}R(K) + \frac{1}{4}h^4\mu_2(K)^2R(f''). \quad (1.17)$$

Consiguiendo el denominador MISE asintótico, siendo una expresión mucho más sencilla de comprender que la expresión dada para el MISE en 1.14.

El cuadrado del sesgo integrado está relacionado asintóticamente con  $h^4$ , lo que implica que para reducir esta cantidad es necesario que  $h$  sea pequeña. Sin

embargo, al reducir  $h$ , aumenta el término principal de la varianza integrada, ya que está proporcionalmente relacionado con  $(nh)^{-1}$ . Por lo tanto, a medida que  $n$  aumenta,  $h$  debe ajustarse de manera que todos los componentes del MISE se reduzcan. Este equilibrio entre la varianza y el sesgo es una representación matemática de la importancia crítica del ancho de banda [11].

Se establece que cuando  $h$  es muy pequeño, la función  $K_h$  tiene numerosos picos (nodos), lo que resulta en una alta variabilidad, ya que en múltiples muestras de  $f$  los picos podrían aparecer en lugares distintos. Sin embargo, este enfoque tiene poco sesgo. Por otro lado, al suavizar más la función, es decir, aumentar  $h$ , la variabilidad disminuye a expensas de introducir sesgo: a medida que  $h$  crece, aumenta el sesgo ya que todas las características se suavizan, pero la varianza disminuye significativamente ya que no se toman en cuenta algunos de los datos. La ventaja del AMISE es que el ancho de banda óptimo, denotado como  $h_{AMISE}$ , tiene una expresión matemática cerrada.

$$h_{AMISE} = \left[ \frac{R(K)}{\mu_2(K)^2 R(f'') n} \right]^{\frac{1}{5}}$$

Para más detalle ver [11].

## 1.4 El estimador de densidad de kernel multivariable

El estimador de densidad *kernel* en  $d$ -dimensiones se define como

$$\bar{f}(x) = n^{-1} \sum_{i=1}^n K_H(x - x_i), \quad (1.18)$$

donde

$$K_H = |H|^{-1/2} K(H^{-1/2}x). \quad (1.19)$$

Existen dos técnicas comunes para generar *kernels* multivariantes a partir de un *kernel* univariante simétrico  $K(x)$ . El primero es a partir del producto de los *kernels* univariante y el segundo a partir de la rotación  $K(x) \in R^d$  ( $K^S(x)$  es radialmente

simétrico:

$$K^P(x) = \prod_{i=1}^d K(x_i) \quad K^S(x) = s_{k,d} K((x^T x)^{1/2}),$$

donde  $s_{k,d}^{-1} = \int_{\mathbb{R}^d} K_1((x^T x)^{1/2}) dx$ , que asegura que  $K^S(x)$  se integre a uno.

Una simplificación de 1.18 se puede obtener restringiendo la forma de  $H$ . Sea  $\mathcal{F}$  la clase de matrices  $d \times d$  simétricas definidas positivamente y  $H \in D$ , donde  $D \subseteq \mathbb{F}$  y es la subclase de matrices  $d \times d$  definidas diagonales positivas. Luego, Sea  $H \in D$ ,  $H = \text{diag}(h_1^2, \dots, h_d^2)$ . Para  $H \in D$  el estimador del kernel se puede escribir como

$$\bar{f}(x) = n^{-1} \left( \prod_{s=1}^d h_s \right)^{-1} \sum_{i=1}^n K \left( \frac{x_1 - x_{i1}}{h_1}, \dots, \frac{x_d - x_{id}}{h_d} \right)$$

Una forma más simplificada se obtiene a través de la restricción  $H \in S$ , donde  $S = \{h^2 I : h > 0\}$ , y  $I$  es la matriz identidad. Con esto, se resulta en el estimador kernel con un único ancho de banda.

$$\bar{f}(x) = n^{-1} h^{-d} \sum_{i=1}^n K \left( \frac{(x - x_i)}{h} \right). \quad (1.20)$$

Al igual que en el caso univariante, se necesita de un enfoque metodológico para la evaluación matemática del rendimiento de un estimador de densidad kernel multivariado.

### 1.4.1 Aproximaciones asintóticas MISE

Es posible, como en el caso univariante, obtener una aproximación asintótica simple al error cuadrático medio integrado (**MISE**) de un estimador de densidad kernel multivariado bajo ciertas suposiciones de suavidad en la densidad  $f$ . Estas suposiciones son necesarias para utilizar una versión multivariada del teorema de Taylor.

**Teorema 1.2**

(Teorema de Taylor multivariante) Sea  $g$  una función  $d$ -variante y  $\alpha_n$  una sucesión de vectores  $d \times 1$  con todos sus componentes tendiendo a cero. Sea, además,  $\mathcal{D}_g(x)$  sea el vector de derivadas parciales de primer orden de  $g$  y  $\mathcal{H}_g(x)$  la matriz Hessiana de  $g$  con entradas  $(i, j)$  igual a  $\frac{\partial^2}{\partial x_i \partial x_j} g(x)$ . Asumiendo que todas las entradas de  $\mathcal{H}_g(x)$  son continuas en un vecindario de  $x$ , luego

$$g(x + \alpha_n) = g(x) + \alpha_n^T \mathcal{D}_g(x) + \frac{1}{2} \alpha_n^T \mathcal{H}_g(x) \alpha_n + o(\alpha_n^T \alpha_n).$$

Para las aproximaciones asintóticas del error cuadrático medio integrado (MISE), se realiza las siguientes suposiciones sobre  $f$ ,  $H$  y  $K$ :

1. Cada entrada de  $\mathcal{H}_f(x)$  es denotada como una función  $f(t)$  que es continua a trozos y su integral al cuadrado,  $\int |f(t)|^2 dt$ , es finita.
2.  $\mathbf{H} = \mathbf{H}_n$  es una sucesión de matrices de *bandwidth* donde  $n^{-1} |\mathbf{H}|^{-1/2}$  y todas las entradas de  $\mathbf{H}$  tienden a cero cuando  $n$  se acerca a infinito. Además, se asume que la relación entre los valores propios más grandes y más pequeños de  $\mathbf{H}$  está acotada para todos los valores de  $n$ .
3.  $K$  es una función *kernel*  $d$ -variable que está acotada y tiene soporte compacto, y cumple con las siguientes condiciones

$$\int K(z) dz = 1 \qquad \int z K(z) dz = 0,$$

y

$$\int z z^T K(z) dz = \mu_2(K) I,$$

donde  $\mu_2(K) = \int z_i^2 K(z) dz$  es independiente de  $i$ .

Se puede notar que todas las densidades de probabilidad esféricamente simétricas con soporte compacto y los *kernel* de productos basados en densidades univariantes simétricas con soporte compacto cumplen la condición 3. Además, se puede observar que el supuesto de soporte compacto de  $K$  puede ser eliminado al imponer condiciones más complejas sobre  $f$ .

## Capítulo 1. Fundamentos teóricos

---

Para introducirse en las aproximaciones de MISE, al igual que en la sección 1.3.4, se establece la esperanza de  $\bar{f}(x)$ , utilizando la versión multivariada del teorema de Taylor.

$$\begin{aligned}
 \mathbb{E}\bar{f}(x) &= \int K_{\mathbf{H}}(x-y)f(y)dy \\
 &= \int K(\mathbf{z})f(x-\mathbf{H}^{-1/2}\mathbf{z})d\mathbf{z} \\
 &= \int K(\mathbf{z})(f(x)-(\mathbf{H}^{-1/2}\mathbf{z})^T\mathcal{D}_f(x)+\frac{1}{2}(\mathbf{H}^{1/2}\mathbf{z})^T\mathcal{H}_f(x)(\mathbf{H}^{1/2}\mathbf{z}))d\mathbf{z}+o(\text{tr}(\mathbf{H})) \\
 &= f(x)\int \mathbf{z}^T\mathbf{H}^{1/2}\mathcal{D}_f(x)K(\mathbf{z})d\mathbf{z}+\frac{1}{2}\int \mathbf{z}^T\mathbf{H}^{1/2}\mathcal{H}_f(x)\mathbf{H}^{1/2}\mathbf{z}K(\mathbf{z})d\mathbf{z}+o(\text{tr}(\mathbf{H})) \\
 &= f(x)+\frac{1}{2}\text{tr}(\mathbf{H}^{1/2}\mathcal{H}_f(x)\mathbf{H}^{1/2})\int \mathbf{z}\mathbf{z}^TK(\mathbf{z})d\mathbf{z}+o(\text{tr}(\mathbf{H})) \\
 &= f(x)+\frac{1}{2}\mu_2(K)\text{tr}(\mathbf{H}\mathcal{H}_f(x))+o(\text{tr}(\mathbf{H})).
 \end{aligned}$$

La expresión del sesgo viene dado por

$$\mathbb{E}\bar{f}(x)-f(x)\approx\frac{1}{2}\mu_2(K)\text{tr}(\mathbf{H}\mathcal{H}_f(x)). \quad (1.21)$$

Y la varianza de  $\bar{f}(x)$  por

$$\begin{aligned}
 \text{Var}\bar{f}(x) &= n^{-1}\left(|\mathbf{H}|^{-1/2}\int K(\mathbf{z})^2f(x-\mathbf{H}^{1/2}\mathbf{z})d\mathbf{z}-\left(\int K(\mathbf{z})f(x-\mathbf{H}^{1/2}\mathbf{z})d\mathbf{z}\right)^2\right) \\
 &= n^{-1}|\mathbf{H}|^{-1/2}R(K)+o(n^{-1}|H|^{-1/2}),
 \end{aligned} \quad (1.22)$$

donde  $R(K)=\int K(\mathbf{z}^2)d\mathbf{z}$ .

De acuerdo a las suposiciones de integrabilidad en las condiciones 1 y 3, se puede obtener el AMISE (Error Cuadrático Medio Integrado Asintótico) del estimador de densidad kernel multivariante de la siguiente manera:

$$AMISE(\bar{f}(x))=n^{-1}|H|^{-1/2}R(K)+\frac{1}{4}\mu_2(K)^2\int \text{tr}^2(\mathbf{H}\mathcal{H}_f(x))dx. \quad (1.23)$$

Se puede expandir el término de sesgo al cuadrado integrado utilizando los resultados

- Si  $\mathbf{A}$  es una matriz cuadrada, la traza de  $\mathbf{A}$ , representada como  $\text{tr}\mathbf{A}$ , se define como la suma de las entradas diagonales de  $\mathbf{A}$ . La traza tiene la propiedad de que

$$\text{tr}(\mathbf{AB})=\text{tr}(\mathbf{BA}).$$

#### 1.4. El estimador de densidad de kernel multivariable

---

Para una matriz  $\mathbf{A}$  de tamaño  $d \times d$ , el vector de  $\mathbf{A}$ , llamado  $\text{vec } \mathbf{A}$ , es el resultado de apilar las columnas de  $\mathbf{A}$  una debajo de otra en un vector de tamaño  $d^2 \times 1$ , en orden de izquierda a derecha. Por otro lado, el *vector-half* de  $\mathbf{A}$ , conocido como  $\text{vech } \mathbf{A}$ , se obtiene a partir de  $\text{vec } \mathbf{A}$  eliminando las entradas diagonales superiores de  $\mathbf{A}$ , resultando en un vector de tamaño  $\frac{1}{2}d(d+1) \times 1$ .

