

Clusterización Funcional Espacial

Nosotros

2022-09-20

Contents

Prefacio	5
1 Introducción	7
1.1 Motivación de datos funcionales espaciales	8
1.2 Por qué es importante el análisis de datos	8
2 Estadística espacial	9
2.1 Preliminares	9
2.2 Clases de Datos Espaciales	10
2.3 Geoestadística	11
2.4 Variograma	21
2.5 Geoestadística Espacio Temporal	62
3 Introducción al análisis de datos funcionales (FDA)	73
3.1 Espacios de Hilbert	74
3.2 Campos aleatorios funcionales espaciales	77
3.3 Campos aleatorios funcionales espaciales multivariantes	77
3.4 Reducción de Dimensionalidad	77
3.5 Detección de atípicos	87
3.6 ANOVA Para datos funcionales	92
4 Introducción a la estadística esepacial funcional	97
4.1 Proceso Espacial Funcional	98
4.2 Datos funcionales espaciales	98

4.3	Variable Regionalizada Funcional	98
4.4	Función de Distribución Conjunta Funcional	98
4.5	Función de Media Funcional	98
4.6	Función de varianza Funcional	98
4.7	Función de Covarianza Funcional	98
4.8	Función de correlación o Correlograma Funcional	98
4.9	Función de semi-variograma Funcional	98
4.10	Estacionariedad de Funciones Aleatorias	98
4.11	No estacionariedad de Funciones Aleatorias	98
4.12	Correlación Espacial Funcional	98
4.13	Índices de correlación espacial	98
5	Agrupación	99
5.1	Algoritmos de agrupación	99
5.2	Agrupación: Datos Multivariantes (apéndice)	99
5.3	Agrupación: Datos Funcionales	99
5.4	Agrupación: Datos Espaciales	104
5.5	Agrupación: Datos funcionales con correlación espacial	107
6	Aplicaciones	113
6.1	Aplicación 1	113
6.2	Aplicación 2	113
6.3	Aplicación 3	113
6.4	Aplicación 4	113
7	Apéndice	115
7.1	Introducción a la simulación geoestadística condicional y no condicional	115
7.2	Estudio de simulación (referencia al artículo)	115
7.3	Índices	115

Prefacio

La dependencia espacial en datos medio ambientales es un criterio influyente en procesos de agrupación, dado que los resultados obtenidos brindan información relevante. Como los métodos clásicos no consideran la dependencia espacial, al considerar esta estructura se producen resultados inesperados, proporcionando agrupaciones de curvas que pueden no ser similares en forma y/o comportamiento. En este trabajo se realiza la agrupación mediante el método k-medias modificado para datos funcionales correlacionados espacialmente aplicado a datos del Índice de Vegetación de Diferencia Normalizada (NDVI) de los páramos del Ecuador. Para esto se implementan índices de calidad que permitan obtener el número adecuado de grupos. Con base en la metodología desarrollada en el método de clasificación jerárquico para datos funcionales con correlación espacial, y dado que los datos funcionales pertenecen al espacio de Hilbert de funciones cuadrado-integrables; se desarrolla el análisis considerando la distancia entre curvas a través de la norma \mathcal{L}^2 , obteniendo una representación reducida de los datos a través de una base finita de tipo Fourier. Luego, se calcula el variograma empírico y se ajusta a un modelo teórico para así ponderar la matriz de distancia entre las curvas por el trazo-variograma y variograma multivariado calculado con los coeficientes de las funciones base, y esta matriz se utiliza para la agrupación de datos funcionales correlacionados espacialmente. Para la validación del método se realizaron varios escenarios de simulación, obteniéndose buenos resultados de correcta clasificación y se complementa con un caso de aplicación a datos del NDVI, donde se obtuvieron cinco regiones distribuidas latitudinalmente.

Chapter 1

Introducción

Los métodos de agrupación tienen por objetivo identificar grupos homogéneos de observaciones que representan la realización de alguna variable aleatoria X (?), y en los cuales los miembros dentro de los grupos tienden a ser similares y distintos de otros miembros de otros grupos (?). Estos métodos también son conocidos como métodos de clasificación no supervisada, pues las observaciones no tienen asociadas ninguna etiqueta de grupo o categoría de forma previa (?); por lo cual, este tipo de métodos se utilizan con frecuencia para detectar patrones especiales en una base de datos, los cuales pueden tener interpretaciones convenientes por parte del usuario (?).

Uno de los métodos más populares de agrupación es el algoritmo k-medias, pues desde 1955 se ha venido utilizando y adaptando a miles de problemas en diferentes campos de las ciencias (?). Muchas de estas adaptaciones asumen que las observaciones de tipo escalar, espacial y multivariante, pueden ser representadas como puntos en un espacio Euclídeo de dimensión finita, pero cuando la dimensión del objeto es sumamente grande las realizaciones de la variable aleatoria X pueden tomar valores en un espacio infinito dimensional (?); este último tipo de dato es muy común, ya que hoy en día se pueden realizar mediciones en cada instante de tiempo y es conocido como dato funcional. Es así que, este problema se puede abordar desde el análisis de datos funcionales, pues su objetivo es inferir la estructura de los datos trabajando con su naturaleza de dimensión infinita sobre espacios de Hilbert (?).

Dado que el campo de aplicación es amplio para el uso de estos métodos, nos centraremos en el área medioambiental, pues en esta área se trabaja con datos temporales, los cuales según (?), también son considerados como datos funcionales. Aplicar un método de agrupación, omitiendo su estructura espacial, dará como resultado grupos en los cuales las curvas pueden no ser similares en forma y/o comportamiento (?). Para abordar este problema, hasta el momento se han propuesto métodos de agrupación como el jerárquico (?) y el dinámico (?), los cuales usan el trazo-variograma como medida de dependencia espacial,

con la diferencia de que en el método jerárquico se realiza el cálculo de manera global, mientras que en el método dinámico se lo realiza de manera local; así, el método jerárquico usa la distancia euclidiana entre las curvas, mientras que el método dinámico usa la distancia euclidiana al cuadrado entre las curvas dentro de cada grupo (?). Al notar la escasez de opciones de métodos, así como la no implementación del método k-medias, contar con más opciones de agrupación que trabajen con este tipo de datos serán de gran ayuda; es por ello que en este trabajo se implementará el método k-medias para datos funcionales con correlación espacial, para lo cual se usará la metodología aplicada en (?), y para evaluar el desempeño del método propuesto, se realizarán varios escenarios de simulación, y posteriormente aplicar este método a los datos del NDVI de los páramos del Ecuador.

1.1 Motivación de datos funcionales espaciales

Cosas...

1.2 Por qué es importante el análisis de datos

cosas...

1.2.1 Data 1(H)

1.2.2 Data 2(C)

1.2.3 Data 3(N)

Chapter 2

Estadística espacial

2.1 Preliminares

2.1.1 Función Aleatoria

Dado un dominio $D \subset \mathbb{R}^n$ y un espacio de probabilidad (Ω, \mathcal{A}, P) , una función aleatoria es una función de dos variables $Z(x, \omega)$ tal que para cada $x \in D$, $Z(x, \cdot)$ es una variable aleatoria en (Ω, \mathcal{A}, P) . Por otro lado, $Z(\cdot, \omega)$, definido en D como la sección de la función aleatoria en $\omega \in \Omega$ es una realización de la función aleatoria. La función aleatoria $Z(x, \omega)$ se denota como $Z(x)$ y una realización es representada por $z(x)$ (?).

2.1.2 Proceso Estocástico

Según (?), una función aleatoria es llamada proceso estocástico cuando x varia en un espacio unidimensional como el tiempo; por otro lado, si x varia en más de una dimensión es conocido como campo aleatorio.

2.1.3 Proceso Espacial

Sea Z la variable aleatoria de interés y s su ubicación espacia. El proceso espacial es el proceso estocástico

$$\{Z(s) : s \in D\} \tag{2.1}$$

donde $D \subset \mathbb{R}^d$ es el conjunto formado por todas las ubicaciones espaciales (?).

2.1.4 Proceso Temporal

Sea Z la variable aleatoria de interés y t el tiempo. El proceso temporal es el proceso estocástico

$$\{Z(t) : t \in D\} \quad (2.2)$$

donde $D \subset \mathbb{R}$ es el conjunto formado por todos los instantes de tiempo (?).

2.1.5 Proceso Espacio Temporal

Sea Z la variable aleatoria de interés, s su ubicación espacial y t el tiempo de ocurrencia. Un proceso espacio-temporal es un proceso estocástico

$$\{Z(s, t) : (s, t) \in D \times D'\} \quad (2.3)$$

donde el conjunto $D \subset \mathbb{R}^d$ es el conjunto de todas las ubicaciones espaciales y el conjunto $D' \subset \mathbb{R}$ es el conjunto de todos los instantes de tiempo. Los conjuntos D y D' pueden ser continuos o discretos, fijos o aleatorios (?).

2.2 Clases de Datos Espaciales

Dependiendo de las características del dominio espacial D del proceso estocástico de interés, se tienen los siguientes tipos de datos espaciales principales, los cuales son geoestadística, datos de área o *lattice*, procesos o patrones espaciales puntuales y datos georreferenciados.

2.2.1 Geoestadística

Es el conjunto de métodos aplicados a datos espaciales, donde las ubicaciones espaciales s provienen de un conjunto continuo y fijo $D \subset \mathbb{R}^d$; cabe resaltar que el propósito de la geoestadística es la interpolación. Además, si el conjunto D no es continuo se puede llegar a obtener predicciones carentes de sentido, y se considera D fijo pues los puntos en el espacio se seleccionan a conveniencia o bajo un esquema de muestreo probabilístico (?).

2.2.2 Datos de Área

En este tipo de datos las ubicaciones espaciales pertenecen a un conjunto discreto contable y fijo $D \subset \mathbb{R}^d$. Las ubicaciones pueden estar espaciadas regular o irregularmente, usualmente son irregulares pues sus separaciones no siguen un patrón predecible(?).

2.2.3 Patrones Puntuales

En este caso, las ubicaciones espaciales pertenecen a un conjunto discreto o continuo y aleatorio $D \subset \mathbb{R}^d$. Con este tipo de dato se realizan análisis con el propósito de determinar si la distribución de las mediciones en la región es aleatoria, agregada o uniforme (?).

2.2.4 Datos Georreferenciados

En este caso, las variables de interés tienen asociadas las coordenadas de las ubicaciones donde fueron medidas, las que pueden ser geográficas, planas o cartesianas (?).

2.3 Geoestadística

La palabra geoestadística fue establecida por Hart en el año de 1954. El significado de esta palabra se lo puede deducir a partir de su prefijo geo, que hace referencia a la tierra; por lo tanto, la geoestadística es una rama de la estadística enfocada en el análisis y la modelización de la variabilidad espacial de fenómenos que ocurren dentro de una zona geográfica; además, es considerada como una disciplina híbrida entre Geología, Matemática, Estadística y Minería de datos (?).

2.3.1 Variable Regionalizada

Una variable regionalizada es un proceso estocástico con dominio continuo contenido en un espacio euclidiano d -dimensional $\{Z(s) : s \in D \subset \mathbb{R}^d\}$.

En el caso de que las mediciones sean realizadas en una superficie; es decir, $d = 2$, $Z(s)$ puede asociarse a la variable aleatoria ligada a ese punto del plano, s representa las coordenadas geográficas y Z la variable en cada una de ellas. La medición de esta variable aleatoria puede representar la magnitud de una variable ambiental medida en un conjunto de coordenadas en la región de estudio (?).

Un proceso estocástico espacial se caracteriza por su distribución de probabilidad de dimensión finita; es decir, la distribución de probabilidad conjunta de un conjunto de variables $Z(s_1), \dots, Z(s_k)$ para todo $k \in \mathbb{N}$ y para todos los puntos $s_1, \dots, s_k \in D$ (?).

2.3.2 Función de Distribución Conjunta

Considérese una función/campo aleatorio $Z = \{Z(s) : s \in D\}$ y k ubicaciones espaciales $s_1, \dots, s_k \in D$. El vector aleatorio $\{Z(s_1), \dots, Z(s_k)\}$ está caracterizado por su función de distribución conjunta:

$$F_{s_1, \dots, s_k}(z_1, \dots, z_k) = P[Z(s_1) \leq z_1, \dots, Z(s_k) \leq z_k] \quad (2.4)$$

2.3.3 Función de Media

La esperanza o el momento de primer orden de un campo aleatorio o proceso estocástico espacial $Z(s)$, es una función no aleatoria de s

$$E[Z(s)] = \mu(s) \quad \text{donde} \quad \mu(s_i) = E(Z(s_i)) \quad i = 1, \dots, k \quad (2.5)$$

En cada sitio s dado, $\mu(s)$ representa la media alrededor de la cual se distribuyen los valores tomados por las realizaciones de la funciones aleatoria (?). Si la esperanza del campo aleatorio varía con la ubicación espacial, se le conoce como deriva o *drift* del campo aleatorio.

En geoestadística existen tres elementos de segundo orden, los cuales son: varianza, covarianza y variograma.

EJEMPLO DE MEDIA CON CALCIO, AQUIFER, NITRATO(MARTA)

2.3.4 Función de varianza

La varianza de un campo aleatorio o proceso estocástico espacial $Z(s)$ respecto a $\mu(s)$, está definida por:

$$V(s) = \sigma^2(s) = \text{Var}[Z(s)] = E[(Z(s) - \mu(s))^2] \quad \text{donde} \quad V(s_i) = \sigma^2(s_i) = V(Z(s_i)) \quad i = 1, \dots, k \quad (2.6)$$

EJEMPLO DE VARIANZA CON CALCIO, AQUIFER, NITRATO(MARTA)

2.3.5 Función de Covarianza

Sean $Z(s_i)$ y $Z(s_j)$ dos variables aleatorias de un proceso estocástico espacial, la covarianza es una función de separación espacial de s_i y s_j y está definida por:

$$C(s_i, s_j) = C(s_i - s_j) \quad (2.7)$$

$$(2.8)$$

$$C(s_i, s_j) = C(Z(s_i), Z(s_j)) \quad (2.9)$$

$$(2.10)$$

$$C(s_i, s_j) = E[(Z(s_i) - \mu(s_i))(Z(s_j) - \mu(s_j))] \quad (2.11)$$

$$(2.12)$$

$$C(s_i, s_j) = E[Z(s_i)Z(s_j)] - \mu(s_i)\mu(s_j) \quad (2.13)$$

Nótese que:

$$C(s, s) = C(0) = \sigma^2(s) \quad (2.14)$$

EJEMPLO DE COVARIANZA CON CALCIO, AQUIFER, NITRATO(MARTA)

2.3.6 Función de correlación o Correlograma

La correlación lineal de dos variables aleatorias de un proceso estocástico espacial $Z(s_i)$ y $Z(s_j)$, está definida por:

$$\rho(s_i, s_j) = \frac{C(s_i, s_j)}{\sigma(s_i)\sigma(s_j)} \quad (2.15)$$

2.3.7 Función de semi-variograma

El semi-variograma entre dos variables aleatorias de un proceso estocástico espacial $Z(s_i)$ y $Z(s_j)$ está dado por:

$$\gamma(s_i, s_j) = \gamma(s_i - s_j) = \frac{1}{2}V[Z(s_i) - Z(s_j)], \quad \forall s_i, s_j \in D \quad (2.16)$$

EJEMPLO DE VARIOG CON CALCIO, AQUIFER, NITRATO(MARTA)

2.3.8 Estacionariedad de Funciones Aleatorias

Una variable aleatoria regionalizada vista como una realización de un proceso estocástico espacial o función aleatoria $\{Z(s) : s \in D\}$, probabilísticamente adquiere sentido cuando es posible inferir toda o parte de la ley de probabilidad del proceso estocástico o campo aleatorio. Es claro que inferir la ley de probabilidad cuando solo se tiene una realización del proceso estocástico espacial o función aleatoria sería imposible; para poder hacer inferencia consistente, se necesitan de varias realizaciones, pero en la realidad solo existe una, inclusive solo una parte de las realizaciones está disponible en las ubicaciones de la muestra; para dar solución a este problema, se introduce el supuesto de estacionariedad, que significa que la ley de probabilidad espacial del proceso estocástico espacial o función aleatoria, o parte de este, es invariante respecto a traslaciones; es decir, las propiedades probabilísticas de un conjunto de observaciones no dependen de las ubicaciones donde fueron medidas, solo dependen de su separación; además, como se mencionó, las realizaciones no son independientes (?). Así, se distinguen tres tipos de estacionariedad:

- Estacionariedad fuerte o de primer orden.
- Estacionariedad débil o de segundo orden.
- Estacionariedad intrínseca .

Estacionariedad Fuerte o de primer orden

Este supuesto se refiere a que el proceso alcanza un estado de equilibrio. Formalmente, una variable regionalizada es estacionaria si su función de distribución no varía con la traslación del vector \mathbf{h} ; esto es, si para:

$$Z(s) = [Z(s_1), \dots, Z(s_n)]' \quad (2.17)$$

$$(2.18)$$

$$Z(s + \mathbf{h}) = [Z(s_1 + \mathbf{h}), \dots, Z(s_n + \mathbf{h})]' \quad (2.19)$$

se tiene que:

$$F_{Z_1, \dots, Z_n}(s_1, \dots, s_n) = F_{Z_1, \dots, Z_n}((s_1 + \mathbf{h}), \dots, (s_n + \mathbf{h})) \quad (2.20)$$

Nótese que esta condición es muy restrictiva, por lo que normalmente se relaja a condiciones de segundo orden, las que limitan el supuesto de estacionariedad a los dos primeros momentos del proceso estocástico o campo aleatorio (?).

EJEMPLOS (MARTA)??

Estacionariedad de Segundo Orden

Sea $\{Z(s) : s \in D\}$ un proceso estocástico o campo aleatorio, se dice débilmente estacionario o de segundo orden si tiene momentos de segundo orden finitos; es decir, que la covarianza existe y se verifica que:

- La esperanza existe, es constante a través del dominio D , y no depende de la ubicación s ; es decir:

$$E(Z(s)) = \mu(s) = \mu \quad (2.21)$$

- La covarianza existe para todo par de variables aleatorias, $Z(s)$ y $Z(s + \mathbf{h})$, y solo depende del vector de separación \mathbf{h} entre las ubicaciones s y $s + \mathbf{h}$; es decir, depende de la dirección y la distancia de separación, y no de sus ubicaciones absolutas y se tiene que:

$$\text{Cov}(Z(s), Z(s + \mathbf{h})) = C(\mathbf{h}), \forall s \in D \text{ y } \mathbf{h} \in \mathbb{R}^d \quad (2.22)$$

La estacionariedad de la covarianza implica que la varianza $\text{Var}(Z(s))$ existe, es finita y no depende de s ; es decir:

$$\text{Var}(Z(s)) = \sigma^2(s) = C(s, s) = C(s - s) = C(0) = \sigma^2 \quad (2.23)$$

La estacionariedad de segundo orden implica la siguiente relación entre la función de semivarianza y la de autocovarianza:

$$\gamma(\mathbf{h}) = \frac{1}{2}V(Z(s + \mathbf{h}) - Z(s)), \quad \text{con } \gamma(\mathbf{h}) = \gamma(s + \mathbf{h}, s) \quad (2.24)$$

$$= \frac{1}{2}(V(Z(s + \mathbf{h})) + V(Z(s)) + 2C(Z(s + \mathbf{h}), Z(s))) \quad (2.25)$$

$$= \frac{1}{2}(C(0) + C(0) + 2C(\mathbf{h})) \quad (2.26)$$

$$= C(0) - C(\mathbf{h}) \quad (2.27)$$

Nótese que si un proceso estocástico o campo aleatorio es estacionario de segundo orden, entonces también será un proceso estocástico estacionariamente fuerte; sin embargo, esto no ocurre en el sentido contrario.

Se debe tener en cuenta que la estacionariedad de segundo orden implica la existencia de la varianza del proceso estocástico o campo aleatorio. Existen

fenómenos con varianza infinita lo que imposibilita su modelización utilizando procesos de este tipo. Sin embargo, existen casos en los que los incrementos, $Z(s + \mathbf{h}) - Z(s)$, tienen varianza finita y que además son un proceso estacionario de segundo orden; este tipo de proceso estocástico o campo aleatorio se conoce como un proceso con estacionariedad intrínseca (?).

EJEMPLOS (MARTA)??

Estacionariedad Intrínseca

Sea $\{Z(s) : s \in D\}$ un proceso estocástico o campo aleatorio; se dice que es un proceso con estacionariedad intrínseca, si:

- $Z(s)$ tiene esperanza finita y constante para todo punto en el dominio, lo cual implica que la esperanza de los incrementos sea cero.

$$E(Z(s + \mathbf{h}) - Z(s)) = 0$$

- Para cualquier vector \mathbf{h} , la varianza del incremento está definida y es una función única de la distancia.

$$V(Z(s + \mathbf{h}) - Z(s)) = E(Z(s + \mathbf{h}) - Z(s))^2 = \gamma(\mathbf{h})$$

EJEMPLOS (MARTA)??

2.3.9 No estacionariedad de Funciones Aleatorias

Sea $\{Z(s) : s \in D\}$ un proceso estocástico o campo aleatorio para el cual la media y/o la función de covarianza depende de la ubicación espacial; es decir, no son traslacionalmente invariantes; se dice que $Z(s)$ es una función aleatoria no estacionaria.

Cuando el proceso estocástico o campo aleatorio $\{Z(s) : s \in D\}$ tiene media no constante y varía con la ubicación espacial, y sus incrementos de primer orden $Z(s + \mathbf{h}) - Z(s)$ son no estacionarios, se dice que la función aleatoria es no intrínseca (?).