- De acuerdo a la notación anterior, otro resultado necesario es:

$$\text{tr}(\mathbf{A}^T \mathbf{B}) = (\text{vec}^T \mathbf{A})(\text{vec } \mathbf{B})$$

Retomando el término del sesgo al cuadrado

$$\begin{aligned} \int \text{tr}^2(\mathbf{H}\mathcal{H}_f(x))dx &= \int (\text{vec}^T \mathbf{H}) (\text{vec } \mathcal{H}_f(x)) (\text{vec}^T \mathcal{H}_f(x)) (\text{vec } \mathbf{H}) dx \\ &= \int (\text{vech}^T \mathbf{H}) \mathbf{D}_d^T (\text{vec } \mathcal{H}_f(x)) (\text{vech}^T \mathcal{H}_f(x)) \mathbf{D}_d(\mathbf{H}) dx \\ &= (\text{vech}^T \mathbf{H}) \Psi_{\mathcal{F}} (\text{vech } \mathbf{H}), \end{aligned}$$

donde  $\Psi_{\mathcal{F}}$  es una matriz de tamaño  $\frac{1}{2}d(d+1) \times \frac{1}{2}d(d+1)$  dado por

$$\Psi_{\mathcal{F}} = \int \text{vech}(2\mathcal{H}_f(x) - dg\mathcal{H}_f(x)) \times \text{vech}^T(2\mathcal{H}_f(x) - dg\mathcal{H}_f(x)) dx,$$

donde  $dg\mathbf{A}$  es igual a  $\mathbf{A}$ . Así, una forma alternativa de 1.23 es

$$AMISE(\bar{f}(x)) = n^{-1} |\mathbf{H}|^{-1/2} R(K) + \frac{1}{4} \mu_2(K)^2 (\text{vech}^T \mathbf{H}) \Psi_{\mathcal{F}} (\text{vech } \mathbf{H}) \quad (1.24)$$

La información a detalle se encuentra en [11]

# 2

---

## *Mean Shift*

---

Un espacio de características, se refiere a la representación de los datos en un espacio multidimensional donde cada dimensión representa una característica particular de los datos. En el contexto de la estimación de densidad no paramétrica mediante enfoque de agrupamiento, esta representación del espacio de características se interpreta como la función de densidad de probabilidad empírica del parámetro que se representa. Las regiones densas en este espacio de características corresponden a los máximos locales de la función de densidad, es decir, a los modos de la densidad desconocida que son puntos donde la densidad de probabilidad es relativamente alta en comparación con los puntos alrededores del espacio. Las áreas que albergan una considerable concentración de datos en el espacio de características coinciden con los picos locales presentes en la función de densidad. En esta sección se aborda todo los tópicos anteriores vistas en el capítulo 1 y su funcionamiento en relación al procedimiento del método Mean Shift.

### **2.1 Descripción de Mean Shift**

El Mean Shift es un proceso iterativo que busca converger hacia un máximo local al estimar el gradiente normalizado de la función de densidad en cada punto. Mediante múltiples iteraciones desde distintos puntos de partida, se pueden identificar las modas de la densidad, es decir, las regiones donde se concentra una mayor densidad de datos.

### 2.1.1 Estimación de la función de densidad de probabilidad

La obtención de la función de densidad se da mediante el uso de un método de estimación basado en funciones *kernel* conocida como *Windows Parzen* y de la idea de la sección multivariante de un estimador en 1.4.

Sea  $x_1, \dots, x_n$  valores en una región  $d$ -dimensional, la estimación de densidad en este caso multivariado con *kernel*  $K(x)$  y una matriz *bandwidth*  $d \times d$  simétricamente positiva, viene dado por:

$$\bar{f}(x) = \frac{1}{2} \sum_{i=1}^n K_H(x - x_i), \quad (2.1)$$

donde

$$K_H(x) = |H|^{-1/2} K(H^{-1/2}x)$$

y  $K$  es un función *kernel* y  $H$  viene definida de la forma de la estimación de densidad en 1.20

$$\bar{f}(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right). \quad (2.2)$$

#### Función kernel y sus propiedades

##### Definición 2.1

En el espacio euclidiano de  $n$  dimensiones,  $R^n$ , podemos representar la  $i$  -ésima componente de  $x$  en  $X$  como  $x_i$ . La norma de  $x \in X$  es no negativo  $\|x\|$  tal que  $\|x\|^2 = \sum_{i=1}^n |x_i|^2$ . El producto interno de  $x$  y  $y \in X$  es  $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$ . Una función  $K : X \rightarrow R$  se dice que es una función *kernel* si existe un perfil  $k : [0, \infty]$ , tal que

$$K(x) = k(\|x\|^2)$$

y cumplen

- $k$  es no negativo.
- $k$  es no creciente,  $k(a) \leq k(b)$  si  $a < b$ .
- $k$  es continua a trozos y  $\int_0^\infty k(r) dr < \infty$ .

Los *kernels* multivariados que se utilizan en el estimador de densidad son simétricos en todas las direcciones y se pueden describir por su forma o perfil, y satisfacen las siguientes propiedades

- $\int_{\mathbb{R}^d} K(x) dx = 1$
- $\int_{\mathbb{R}^d} xx^T K(x) dx = c_K \mathbf{I}$
- $\int_{\mathbb{R}^d} xK(x) dx = 0$
- $\lim_{\|x\|^d \rightarrow \infty} \|x\|^d K(x) = 0$

### Tipos de *kernels*

Resulta útil mencionar dos tipos de *kernels*, como los siguientes:

*kernel Gaussiano*

$$K_N(x) = (2\pi)^{-d/2} \exp\left(-\frac{1}{2}\|x\|^2\right)$$

donde su perfil viene dado por

$$k_k = \exp\left(-\frac{1}{2}\|x\|^2\right)$$

*kernel de Epanechnikov*

$$K_E(\mathbf{x}) = \begin{cases} \frac{1}{2}c_d^{-1}(d+2)(1-\|\mathbf{x}\|^2) & \text{si } \|\mathbf{x}\| \leq 1 \\ 0 & \text{caso contrario.} \end{cases} \quad (2.3)$$

donde  $c_d$  es el volumen de la esfera unitaria, cuyo perfil es

$$k_E(x) = \begin{cases} 1-x & \text{si } 0 \leq x \leq 1 \\ 0 & \text{si } x > 1. \end{cases} \quad (2.4)$$

Algo importante a destacar es cuán fiable resulta ser el estimador como en 1.3.4, y resulta seguir el mismo enfoque, la calidad del estimador del *kernel* de densidad viene dado por MISE entre la densidad y su estimación integrado sobre todo. Sin embargo, en la práctica, solo se puede calcular una aproximación asintótica de esta medida (AMISE) 1.4.1. Bajo la suposición de que el número de puntos de datos tiende a infinito ( $n \rightarrow \infty$ ) y que el ancho de banda disminuye a una velocidad menor que  $n^{-1}$ , tanto para los datos univariados como para los multivariados, el

núcleo de Epanechnikov minimiza la medida AMISE y tiene el perfil 2.4 lo que produce el núcleo radialmente simétrico 2.3.

Haciendo un cambio de notación de la expresión en 2.1.1 y de acuerdo a 2.2 se tiene

$$\bar{f}_{h,K}(x) = \frac{c_{k,d}}{nh^d} \sum_{i=1}^n k \left( \left\| \frac{x - x_i}{h} \right\|^2 \right). \quad (2.5)$$

Para analizar un espacio de características con una densidad subyacente  $f(x)$ , el primer paso es identificar las modas de esta densidad. Estas modas se encuentran en los puntos donde el gradiente es igual a cero, es decir, donde  $\nabla f(x) = 0$ . El procedimiento de cambio medio es una manera elegante de localizar estos puntos sin necesidad de estimar la densidad.

### 2.1.2 Estimación del gradiente de densidad

Se necesita localizar las modas de la estimación de densidad. Para esto tomamos el gradiente de la ecuación 2.5

$$\begin{aligned} \nabla \bar{f}_{h,K}(x) &= \frac{c_{k,d}}{nh^d} \sum_{i=1}^n k \left( \left\| \frac{x - x_i}{h} \right\|^2 \right) \\ &= \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^n (x - x_i) k' \left( \left\| \frac{x - x_i}{h} \right\|^2 \right) \end{aligned} \quad (2.6)$$

Se define la función

$$g(x) = -k'(x), \quad (2.7)$$

suponiendo que  $k$  es diferenciable en  $[0, \infty)$  excepto en un conjunto finito de puntos. La cual se puede apreciar que cumple con las condiciones suficientes de un *kernel*

Con un cambio de notación mediante el nuevo perfil definido se tiene:

$$G(x) = c_{g,d} g(\|x\|^2), \quad (2.8)$$

donde  $c_{c,g}$  es el término de normalización.

Ahora, introduciendo 2.7 en la ecuación anterior se obtiene:

$$\begin{aligned}
 \nabla \bar{f}_{h,G} &= \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^n (x_i - x) g \left( \left\| \frac{x - x_i}{h} \right\|^2 \right) \\
 &= \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^n \left[ x_i g \left( \left\| \frac{x - x_i}{h} \right\|^2 \right) - x g \left( \left\| \frac{x - x_i}{h} \right\|^2 \right) \right] \frac{\sum_{i=1}^n g \left( \left\| \frac{x - x_i}{h} \right\|^2 \right)}{\sum_{i=1}^n g \left( \left\| \frac{x - x_i}{h} \right\|^2 \right)} \\
 &= \frac{2c_{k,d}}{nh^{d+2}} \left[ \frac{\sum_{i=1}^n x_i g \left( \left\| \frac{x - x_i}{h} \right\|^2 \right)}{\sum_{i=1}^n g \left( \left\| \frac{x - x_i}{h} \right\|^2 \right)} - \frac{x g \left( \left\| \frac{x - x_i}{h} \right\|^2 \right)}{\sum_{i=1}^n g \left( \left\| \frac{x - x_i}{h} \right\|^2 \right)} \right] \sum_{i=1}^n g \left( \left\| \frac{x - x_i}{h} \right\|^2 \right) \\
 &= \frac{2c_{k,d}}{nh^{d+2}} \left[ \sum_{i=1}^n g \left( \left\| \frac{x - x_i}{h} \right\|^2 \right) \right] \left[ \frac{\sum_{i=1}^n x_i g \left( \left\| \frac{x - x_i}{h} \right\|^2 \right)}{\sum_{i=1}^n g \left( \left\| \frac{x - x_i}{h} \right\|^2 \right)} - x \right].
 \end{aligned} \tag{2.9}$$

Donde  $\sum_{i=1}^n g \left( \left\| \frac{x - x_i}{h} \right\|^2 \right)$  es asumido como un número positivo. De 2.9 se tiene que la estimación de densidad de  $x$  calculada con el *kernel*  $G$  viene por

$$\bar{f}_{h,G} = \frac{c_{g,d}}{nh^d} \sum_{i=1}^n g \left( \left\| \frac{x - x_i}{h} \right\|^2 \right), \tag{2.10}$$

y el vector Mean Shift con *kernel*  $G$  por:

$$m_{h,G} = \frac{\sum_{i=1}^n x_i g \left( \left\| \frac{x - x_i}{h} \right\|^2 \right)}{\sum_{i=1}^n g \left( \left\| \frac{x - x_i}{h} \right\|^2 \right)} - x, \tag{2.11}$$

el procedimiento del mean shift se obtiene con la diferencia calculando el promedio de los pesos de los datos utilizando el *kernel*  $G$  que se encuentran dentro de una vecindad definida por una región centrada en  $x$ .

De 2.9, 2.10 y 2.11 se obtiene:

$$m_{h,G} = \frac{c_{k,d}}{2c_{g,d}} h^2 \frac{\nabla \bar{f}_{n,K}(x)}{\bar{f}_{h,G}(x)}. \tag{2.12}$$

Nótese que el vector de Mean Shift se define como una estimación normalizada del gradiente de densidad de la muestra. La normalización, que se realiza utilizando la función de densidad con el *kernel*  $G$ , permite dar pasos más largos en regiones donde la densidad es baja y pasos más pequeños cerca de los picos de la función de densidad.

## 2.2 Criterio de convergencia de Mean Shift

### 2.2.1 Convergencia mediante el *Kernel Epanechnikov*

Supongamos que  $\{y_j\}_{j=1,2,\dots}$  es la secuencia de puntos obtenidos a través de las iteraciones de Mean Shift, es decir,

$$y_{j+1} = \frac{1}{n_j} \sum_{x_i \in S_h(y_j)} x_i,$$

done  $n_j$  representa el número de puntos  $x_i$  que se encuentran dentro del espacio definido por la hiperesfera  $S_h(y_j)$ , la cual está centrada en  $y_j$  y tiene un ancho  $h$ .

#### Teorema 2.1

Sea  $\bar{f}_E = \{\bar{f}_E(j)\}_{j=1,2,\dots}$  la sucesión de densidades resultantes del uso del *kernel* de Epanechnikov en los puntos  $\{y_j\}_{j=1,2,\dots}$  definido como la sucesión de puntos del procedimiento Mean Shift con un *kernel* Uniforme. Entonces la sucesión de densidades es convergente.

*Demostración.* Se tiene un cantidad  $n$  de de valores  $x_i$  de la muestra, por tanto la sucesión de densidad  $\bar{f}_E$  está acotada superiormente.

Si se demuestra que  $\bar{f}_E$  es estrictamente creciente, la sucesión converge.