EJEMPLOS (BUSCAR)??

2.3.10 Isotropía

Identificar la estacionariedad en un campo temporal es fácil, ya que solo existe una dirección de variación. Por otro lado, en el campo espacial existen múltiples direcciones, por lo que se debe asumir que el fenómeno es estacionario en todas ellas. Cuando la esperanza de la variable cambie respecto a las direcciones o

cuando la covarianza o correlación dependan del sentido en el que se determinen, entonces se dirá que no se tiene estacionariedad (?).

Ahora bien, si $C(\mathbf{h})$ y/o $\gamma(\mathbf{h})$, son funciones únicas de la magnitud $h = \|\mathbf{h}\|$

$$Cov(Z(s), Z(s + \mathbf{h})) = C(\|\mathbf{h}\|) = C(h) \quad (2.28)$$

$$\frac{1}{2}Var(Z(s + \mathbf{h}) - Z(s)) = \gamma(\|\mathbf{h}\|) = \gamma(h) \quad (2.29)$$

el proceso posee función de covarianza y/o semivarianza isotrópica. La estacionariedad posibilita combinar pares de datos con la misma diferencia en magnitud de separación de las coordenadas, y si los vectores de diferencias tienen la posibilidad de ser reemplazados con distancias escalares, entonces el campo se dice isotrópico (?). Si la correlación entre los datos no es dependiente de la dirección en la que esta se calcule se plantea que el fenómeno es isotrópico (?); en caso opuesto un campo aleatorio que es estacionario pero no isotrópico se desenvuelve de forma distinta según las diferentes direcciones del espacio; por lo cual, no es suficiente conocer cuánto permanecen distanciadas las ubicaciones, sino además es necesario conocer la orientación de esa distancia; a dichos campos se los conoce como campos aleatorios anisotrópicos (?).

EJEMPLOS AQUIFER (SIN TRENDS)

2.3.11 Anisotropía

Buscar teoría y poner lo de las transformaciones con funciones trigonométricas

2.3.12 Correlación Espacial

El análisis estructural es la primera etapa en el desarrollo de un análisis geostatístico; a partir de este se puede expresar la estructura de dependencia espacial o correlación que existe entre los datos medidos de una variable mediante funciones como el variograma y/o covariograma (?).

Covariograma

Como se definió anteriormente, la función de covarianza de un proceso estocástico espacial o campo aleatorio esta dada por:

$$C(s_i, s_j) = C(Z(s_i), Z(s_j)) \quad (2.30)$$

$$= E[(Z(s_i) - \mu(s_i))(Z(s_j) - \mu(s_j))], \forall s_i, s_j \in D \quad (2.31)$$

Bajo la hipótesis de estacionariedad de segundo orden, la función de covarianza tiene las siguientes propiedades:

- Solo depende del vector de separación \mathbf{h} entre las ubicaciones espaciales:

$$C(h) = E[(Z(s + \mathbf{h}) - \mu)(Z(s) - \mu)], \quad \forall s, s + \mathbf{h} \in D \subset \mathbb{R}^d \quad (2.32)$$

donde μ es la media constante del proceso estocástico espacial o campo aleatorio. El covariograma muestra el comportamiento de la correlación entre $Z(s)$ y $Z(s + \mathbf{h})$. Cuando la función de covarianza solo depende de la distancia entre las ubicaciones s y $s + \mathbf{h}$ se conoce como **proceso isotrópico**. Cuando depende de la distancia y de la dirección del vector de separación \mathbf{h} se conoce como un **proceso anisotrópico**.

- Está acotado por la varianza del proceso estocástico espacial o campo aleatorio en el origen; es decir:

$$|C(h)| \leq C(0) = \text{Var}(Z(s)) \quad (2.33)$$

- Es una función par; es decir:

$$C(h) = C(-h) \quad (2.34)$$

- Es una función definida positiva, en términos de diferencia entre coordenadas $s_i - s_j = \mathbf{h}$:

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(s_i - s_j) \geq 0, \quad \forall n \in \mathbb{N}^*, \forall \lambda_1, \dots, \lambda_n \in \mathbb{R}, \forall s_1, \dots, s_n \in D \subset \mathbb{R}^d \quad (2.35)$$

Esta condición es la más importante y necesaria para que una función de covarianza esté bien definida (?).

- Si $C_k(\mathbf{h}), \forall k \in \mathbb{N}$ son funciones de covarianza en \mathbb{R}^d y:

$$\lim_{k \rightarrow \infty} C_k(\mathbf{h}) = C(\mathbf{h}), \quad \forall \mathbf{h} \in \mathbb{R}^d \quad (2.36)$$

entonces, $C(\mathbf{h})$ es una función de covarianza en \mathbb{R}^d , teniendo en cuenta que este límite existe $\forall \mathbf{h}$.

- Toda combinación lineal de funciones de covarianza con coeficientes positivos, también es una función de covarianza.
- El producto de funciones de covarianza es también una función de covarianza.

Semivariograma

El semivariograma es el instrumento más usado para describir la dependencia espacial de variables regionalizadas, pues este cubre un mayor espectro de estas variables en comparación de la función de covarianza, incluyendo la estacionariedad intrínseca de funciones aleatorias, en donde la covarianza no puede ser definida (?).

Como se definió anteriormente, el semivariograma está dado por:

$$\gamma(s_i - s_j) = \frac{1}{2} \text{Var}(Z(s_i) - Z(s_j)), \forall s_i, s_j \in D \quad (2.37)$$

Bajo el supuesto de estacionariedad de segundo orden o de hipótesis intrínseca (sin media/drift), se escribe como:

$$\gamma(h) = \frac{1}{2} \text{Var}(Z(s + \mathbf{h}) - Z(s)) = \frac{1}{2} E((Z(s + \mathbf{h}) - Z(s))^2) \quad (2.38)$$

la cual muestra como la disimilitud entre $Z(s + \mathbf{h})$ y $Z(s)$ se desenvuelve con la distancia h .

Como se vio anteriormente, bajo el supuesto de estacionariedad de segundo orden, el semivariograma y el covariograma verifican la siguiente relación:

$$C(h) = C(0) - \gamma(h) \quad (2.39)$$

En la práctica se utiliza el semivariograma en lugar del covariograma puesto que este no necesita conocer la media de la función aleatoria (?).

Para que una función de semivariograma sea válido necesita verificar las siguientes propiedades teóricas:

- Por definición $\gamma(0) = 0$, aunque es común que presente una discontinuidad en el origen; esta discontinuidad se la conoce como efecto pepita.
- Es una función par; es decir:

$$\gamma(h) = \gamma(-h) \quad (2.40)$$

- Siempre toma valores mayores o iguales que cero:

$$\gamma(h) \geq 0 \quad (2.41)$$

- Bajo estacionariedad intrínseca, es una función condicionalmente negativa; es decir:

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(s_i - s_j) \leq 0, \forall n \in \mathbb{N}^*, \forall \lambda_1, \dots, \lambda_n \in \mathbb{R} : \sum_{i=1}^n \lambda_i = 0, \forall s_1, \dots, s_n \in D \subset \mathbb{R}^d \quad (2.42)$$

- El semivariograma de una proceso estocástico o campo aleatorio estacionario de segundo orden es finito; es decir, su comportamiento tiende a una línea horizontal a medida que se incrementa la separación de las coordenadas; sin embargo, el de una función aleatoria intrínsecamente estacionaria sin media o *drift* puede crecer al infinito; es decir, el semivariograma no se estabiliza; en este caso el semivariograma puede ser al menos intrínsecamente estacionario si crece más lento que una ecuación de segundo grado; es decir:

$$\lim_{|h| \rightarrow \infty} \frac{\gamma(h)}{|h|^2} = 0 \quad (2.43)$$

El semivariograma de un proceso estocástico espacial estacionario de segundo orden depende de los siguientes parámetros (ver Figura 2.1):

- Silla: Es la asíntota superior del semivariograma. Solamente los procesos estacionarios de segundo orden tienen silla; en estos casos la silla es de $C(0) = \sigma^2$.
- Rango: Es la distancia sobre la cual los puntos ya no ejercen influencia sobre otros; es decir, son independientes. Los puntos que se encuentran separados por una distancia menor o igual al rango se consideran espacialmente correlacionados, mientras que los puntos separados por una distancia mayor se consideran independientes (?).
- Efecto Pepita: Es una discontinuidad en el origen, $\gamma(h) \rightarrow c_0$ cuando $h \rightarrow 0$.

Existen dos causas principales para que se produzca este efecto. La primera es la existencia de error de medición (*EM*); sucede una vez que no es viable repetir una medición en la ubicación s sin error, evidenciando su variabilidad. La segunda causa es el efecto de micro-escala (*MS*); el cual se produce ya que existe un proceso espacial que opera a distancias más pequeñas de las que fueron consideradas en el muestreo. Si *ES* y *MS* son diferentes de cero, el semivariograma presentará una discontinuidad puntual en el origen, conocido como efecto pepita y representado de la siguiente manera (?):

$$c_0 = \sigma_{EM}^2 + \sigma_{MS}^2 \quad (2.44)$$

- Silla Parcial: Si un semivariograma tiene efecto pepita c_0 y silla $C(0)$, la diferencia $c_p = C(0) - c_0$ es la silla parcial del semivariograma, la cual representa la varianza de la variable $Z(s)$ sin el efecto pepita.

La relación entre el covariograma y el semivariograma se modifica para tomar en cuenta el efecto pepita de la siguiente manera:

$$\gamma(h) = c_0 + c_p \gamma'(h) \quad (2.45)$$

$$C(h) = \begin{cases} c_p(1 - \gamma'(h)) & \text{si } h > 0 \\ c_0 + c_p & \text{si } h = 0 \end{cases} \quad (2.46)$$

donde $\gamma'(h)$ es el modelo de variograma, c_p es la silla parcial y c_0 es el efecto pepita (?).

A partir de este punto se omitirá el prefijo “semi” y se tratará al semivariograma solo por variograma.

2.4 Variograma

En la práctica, se trabaja con las realizaciones del proceso estocástico o campo aleatorio en estudio; es decir, se cuenta con un conjunto de datos georeferenciados en un dominio dado. Utilizando estos datos, se puede inferir la estructura de dependencia espacial del fenómeno (?).

Por lo general, la descripción de la distribución espacial se limita a los primeros momentos. La esperanza o el momento de primer orden involucra un solo sitio a la vez y no proporciona información sobre la dependencia espacial. Por otro lado, la covarianza, correlograma y variograma o momentos de segundo orden están definidos mediante un par ubicaciones; así, estos momentos proporcionan información de la continuidad y dependencia espacial de la variable regionalizada (?).

2.4.1 Disimilitud contra Separación

La variabilidad de una variable regionalizada $Z(s)$, se mide en diferentes escalas y se calcula a través de la disimilitud entre los valores z_{s1} y z_{s2} ubicados en los puntos s_1 y s_2 del dominio espacial D . La medida de disimilitud $\hat{\gamma}$ de estos valores está dada por:

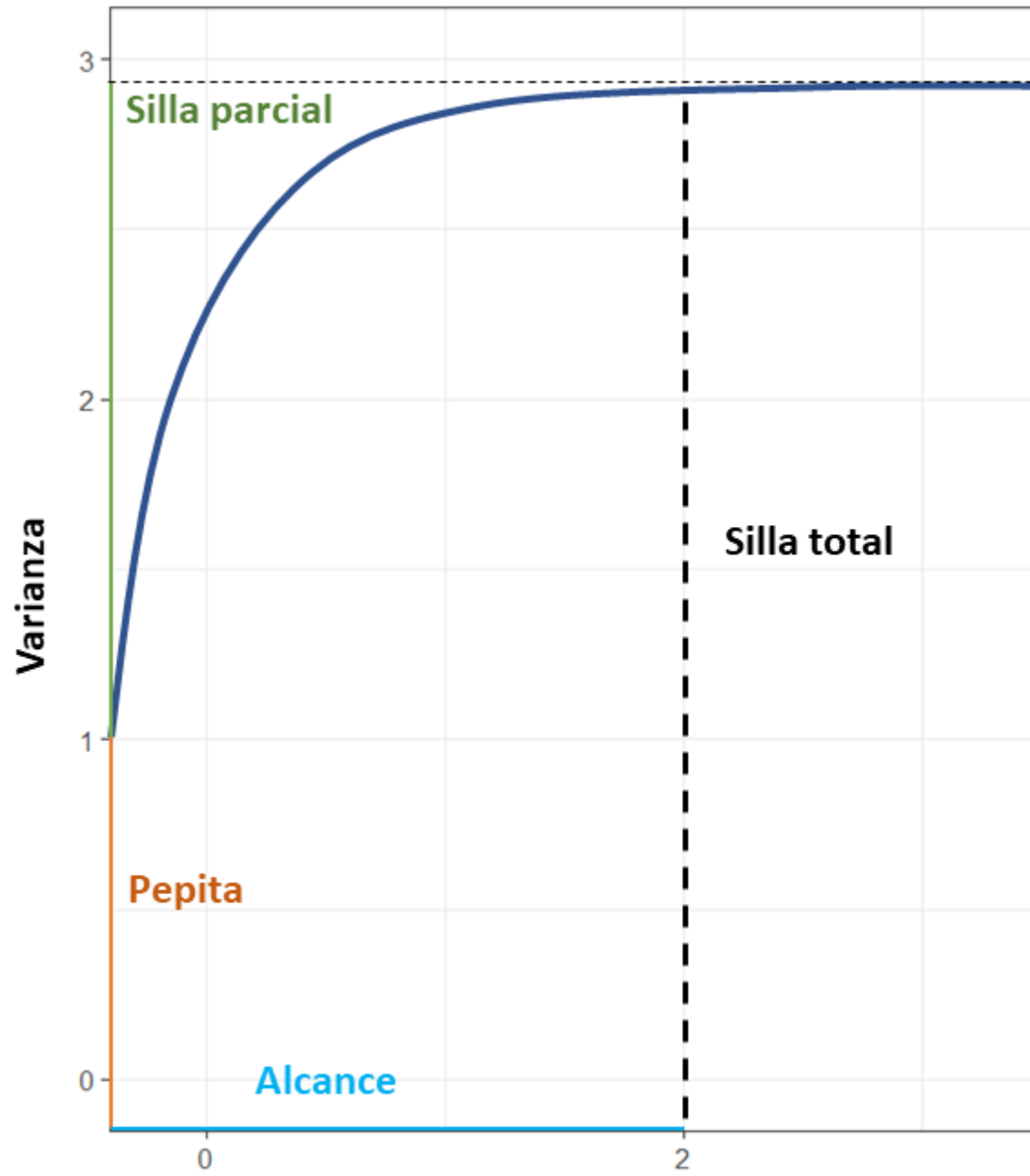


Figure 2.1: Partes de un variograma

$$\hat{\gamma}(s_1, s_2) = \frac{(z_{s1} - z_{s2})^2}{2} \quad (2.47)$$

Los dos puntos s_1 y s_2 en el espacio geográfico pueden unirse por un vector $\mathbf{h} = s_1 - s_2$ como se muestra en la Figura 2.2.

Si $\hat{\gamma}$ es dependiente de la separación y orientación del par de ubicaciones descritos por el vector \mathbf{h} , se tiene que (?):

$$\hat{\gamma}(h) = \frac{1}{2}(z(s_1 + \mathbf{h}) - z(s_1))^2 \quad (2.48)$$

La disimilitud es simétrica con respecto de h , pues es una cantidad al cuadrado, por lo que, su representación será mostrada utilizando todos los pares de datos de la muestra en el conjunto D . El gráfico de la disimilitud $\hat{\gamma}$ contra la separación espacial absoluta h , es llamado nube del variograma (?).

La disimilitud con frecuencia se incrementa con la distancia, pues los puntos cercanos tienden a ser similares (?) (ver Figura 2.3).

2.4.2 Variograma Empírico

Teniendo en cuenta que el variograma $\gamma(\mathbf{h})$ es la varianza de la variable de incrementos $Z(s + \mathbf{h}) - Z(s)$, el estimador clásico se basa en la estimación de esta varianza mediante el método de momentos y está definido para un vector de separación \mathbf{h} de la siguiente manera:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{N(\mathbf{h})} (Z(s + \mathbf{h}) - Z(s))^2, \quad \mathbf{h} \in \mathbb{R}^d \quad (2.49)$$

donde:

$$N(\mathbf{h}) = \{(s_i, s_j) : s_i - s_j = \mathbf{h}, i, j = 1, \dots, n\} \quad (2.50)$$

siendo $N(\mathbf{h})$ el conjunto de todos los pares de ubicaciones cuya separación corresponde a un vector \mathbf{h} y $|N(\mathbf{h})|$ es la cardinalidad de $N(\mathbf{h})$. Nótese que $N(-\mathbf{h}) \neq N(\mathbf{h})$, aunque $\hat{\gamma}(-\mathbf{h}) = \hat{\gamma}(\mathbf{h})$, y la media μ no necesita ser estimada.

Al ser un estimador de la varianza muestral es sensible a datos atípicos; es por ello que (?) propuso estimadores resistentes a datos atípicos, asumiendo normalidad marginal del campo aleatorio (?); sin embargo, la diferencia entre estos estimadores es pequeña (?).

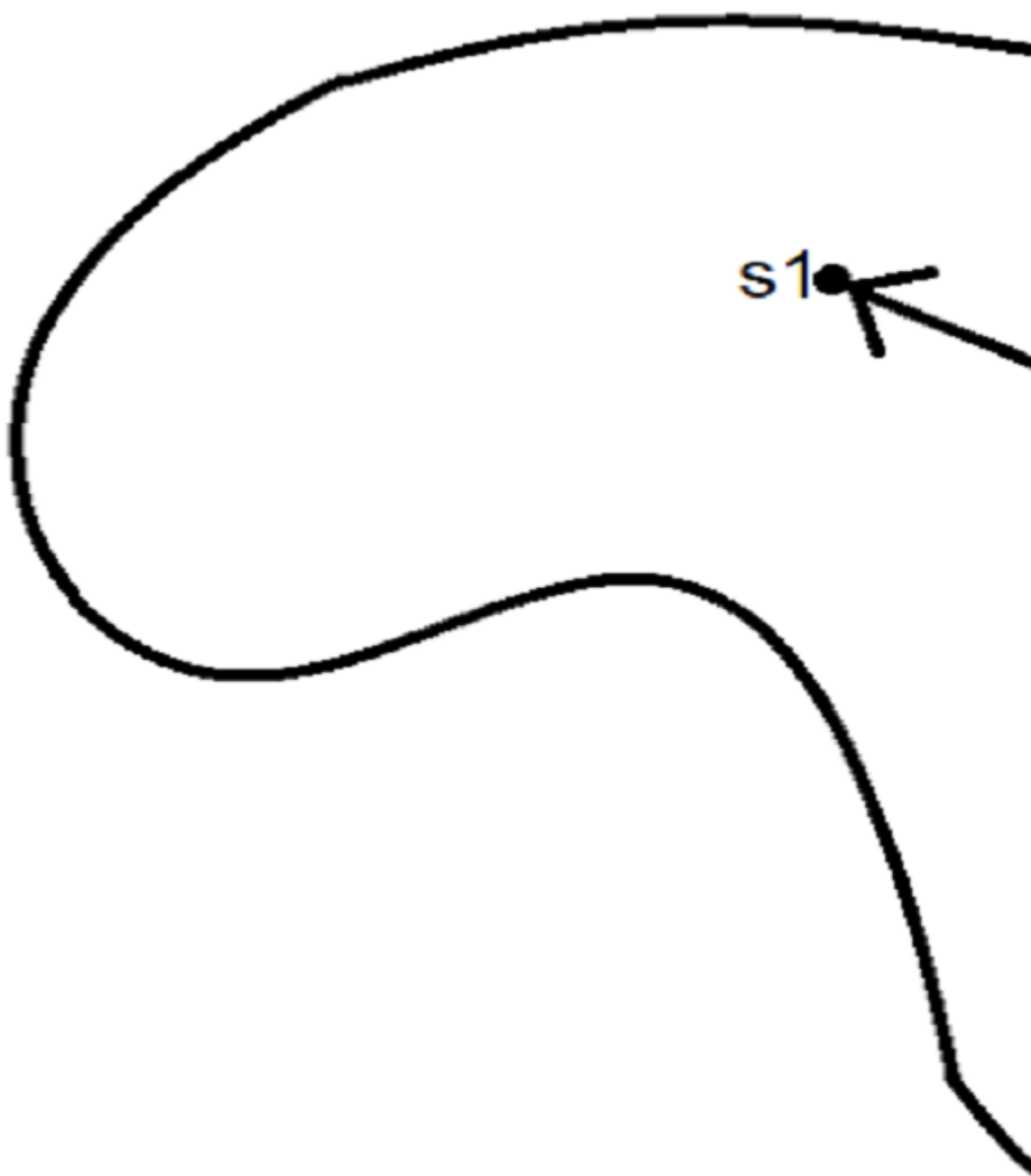


Figure 2.2: Distancia

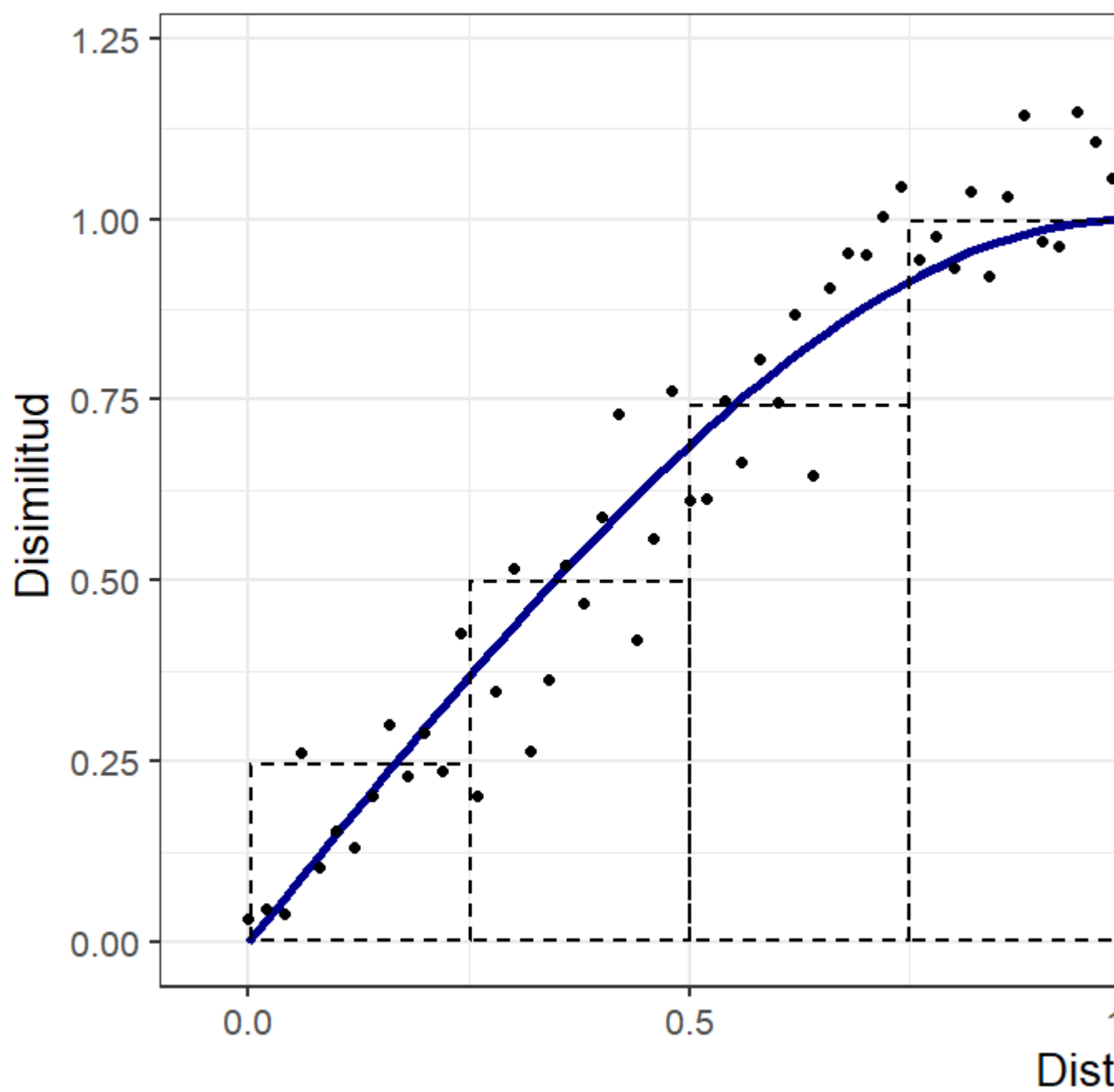


Figure 2.3: Disimilitud respecto de h.

Tolerancia en los parámetros de cálculo

En la práctica, los datos están distribuidos irregularmente en el dominio espacial D , por lo que el número de pares $|N(\mathbf{h})|$ que se utilizan en el cálculo del variograma empírico, $\hat{\gamma}(\mathbf{h})$ para un vector \mathbf{h} dado, es por lo general escaso; dando como resultado una estimación del variograma empírico errónea, provocando que sea imposible de interpretarlo y modelarlo. Para corregir esto, se suelen añadir algunas tolerancias de cálculo sobre las distancias y direcciones, teniendo así que:

$$\hat{\gamma}^+(\mathbf{h}) = \frac{1}{2|N^+(\mathbf{h})|} \sum_{N^+(\mathbf{h})} (Z(s + \mathbf{h}) - Z(s))^2 \quad (2.51)$$

$$\text{donde } N^+ = \{(s_i, s_j) : s_i - s_j \in T(\mathbf{h})\} = \bigcup_{\mathbf{h}' \in T(\mathbf{h})} N(\mathbf{h}')$$

Siendo $T(\mathbf{h})$ una región de tolerancia alrededor de \mathbf{h} ; es decir, $T(\mathbf{h}) = [\mathbf{h} - \Delta\mathbf{h}, \mathbf{h} + \Delta\mathbf{h}]$ en el caso unidimensional. En el caso de dimensión dos o tres, existen tolerancias sobre la longitud de \mathbf{h} y sobre su orientación (?) (véase Figura 2.4):

El ancho de banda limita la división del cono de tolerancia a una expansión máxima. En el espacio de tres dimensiones, se introducen dos anchos de banda, horizontal y vertical (?).

Propiedades del variograma empírico

- El variograma empírico $\hat{\gamma}(\mathbf{h})$ es un estimador insesgado del variograma teórico (?):

$$E(\hat{\gamma}(\mathbf{h})) = \gamma(\mathbf{h}) \quad (2.52)$$

- La varianza relativa es un indicador de robustez:

$$\frac{Var(\hat{\gamma}(\mathbf{h}))}{(\gamma(\mathbf{h}))^2} \quad (2.53)$$

A medida que más alta sea dicha varianza, más susceptible es el variograma empírico de fluctuar en torno a su valor esperado, y más difícil se vuelve la inferencia estadística.

Los principales factores que influyen en la varianza relativa son (?):

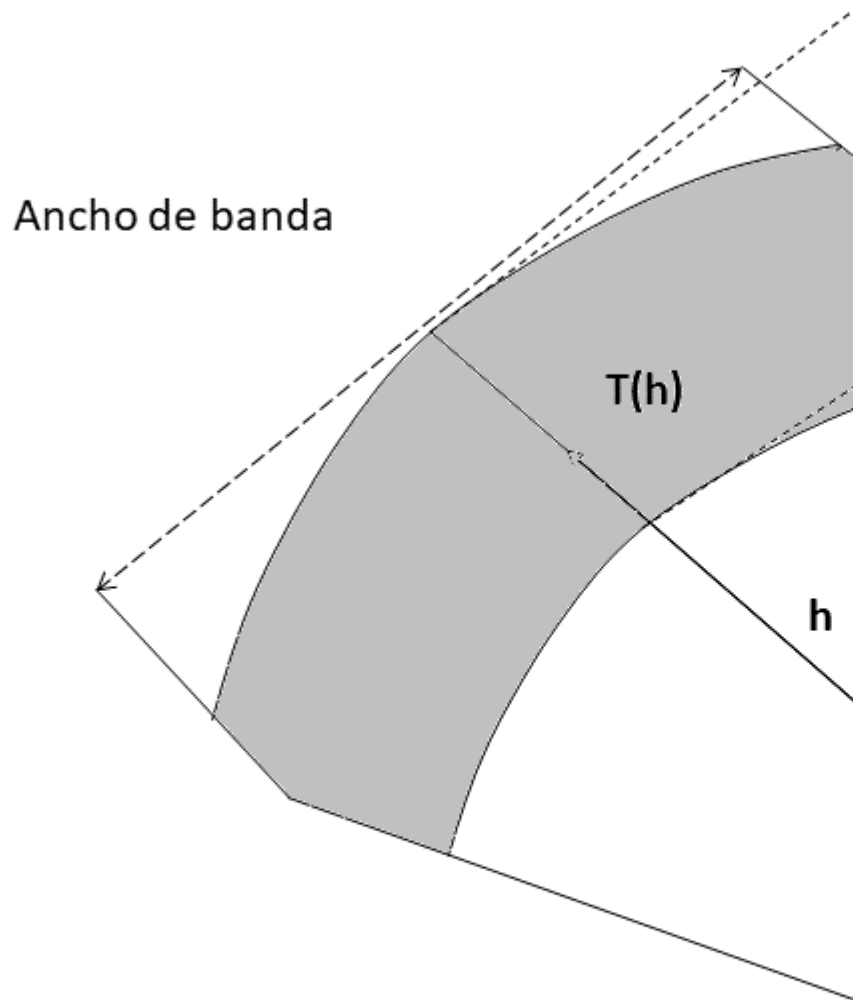


Figure 2.4: Región de tolerancia $T(extbfh)$ alrededor del vector $extbfh$

- Norma del vector \mathbf{h} : la varianza relativa crece cuando la distancia aumenta.
 - La irregularidad de la malla de muestreo, tiene la posibilidad de causar grandes variaciones en el cálculo del variograma empírico, incluso a distancias pequeñas.
 - A medida que más reducido es el número de pares de datos, más grandes son las variaciones.
 - La existencia de datos atípicos, puesto que en el cálculo del variograma empírico, se elevan los valores al cuadrado.
- Las direcciones de cálculo del variograma empírico deben tener en consideración la anisotropía de la variable regionalizada; en la situación de isotropía, donde los variogramas direccionales son semejantes, se puede tener en cuenta un variograma omnidireccional, determinado por:

$$\bar{\gamma}^+(h) = \frac{1}{2|N^+(\mathbf{h})|} \sum_{N^+(\mathbf{h})} (Z(s + \mathbf{h}) - Z(s))^2 \quad (2.54)$$

donde: $|N^+(\mathbf{h})| = \{(s_i, s_j) : s_i - s_j \approx \mathbf{h}\} (?)$.

- Si el variograma presenta comportamiento anisotrópico es necesario realizar transformaciones de las coordenadas que permitan obtener variables regionalizadas anisotrópicas a partir de los modelos isotrópicos (?).

Nube variográfica

Teniendo presente que bajo el supuesto de estacionariedad, los valores $\hat{\gamma}(s_i, s_j)$ son estimadores insesgados de los valores que corresponden $\gamma(s_i, s_j)$. La recopilación de pares de distancias y sus correspondientes valores de variograma, $\{(\mathbf{h}, \gamma(\mathbf{h})) : \mathbf{h} = s_i - s_j\}$, es conocido como variograma empírico y su gráfico como nube variográfica. Para mejorar la conducta del variograma empírico como un estimador del variograma teórico $\gamma(\mathbf{h})$ es necesario aplicar un tipo de suavización, puesto que se espera que la función $\gamma(\mathbf{h})$ varíe suavemente en función de \mathbf{h} , por lo cual se disminuye la varianza sin añadir sesgo llevando a cabo un promedio de los valores de $\hat{\gamma}$ sobre rangos o *bins* adecuados de distancia entre puntos s_i, s_j (?).

Si el diseño de muestreo es una grilla regular la suavización puede ser lograda sin introducir sesgo, simplemente promediando todos los valores $\hat{\gamma}(\mathbf{h})$ para cada \mathbf{h} diferente. Sin embargo, si el diseño es irregular, para un rango de ancho m se define un variograma γ_k , para un entero positivo k , como el promedio de todos los $\hat{\gamma}(\mathbf{h})$ para el cual la respectiva distancia h satisfaga $(k-1)m < h \leq km$; entonces, γ_k es aproximadamente una estimación insesgada de $\gamma(h_k)$; por conveniencia se adopta $h_k = (k-0.5)m$, que es el punto medio del intervalo respectivo (?).

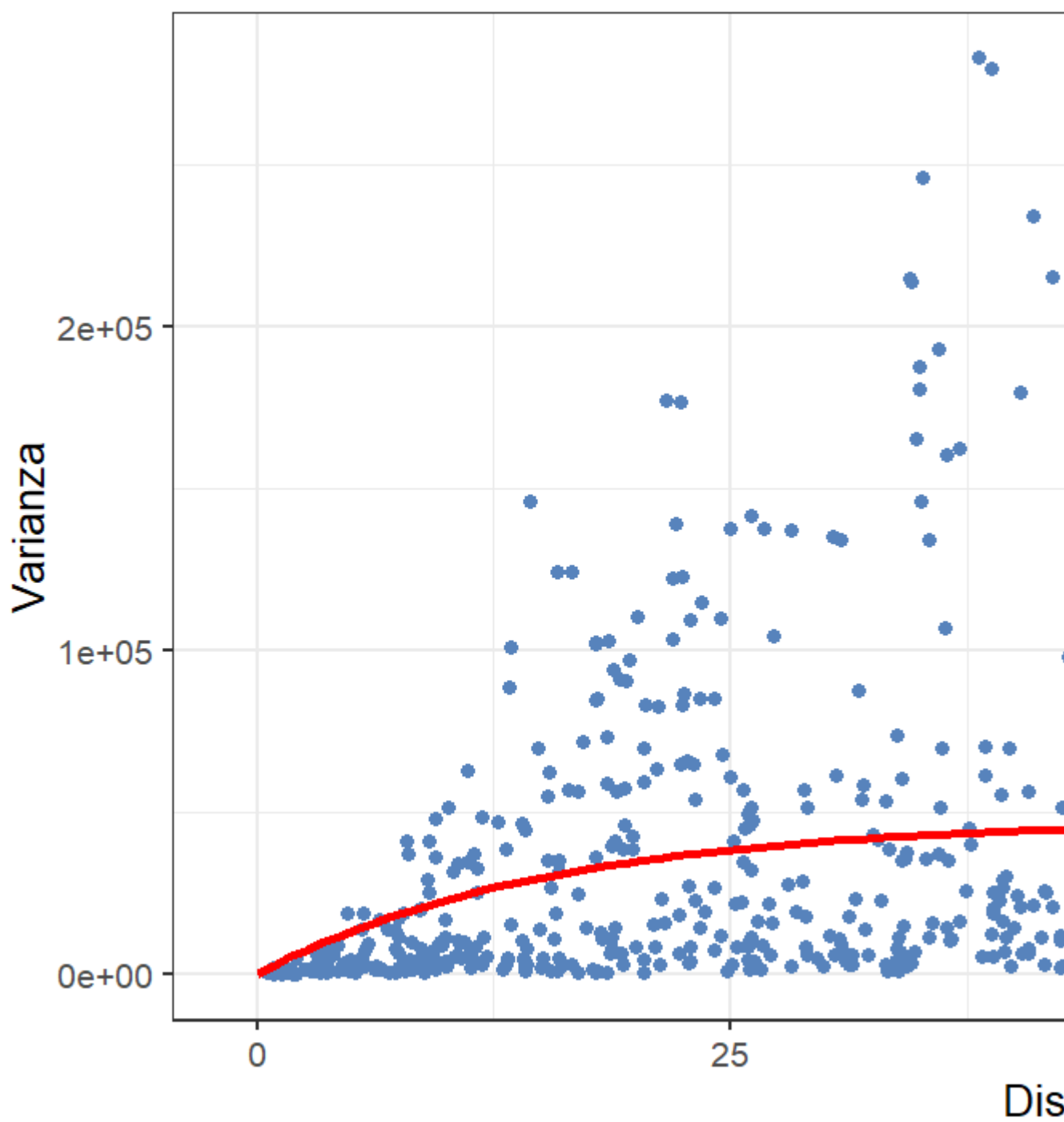


Figure 2.5: Nube variográfica

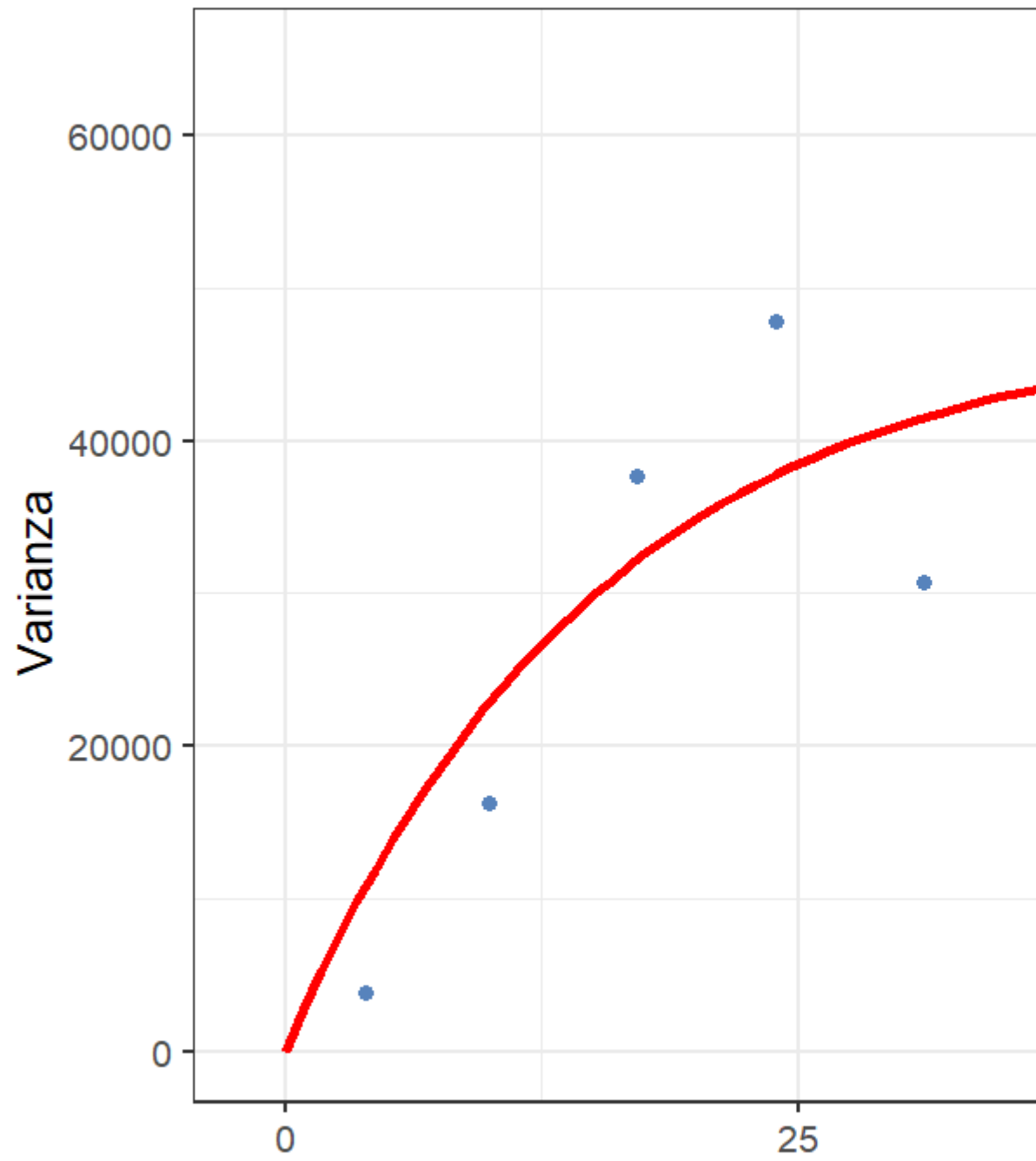


Figure 2.6: Bin variograma

2.4.3 Estimación del variograma (Ejemplos)*

Se sabe que no se puede utilizar de manera directa el variograma empírico, puesto que está definido para ciertas distancias y direcciones. Por otro lado, debido a los parámetros de tolerancia, está sujeto a aproximaciones, puesto que en el proceso de cálculo se utiliza un número limitado de datos. Para dar solución a esto, se ajusta un modelo teórico de variograma a partir del variograma empírico, siendo esta etapa la más significativa de todo análisis geoestadístico, puesto que nos da información de la continuidad espacial de la variable en estudio (?).

Una función de variograma teórico es ajustada a la sucesión de disimilitudes medias; nótese que este ajuste involucra una interpretación del comportamiento en el origen y el comportamiento para largas distancias, más allá del rango del variograma empírico. El ajuste puede llevarse a cabo de manera empírica mediante criterio experto, puesto que en la mayoría de casos prácticos no es fundamental el que tan bien la función de variograma se ajuste a la secuencia de puntos; lo más relevante es el tipo de continuidad que se asume para la variable regionalizada y la hipótesis de estacionariedad asociada al proceso estocástico espacial o campo aleatorio; estas suposiciones sirven de guía en la selección de una función de variograma teórico (?).

Comportamiento en el origen

Si las distancias son cercanas a cero; es decir, la variable regionalizada es regular en el espacio, más regular será el variograma en el origen. Generalmente, se suelen diferenciar tres tipos de comportamiento para el variograma en el origen (ver Figura 2.7) (?):

- Parabólico: Corresponde a una variable regionalizada bastante regular en el espacio.
- Lineal: Corresponde a una variable regionalizada continua, pero no tan regular.
- Discontinuo: Corresponde a una variable regionalizada más errática; es decir, con discontinuidades en la distribución espacial de sus valores. La diferencia entre dos datos bastante cercanos es grande, lo que provoca el efecto pepita.

Comportamiento para distancias muy grades

El variograma habitualmente crece a partir del origen y se estabiliza a partir de una distancia ϕ , alrededor de una silla σ^2 ; en este caso se conoce que esa silla es igual a la varianza a priori $C(0) = \sigma^2$.