Entonces,

Se define  $n_j, n'_j$  y  $n''_j$  como  $n_j = n'_j + n''_j$  el número de puntos que caen dentro de la ventana d-dimensional denominada como  $S_h(y_j)$ . Se define como el origen a  $y_j$  con lo que se puede notar  $\|y_j - x_i\|^2 = \|x_i\|^2$

$$S'_h(y_j) = S_h(y_j) - S''_h(y_j)$$

$$S''_h(y_j) = S_h(y_j) \cap S_h(y_{j+1}),$$

como se muestra en la gráfica 2.1.

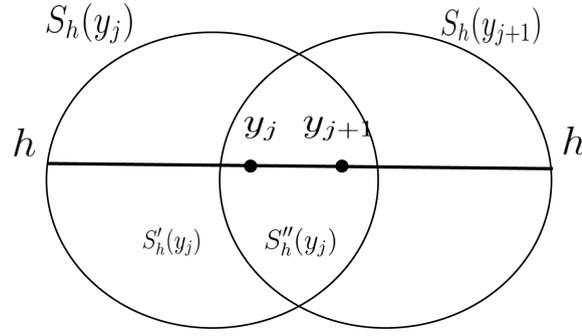


Figura 2.1: Ventanas d-dimensionales utilizadas en dos iteraciones sucesivas de Mean Shift

De la hipótesis,

$$\begin{aligned} \bar{f}_E(j) = \bar{f}_E(y_j) &= \frac{1}{nh^d} \sum_{x_i \in S_h(y_j)} K_E \left( \frac{y_j - x_i}{h} \right) \\ &= \frac{d+2}{2n(h^d c_d)} \sum_{x_i \in S_h(y_j)} \left( 1 - \frac{\|x_i\|^2}{h^2} \right) \end{aligned}$$

Dado que  $K_E$  es no negativo y  $S_h''(y_j) \cap S_h(y_{j+1})$  se tiene

$$\begin{aligned} \bar{f}_E(j+1) = \bar{f}_E(y_{j+1}) &= \frac{d+2}{2n(h^d c_d)} \sum_{x_i \in S_h(y_{j+1})} \left( 1 - \frac{\|y_{j+1} - x_i\|^2}{h^2} \right) \\ &\geq \frac{d+2}{2n(h^d c_d)} \sum_{x_i \in S_h''(y_j)} \left( 1 - \frac{\|y_{j+1} - x_i\|^2}{h^2} \right), \end{aligned}$$

de  $n_j = n'_j + n''_j$  se sabe que  $n'_j = n_j - n''_j$

$$\begin{aligned} \bar{E}(j+1) - \bar{f}_E(j) &= \frac{d+2}{2n(h^d c_d)} \left[ \sum_{x_i \in S_h''(y_j)} \left( 1 - \frac{\|y_{j+1} - x_i\|^2}{h^2} \right) - \sum_{x_i \in S_h(y_j)} \left( 1 - \frac{\|x_i\|^2}{h^2} \right) \right] \\ &= \frac{d+2}{2n(h^d c_d)} \left[ \sum_{x_i \in S_h''(y_j)} \frac{\|y_{j+1} - x_i\|^2}{h^2} + \sum_{x_i \in S_h(y_j)} \frac{\|x_i\|^2}{h^2} + n''_j - n_j \right] \\ &= \frac{d+2}{2n(h^{d+2} c_d)} \left[ \sum_{x_i \in S_h(y_i)} \|x_i\|^2 - \sum_{x_i \in S_h''(y_j)} \|y_{j+1} - x_i\|^2 - n'_j h^2 \right], \end{aligned}$$

por definición  $\|y_{j+1} - x_i\| \geq h^2$  para todo  $x_i \in S'_h(y_j)$ , por tanto  $\sum_{x_i \in S'_h(y_j)} \|y_{j+1} - x_i\|^2 \geq n'_j h^2$ , entonces

$$\begin{aligned}
 &\geq \frac{d+2}{2n(h^{d+2}c_d)} \left[ \sum_{x_i \in S_h(y_j)} \|x_i\|^2 - \sum_{x_i \in S''_h(y_j)} \|y_{j+1} - x_i\|^2 - \sum_{x_i \in S'_h(y_j)} \|y_{j+1} - x_i\|^2 \right] \\
 &= \frac{d+2}{2n(h^{d+2}c_d)} \left[ \sum_{x_i \in S_h(y_j)} \|x_i\|^2 - \sum_{x_i \in S_h(y_j)} \|y_{j+1} - x_i\|^2 \right] \\
 &= \frac{d+2}{2n(h^{d+2}c_d)} \left[ \sum_{x_i \in S_h(y_j)} \|x_i\|^2 - \sum_{x_i \in S_h(y_j)} \|x_i\|^2 + \|y_{j+1}\|^2 - 2y_{j+1}^t x_i \right] \\
 &= \frac{d+2}{2n(h^{d+2}c_d)} \left[ \sum_{x_i \in S_h(y_j)} 2y_{j+1}^t x_i - \sum_{x_i \in S_h(y_j)} \|y_{j+1}\|^2 \right] \\
 &= \frac{d+2}{2n(h^{d+2}c_d)} \left[ 2y_{j+1}^t \sum_{x_i \in S_h(y_j)} x_i - n_j \|y_{j+1}\|^2 \right],
 \end{aligned}$$

de  $y_{j+1} = \frac{1}{n_j} \sum_{x_i \in S_h(y_j)} x_i$ , se obtiene  $n_j y_{j+1} = \sum_{x_i \in S_h(y_j)} x_i$  y reemplazando en la expresión anterior

$$\begin{aligned}
 &= \frac{d+2}{2n(h^{d+2}c_d)} \left[ 2n_j \|y_{j+1}\|^2 - n_j \|y_{j+1}\|^2 \right] \\
 &= \frac{d+2}{2n(h^{d+2}c_d)} n_j \|y_{j+1}\|^2,
 \end{aligned}$$

la expresión obtenida es estrictamente positiva, con lo cual  $\bar{f}_E$  es acotada y monótonamente estricta creciente y se prueba que  $y_j$  converge.  $\square$

## 2.2.2 Criterio suficiente de convergencia de Mean Shift

Una condición suficiente de convergencia del método para cualquier tipo de *kernels* viene dado por el siguiente teorema.

### Teorema 2.2

Para todo *kernel*  $K$  convexo con perfil monótonamente decreciente. Las sucesiones  $\{y_j\}_{j=1,2,\dots}$  y  $\{\bar{f}_{h,K}(j)\}_{j=1,2,\dots}$  convergen y  $\{\bar{f}_{h,K}(j)\}_{j=1,2,\dots}$  es monótonamente creciente.

*Demostración.* La sucesión de densidades estimadas es acotada superiormente. Se mostrará que  $\bar{f}_{h,K}$  es monótona creciente

$$\bar{f}_{h,K}(j+1) - \bar{f}_{h,K}(j) = \frac{1}{nh^d} \sum_{i=1}^n \left[ k \left( \left\| \frac{y_{j+1} - x_i}{h} \right\|^2 \right) - k \left( \left\| \frac{x_i}{h} \right\|^2 \right) \right],$$

dado que el *kernel* es convexo, se cumple la desigualdad  $k(x_2) \geq k(x_1) + k'(x_1)(x_2 - x_1)$  para cualquier  $x_1, x_2 \in [0, \infty]$ . Si definimos  $-k' = g$ , entonces podemos reescribir la desigualdad como  $k(x_2) - k(x_1) \geq g(x_1)(x_1 - x_2)$ . Al reemplazar los valores, obtenemos:

$$\begin{aligned} \bar{f}_{h,K}(j+1) - \bar{f}_{h,K}(j) &\geq \frac{1}{nh^{d+2}} \sum_{i=1}^n g \left( \left\| \frac{x_i}{h} \right\|^2 \right) \left[ \|x_i^2\| - \|y_{j+1} - x_i\|^2 \right] \\ &= \frac{1}{nh^{d+2}} \sum_{i=1}^n g \left( \left\| \frac{x_i}{h} \right\|^2 \right) \left[ \|x_i\|^2 - \|x\|^2 - \|y_{j+1}^t\|^2 + 2y_{j+1}^t x_i \right] \\ &= \frac{1}{nh^{d+2}} \sum_{i=1}^n g \left( \left\| \frac{x_i}{h} \right\|^2 \right) \left[ 2y_{j+1}^t x_i - \|y_{j+1}\|^2 \right] \\ &= \frac{1}{nh^{d+2}} \left[ 2y_{j+1}^t \sum_{i=1}^n x_i g \left( \left\| \frac{x_i}{h} \right\|^2 \right) - \|y_{j+1}\|^2 \sum_{i=1}^n g \left( \left\| \frac{x_i}{h} \right\|^2 \right) \right] \\ &= \frac{1}{nh^{d+2}} \left[ 2y_{j+1}^t y_{j+1} \sum_{i=1}^n g \left( \left\| \frac{x_i}{h} \right\|^2 \right) - \|y_{j+1}\|^2 \sum_{i=1}^n g \left( \left\| \frac{x_i}{h} \right\|^2 \right) \right] \\ &= \frac{1}{nh^{d+2}} \|y_{j+1}\|^2 \sum_{i=1}^n g \left( \left\| \frac{x_i}{h} \right\|^2 \right). \end{aligned}$$

Dado que  $k$  es monótonamente decreciente, podemos afirmar que  $-k' = g$  será positiva. Por lo tanto, la expresión  $\sum_{i=1}^n g \left( \left\| \frac{x_i}{h} \right\|^2 \right)$  también será positiva. Esta expresión se vuelve positiva siempre que  $y_j \neq y_{j+1}$ , lo cual demuestra la convergencia de la secuencia de densidades hacia la moda.

Para demostrar la convergencia en la secuencia  $y_{j=1 \dots n'}$ , se puede reescribir la ecuación anterior sin asumir que  $y_j = 0$ .

$$\bar{f}_{h,K}(j+1) - \bar{f}_{h,K}(j) \geq \frac{1}{nh^{d+2}} \|y_{j+1} - y_j\|^2 \sum_{i=1}^n g \left( \left\| \frac{y_j - x_i}{h} \right\|^2 \right)$$

Luego:

$$\begin{aligned}
 \bar{f}_{h,k}(j+1) - \bar{f}_{h,k}(j) &\geq \|y_{j+1} - y_j\|^2 M \\
 \bar{f}_{h,k}(j+2) - \bar{f}_{h,k}(j+1) &\geq \|y_{j+1+2} - y_{j+1}\|^2 M \\
 \bar{f}_{h,k}(j+3) - \bar{f}_{h,k}(j+2) &\geq \|y_{j+1+3} - y_{j+2}\|^2 M \\
 &\vdots \\
 \bar{f}_{h,k}(j+n) - \bar{f}_{h,k}(j+n-1) &\geq \|y_{j+1+n} - y_{j+n-1}\|^2 M,
 \end{aligned}$$

Sabemos que esta desigualdad se cumple para términos consecutivos  $j$  y  $j+1$ , y donde el valor mínimo de  $\frac{1}{nh^{d+2}} \sum_{i=1}^n g\left(\left\|\frac{y_j - x_i}{h}\right\|^2\right)$  se representa como  $M$ . Añadiendo los términos a la izquierda y derecha, se obtiene:

$$\bar{f}_{h,k}(j+n) - \bar{f}_{h,k}(j) \geq \|y_{j+n} - y_j\|^2 M.$$

Dado que hemos demostrado la convergencia de la sucesión  $\{f_{b,k}(j)\}$ , podemos concluir que la sucesión  $y_j$  con  $j = 1, 2, \dots$  es una sucesión de Cauchy y, por consiguiente, converge. □

### 2.2.3 Propiedad de una trayectoria suave

Una trayectoria suave del método se establece de acuerdo al ángulo entre dos vectores Mean shift, el cual debe ser menor a 90 grados. Esto viene enunciado por el siguiente lema.