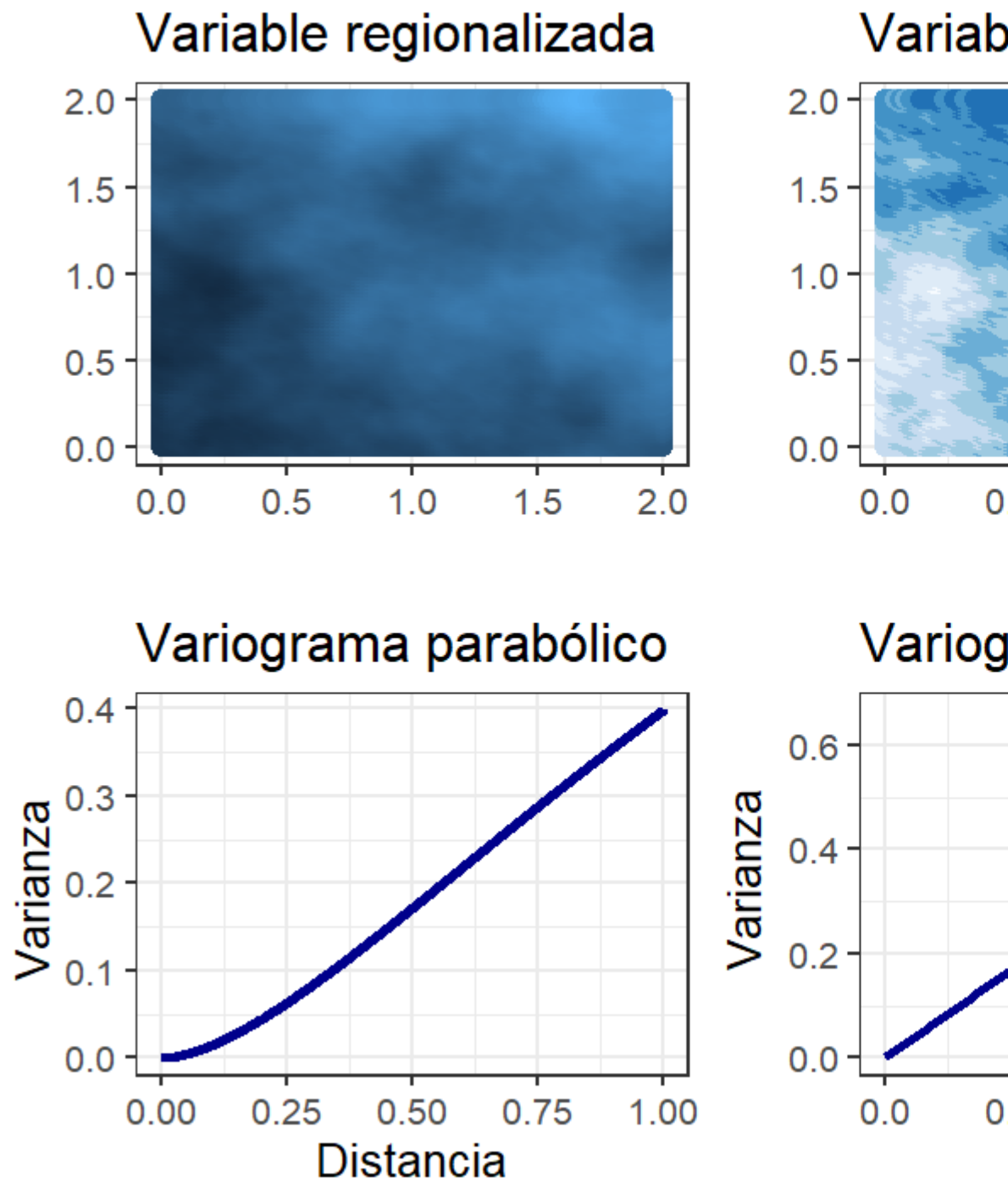


Figure 2.7: Comportamientos del variograma en el origen.

Como se vio anteriormente, las variables aleatorias $Z(s)$ y $Z(s + \mathbf{h})$ están correlacionadas, si la longitud del vector de separación \mathbf{h} es inferior a la distancia ϕ o alcance; más allá de $|\mathbf{h}| = \phi$, el variograma es constante e igual a su silla, y las variables $Z(s)$ y $Z(s + \mathbf{h})$ son independientes. Los variogramas con silla, se denominan *modelos de transición* (ver Figura 2.8) (?).

Existen casos en los cuales el variograma crece infinitamente una vez la distancia se incrementa y no posee ni silla ni rango; en este caso, la varianza es infinita $C(0) = \infty$ y no existe la función de covarianza ni el correlograma. Cuando se tiene ausencia de silla, es posible que sea una consecuencia del efecto de micro-escala (?).

2.4.4 Criterios para la estimación y ajuste del variograma

Tendencia:

Es fundamental verificar que los datos no presenten tendencia, puesto que cuando la media del proceso $\mu(s)$ no es constante; es decir, el proceso no es estacionario, el variograma empírico calculado a partir de las observaciones es erróneo; este comportamiento de tendencia se puede comprobar mediante un gráfico de dispersión de la variable de respuesta frente a cada coordenada espacial (ver Figura 2.9). En caso de que se evidencie tendencia se sugiere que se incluya un modelo de superficie de tendencia para la media que varía espacialmente, cuando una superficie de tendencia es incluida en el modelo, las dos coordenadas espaciales deben contribuir en este, puesto que la orientación de la región de estudio es arbitraria. En la práctica; al trabajar con datos geográficos de gran escala, se puede esperar que ciertas variables relacionadas a un ambiente físico muestren dependencia en la latitud (?).

Cuando la función de media no es constante, el variograma empírico atribuye de manera errónea la variación inducida por esta a la estructura de covarianza de gran escala en el proceso no observado $Z(s)$. Una solución es estimar $\mu(s)$ mediante un modelo de superficie de tendencia o, si la información de la covariable está disponible, mediante un modelo general de regresión, y utilizar los residuos $R_i = Z_i - \hat{\mu}(s_i)$ en lugar de las observaciones para el cálculo del variograma empírico (?). Este comportamiento se puede visualizar en la Figura 2.10.

Establecer la variable correcta:

En la estimación del variograma influye la distribución de los datos, pues estos pueden estar sesgados o tener valores extremos altos o bajos y por tanto su variograma calculado suele presentar un comportamiento errático. Para abordar este problema es recomendable transformar los datos a un espacio Normal o Gaussiano (?). Ver Figura 2.11.

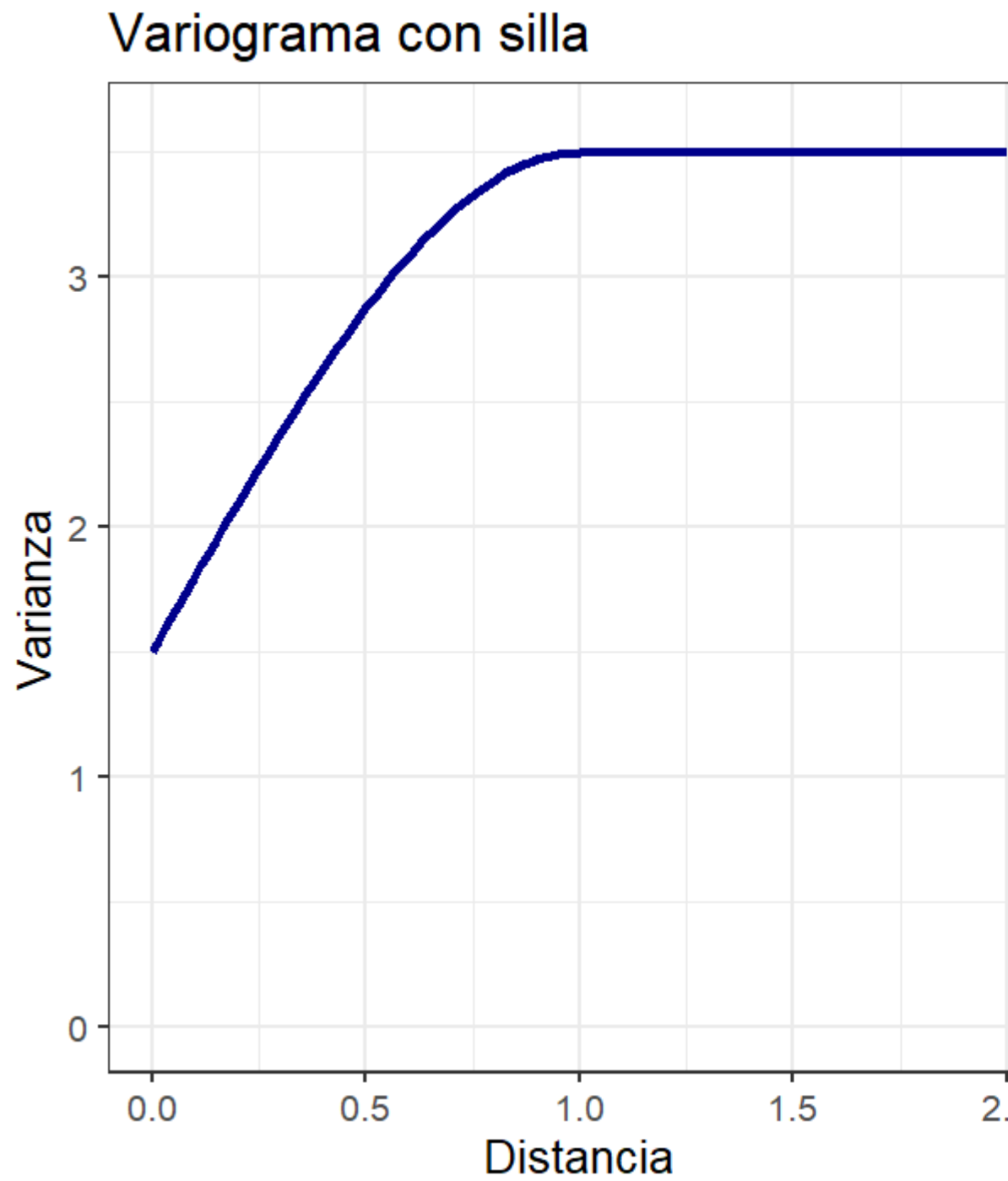


Figure 2.8: Lado izquierdo: Variograma con silla y alcance. Lado derecho: Variograma sin silla y alcance.

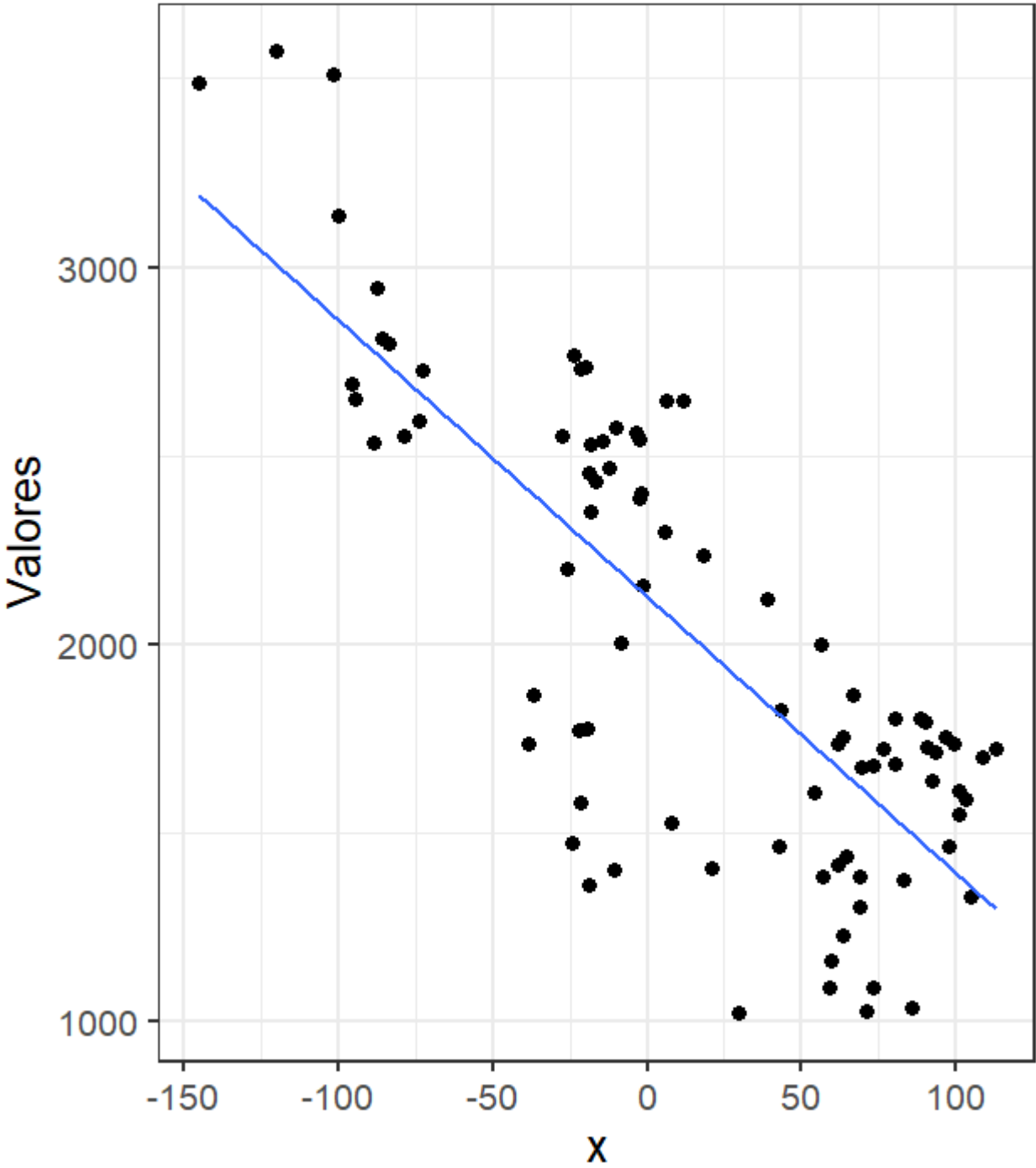


Figure 2.9: Tendencia

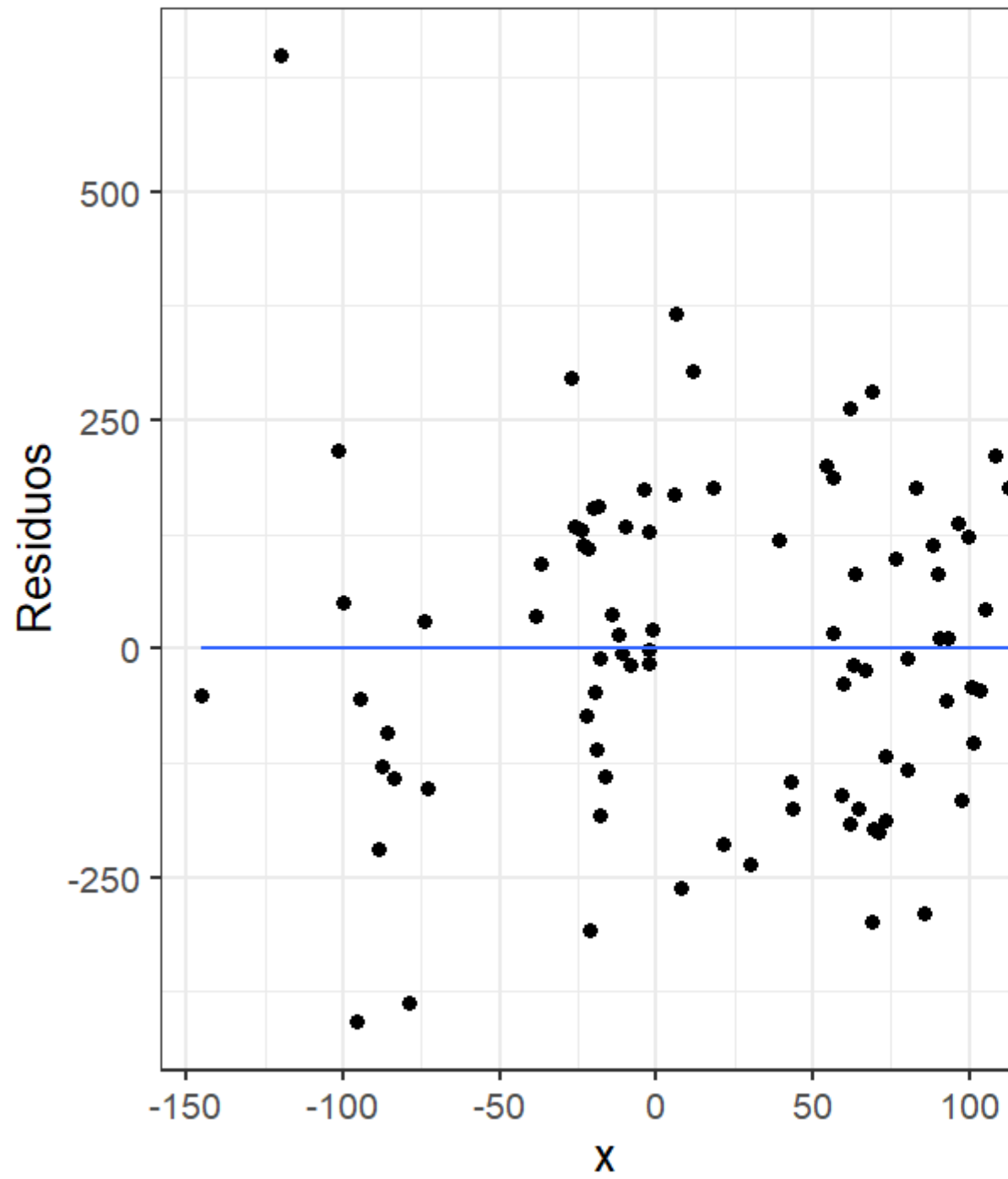


Figure 2.10: Tendencia

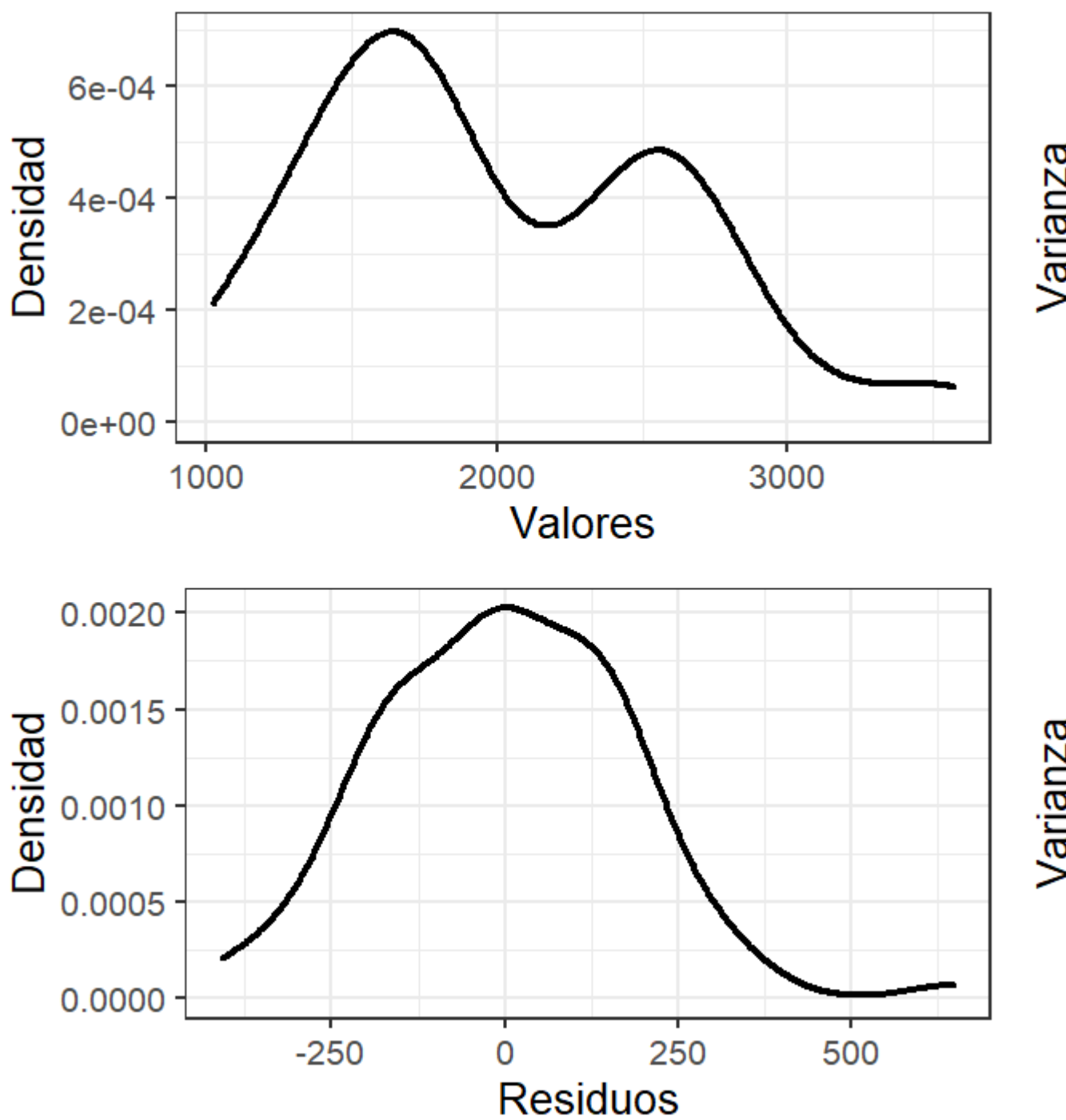


Figure 2.11: Arriba: Variograma empírico calculado a partir de los datos originales. Abajo: Variograma empírico calculado a partir de los residuos.

Verificación de isotropía:

Como bien se conoce, se debe verificar la estacionariedad de los datos, puesto que este influye de manera directa en la estimación del variograma empírico; teniendo en cuenta que un proceso espacial puede estar definido en varias direcciones, este criterio se debe verificar para todas estas. No obstante, en la práctica la isotropía se estudia por medio del cálculo de funciones de autocovarianza o de semivarianza muestrales en algunas direcciones que habitualmente son: $0^\circ, 45^\circ, 90^\circ$ y 135° , ver Figura. 2.12.

Con la verificación de los puntos anteriores, se procede al cálculo del variograma empírico y su posterior ajuste de un modelo teórico por medio de los métodos que se exponen a continuación.