#### **Lema 2.1**

la suma ponderada de las proyecciones de  $(y_{j+1} - x_i)$  en la dirección  $(0, y_{j+1})$  es negativa cuando los pesos están dados por un kernel normal centrado en  $y_{j+1}$  es decir que

$$\sum_{i=1}^n (\|y_{j+1}\|^2 - y_{j+1}^t x_i) \exp\left(-\left\|\frac{y_{j+1} - x_i}{h}\right\|^2\right) \leq 0.$$

*Demostración.* De la demostración encontrada en el apéndice del trabajo de Comaniciu, se tiene la desigualdad

$$\|y_{j+1}\|^2 \leq y_{j+1}^t \frac{\sum_{i=1}^n x_i \exp\left(-\left\|\frac{y_{j+1}-x_i}{h}\right\|^2\right)}{\sum_{i=1}^n \exp\left(-\left\|\frac{y_{j+1}-x_i}{h}\right\|^2\right)} = y_{j+1}^t y_{j+2}$$

Teniendo en cuenta que  $v$  y  $w$  son dos vectores, se sabe que  $\frac{vw}{\|v\|\|w\|} = \cos(\theta)$  y por lo tanto  $\theta \leq 90 \rightarrow \cos(\theta) > 0$ , es decir que la distancia entre los vectores es menor a  $90^\circ$  si  $vw > 0$ .

$$\begin{aligned} y_{j+1}^t (y_{j+2} - y_{j+1}) &\geq 0 \\ y_{j+1}^t y_{j+2} - y_{j+1}^t y_{j+1} &\geq 0 \\ y_{j+1}^t y_{j+2} &\geq y_{j+1}^t y_{j+1} \\ y_{j+1} y_{j+2} &\geq \|y_{j+1}\|^2. \end{aligned}$$

□

### 2.2.4 Elección del Bandwidth

El bandwidth como único parámetro del método Mean Shift resulta determinante a la hora del funcionamiento correcto del agrupamiento o clustering. Comaniciu en su trabajo menciona cuatro diferentes técnicas para la selección del parámetro bandwidth  $h$  utilizando la función kernel en el desarrollo del Mean Shift:

- Elección estadística: se refiere al bandwidth que logra el mejor equilibrio entre el sesgo y la varianza del estimador para todos los valores de  $x$  en el espacio  $R^d$ . Este ancho de banda minimiza la medida AMISE, que es la diferencia cuadrática entre la estimación y el valor real.
- Estabilidad de descomposición: el ancho de banda es el promedio del rango en el cual se obtuvo el mismo número de grupos (clusters) en diferentes experimentos.
- Bandwidth máximo: es el mejor ancho de banda que maximiza una función objetivo que evalúa la calidad de la descomposición. Esta función

objetivo generalmente analiza la variabilidad entre grupos (inter-cluster) o la conectividad aislada entre ellos.

- Información de nivel superior: la descomposición en varios casos depende de la tarea a realizar, resulta útil el bandwidth de la información proporcionada por el usuario o de niveles superiores de la tarea.

En las técnicas no paramétricas, una inadecuada elección del parámetro bandwidth puede ocasionar problemas en la detección de las modas significativas, causando que las agrupaciones se unan entre sí o se generen otras no representativas.

En el trabajo realizado por Chacon se establece que “el problema de la selección del ancho de banda para la agrupación por Mean shift, aunque relacionado, es diferente del de la selección del bandwidth para la estimación del gradiente de densidad y presenta sus propias peculiaridades que sin duda se merecen ser estudiadas con más detalle”

# 3

---

## *Implementación y Resultados*

---

En esta sección, se detalla la implementación del método, abarcando tanto el manejo de los datos empleados como el software utilizado. Además, se presenta y discute el agrupamiento obtenido para mejorar la comprensión del proceso. Cabe mencionar que la implementación se realiza en dos instancias, el primero con un enfoque en datos de temperatura en la Región Andina del Ecuador, y el segundo con uso de datos de una estación meteorológica del Cantón Riobamba de la Provincia de Chimborazo, Ecuador.

### **3.1 Implementación Región Andina del Ecuador**

#### **3.1.1 Base de datos**

Se utilizó una base de datos de los modelos de Circulación atmosférica CNRM Centre National de Recherches Meteorológicas (CNRM) y Cerfacs de la sexta fase del Proyecto de Intercomparación de Modelos Acoplados 6 (CMIP6) para la región Sierra del Ecuador. Los datos fueron proporcionados por el Grupo de Investigación de Energías Alternativas y Ambiente GEAA, para un periodo 2020-2050, con una resolución de  $25\text{km} \times 25\text{km}$ .

Los datos fueron estructurados en una matriz de  $10954 \times 10$ , debido a que se trata de un arreglo de malla en diferentes ubicaciones, cada una de las columnas se considera como una componente como se observa en la Figura 3.2, los valores siguen una distribución normal. Además tienen una distribución normal estandarizada, es decir, poseen una media de cero y una desviación estandar de

### 3.1. Implementación Región Andina del Ecuador

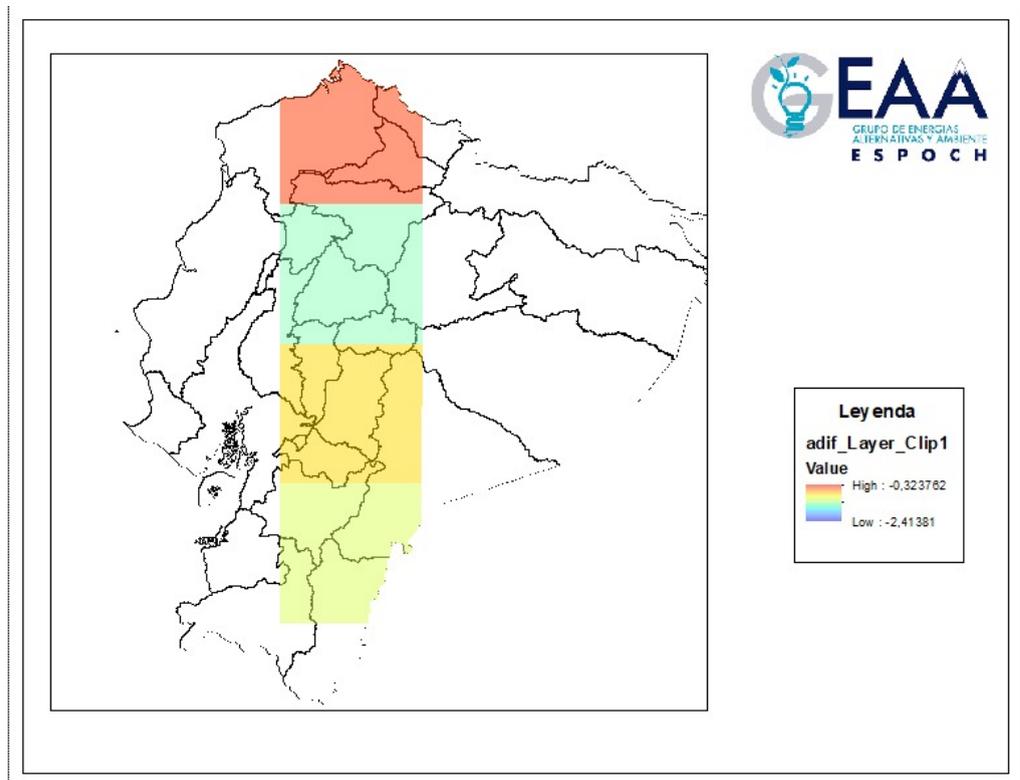


Figura 3.1: Proyección climática CNRM para la región Andina del Ecuador en el período 2020-2050.

uno.

El Análisis de Componente Principales (PCA) permite convertir un conjunto de muestras de variables correlacionadas en un conjunto de variables no correlacionadas linealmente. Este método crea nuevos ejes ordenados, el primer eje contiene la información más importante, el segundo la información que ocupa el segundo de puesto en forma de importancia y así sucesivamente. Habitualmente se incorpora el número de ejes necesarios para mantener la proporción de información en un 90 % y descartar el resto de los ejes. Mediante este proceso se redujo la dimensionalidad de la matriz inicial a 3 componentes principales con una variabilidad del 89.5 % como se visualiza en la Figura 3.3. Además, los algoritmos Machine Learning consiguen mejores niveles de calidad con las salidas producidas por el PCA.

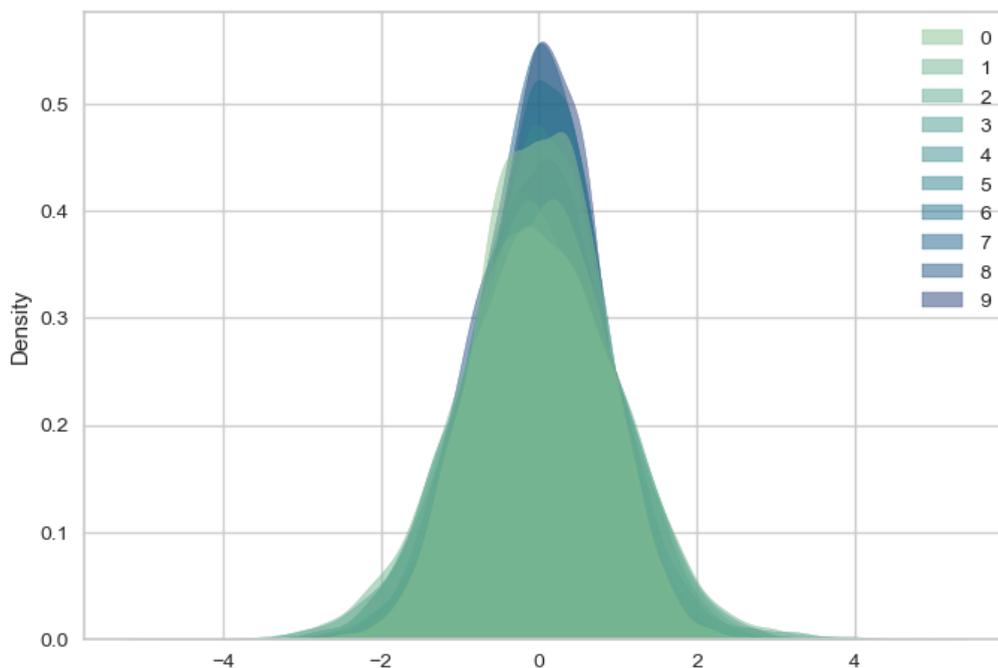


Figura 3.2: Distribución de los datos de temperatura mínima

### 3.1.2 Implementación del método

La implementación del método Mean shift se realizó para el caso unidimensional de la variable temperatura mínima mediante el lenguaje de programación Python (<https://github.com/geaaepoch>).

Se importó la biblioteca sklearn, una herramienta que facilita el análisis de datos, desarrollo e implementación de algoritmos Machine Learning. En este caso se trabajó con `sklearn.cluster.MeanShift`, donde el parámetro de mayor relevancia es el ancho de banda (`bandwidth`), especificada a continuación.

#### Elección del ancho de banda (`bandwidth`)

El ancho de banda posee un rol que influye en la estimación de la función de densidad de probabilidad determinando la forma en que los datos se acoplan a una estimación de la distribución real desconocida como se muestra en la figura [3.4](#).

### 3.1. Implementación Región Andina del Ecuador

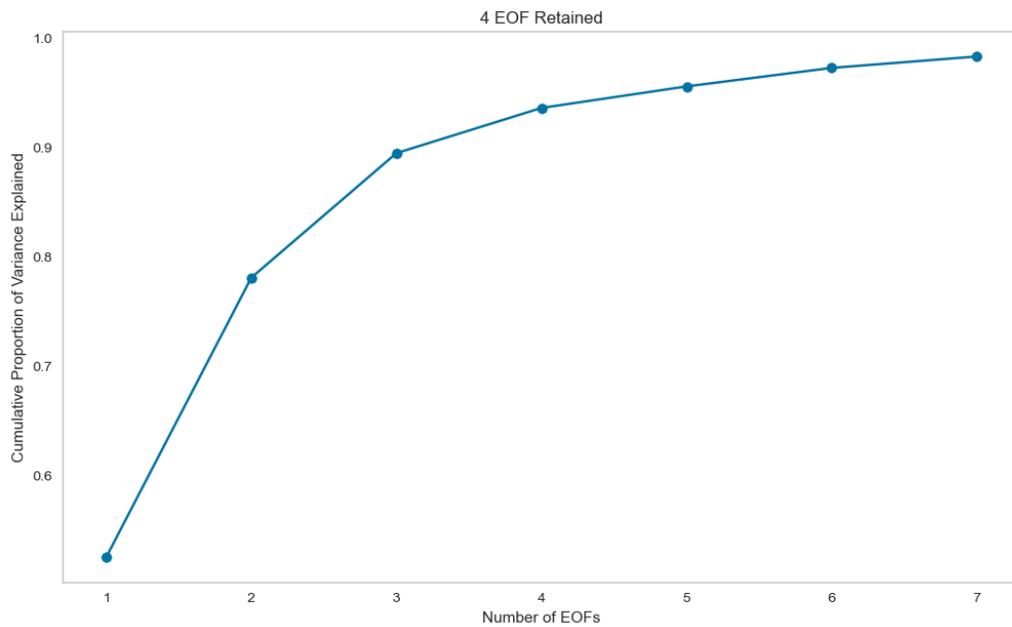


Figura 3.3: Análisis de componentes principales (PCA)

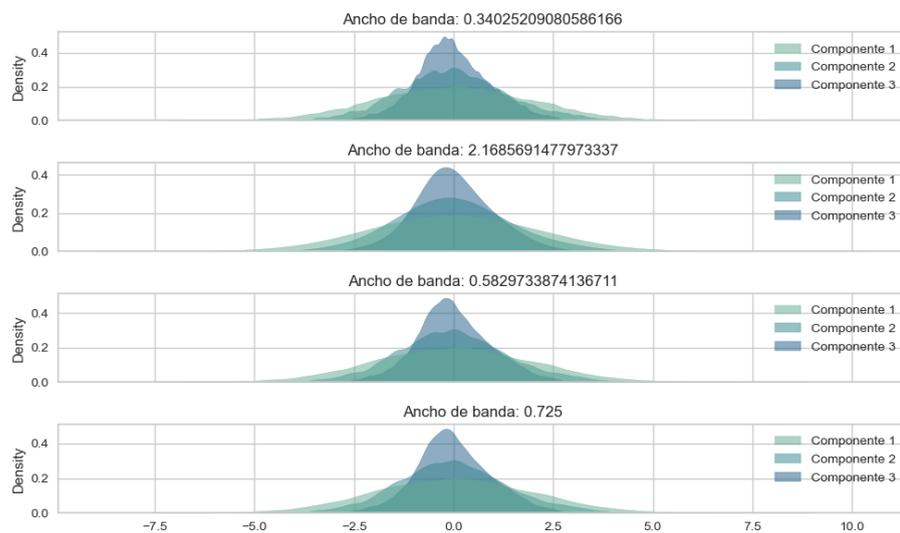


Figura 3.4: Suavidad de la distribución de los datos para cadaa valor de *bandwidth*

En este caso, el valor que toma el parámetro de ancho de banda se maneja en un rango de 0.3 a 2 debido a la inestabilidad de los clusteres resultantes una vez implementado el método. Este valor se escogió a través de tres técnicas de

estimación del bandwidth. La primera, la regla de Silverman, que se desprende del cálculo AMISE y ve la relación directa entre el valor del rango intercuartil y la desviación estándar de los datos. El segundo, la función empleada en el algoritmo de la biblioteca sklearn, la cual se establece a partir de la distancia de los puntos pares en los cuartiles definidos como entrada y por último la validación cruzada, en la cual se estima un valor óptimo dentro de un rango de cifras pertenecientes al bandwidth. Los resultados de este parámetro establecen que la regla de Silverman varía en décimas por el valor de la validación cruzada, mientras que por la distancia entre pares de puntos, varía en unidades. Ver Tabla 3.1.

Regla de silverman	Validación cruzada	Distancia entre pares
0.34	0.73	1.83

Tabla 3.1: Valores del parámetro *bandwidth*

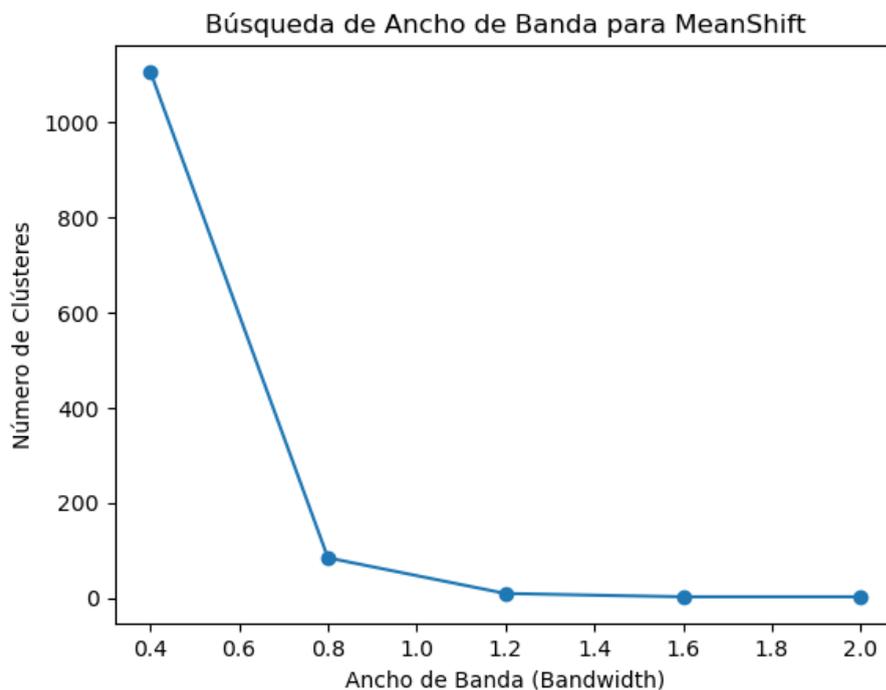


Figura 3.5: Relación existente del valor del bandwidth y el número de clusters

### 3.1. Implementación Región Andina del Ecuador

Una vez establecidos los distintos valores de bandwidth, se procede a ver el desempeño de cada uno de ellos en el algoritmo. Los resultados nos reflejan agrupaciones sumamente numerosas para el ancho de banda de 0.34 y 0.73, mientras que para 1.83 se obtuvo 4 grupos. Ver Tabla 3.2

Bandwidth utilizado	# de clusters
Regla Silverman	1441
Validación Cruzada	123
Distancia pares	4

Tabla 3.2: Grupos formados a partir del ancho de banda

Existe una relación directa entre el valor de bandwidth y el número de cluster como se evidencia en los resultados expuestos y en la Figura 3.5

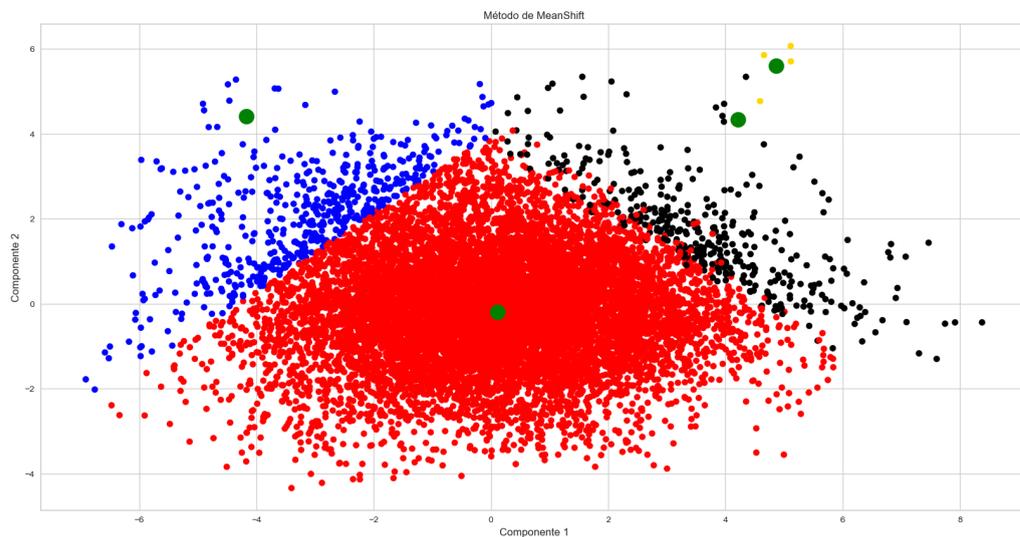


Figura 3.6: Agrupación resultante con un bandwidth de 1.38

La visualización proyectado en un espacio tridimensional resulta de la siguiente forma

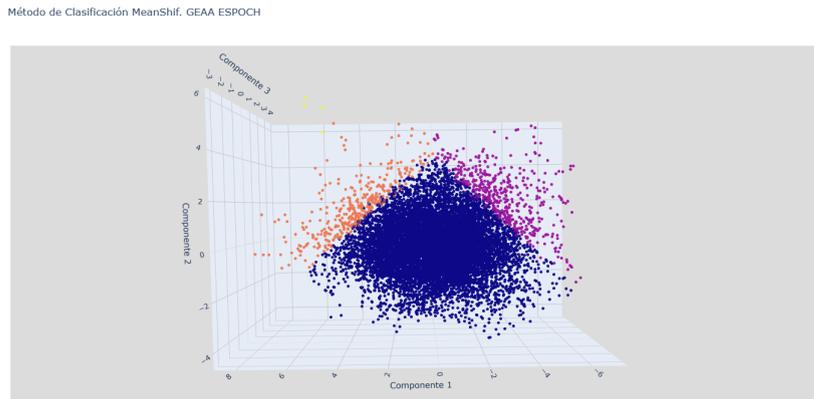


Figura 3.7: Proyección de los cluster en un plano tridimensional

### 3.1.3 Análisis por clusters

Se realizó la clasificación de las agrupaciones, es decir, se cuantificó las etiquetas obtenidas en cada cluster. En la Tabla 3.3 se observa de mayor a menor la ocurrencia de datos en cada agrupación.

Cluster	Frecuencia
0	10277
1	400
2	273
3	4

Tabla 3.3: Frecuencia de datos en cada cluster

Se graficó la estimación de *kernel* de los datos por cada cluster para describir la distribución por cada componente.

### 3.1. Implementación Región Andina del Ecuador

---

En el cluster 0 las componentes presenta una curva simétrica con respecto a la media y una desviación estándar baja. En el componente 1 se refleja una mayor dispersión de los datos con respecto a los componentes 2 y 3 como se observa en la Figura 3.8.

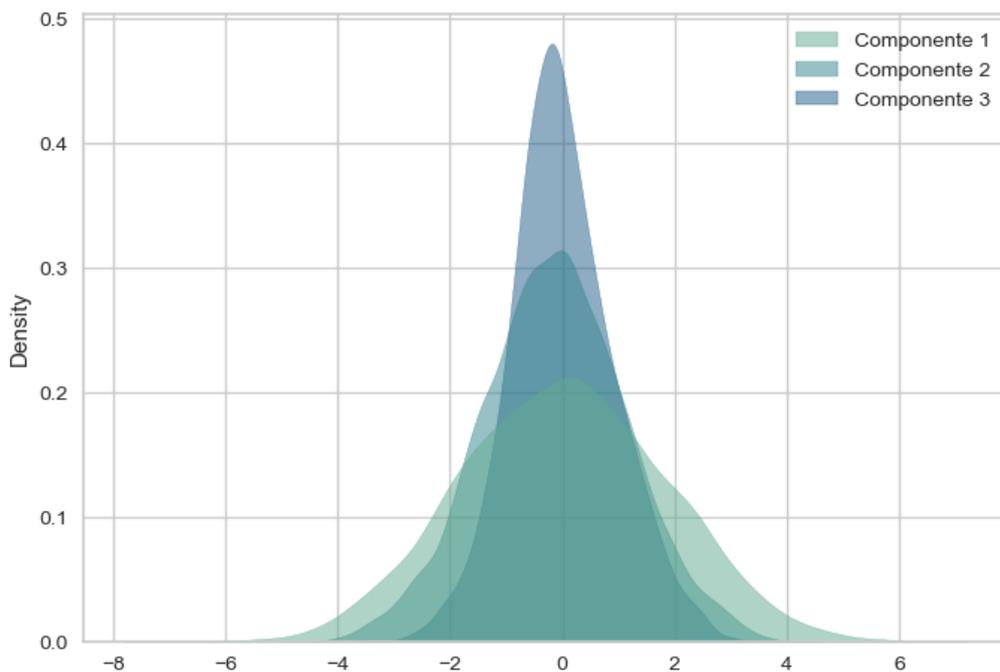


Figura 3.8: Distribución Cluster 0

En el cluster 1 se presentó una asimetría en los componentes 1 y 2 alejados de la media como se visualiza en la Figura 3.9. El componente 3 refleja una curva cercana a la media pero no simétrica.

En el cluster 2, los valores están dispersos en los 3 componentes siendo el componente 1 el que presenta una mayor dispersión respecto a los componentes 2 y 3, estas componentes se encuentran alejadas de la media y experimentan mayor desviación estándar, como se observa en la Figura 3.10.