2.4.5 Métodos de estimación teórica más utilizados

Los parámetros de la función de covarianza pueden ser estimados por medio de los métodos de Máxima Verosimilitud (ML) y Máxima Verosimilitud restringida (REML), los cuales requieren la especificación de la distribución del vector $Z = (Z(s_1), \dots, Z(s_n))$; generalmente, se asume normalidad multivariada. Para el caso ML, se tiene que:

$$Z \sim N_n(X\beta, \Sigma(\theta)) \quad (2.55)$$

donde $\Sigma(\theta) = Cov(Z)$ es una matriz de dimensión $n \times n$ y X es una matriz de dimensión $n \times q$ con $q < n$, de variables explicativas, dentro de las cuales comúnmente se encuentran las coordenadas geográficas; de donde el negativo de la función de logverosimil es:

$$L(\beta, \theta) = \frac{n}{2} \log(2\pi) + \frac{1}{2} \log |\Sigma(\theta)| + \frac{1}{2} (Z - X\beta)' \Sigma^{-1}(\theta) (Z - X\beta) \quad (2.56)$$

El elemento i, j de la matriz $\Sigma(\theta)$ corresponde a la covarianza espacial entre las variables $Z(s_i)$ y $Z(s_j)$; esto es:

$$Cov(Z(s_i), Z(s_j); \theta) = C(s_i - s_j; \theta) = C(\mathbf{h}; \theta) \quad (2.57)$$

El estimador $\hat{\theta}$ es sesgado pero asintóticamente eficiente; no obstante, al trabajar con una muestra grande se debe tener en consideración que debido a su conducta iterativa, se realizarán grandes cantidades de operaciones computacionales debido al cálculo del determinante y de la inversa de la matriz de covarianza (?).

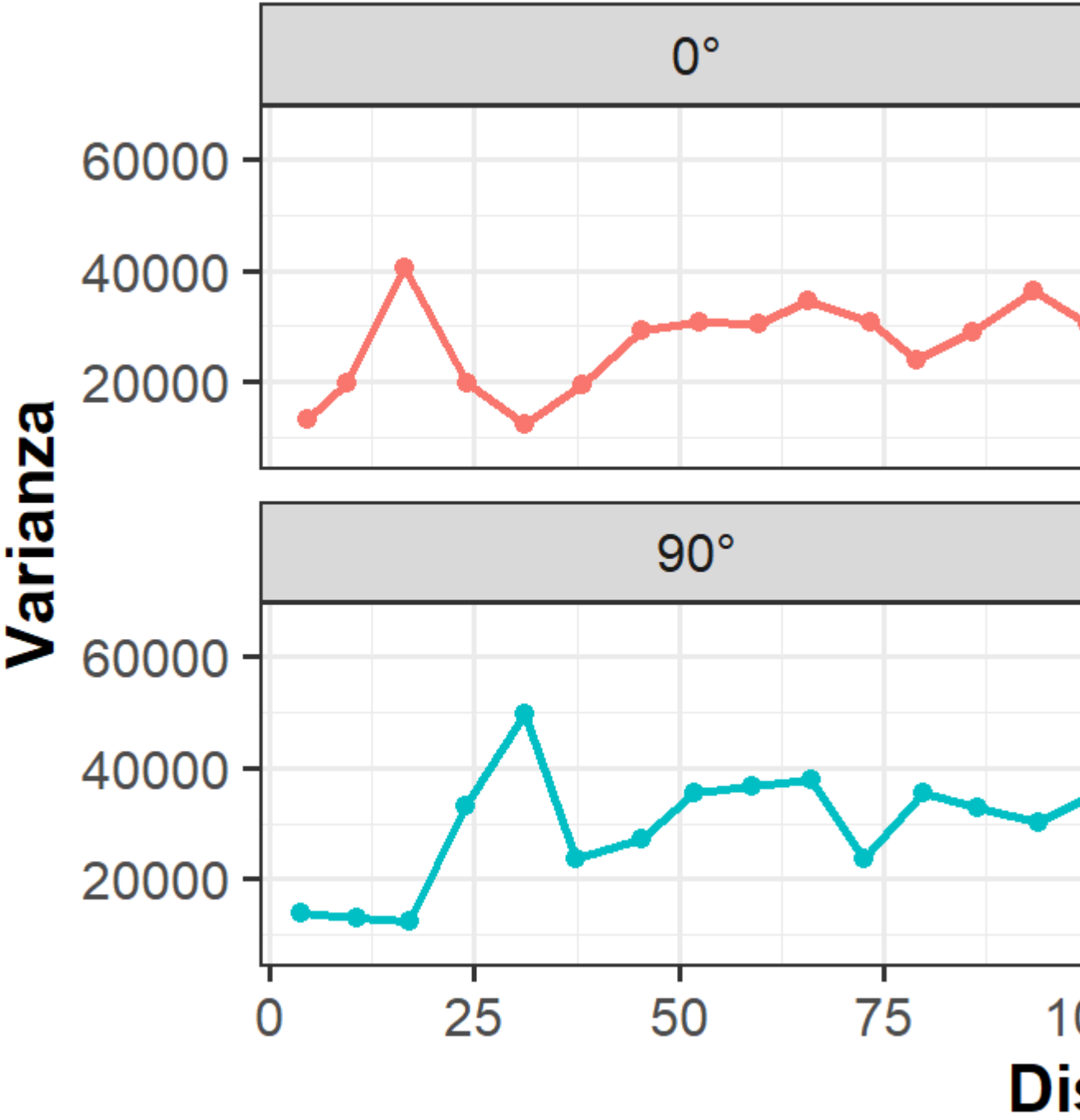


Figure 2.12: Comportamiento

Estimador REML

El estimador REML es una modificación del estimador ML propuesto para disminuir el sesgo, el cual reemplaza la maximización de la verosimilitud del vector Z , por la del vector $A'Z$ tal que: $E(A'Z) = 0$, donde A es una matriz de dimensión $n \times (n - p)$, de rango columna completo.

Con esta modificación y dado que

$$\text{Var}(A'Z) = A'\Sigma(\theta)A \quad (2.58)$$

el negativo de la función de logverosímil queda:

$$L(\theta) = \frac{n}{2} \log(2\pi) + \frac{1}{2} \log |A'\Sigma(\theta)A| + \frac{1}{2} Z'(A'\Sigma(\theta)A)^{-1} A'Z \quad (2.59)$$

Esta función es dependiente únicamente de θ ; de esta manera este método no usa la modelización de la superficie de tendencia, sino que se basa directamente en un vector de incrementos de media 0. No obstante, pese a minimizar el sesgo en las estimaciones de θ , aun interviene un costo computacional elevado (?).

Verosimilitud Compuesta (CL)

El método CL que se utiliza para estimar θ , involucra la suma de componentes individuales de funciones logverosimilitud, correspondiente a las distribuciones marginales de las variables de interés. Por consiguiente, la distribución multivariada de Z no es necesaria, puesto que se basa en las distribuciones marginales $f(Z(s_i); \theta)$ y se asume la existencia del gradiente y de la matriz Hessiana de f .

Ahora bien, se suponen conocidas $f(Z(s_i); \theta)$, excepto el parámetro θ ; entonces:

$$\log(Z(s_i); \theta) = \ln(f(Z(s_i); \theta)) \quad (2.60)$$

es una función logverosímil y la función de verosimilitud compuesta está definida por:

$$CL(\theta) = \sum_{i=1}^n \log(Z(s_i); \theta) \quad (2.61)$$

Al gradiente de CL, $\nabla CL = CS(\theta)$, se le conoce como función de score compuesta.

Así, los valores estimados de $\hat{\theta}$ se determinan resolviendo el siguiente sistema de ecuaciones:

$$CS(\theta) = \sum_{i=1}^n \nabla \log(Z(s_i); \theta) = 0 \quad (2.62)$$

Una de las razones principales para usar las funciones de verosimilitudes marginales es que aunque al inicio no se cumpla el supuesto requerido, es posible aproximarse a este por medio de alguna transformación (?).

Mínimos cuadrados ponderados

Este método utiliza la matriz de ponderación $W(\theta)$ y suele ser expresado en términos del variograma o de la covarianza, gracias a su equivalencia en procesos estacionarios de segundo orden. Para el caso espacio-temporal, se estima θ para un variograma minimizando la siguiente expresión:

$$(2\hat{\gamma} - 2\gamma(\theta))' W^{-1}(\theta) (2\hat{\gamma} - 2\gamma(\theta)) \quad (2.63)$$

donde la matriz de ponderación $W(\theta)$ está dada por:

$$W(\theta) = \text{Diag}(\text{Var}(2\hat{\gamma}(\mathbf{h}_k))) \approx \text{Diag} \left(\frac{2(2\gamma(\mathbf{h}_k|\theta))^2}{N(\mathbf{h}_k)} \right) \quad (2.64)$$

con:

$$N(\mathbf{h}_k) = \{(i, j) : s_i - s_j = \mathbf{h}_k\} \quad (2.65)$$

Para las ubicaciones $i, j = 1, \dots, n$, que producen los primeros k rezagos espaciales $k = 1, \dots, K$; generalmente se utilizan los rezagos espaciales hasta la mitad de la distancia máxima entre cualquier par de ubicaciones, debido a que para ubicaciones bastante separadas disminuye notoriamente la cantidad de puntos incluidos en la estimación del variograma. La aproximación de $\text{Var}(2\hat{\gamma}(\mathbf{h}_k|\theta))$, se obtiene bajo el supuesto de que $Z(s) \sim N(\mu; \sigma^2)$, $\forall s \in \mathbb{R}^d$, y que por consiguiente:

$$(Z(s + \mathbf{h}) - Z(s))^2 \approx 2\gamma(\mathbf{h})\chi_1^2 \quad (2.66)$$

El inconveniente que presenta este método es la necesidad de definir clases de rezagos para realizar una estimación empírica de la covarianza o del variograma, pues si no se poseen muchos datos la cantidad de estos en cada una de las clases disminuye, de tal forma que para los primeros rezagos se pueden tener suficientes datos para la estimación de cada $\hat{\gamma}(\mathbf{h}_k)$; sin embargo, para los últimos rezagos podrían no existir suficientes datos para la estimación (?).

Modelo lineal de regionalización

En la práctica, puede suceder que el variograma empírico no presente una apariencia o forma simple de ser modelizado por un modelo teórico; no obstante, esto no es un inconveniente, puesto que existe la posibilidad de combinar modelos teóricos de variograma, obteniendo nuevos modelos más complejos.

Teniendo presente que un fenómeno regionalizado podría ser considerado como la suma de diversos subfenómenos independientes, el modelo lineal de regionalización construye un campo aleatorio $Z(s)$ como una combinación lineal de L campos aleatorios independientes, cada uno con media 0 y función de covarianza $C_l(h)$, mutuamente independientes (?).

Así, sea Y_l un subcampo aleatorio de Z para todo $l = 1, \dots, L$, se tiene que:

$$Z(s) = \sum_{l=1}^L a_l Y_l(s) + \mu \quad (2.67)$$

con:

$$E(Z(s)) = \mu \quad (2.68)$$

$$E(Y_l(s)) = 0 \quad (2.69)$$

$$Cov(Y_l(s), Y_{l'}(s + \mathbf{h})) = \begin{cases} C_l(h) & \text{si } l = l' \\ 0 & \text{en otro caso.} \end{cases} \quad (2.70)$$

Esto implica que la función de covarianza de $Z(s)$ es una combinación lineal de las L funciones $C_l(h)$ y bajo el supuesto de independencia entre los Y , se tiene que:

$$C(h) = Cov(Z(s), Z(s + \mathbf{h})) \quad (2.71)$$

$$= \sum_{l=0}^L \sum_{l'=0}^L Cov((Y_l(s), Y_{l'}(s + \mathbf{h}))) \quad (2.72)$$

$$= \sum_{l=0}^L a_l a_l C_l(h) \quad (2.73)$$

$$= \sum_{l=0}^L a_l^2 C_l(h) \quad (2.74)$$

por lo tanto, el modelo de covarianza $C(h)$ es:

$$C(h) = \sum_{l=0}^L b_l C_l(h), \text{ con } b_l = a_l^2 \geq 0 \quad (2.75)$$

De esta forma, como $b_l > 0$ es siempre positivo, y es la silla del modelo básico de covarianza $C_l(h)$. Las condiciones suficientes para que $C(h)$ sea un modelo válido de covarianza son:

- Las funciones $C_l(h)$ son modelos de covarianza válidos.
- $b_l > 0, \forall l = 1, \dots, L$

En términos de variogramas, se tiene que:

$$E[(Y_l(s) - Y_l(s + \mathbf{h}))(Y_{l'}(s) - Y_{l'}(s + \mathbf{h}))] = \begin{cases} \gamma'_l(h) & \text{si } l = l' \\ 0 & \text{en otro caso.} \end{cases} \quad (2.76)$$

Entonces,

$$\gamma(h) = \frac{1}{2} E[Z(s + \mathbf{h}) - Z(s)]^2 = \sum_{l=0}^L b_l \gamma'_l(h) \quad (2.77)$$

con $b_l = (a_l)^2 \geq 0$, donde b_l es la contribución en varianza del correspondiente variograma $\gamma'_l(h)$ (?).

GRAFICOSY EJEMPLOS

2.4.6 Modelos teóricos de variograma

En esta sección se presentan las funciones de variograma más comunes, que se definen para el caso isotrópico de funciones aleatorias. Para la representación gráfica de la función de variograma se hace uso de la relación $\gamma(h) = C(0) - C(h)$. Estas funciones se pueden clasificar de la siguiente manera:

- Variogramas con silla o variogramas de transición.
- Variogramas con silla y efecto hueco.
- Variogramas sin silla.

Variogramas con silla

Modelo esférico

Este modelo es válido solo en $\mathbb{R}^1, \mathbb{R}^2, \mathbb{R}^3$, y está definido por:

$$\gamma(\|\mathbf{h}\|; \theta) = \begin{cases} \sigma^2 \left(1.5 \frac{\|\mathbf{h}\|}{\phi} - 0.5 \left(\frac{\|\mathbf{h}\|}{\phi} \right)^3 \right) & \text{si } \|\mathbf{h}\| \leq \phi \\ \sigma^2 & \text{si } \|\mathbf{h}\| > \phi \end{cases} \quad (2.78)$$

con $\theta = (\sigma^2, \phi)$, donde $\sigma^2 = C(0)$ es el valor del variograma donde alcanza la silla, y ϕ es el rango.

Este variograma presenta un comportamiento lineal cerca del origen, lo cual sugiere continuidad pero con cierto grado de irregularidad en el proceso estocástico espacial o campo aleatorio. Sin embargo, referente a su comportamiento en distancias largas, alcanza su silla en $\|\mathbf{h}\| = \phi$. El uso habitual de este modelo en la práctica es gracias a su rango bien definido, su representación polinomial simple y su validez en $\mathbb{R}^1, \mathbb{R}^2, \mathbb{R}^3$. La principal razón para el uso de este es que se comporta de forma casi lineal hasta que alcanza su silla, y presenta estabilidad en una gran variedad de observaciones.

Modelo de efecto pepita puro

Este variograma refleja la ausencia de dependencia espacial en el proceso estocástico o campo aleatorio; este modelo puede ser visto como un caso particular del modelo esférico si $\phi \rightarrow 0$:

$$\gamma(\|\mathbf{h}\|; \theta) = \begin{cases} \sigma^2 & \text{si } \|\mathbf{h}\| = 0 \\ 0 & \text{si } \|\mathbf{h}\| > 0 \end{cases} \quad (2.79)$$

con $\theta = \sigma^2$.

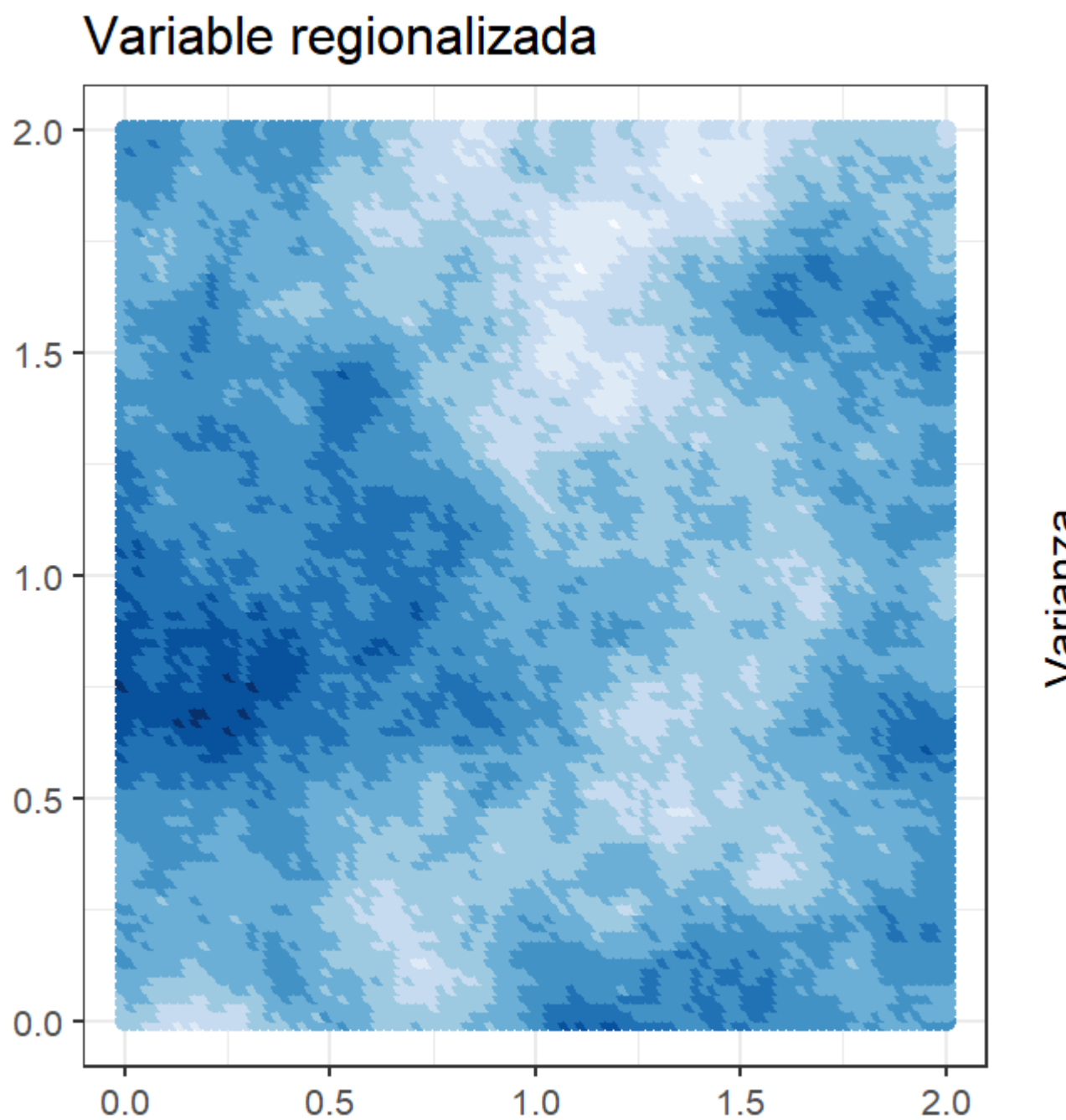


Figure 2.13: Izquierda: Variable regionalizada. Derecha: Variograma esférico.

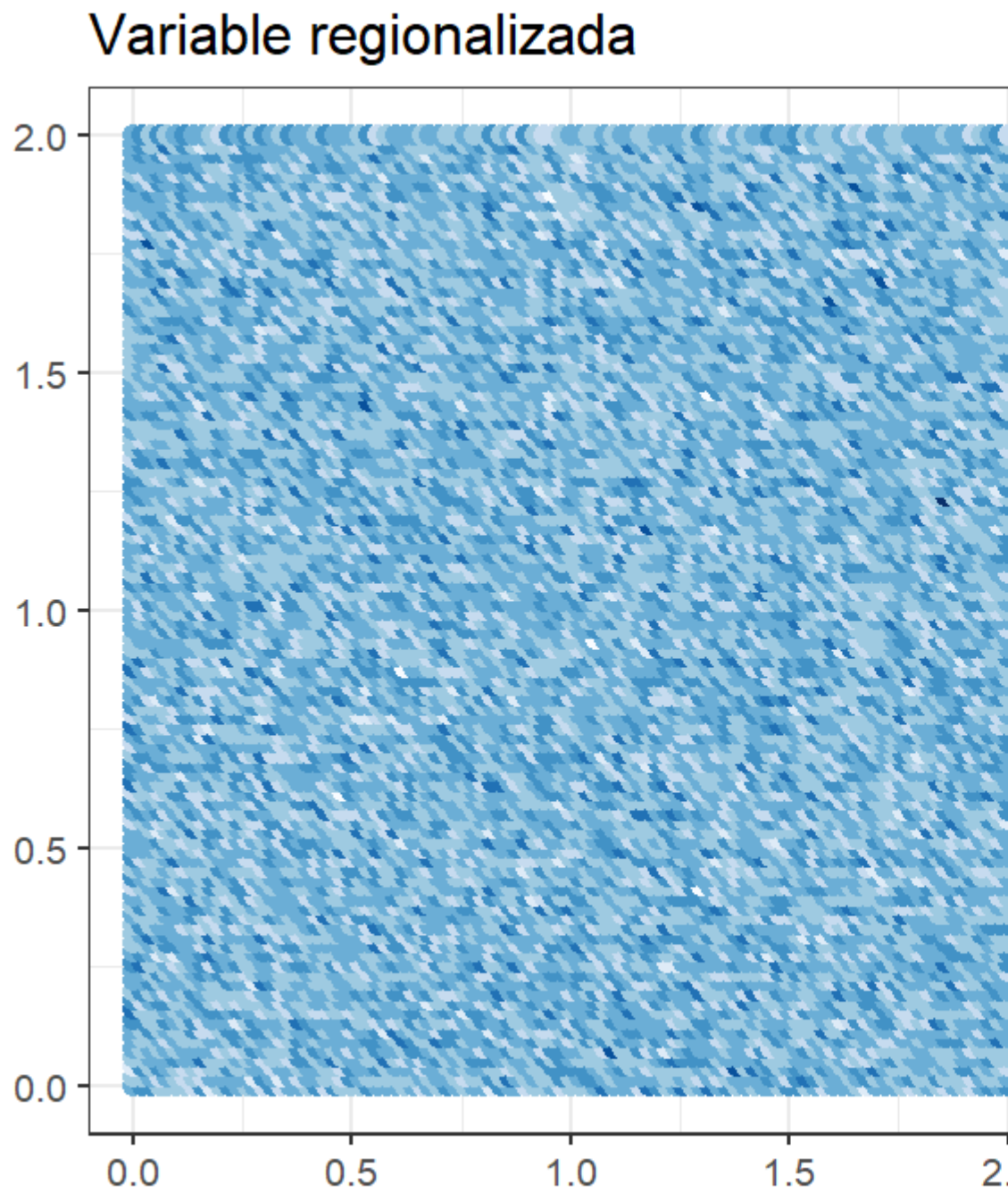


Figure 2.14: Izquierda: Variable regionalizada. Derecha: Variograma pepita puro.