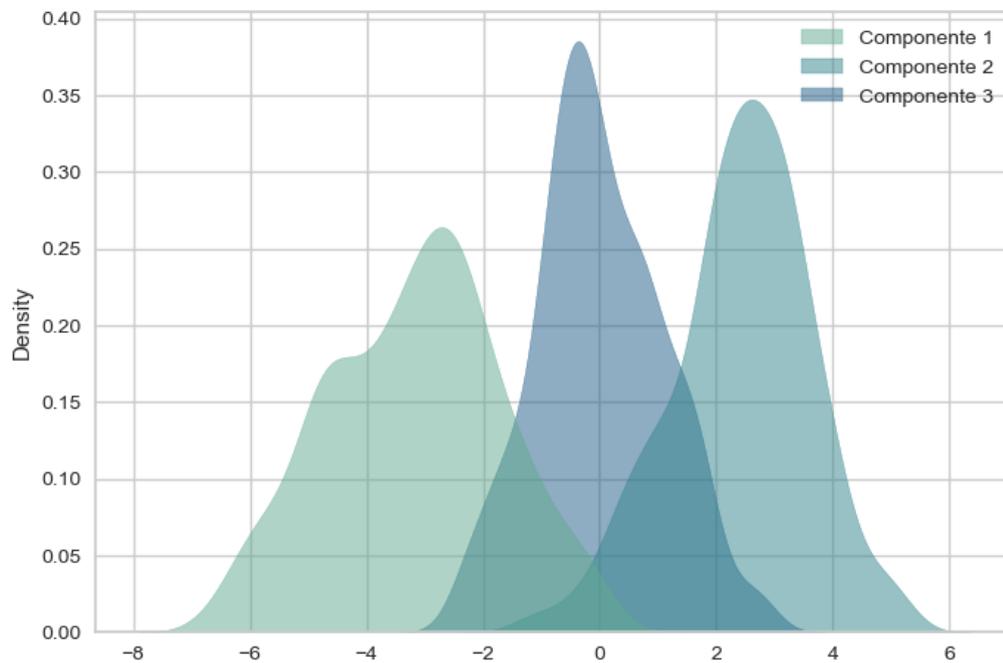


Figura 3.9: Distribución Cluster 1

En el cluster 3, ninguno de los componentes se encuentra en la media. El componente 3 presenta valores de desviación estándar negativo mientras que las componentes 1 y 2 presentan valores positivos como lo muestra la Figura 3.11

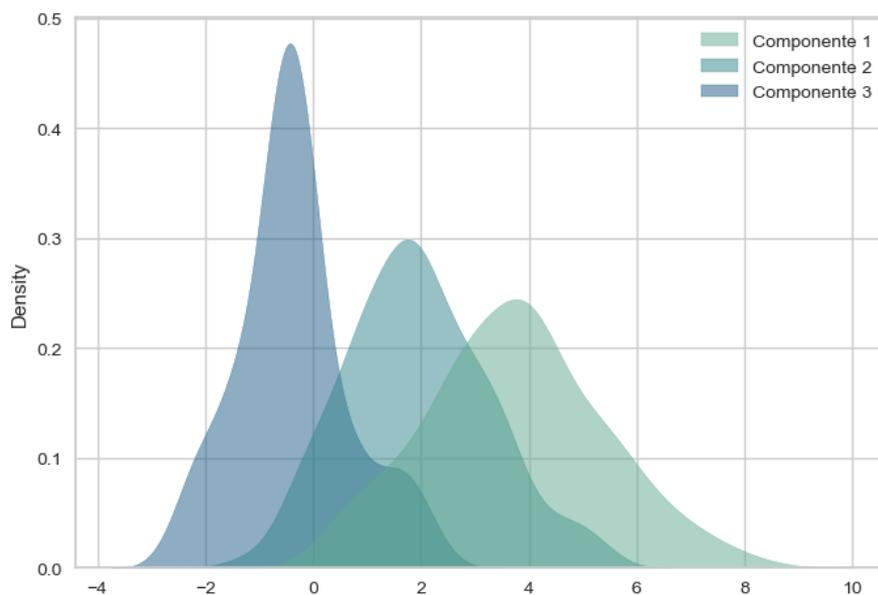


Figura 3.10: Distribución Cluster 2

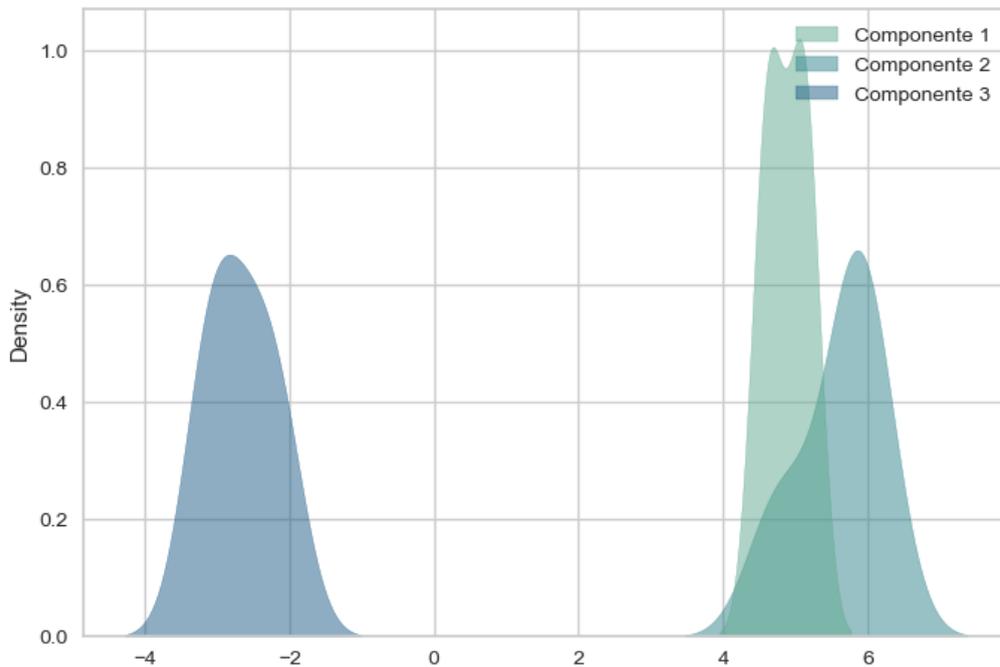


Figura 3.11: Distribución Cluster 3

Esto significa que en los próximos 50 años, (2020-2050), el modelo de cambio climático CRNM esta reproduciendo una realidad en la región andina del Ecuador, en la que el 93.81 % de los datos de temperatura mínima no experimenta variación, se encuentran con valores cercanos a la media y una desviación estándar mínima. Los eventos extremos y anomalías se presentan en bajo porcentaje y esta caracterizados en los cluster 1,2 y 3.

### 3.2 Implementación en una estación meteorológica

Se realizó una nueva implementación con la misma metodología para una estación meteorológica en la Provincia de Chimborazo, específicamente, de la Estación Meteorológica ALAO de la parroquia Pungalá en el cantón Riobamba con coordenadas -1.868806 y 78.54166432318941 en longitud e latitud respectivamente, y a una altura de 3064 metros sobre el nivel del mar.

#### 3.2.1 Base de datos

Los datos diarios de temperatura mínima fueron proporcionados por el Grupo de Investigación de Energías Alternativas y Ambiente GEAA. Estos datos pasaron por un procesamiento previo de imputación, relleno y homogenización. Se analizó la variabilidad atmosférica en dos años utilizando la estación meteorológica de ALAO debido a que no se encuentran datos faltantes para el año 2015-2016.

#### 3.2.2 Implementación del método

Se calculó el ancho de banda (bandwidth) con la Regla de Silverman, la validación cruzada y la distancia entre pares, como se ve en la tabla 3.4. Se escogió la distancia entre pares como el mejor parámetro debido a que estima un número de clusters coherentes.

Regla de silverman	Validación cruzada	Distancia entre pares
0.61	0.71	1.82

Tabla 3.4: Valores del parámetro *bandwidth*

### 3.2. Implementación en una estación meteorológica

En la figura 3.12, se observa los datos correspondientes a los años 2015 y 2016 con las etiquetas ALAO15 y ALAO16 respectivamente. Estos datos se suavizan a través de una estimación de densidad por medio de un *kernel* normal, la influencia del parámetro ancho de banda de las tres técnicas se visualizan en la suavidad de la curva. El valor de 1.82 presenta un mejor suavizado de los datos en comparación con la validación cruzada y la regla de silverman.

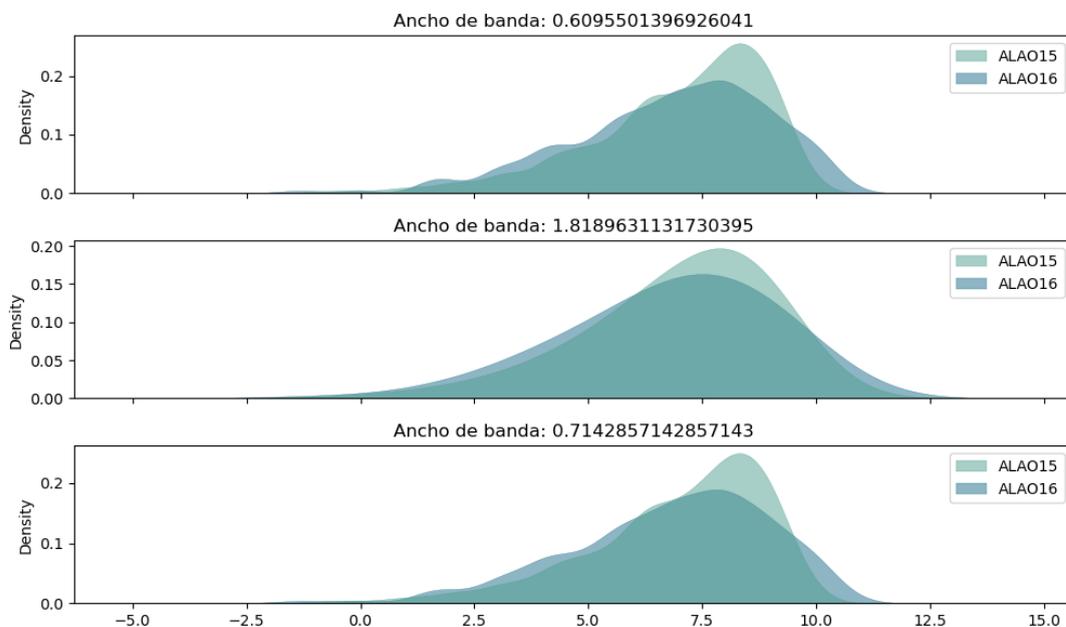


Figura 3.12: Estimación mediante un kernel normal

La relación entre el ancho de banda y los números de cluster se presentan en la Figura 3.13, evidenciado la relación inversamente proporcional entre el número de clústeres y el valor del *bandwidth*.

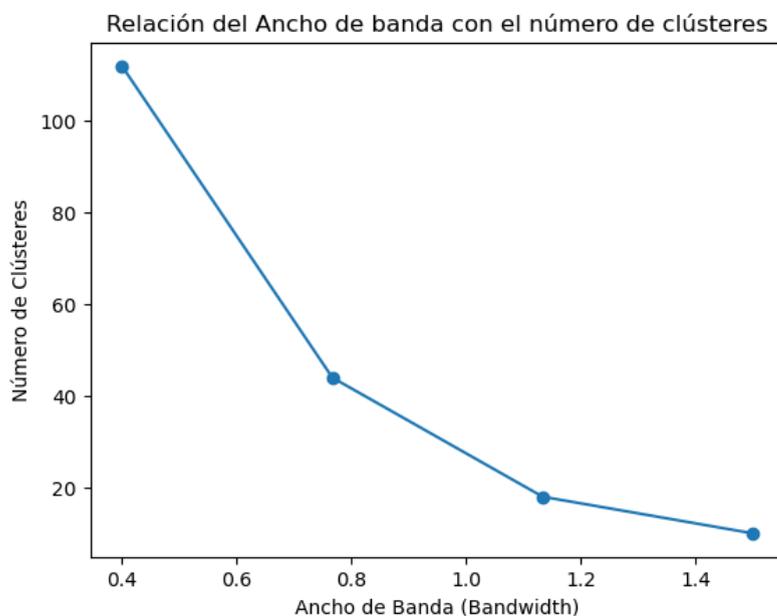


Figura 3.13: Relación Bandwidth-Clusters

El método Mean Shift busca iterativamente mover cada punto hacia una región de mayor densidad, es decir, hacia la moda más cercana en el conjunto de datos. El proceso que realiza el algoritmo de scikitlearn (<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.MeanShift.html>) se trata de una implementación de una versión simplificada y eficiente del algoritmo Mean Shift.