Se debe tener en cuenta que el modelo esférico corresponde a una función aleatoria continua, mientras que el modelo de efecto pepita puro, a una discontinuidad.

Modelo exponencial

Este modelo es válido en \mathbb{R}^d , $d \geq 1$, y está definido por:

$$\gamma(\|\mathbf{h}\|; \theta) = \sigma^2 \left(1 - \exp \left(-\frac{\|\mathbf{h}\|}{\phi} \right) \right), \quad \theta = (\sigma^2, \phi) \quad (2.80)$$

El modelo exponencial refleja un comportamiento lineal cerca del origen, lo cual indica continuidad pero con un cierto grado de irregularidad en la función aleatoria. Por otro lado, en cuanto a su comportamiento en largas distancias, este alcanza su silla solo asintóticamente cuando $\|\mathbf{h}\| \rightarrow \infty$. El valor del rango es igual a la distancia para el cual el variograma toma un valor igual al 95% de la silla.

Modelo Gaussiano

Este modelo es válido en \mathbb{R}^d , $d \geq 1$, y está definido por:

$$\gamma(\|\mathbf{h}\|; \theta) = \sigma^2 \left(1 - \exp \left(-\frac{\|\mathbf{h}\|^2}{\phi^2} \right) \right), \quad \theta = (\sigma^2, \phi) \quad (2.81)$$

La principal característica de este modelo es su forma parabólica cerca al origen; la dependencia espacial se desvanece solo en una distancia que tiende a infinito; por consiguiente, este modelo es considerado poco realista en la práctica.

Modelo Cúbico

Este modelo es válido en $\mathbb{R}^1, \mathbb{R}^2, \mathbb{R}^3$; generalmente es similar al modelo Gaussiano, puesto que también presenta un comportamiento parabólico cerca del origen. Sin embargo, este modelo alcanza una silla plana a una distancia ϕ , y está definido por:

$$\gamma(\|\mathbf{h}\|; \theta) = \begin{cases} \sigma^2 \left(7 \frac{\|\mathbf{h}\|^2}{\phi^2} - \frac{35\|\mathbf{h}\|^3}{4\phi^3} + \frac{7\|\mathbf{h}\|^5}{2\phi^5} - \frac{3\|\mathbf{h}\|^7}{4\phi^7} \right) & \text{si } 0 \leq \|\mathbf{h}\| \leq \phi \\ \sigma^2 & \text{si } \|\mathbf{h}\| > \phi \end{cases} \quad (2.82)$$

con $\theta = (\sigma^2, \phi)$.

Modelo Estable

Este modelo es válido en \mathbb{R}^d , $d \geq 1$ y está definido por:

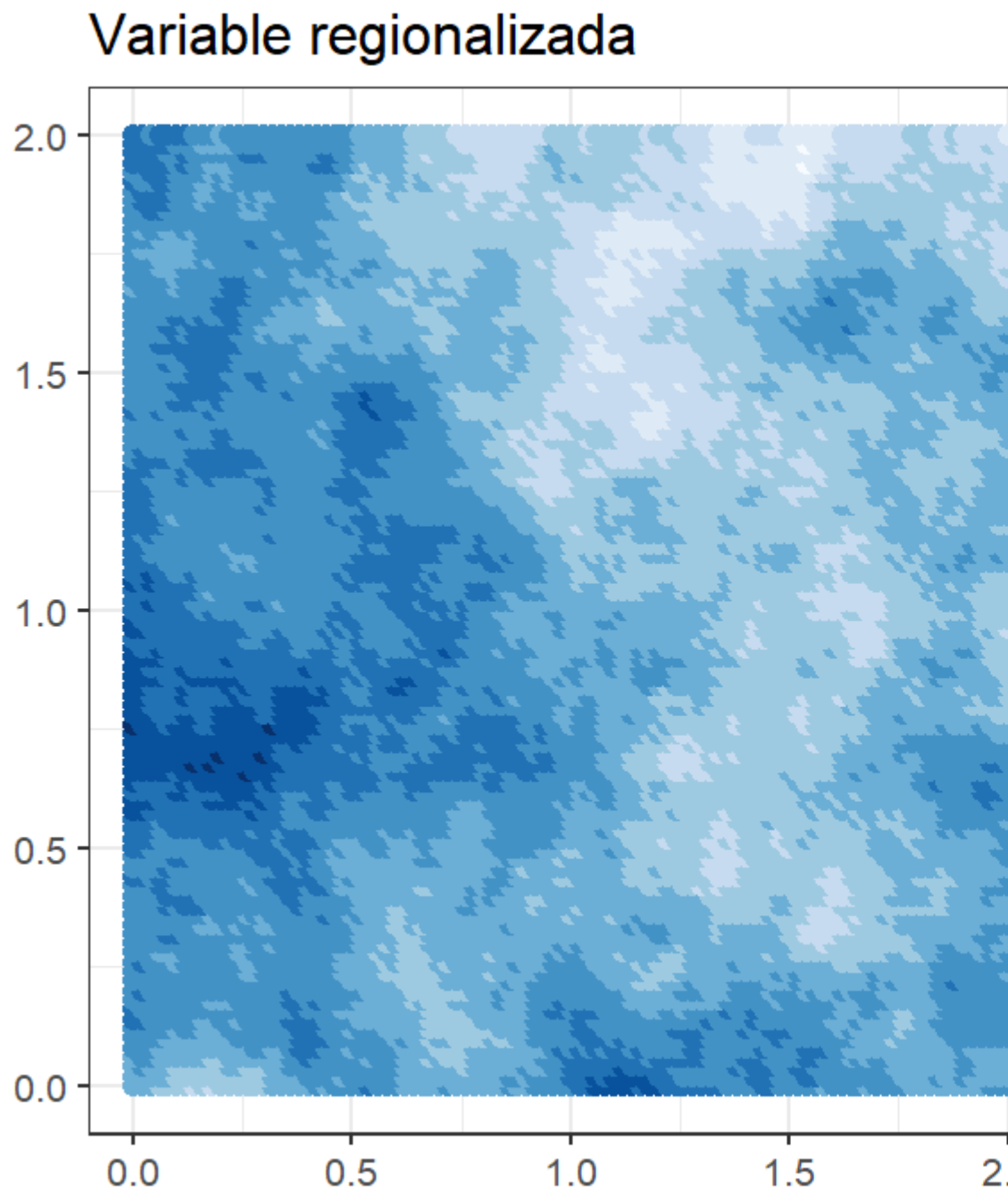


Figure 2.15: Izquierda: Variable regionalizada. Derecha: Variograma exponencial.

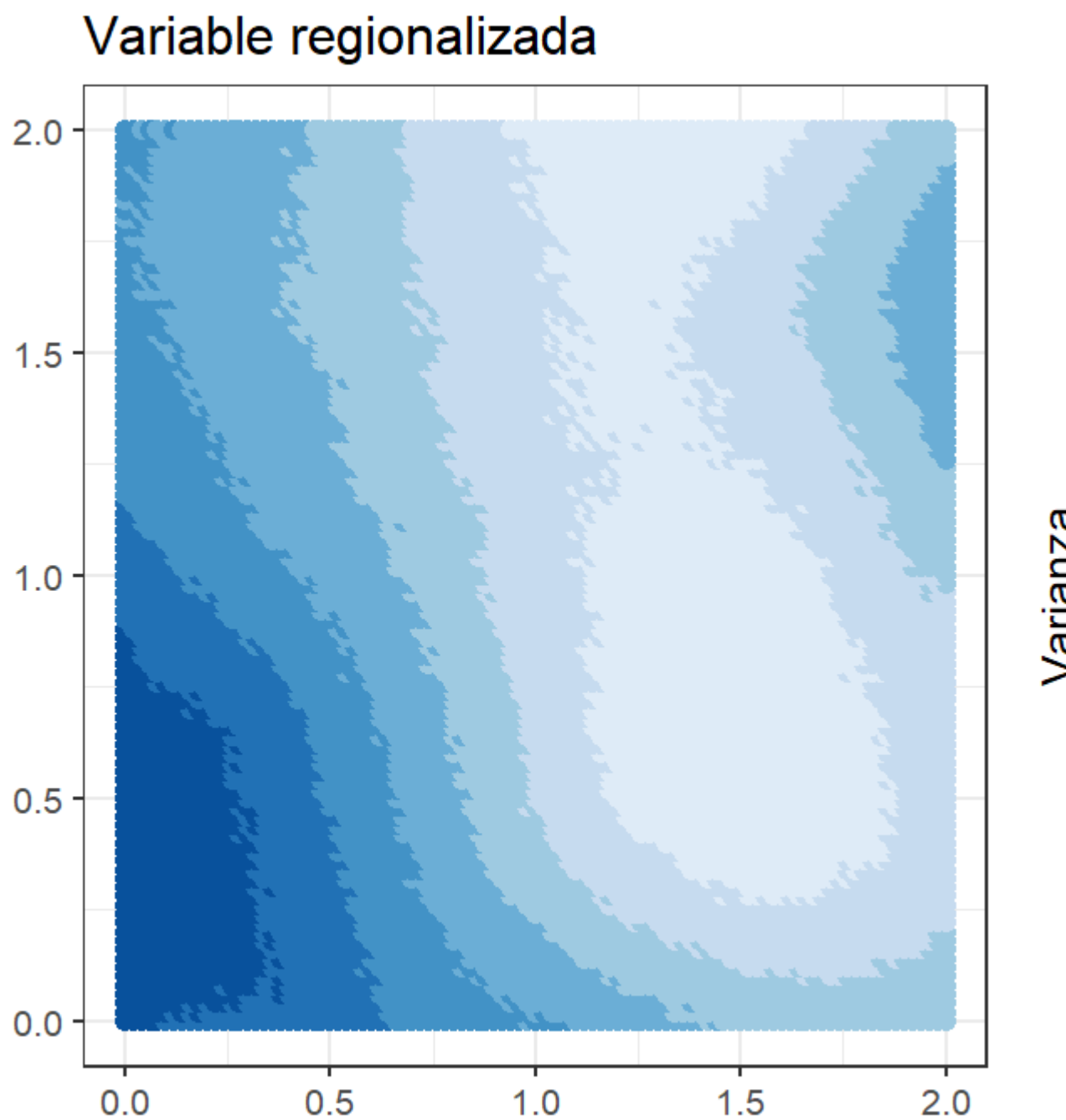


Figure 2.16: Izquierda: Variable regionalizada. Derecha: Variograma Gaussiano.

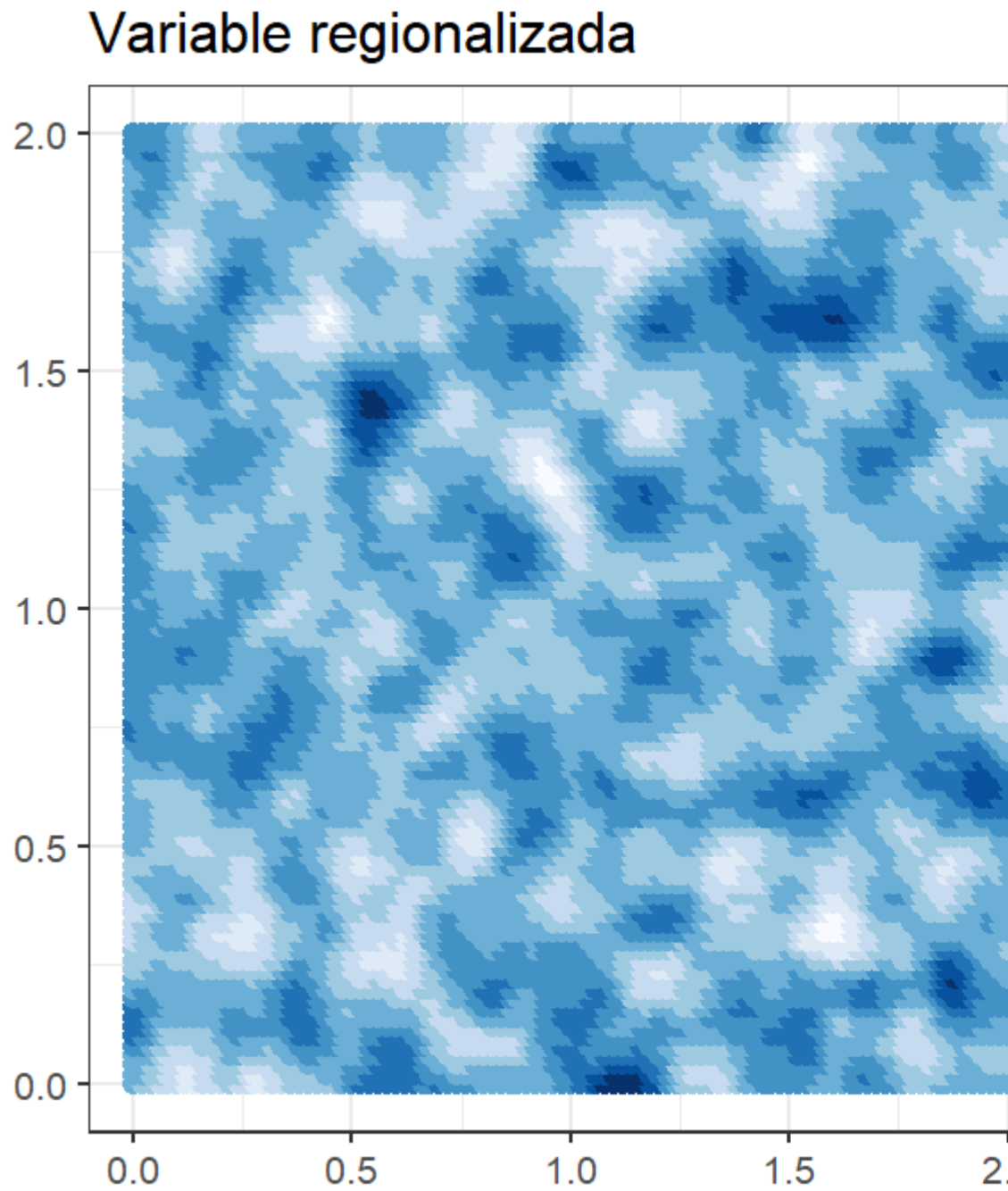


Figure 2.17: Izquierda: Variable regionalizada. Derecha: Variograma cúbico.

$$\gamma(\|\mathbf{h}\|; \theta) = \sigma^2 \left(1 - \exp \left(- \left(\frac{\|\mathbf{h}\|}{\phi} \right)^\alpha \right) \right), \quad 0 < \alpha \leq 2, \quad \theta = (\sigma^2, \phi) \quad (2.83)$$

Notar que para $\alpha = 1$, se obtiene el modelo exponencial y con $\alpha = 2$, se convierte en un modelo Gaussiano.

Modelo de Cauchy Generalizado

Este modelo es válido en \mathbb{R}^d , $d \geq 1$, y está definido por:

$$\gamma(\|\mathbf{h}\|; \theta) = \sigma^2 \left(1 - \frac{1}{\left(1 + \left(\frac{\|\mathbf{h}\|}{\phi} \right)^2 \right)^\alpha} \right), \quad \theta = (\sigma^2, \phi) \quad (2.84)$$

Si $\alpha = 1$ es conocido como el modelo de Cauchy.

Este modelo muestra un comportamiento parabólico cerca del origen y si $\alpha < 2$ alcanza la silla lentamente.

Modelo de K-Bessel o Matérn

Este modelo es válido en \mathbb{R}^d , $d \geq 1$, y esta definido por:

$$\gamma(\|\mathbf{h}\|; \theta) = \sigma^2 \left(1 - \frac{1}{2^{\alpha-1} \Gamma(\alpha)} \left(\frac{\|\mathbf{h}\|}{\phi} \right)^\alpha K_\alpha \left(\frac{\|\mathbf{h}\|}{\phi} \right) \right), \quad \alpha > 0, \quad \theta = (\sigma^2, \phi) \quad (2.85)$$

donde K_α es la función de Bessel modificada de segundo tipo de orden α , y que está definida por:

$$K_\alpha(\nu) = \frac{\pi}{2 \sin(\alpha\pi)} \left(\sum_{k=0}^{\infty} \frac{1}{k! \Gamma(-\alpha + k + 1)} \left(\frac{\nu}{2} \right)^{2k-\alpha} - \sum_{k=0}^{\infty} \frac{1}{k! \Gamma(\alpha + k + 1)} \left(\frac{\nu}{2} \right)^{2k+\alpha} \right) \quad (2.86)$$

El modelo K-Bessel puede tener cualquier tipo de comportamiento cerca del origen. Para $\alpha = 1/2$, se obtiene el modelo exponencial.

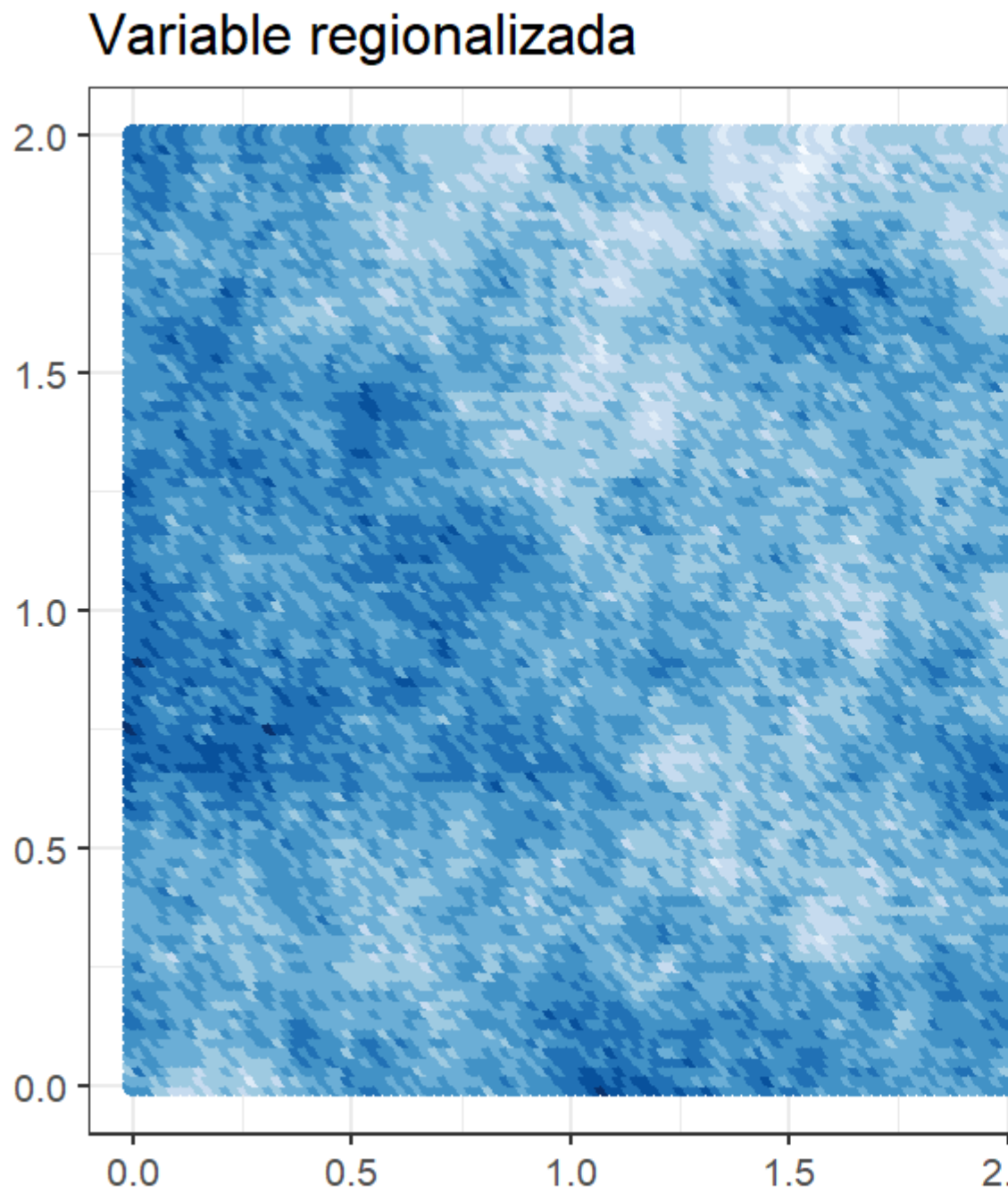


Figure 2.18: Izquierda: Variable regionalizada. Derecha: Variograma estable.