Bandwidth utilizado	Núm. de clusters
Regla Silverman	54
Validación Cruzada	42
Distancia pares	5

Tabla 3.5: Grupos formados a partir del ancho de banda

Mediante las tres técnicas del parámetro de bandwidth se estableció el números de clusteres como se observa en la Tabla 3.13. Los resultados marcan agrupaciones muy altas con la regla de silverman y la validación cruzada, con valores de 54 y 42 agrupaciones respectivamente. El método de distancia entre

pares (badwidthestimate) estableció un número coherente e interpretable de 5 agrupaciones utilizando todos los datos y el cuantil de 0.15. Este resultado se visualiza en la Figura 3.14.

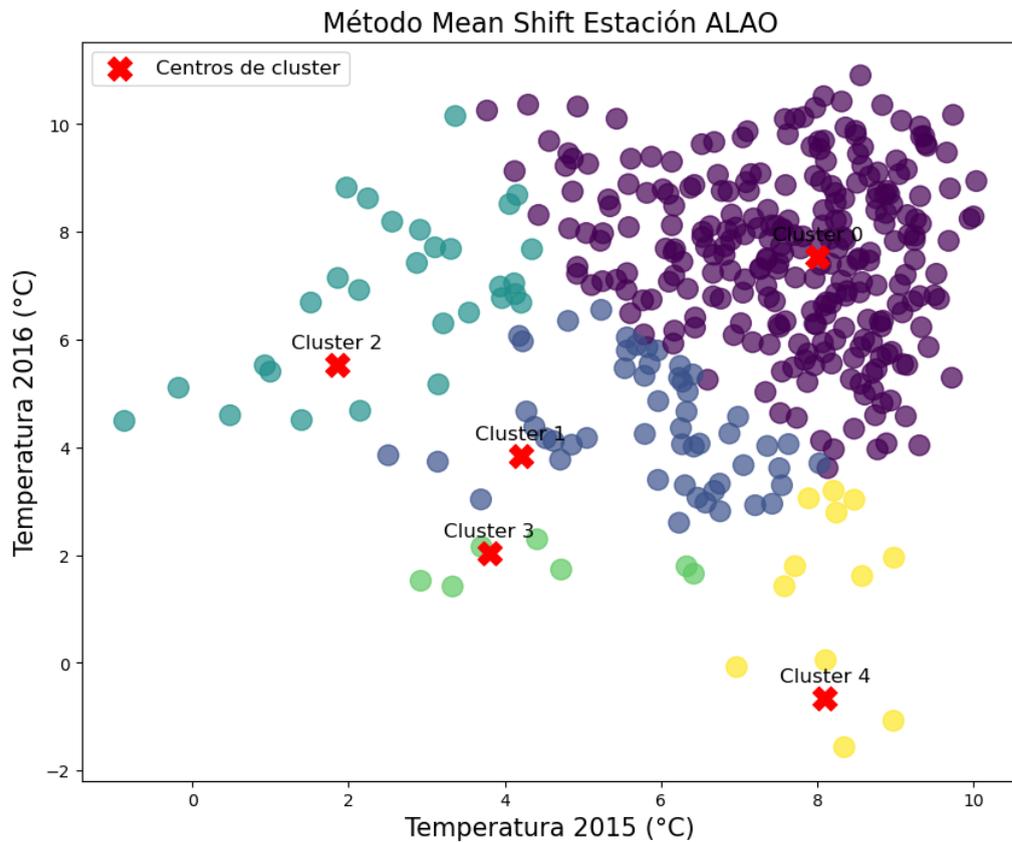


Figura 3.14: Temperatura mínima

#### 3.2.3 Análisis por cluster

Se cuantificó las etiquetas obtenidas en cada cluster. En la tabla 3.6 se observa la ocurrencia de datos en cada agrupación.

Cluster	Frecuencia
0	265
1	52
2	29
3	7
4	12

Tabla 3.6: Frecuencia de datos en cada cluster

Cluster	0		1		2		3		4	
Años	2015	2016	2015	2016	2015	2016	2015	2016	2015	2016
Promedio	7.67	7.62	5.89	4.41	2.61	6.86	4.55	1.80	8.17	1.35
Desviación	1.34	1.54	1.18	1.07	1.41	1.48	1.39	0.32	0.58	1.65
Mínimo	3.77	3.62	2.51	2.60	-0.87	4.49	2.96	1.42	6.97	-1.57
Máximo	10.04	10.91	8.03	6.55	4.35	10.15	6.42	2.29	8.98	3.20

Tabla 3.7: Análisis por cada cluster

En la Tabla 3.7 se observa en detalle los valores estadísticos por cada cluster en los años 2015 y 2016. Se dispuso de esta forma para la interpretación y análisis de la temperatura mínima en la región. El análisis se lo realizó debido a su importancia como un factor de causa y riesgo en una variedad de sectores, desde la salud pública hasta la agricultura y la gestión de recursos naturales. Además del efecto en la distribución espacio temporal de la precipitación.

### 3.2. Implementación en una estación meteorológica

En el cluster 0 las temperaturas para ALAO 2015 y 2016 son similares con una diferencia de  $0.05\text{ }^{\circ}\text{C}$  en la media anual, este cluster es el de mayor ocurrencia con 72 % que corresponde a 264 días. Para los dos años de estudio en el mes de marzo no experimento cambios en el valor de la temperatura, mientras que el patrón en los meses de agosto, septiembre y noviembre solo ocurrieron de 12 a 15 días, como se visualiza en la Figura 3.15

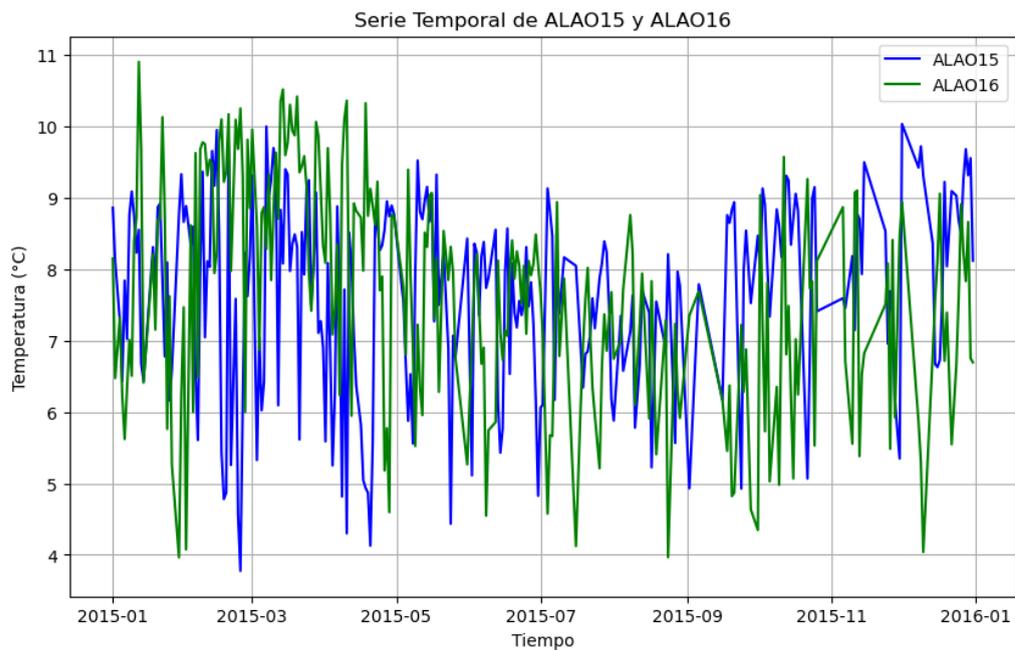


Figura 3.15: Serie temporal cluster 0

En el cluster 1 las temperaturas promedio anual para cada año varían en dos unidades, este cluster presenta un 15 % de ocurrencia diaria que corresponde a 53 días. En los meses de agosto, septiembre y noviembre este patrón se presentó de 7 y 8 días, mientras que en el mes de abril se presenta un solo día. Existen regiones pequeñas donde los datos de temperatura son similares como se observa en la Figura 3.16 en los meses de febrero, mayo, octubre y noviembre.

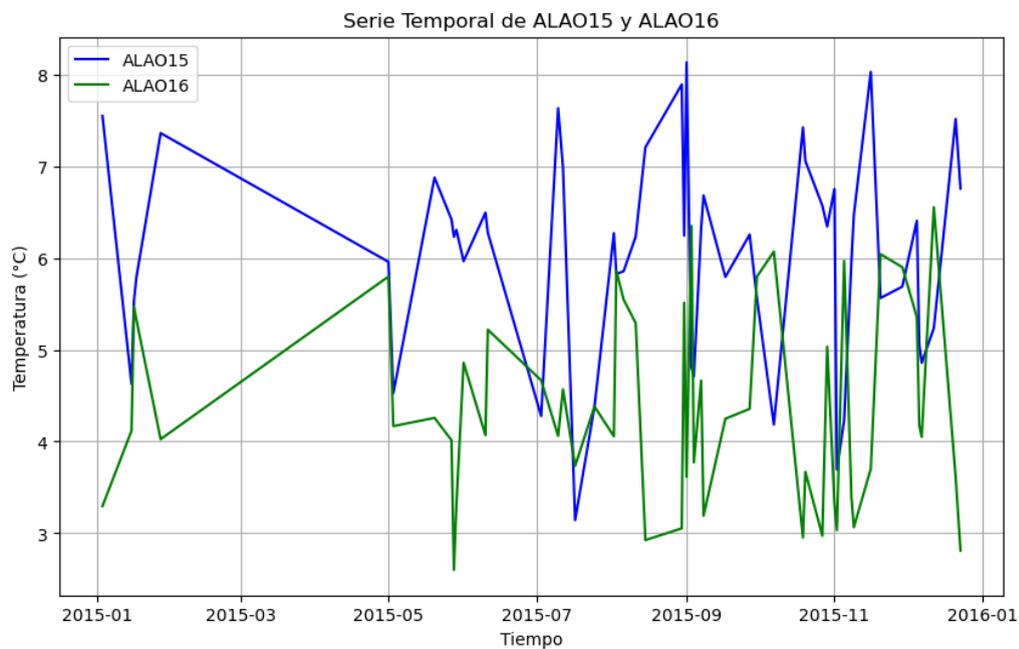


Figura 3.16: Serie temporal cluster 1

En el cluster 2, la frecuencia de ocurrencia es de 8 % que corresponde a 30 días. Se presentó una diferencia de 4°C en la temperatura media anual, es decir, el año 2015 experimentó temperaturas de -0.87°C hasta 4.35°C, mientras que en el año 2016 experimentó superiores de 3.85 a 10.15°C. Los meses de febrero, marzo y abril poseen un solo día de ocurrencia.