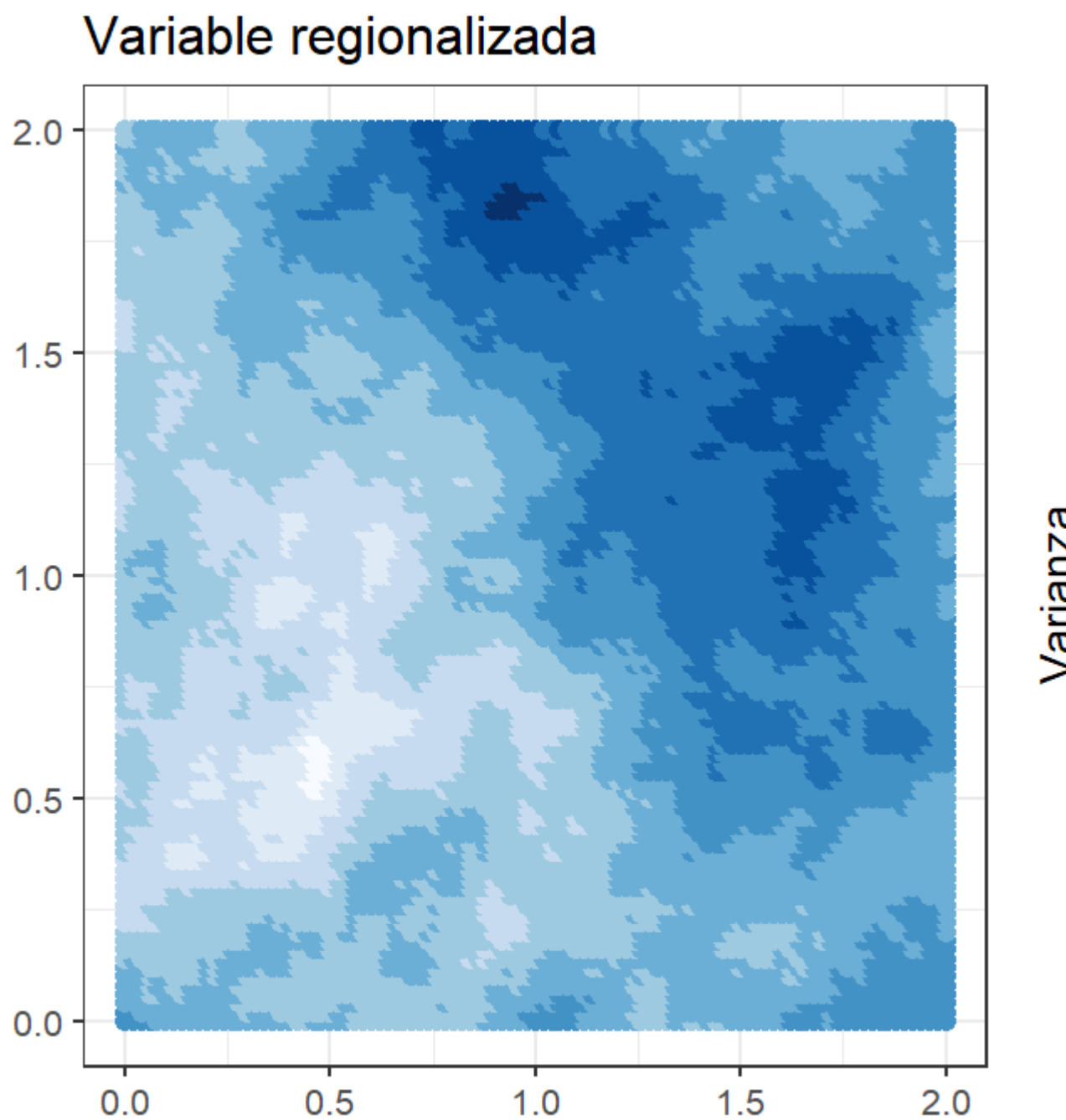


Figure 2.19: Izquierda: Variable regionalizada. Derecha: Variograma Cauchy generalizado.

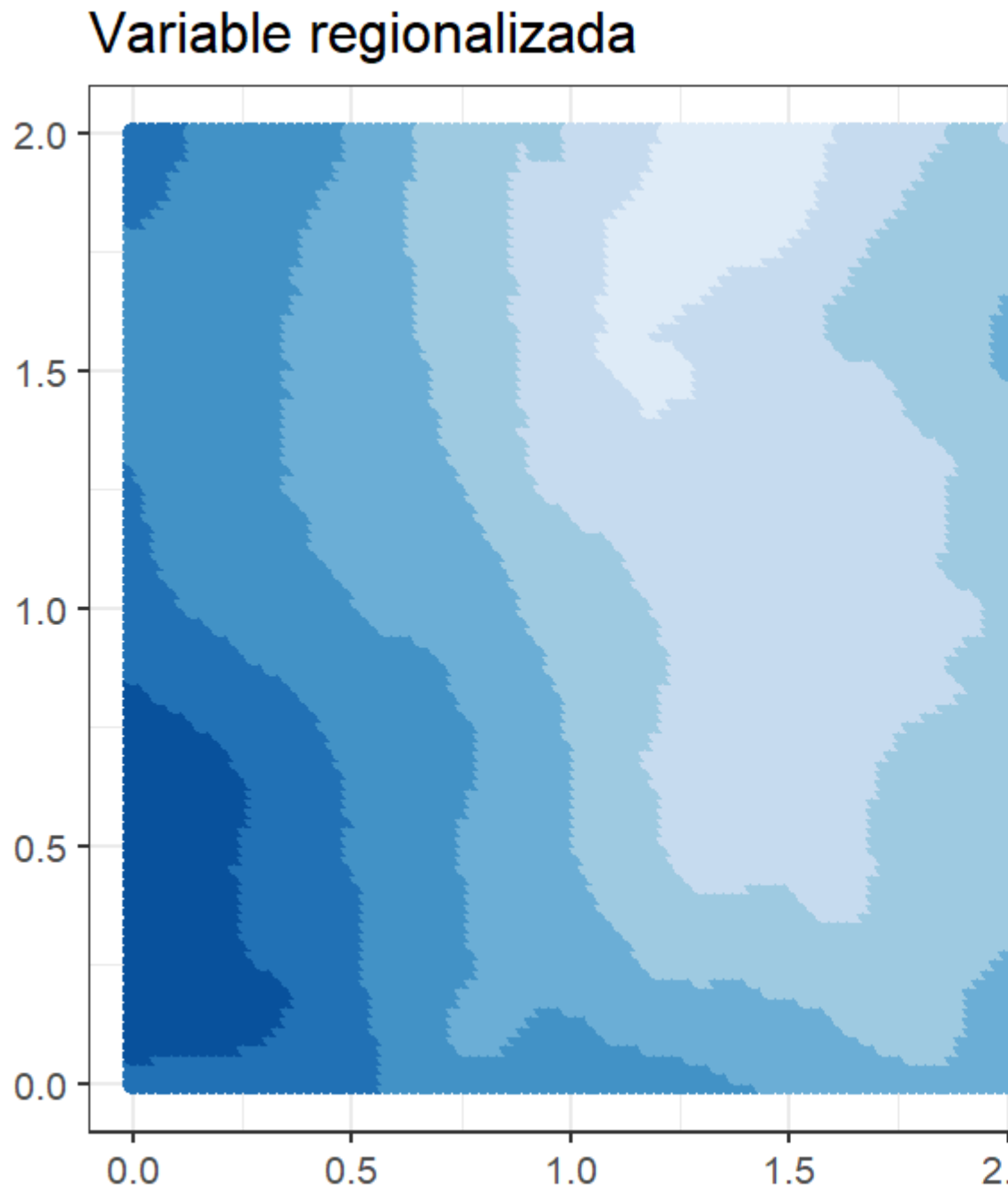


Figure 2.20: Izquierda: Variable regionalizada. Derecha: Variograma Matérn.

Variogramas con efecto hueco

En la práctica la dependencia espacial puede no crecer monótonamente, incluso esta podría ser negativa o presentar alternaciones entre dependencia espacial positiva y negativa. Estos modelos se llaman *modelos de efecto hueco* y se utilizan para definir oscilaciones que involucren un significado físico. Presentan un comportamiento lineal o parabólico cerca del origen, pueden tener o no tener silla, y pueden ser o no periódicos o pseudoperiódicos.

Modelo J-Bessel

Este modelo está definido por:

$$\gamma(\|\mathbf{h}\|; \theta) = \sigma^2 \left(1 - \left(\frac{2\phi}{\|\mathbf{h}\|} \right)^\alpha \Gamma(\alpha + 1) J_\alpha \left(\frac{\|\mathbf{h}\|}{\phi} \right) \right), \quad \theta = (\sigma^2, \phi) \quad (2.87)$$

donde α es el parámetro de forma, ϕ es el parámetro de escala, Γ es la función de Euler, y J_α es la función *J-Bessel* del primer tipo de orden α dada por:

$$J_\alpha(\nu) = \left(\frac{\nu}{2} \right)^2 \sum_{k=0}^{\infty} \frac{-1^k}{k! \Gamma(\alpha + k + 1)} \left(\frac{\nu}{2} \right)^{2k} \quad (2.88)$$

y es válido para $\mathbb{R}^d, d \leq 2(\alpha + 1).$

Modelo seno cardinal

Este modelo es una particularización del modelo J-Bessel para $\alpha = 1/2$, y es uno de los pocos modelos de efecto hueco válidos en \mathbb{R}^3 , y está definido por:

$$\gamma(\|\mathbf{h}\|; \theta) = \sigma^2 \left(1 - \frac{\phi}{\|\mathbf{h}\|} \sin \left(\frac{\|\mathbf{h}\|}{\phi} \right) \right), \quad \theta = (\sigma^2, \phi) \quad (2.89)$$

Variogramas sin silla

Estos modelos corresponden a procesos estocásticos espaciales o campos aleatorios que son intrínsecamente estacionarios, pero no son estacionarios de segundo orden. Estos modelos son no acotados y corresponden a un proceso estocástico espacial o campo aleatorio con capacidad ilimitada para la dispersión espacial y, por consiguiente, ni su covarianza, ni su varianza pueden ser definidas.

Modelo Potencia

Este modelo es válido en $\mathbb{R}^d, d \geq 1$ y está definido por:

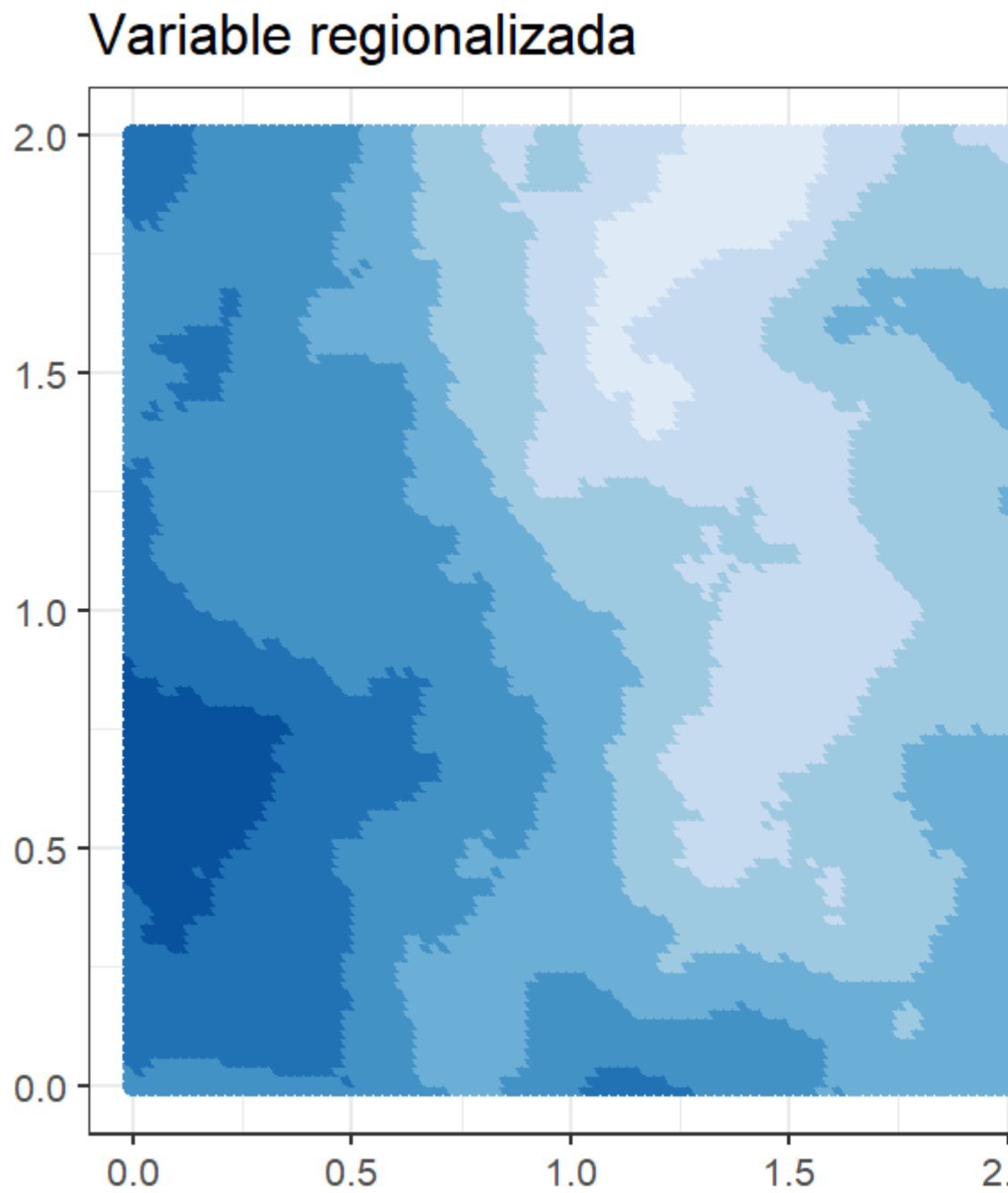


Figure 2.21: Izquierda: Variable regionalizada. Derecha: Variograma J-Bessel.

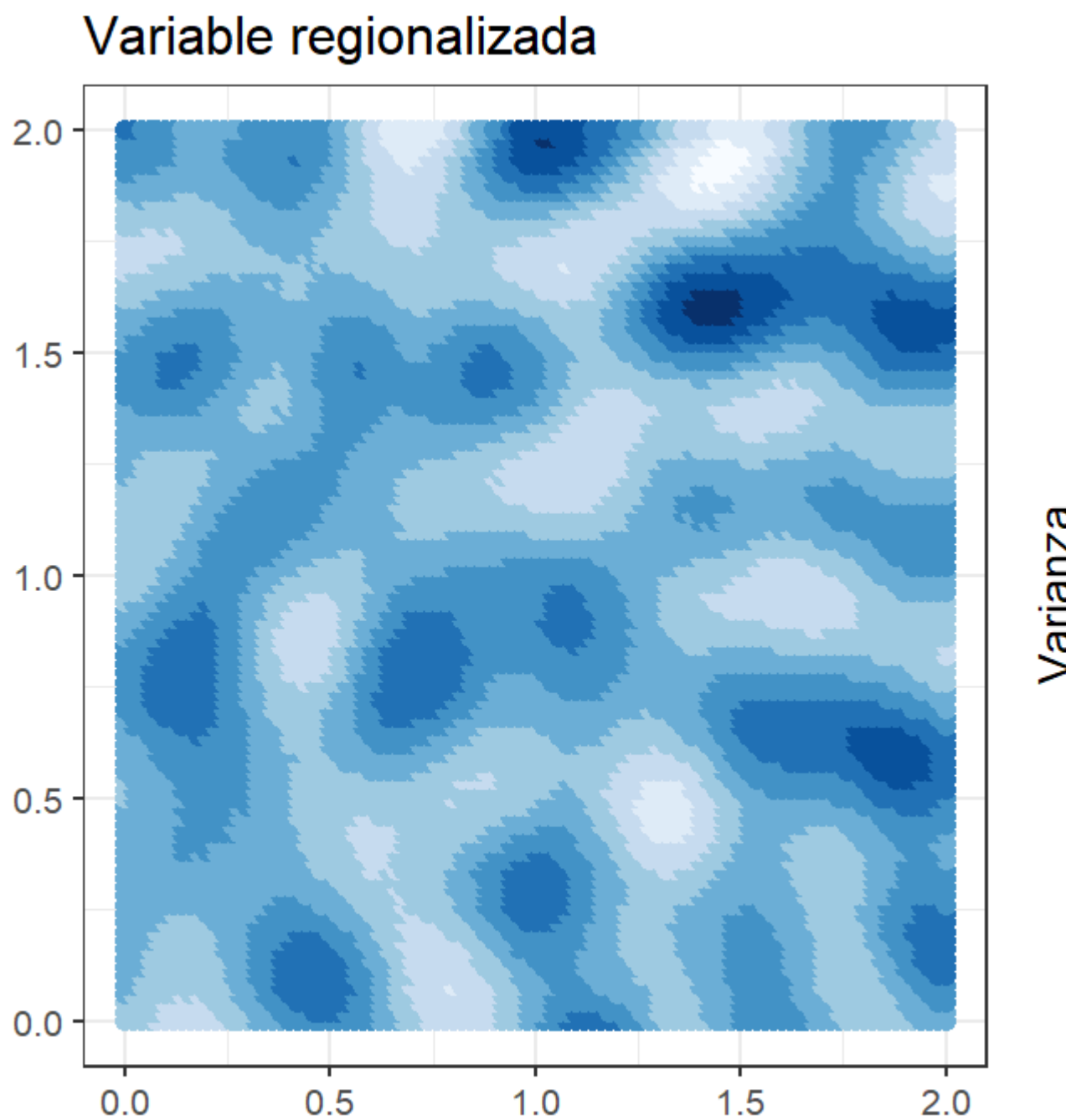


Figure 2.22: Izquierda: Variable regionalizada. Derecha: Variograma seno cardinal.

$$\gamma(\|\mathbf{h}\|) = (\|\mathbf{h}\|)^\alpha, \quad 0 < \alpha < 2,$$

Este variograma no posee ni silla ni rango, sino que crece indefinidamente. Si α es cercano a cero, se dice que es un variograma de efecto pepita, si α es cercano a dos tiene un comportamiento parabólico y si α es igual a uno presenta un comportamiento lineal. Para $\alpha \geq 2$ la condición $\lim_{h \rightarrow \infty} \frac{\gamma(\|\mathbf{h}\|)}{h^2} = 0$ no se satisface; es decir, el modelo no es intrínsecamente estacionario.

Modelo lineal

Este modelo es un caso especial del modelo potencia si $\alpha = 1$, y está asociado con la estacionariedad intrínseca, pero no con la estacionariedad de segundo orden y está definido por:

$$\gamma(\|\mathbf{h}\|) = \begin{cases} 0 & \text{si } \|\mathbf{h}\| = 0 \\ \|\mathbf{h}\| & \text{si } \|\mathbf{h}\| > 0 \end{cases}$$

2.4.7 Caso Multivariante (Ejemplos)

El variograma multivariado fue formalizado en 1991 como una aplicación a procesos estocásticos espaciales o campos aleatorios estacionarios. Para un vector de p variables regionalizadas estacionarias y para toda métrica M , el covariograma multivariado y el variograma multivariado son definidos como:

- Covariograma multivariado:

$$K(h) = E[(Z(s) - \mu)M(Z(s + \mathbf{h}) - \mu)'] \quad (2.90)$$

- Variograma multivariado:

$$\Gamma(h) = E[(Z(s) - Z(s + \mathbf{h}))M(Z(s) - Z(s + \mathbf{h}))'] \quad (2.91)$$

donde $Z(s)$ es un vector fila de p procesos estocásticos espaciales o campos aleatorios estacionarios de segundo orden, $\mu = E[Z(s)]$ y M es una matriz simétrica definida positiva de tamaño $p \times p$ utilizada como métrica en el cálculo de similaridades (?).

Asumiendo estacionariedad de segundo orden, la función de autocovarianza multivariada está relacionada con el variograma multivariado por medio de:

$$K(h) = \Gamma(\infty) - \Gamma(h) \quad (2.92)$$

donde $\Gamma(\infty)$ es la silla del variograma multivariado. Si $\Gamma(0) = 0$ entonces $K(0) = \Gamma(h)$.

El variograma multivariado representa la esperanza matemática de una medida de disimilitud cuadrática, y la función de autocovarianza multivariada representa la esperanza matemática de una medida de similaridad multivariante (?).

La estimación del variograma multivariado se realiza promediando las disimilitudes multivariadas al cuadrado, de forma similar al variograma tradicional; es decir:

$$\Gamma^*(h) = \frac{1}{2|N(\mathbf{h})|} \sum_{N(\mathbf{h})} d_{ij}^2 \quad (2.93)$$

donde d_{ij} es la disimilitud entre las muestras i y j calculadas con una métrica dada y $N(\mathbf{h})$ como se definió anteriormente (?).

Ahora bien, para realizar el ajuste de los modelos teóricos de variogramas y variogramas cruzados, se utiliza el Modelo Lineal de Coregionalización.