### 3.2. Implementación en una estación meteorológica

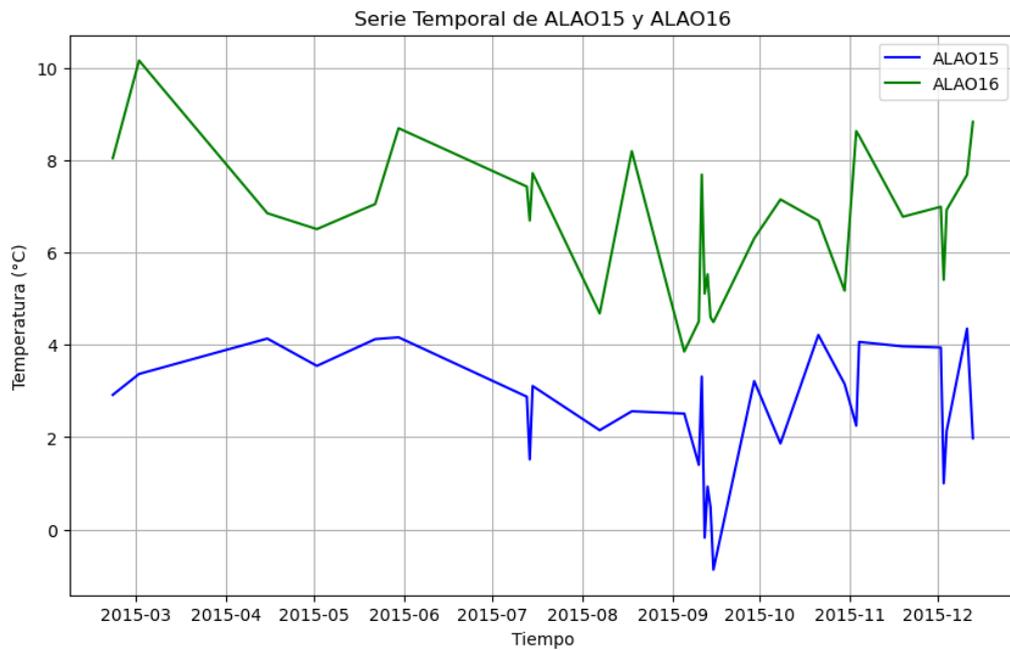


Figura 3.17: Serie temporal cluster 2

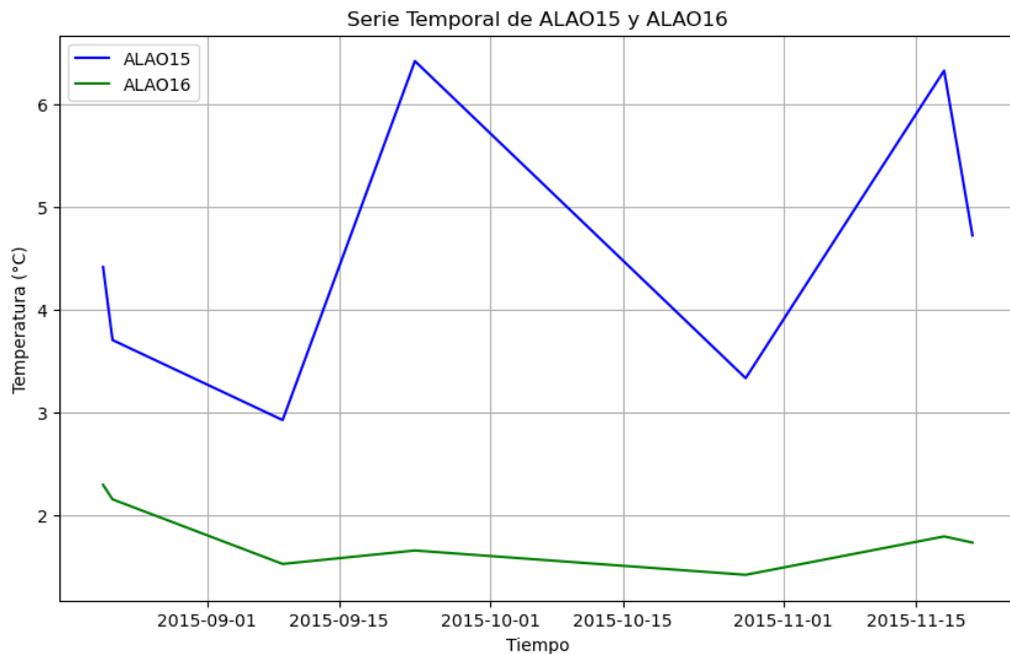


Figura 3.18: Serie temporal cluster 3

En el cluster 3, con frecuencia de ocurrencia del 1.92% que corresponde a 7 días. En el año 2015 existió una variación de temperatura con picos de 6.00 a 2.93°C,

mientras que el año 2015 se mantuvo una tendencia constante desde 2.29 hasta 1.41°C como se observa en la Figura 3.18

Finalmente, se observa que en el cluster 4 con ocurrencia del 3.01% que corresponde a 11 días, existe una variación significativa entre el año 2015 y 2016, la temperatura para el primer año entre 6.96 y 8.52°C mientras que en el segundo presentó temperaturas de -1.56 a 3.20°C, como se indica en la Figura 3.19

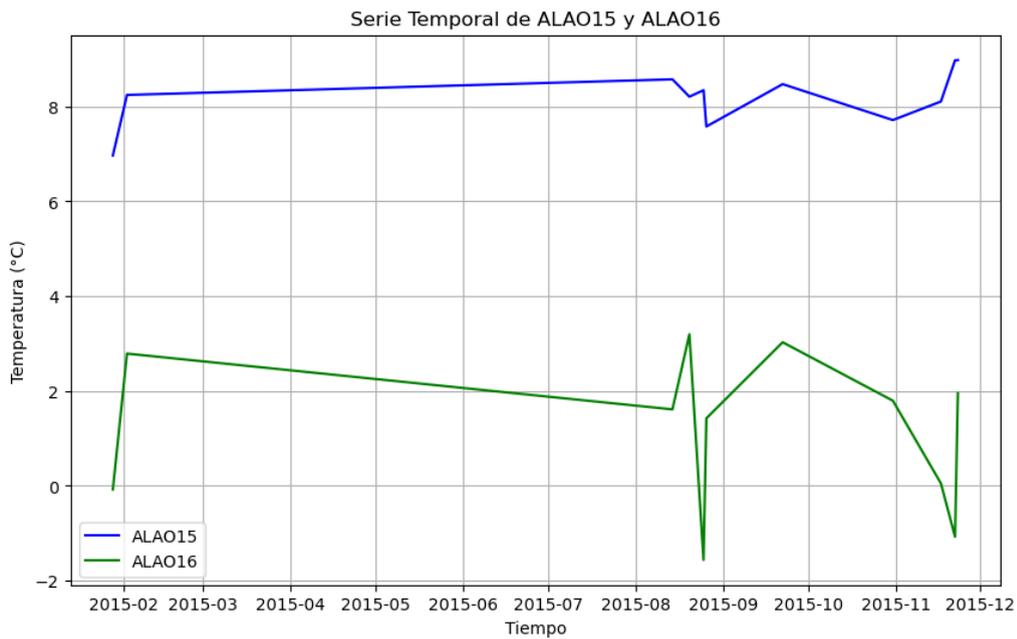


Figura 3.19: Serie temporal cluster 4

# Conclusiones y Recomendaciones

---

## Conclusiones

- A partir del estudio y comprensión del método Mean Shift, se proporcionó una descripción en detalle de los conceptos teóricos y bases matemáticas, destacando las estimaciones de densidad a través de las funciones *kernel* y la capacidad para identificar modas en un conjunto de datos espaciales y asignar puntos a las modas para establecer un cluster.
- Se implementó el método utilizando la biblioteca scikitlearn en el lenguaje de programación Python. El algoritmo encontró puntos con mayor densidad (modas) estableciendo los centroides y agrupando los datos.
- El Mean Shift presentó una limitación asociado al ancho de banda o bandwidth, debido a que resultó sensible a la elección de este parámetro, ya que es inversamente proporcional al número de clusters.
- Se implementó el método con un ancho de banda de 1.38 para el análisis de temperatura mínima en la región andina, utilizando datos del modelo CNRM 2020-2050, obteniendo 4 clusters, donde el 93.87 % de los datos se encuentran en un solo cluster lo que indicó que existe baja variabilidad con datos centrados en la media y una desviación estándar mínima mientras que los extremos y las anomalías se presentaron en un 6.13 % presentando una mayor desviación estándar caracterizados en el cluster 1, 2 y 3.
- La implementación del método en la estación meteorológica de Alao estableció 5 clusters con un ancho de banda de 1.82. Se analizó la temperatura mínima para el año 2015 y 2016. En el cluster 0 que corresponde al 72 % de días se mantuvo en una media de 7.65°C. En el cluster 1 existió una diferencia de 1.17 °C . En

el cluster 2 el año 2016 experimenta un incremento de temperatura de  $4.25^{\circ}\text{C}$  mientras que en el cluster 3 que ocurre 7 días al año existió un decremento de  $2.75^{\circ}\text{C}$ . Además, se encontró una anomalía en el cluster 4 de un decremento de temperatura de  $8.17^{\circ}\text{C}$  a  $1.25^{\circ}\text{C}$  que ocurre 11 días al año.

## Recomendaciones

- Se recomienda realizar la implementación enfocándose en la parte espacial del algoritmo, estableciendo la estimación de densidad a través de un kernel multivariado, dado que la aplicación realizada fue un caso particular de forma unidimensional.
- Realizar un estudio exhaustivo y pertinente de la elección del parámetro del ancho de banda (bandwidth), dado que este tema sigue en un campo de investigaciones activa.
- La implementación realizada fue mediante el lenguaje de programación Python, sin embargo se puede incursionar la implementación en otros lenguajes como Matlab y C++

# Bibliografía

---

- [1] **CARREIRA, Miguel.** “A review of mean-shift algorithms for clustering”. [en línea], 2015, (USA) [Consulta: 20 de octubre 2023]. Disponible en:  
<https://arxiv.org/abs/1503.00687>
- [2] **COMANICIU, Dorin & MEER; Peter.** “Mean Shift: A Robust approach Toward Feature Space Analysis”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* [en línea], 2002, (USA), vol. 24(5), págs. 603-619 [Consulta: 01 junio 2023]. Disponible en:  
<http://dx.doi.org/10.1109/34.1000236>
- [3] **CHACÓN, José & MONFORT, Pablo.** “A comparison of bandwidth selectors for mean shift clustering”. [en línea], 2013, (España) [Consulta: 20 noviembre 2023]. Disponible en:  
<https://arxiv.org/abs/1310.7855>
- [4] **CHENG, Yizong.** “Mean shift, mode seeking, and clustering”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* [en línea], 1995, (USA), vol. 17(8), 790-799. [Consulta: 10 junio 2023]. Disponible en:  
<http://dx.doi.org/10.1109/34.400568>
- [5] **COMANICIU, D; et al.** “The variable bandwidth mean shift and data-driven scale selection”. *Proceedings Eighth IEEE International Conference on Computer Vision* [en línea], 2001, (Canada), vol. 1, págs. 438-445. [Consulta: 30 octubre 2023]. Disponible en:  
<https://dx.doi.org/10.1109/ICCV.2001.937550>.
- [6] **DUDA, Richard; et al.** *Pattern Classificaton*. 2da ed. Jhon Wiley & Sons inc, 2001, págs. 161-172.

## Bibliografía

---

- [7] **EVERITT, Brian; et al.** *Cluster Analysis*. 5ta ed. Jhon Wiley & Sons inc, 2011, págs. 01-13.
- [8] **FASHING, Mark & TOMASI, Carlo.** “Mean Shift Is a Bound Optimization”. [en línea], 2005, vol. 27, págs. 471-474.[Consultado en: 18 junio 2023]. Disponible:  
<https://doi.org/10.1109/TPAMI.2005.59>
- [9] **FUKUNAGA, K & HOSTETLER.** “The estimation of the gradient of a density function, with applications in pattern recognition”. *IEEE Transactions on Information Theory*[en línea], 1975,,vol. 21, págs. 32-40.[CONSULTA:18 junio 2023]. Disponible en:  
<https://dx.doi.org/10.1109/TIT.1975.1055330>.
- [10] **MEER, Peter & COMANICIU, Dorin.** “Mean Shift Analysis and Applications”. *Computer Vision, IEEE International Conference on* [en línea],1999,(USA), vol. 2, 1197. [Consulta: 30 octubre 2023]. Disponible en:  
<http://dx.doi.org/10.1109/ICCV.1999.790416>
- [11] **WAND, M.P & JONES, M.C** *Kernel Smoothing*[en línea].New York-USA:Springer-science and business media, B.V. ,1995.[Consulta: 06 noviembre 2023]. Disponible en:  
<https://doi.org/10.1201/b14876>
- [12] **STORK, David; et al..** *Pattern classification* [en línea].2da ed.USA: Wiley editorial, 2000.[Consulta: 06 noviembre 2023]. Disponible en:  
[https://books.google.com/books/about/Pattern\\_Classification.html?id=Br33IRC3PkQC](https://books.google.com/books/about/Pattern_Classification.html?id=Br33IRC3PkQC)
- [13] **WU, K.L & YANG, M.S** “Mean shift based clustering”. *Pattern Recognition*[en línea],2007,(Taiwan),vol. 40, págs 3035-3052.[Consulta: 8 noviembre 2023]. ISSN 0031-3203. Disponible en:  
<https://doi.org/10.1016/j.patcog.2007.02.006>.

# Índice alfabético

---

Ampliación Taylor, [22](#)

Distribución de probabilidad, [4](#)

Error cuadrático medio, [19](#)

Error cuadrático medio integrado, [21](#)

Esperanza matemática, [6](#)

Función de densidad, [6](#)

Función kernel, [8](#)

Kernel, [33](#)

Orden  $o$ , [22](#)

Soporte compacto, [11](#)

Variable aleatoria, [3](#)

Variable aleatoria continua, [5](#)

Variable aleatoria discreta, [4](#)