Modelo lineal de coregionalización (MLC)

Para el cálculo del MLC es necesario introducir las siguientes medidas:

- **Covarianza cruzada:**

$$C_{rr'}(\mathbf{h}) = [E(Z_r(s) - E(Z_r(s)))][E(Z_{r'}(s + \mathbf{h}) - E(Z_{r'}(s + \mathbf{h})))] \quad (2.94)$$

- **Correlación cruzada:**

$$\rho_{rr'} = \frac{C_{rr'}(\mathbf{h})}{\sigma_r \sigma_{r'}} \quad (2.95)$$

- **Variograma cruzado:**

$$\gamma_{rr'}(\mathbf{h}) = \frac{1}{2} E [Z_r(s + \mathbf{h}) - Z_r(s)] [Z_{r'}(s + \mathbf{h}) - Z_{r'}(s)] \quad (2.96)$$

- **El pseudo variograma cruzado:**

$$\phi_{rr'} = \frac{1}{2} E [Z_r(s + \mathbf{h}) - Z_{r'}(s)]^2 \quad (2.97)$$

- **Co-dispersión:**

$$v_{rr'}(\mathbf{h}) = \frac{\gamma_{rr'}(\mathbf{h})}{\sqrt{\gamma_{rr}(\mathbf{h})}\sqrt{\gamma_{r'r'}(\mathbf{h})}} \in [-1, 1] \quad (2.98)$$

Al igual que en el caso univariante, el variograma cruzado posee las siguientes propiedades: es invariante bajo traslación, $\gamma_{rr'}(0) = 0$, $\gamma_{rr'}(\mathbf{h}) = \gamma_{r'r'}(-\mathbf{h})$. Además, esta medida es la más utilizada en las aplicaciones geoestadísticas.

Ahora bien, sea $\mathbb{Z} = (Z_1, \dots, Z_p)$ un proceso estocástico o campo aleatorio espacial multiavariante; se tiene que el MLC es una suma proporcional de modelos de covarianza o variogramas. En notación matricial donde $C(\mathbf{h}) = [C_{ij}(\mathbf{h})]$ es una matriz de covarianza de dimensión $p \times p$ y de manera similar $\Gamma(\mathbf{h}) = [\Gamma_{ij}(\mathbf{h})]$, con

$$C(\mathbf{h}) = \sum_{k=1}^L B_k C_k(\mathbf{h}) \quad (2.99)$$

$$\Gamma(\mathbf{h}) = \sum_{k=1}^L B_k \gamma_k(\mathbf{h}) \quad (2.100)$$

donde cada B_k se conoce como *matriz de coregionalización* y:

$$C_{ij}(\mathbf{h}) = \sum_{k=1}^L b_k(i, j) C_k(\mathbf{h}) \quad \forall i, j = 1, \dots, p \quad (2.101)$$

$$\Gamma_{ij}(\mathbf{h}) = \sum_{k=1}^L b_k(i, j) \gamma_k(\mathbf{h}) \quad \forall i, j = 1, \dots, p \quad (2.102)$$

El MLC asume que todos los variogramas simples y cruzados pueden ser expresados como una combinación lineal de modelos básicos (exponencial, Gaussiano, esférico, entre otros) idénticos indexados por k . Una condición suficiente para que el modelo sea válido es que cada una de las matrices B_k sean definidas positivas.

Por construcción del modelo, todas las covarianzas cruzadas son simétricas, y los variogramas cruzados son modelos de variograma válidos.

Ajuste del MLC

El ajuste del MLC puede ser realizado mediante el método de mínimos cuadrados, al igual que en el caso univariante, teniendo en cuenta que la única diferencia es que la silla de cada elemento es reemplazada por una matriz de sillas.

El ajuste es llevado a cabo por el algoritmo propuesto por Goulard (1989) quienes implementaron este algoritmo de manera iterativa para asegurar la positividad de los coeficientes B_k .

Se definen matricialmente el variograma empírico y variograma teórico multivariado como sigue:

$$\Gamma(\hat{\mathbf{h}}) = [\hat{\gamma}_{ij}(\mathbf{h})] \quad (2.103)$$

$$\Gamma(\mathbf{h}) = [\gamma_{ij}(\mathbf{h})] = \sum_k B_k \gamma_k(\mathbf{h}) \quad (2.104)$$

el criterio de bondad de ajuste es la suma de cuadrados ponderados (SCP) de todos los términos de la matriz de errores $\hat{\Gamma}(\mathbf{h}) - \Gamma(\mathbf{h})$ y se suman sobre el conjunto de retardos J utilizados para el ajuste; es decir:

$$SCP = \sum_{k \in J} w(h) \text{traza} \left\{ \left[V^* (\hat{\Gamma}(\mathbf{h}) - \Gamma(\mathbf{h})) \right]^2 \right\} \quad (2.105)$$

Los ponderadores $w(h)$ son positivos y usualmente iguales al número de pares utilizados para la estimación del variograma en el retardo h . La matriz V^* es definida positiva y diseñada para equilibrar la influencia de las variables; por lo general es la diagonal de la matriz o inversa de la matriz de varianzas o la matriz identidad. La idea es minimizar el criterio optimizando un B_k a la vez y repitiendo esto hasta que no haya mejora alguna. El residuo para el ajuste actual menos el k -ésimo término es:

$$d\Gamma_k(\mathbf{h}) = \hat{\Gamma}(\mathbf{h}) - \sum_{u \neq k} B_u \gamma_u(\mathbf{h}) \quad (2.106)$$

En ausencia de restricción de positividad, el ajuste óptimo de $d\Gamma_k$ por $B_k \gamma_k(\mathbf{h})$ se obtiene por medio de la cancelación de la derivada de SCP en relación a B_k

$$\frac{\partial SCP}{\partial B_k} = -2V^* \left[\sum_{h \in J} w(h) \gamma_k(\mathbf{h}) (d\Gamma_k(\mathbf{h}) - \gamma_k(\mathbf{h}) B_k) \right] V^* = 0 \quad (2.107)$$

Como V^* es no singular se tiene que:

$$B_k = \frac{1}{\alpha_k} \sum_{h \in J} w(h) \gamma_k(\mathbf{h}) d\Gamma_k(\mathbf{h}) \quad (2.108)$$

$$(2.109)$$

donde

$$\alpha_k = \sum_{h \in J} w(h) \gamma_k(\mathbf{h})^2 \quad (2.110)$$

La solución restringida $B_k^+ \geq 0$ es la matriz definida positiva cercana a B_k de acuerdo a la norma definida por V^* . Dado que es simétrica, la matriz B_k tiene descomposición de la siguiente forma

$$B_k = U_k \nabla_k U_k' \quad \text{con} \quad U_k' V^* U_k = I_p \quad (2.111)$$

donde U_k es una matriz de vectores propios de $B_k V^*$ y ∇_k es la matriz diagonal de sus valores propios.

Por tanto, la solución restringida es

$$B_k' = U_k \nabla_k^+ U_k' \quad (2.112)$$

donde ∇_k^+ es la matriz ∇_k en el cual los valores propios negativos son reemplazados por ceros. Este algoritmo iterativo siempre converge y la solución no depende del punto de partida (?).

2.5 Geoestadística Espacio Temporal

A través de la geoestadística se han estudiado los campos aleatorios en la naturaleza; sin embargo, existen ramas de la ciencia que presentan aplicaciones prácticas como las ciencias sociales, monitorización de contaminación ambiental, climatología, geología, biología, medicina, arqueología y cualquier otro campo de la ciencia que estudie fenómenos que ocurren tanto en espacio como en tiempo (?). Al considerar la evolución conjunta del proceso espacio-temporal, en lugar de solo considerar su distribución espacial para un determinado instante de tiempo (proceso puramente espacial) o la evolución temporal del proceso sobre una localización determinada (proceso puramente temporal), se obtendrá grandes beneficios en la modelización y predicción de un fenómeno en estudio (?). Es así que considerar ambos elementos por separado nos llevaría a una pérdida de información. En este sentido, la geoestadística espacio-temporal ofrece las herramientas adecuadas para este tipo de estudios. Es por ello que la demanda de modelos covariográficos que describen la evolución de procesos en espacio-tiempo se hace cada vez más notoria.

Los métodos geoestadísticos y espacio-temporales utilizan distancias euclídeas o espacio temporales, por tanto se consideran los métodos basados en distancias para el cálculo entre las observaciones que junto la información proporcionada por el variograma permitirá la generación de mejores pronósticos (?). Por otro lado se debe considerar que en el campo espacio-tiempo también se trabaja con grandes volúmenes de datos debido a la resolución espacial y periodicidad temporal, lo que conlleva al uso de métodos computacionales intensivos.

2.5.1 Conceptos

Se puede considerar a cada ubicación espacio-temporal como un punto en $\mathbb{R}^d \times \mathbb{R}$, con \mathbb{R}^d siendo el espacio euclidiano d -dimensional y \mathbb{R} la dimensión del tiempo. Notemos que ni las escalas ni las unidades de distancia en el tiempo y el espacio son generalmente comparables. Por lo que la dificultad principal en el análisis de procesos espacio-temporales recae en la elección de un modelo de covarianza válido que se ajuste a las observaciones. Bajo hipótesis de regularidad se podrán analizar las observaciones separadas por un vector de distancia espacio temporal (\mathbf{h}, u) (?).

El objetivo de esta sección es introducir aquellos conceptos más relevantes en el análisis de procesos estocásticos espacio-temporales. Muchos de los conceptos tratados serán generalizaciones del caso espacial que se han visto en el capítulo X.

Proceso Espacio Temporal

Sea Z la variable aleatoria de interés, s su ubicación espacial y t el tiempo de ocurrencia. Un proceso espacio-temporal es un proceso estocástico

$$\{Z(s,t):(s,t) \in D \times D'\}$$

donde el conjunto $D \subset \mathbb{R}^d$ es el conjunto de todas las ubicaciones espaciales y el conjunto $D' \subset \mathbb{R}$ es el conjunto de todos los instantes de tiempo. Los conjuntos D y D' pueden ser continuos o discretos, fijos o aleatorios (?).

Siguiendo los conceptos clásicos de geoestadística, la distribución del proceso espacio-temporal es modelado como una distribución Gaussiana. Esta distribución será caracterizada definiendo su momento de primer orden y su momento de segundo orden, es decir, su esperanza y función de covarianza (?).

Propiedad de regularidad

Un campo aleatorio $Z(s, t)$ en $\mathbb{R}^d \times \mathbb{R}$ cumple la propiedad de regularidad si:

$$V(Z(s, t)) < \infty, \forall (s, t) \in \mathbb{R}^d \times \mathbb{R} \quad (2.113)$$

que a su vez asegura la existencia de los dos primeros momentos.

Campo aleatorio Gaussiano

Un campo aleatorio $Z(s, t)$ se dice Gaussiano si el vector aleatorio $\mathbf{Z} = (Z(s_1, t_1), \dots, Z(s_n, t_n))$ para cualquier conjunto de ubicaciones espacio-temporales $(s_1, t_1), \dots, (s_n, t_n)$, sigue una distribución Gaussiana multivariante.

Función de media

La media de un campo aleatorio espacio-temporal está definido por la función

$$E[Z(s, t)] = \mu(s, t) \quad (2.114)$$

Función de covarianza

Para cualquier para de puntos espacio-temporales, la covarianza de un campo aleatorio espacio temporal está definido por la función

$$C[(s_i, t_i), (s_j, t_j)] = C[Z(s_i, t_i), Z(s_j, t_j)] \quad (2.115)$$

Estacionariedad fuerte o estricta

Un campo aleatorio $Z(s, t)$ es estrictamente estacionario si su distribución de probabilidad es traslacionalmente invariante. Es decir que, para dos vectores \mathbf{h} y u

$$Z(s, t) = [Z(s_1, t_1), Z(s_2, t_2), \dots, Z(s_n, t_n)]' \quad (2.116)$$

y

$$Z(s + \mathbf{h}, t + u) = [Z(s_1 + \mathbf{h}, t_1 + u), Z(s_2 + \mathbf{h}, t_2 + u), \dots, Z(s_n + \mathbf{h}, t_n + u)]' \quad (2.117)$$

Tienen la misma función de distribución multivariante, es decir

$$F_{Z_1, \dots, Z_n}[(s_1, t_1), \dots, (s_n, t_n)] = F_{Z_1, \dots, Z_n}[(s_1 + \mathbf{h}, t_1 + u), \dots, (s_n + \mathbf{h}, t_n + u)] \quad (2.118)$$

Estacionariedad de segundo orden

Un campo aleatorio $Z(s, t)$ es estacionario de segundo orden si tiene media constante y la función de covarianza depende solo de \mathbf{h} y u , es decir

$$E[Z(s, t)] = \mu(s, t) = \mu, \quad \forall (s, t) \in \mathbb{R}^2 \times \mathbb{R} \quad (2.119)$$

$$C[Z(s_i, t_i), Z(s_j, t_j)] = C(\mathbf{h}, u), \quad \forall (s, t) \in \mathbb{R}^2 \times \mathbb{R} \quad (2.120)$$

Estacionariedad intrínseca

Un campo aleatorio espacio-temporal $Z(s, t)$ es intrínsecamente estacionario si el proceso de incrementos $Z(s + \mathbf{h}, t + u) - Z(s, t)$ es estacionario de segundo orden en espacio-tiempo para cualquier (\mathbf{h}, t) fijo.

Función de covarianza estacionaria espacial

Un campo aleatorio espacio-temporal $Z(s, t)$ tiene función de covarianza estacionaria espacial si, para cualquier par de puntos $(s_i, t_i), (s_j, t_j) \in \mathbb{R}^d \times \mathbb{R}$ la función de covarianza $C[(s_i, t_i), (s_j, t_j)]$ solo depende de la distancia entre las ubicaciones $(s_i - s_j)$ y los tiempos t_i, t_j .

Función de covarianza estacionaria temporal

Un campo aleatorio espacio-temporal $Z(s, t)$ tiene función de covarianza estacionaria temporal si, para cualquier par de puntos $(s_i, t_i), (s_j, t_j) \in \mathbb{R}^d \times \mathbb{R}$ la función de covarianza $C[(s_i, t_i), (s_j, t_j)]$ solo depende de la distancia entre los tiempos $(t_i - t_j)$ y de las ubicaciones espaciales s_i, s_j .

Función de covarianza estacionaria espacio-temporal

Si el campo aleatorio $Z(s, t)$ tiene función de covarianza en términos de espacio y de tiempo, entonces se dice que tiene función de covarianza estacionaria y puede ser expresada de la siguiente forma

$$C[(s_i, t_i), (s_j, t_j)] = C[\mathbf{h}, u] \quad (2.121)$$

donde $\mathbf{h} = s_i - s_j$ y $u = t_i - t_j$ son las distancias en espacio y en tiempo respectivamente.

Si un proceso espacio-temporal tiene función de covarianza estacionaria, entonces su varianza no depende de la ubicación espao-temporal, pues

$$Var(Z(s, t)) = C(0, 0) = \sigma^2, \quad \forall (s, t) \in \mathbb{R}^d \times \mathbb{R} \quad (2.122)$$

donde $C(0, 0) \geq 0$ se conoce como la varianza a priori del proceso.

Función de covarianza separable

Un campo aleatorio $Z(s, t)$ tiene función de covarianza separable si tiene función de covarianza puramente espacial $C_s(s_i, s_j)$ y función de covarianza puramente temporal $C_t(t_i, t_j)$ tal que

$$C[(s_i, t_i), (s_j, t_j)] = C_s(s_i, s_j)C_t(t_i, t_j) \quad (2.123)$$

para cualquier par de puntos espacio-temporales $(s_i, t_i), (s_j, t_j) \in \mathbb{R}^d \times \mathbb{R}$. Si esta descomposición no es posible, la función de covarianza se llamará **no separable**.

Al descomponer la covarianza de acuerdo con (2.123) permite mayor eficiencia en términos computacionales. Es por esto que los modelos separables han sido ampliamente utilizados en aplicaciones geoestadísticas, inclusive en situaciones donde esta hipótesis no se justifica por la naturaleza del proceso bajo análisis.

Función de covarianza simétrica

Un campo aleatorio espacio-temporal $Z(s, t)$ tiene función de covarianza simétrica si para cualquier par de puntos espacio-temporales $(s_i, t_i), (s_j, t_j) \in \mathbb{R}^d \times \mathbb{R}$ se tiene que

$$C[(s_i, t_i), (s_j, t_j)] = C[(s_j, t_j), (s_i, t_i)]$$

Nótese que la separabilidad es un caso particular de la propiedad de simetría.

En el caso de tener función de covarianza espacio-temporal estacionaria, $\forall (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}$ la propiedad de simetría se puede representar de la siguiente manera

$$C(\mathbf{h}, u) = C(\mathbf{h}, -u) \quad (2.124)$$

$$= C(-\mathbf{h}, u) \quad (2.125)$$

$$= C(-\mathbf{h}, -u) \quad (2.126)$$

Función de covarianza espacialmente isotrópica

Un campo aleatorio espacio-temporal $Z(s, t)$ tiene función de covarianza espacialmente isotrópica si

$$C(\mathbf{h}, u) = C(\|\mathbf{h}\|, u), \quad \forall (s, t) \in \mathbb{R}^d \times \mathbb{R} \quad (2.127)$$

Función de covarianza temporalmente isotrópica

Un campo aleatorio espacio-temporal $Z(s, t)$ tiene función de covarianza temporalmente isotrópica si

$$C(\mathbf{h}, u) = C(\mathbf{h}, |u|), \quad \forall (s, t) \in \mathbb{R}^d \times \mathbb{R} \quad (2.128)$$

Nótese que si la función de covarianza de un campo aleatorio espacio-temporal es estacionaria con respecto a espacio y tiempo, entonces es simétrica (?).

Al igual que en el caso de la geoestadística clásica, es frecuente el uso del variograma para la modelización de la estructura espacio-temporal de un campo aleatorio.

Variograma espacio-temporal

La función de variograma espacio-temporal está definido de la siguiente forma

$$\gamma[(s_i, t_i), (s_j, t_j)] = \frac{1}{2} V[Z(s_i, t_i) - Z(s_j, t_j)] \quad (2.129)$$

En el caso de un campo aleatorio con media constante se tiene que

$$\gamma[(s_i, t_i), (s_j, t_j)] = \frac{1}{2} E[Z(s_i, t_i) - Z(s_j, t_j)]^2 \quad (2.130)$$

Siempre que sea posible definir la función de covarianza y la función de variograma, estas estarán relacionadas de acuerdo a lo siguiente

$$\gamma[(s_i, t_i), (s_j, t_j)] = \frac{1}{2} V[Z(s_i, t_i)] + \frac{1}{2} V[Z(s_j, t_j)] - C[(s_i, t_i), (s_j, t_j)] \quad (2.131)$$

Variograma espacio-temporal intrínsecamente estacionario espacial

Un campo aleatorio $Z(s, t)$ tiene variograma intrínsecamente estacionario con respecto al espacio si, para cualquier par de puntos espacio-temporales $(s_i, t_i), (s_j, t_j) \in \mathbb{R}^d \times \mathbb{R}$, el variograma $\gamma[(s_i, t_i), (s_j, t_j)]$ solo depende de $\mathbf{h} = s_i - s_j$ y los tiempos t_i y t_j .

Variograma espacio-temporal intrínsecamente estacionario temporal

Un campo aleatorio $Z(s, t)$ tiene variograma intrínsecamente estacionario con respecto al tiempo si, para cualquier par de puntos espacio-temporales $(s_i, t_i), (s_j, t_j) \in \mathbb{R}^d \times \mathbb{R}$, el variograma $\gamma[(s_i, t_i), (s_j, t_j)]$ solo depende de $u = t_i - t_j$ y las ubicaciones s_i y s_j .

Variograma espacio-temporal intrínsecamente estacionario

Un campo aleatorio $Z(s, t)$ tiene variograma intrínsecamente estacionario si, su variograma es intrínsecamente estacionario en espacio y tiempo. En este caso, el variograma se puede expresar de la siguiente forma

$$\gamma[(s_i, t_i), (s_j, t_j)] = \gamma(\mathbf{h}, u) \quad (2.132)$$

donde $\mathbf{h} = s_i - s_j$ y $u = t_i - t_j$ representan la distancia en espacio y tiempo respectivamente.

Dado un campo aleatorio estacionario de segundo orden $Z(s, t)$ con función de covarianza $C(\mathbf{h}, u)$, debido a (2.131), se cumple que

$$V[Z(s_i, t_i) - Z(s_j, t_j)] = V[Z(s_i, t_i)] + V[Z(s_j, t_j)] - 2C[(s_i, t_i), (s_j, t_j)] \quad (2.133)$$

$$= 2C(\mathbf{0}, 0) - 2C(s_i - s_j, t_i - t_j) \quad (2.134)$$

por tanto el campo aleatorio $Z(s, t)$ será intrínsecamente estacionario con variograma

$$\gamma(\mathbf{h}, u) = C(\mathbf{0}, 0) - C(\mathbf{h}, u) \quad (2.135)$$

Función de correlación espacio-temporal

Sea $Z(s, t)$ un campo aleatorio espacio-temporal estacionario de segundo orden con varianza a priori $\sigma^2 = C(\mathbf{0}, 0) > 0$. La función de correlación se define como sigue

$$\rho(\mathbf{h}, u) = \frac{C(\mathbf{h}, u)}{C(\mathbf{0}, u)} \quad (2.136)$$

si $\rho(\mathbf{h}, u)$ es función de correlación en $\mathbb{R}^d \times \mathbb{R}$, entonces sus funciones marginales $\rho(\mathbf{0}, u)$ y $\rho(\mathbf{h}, 0)$ serán funciones de correlación espacial en \mathbb{R}^d y temporal en \mathbb{R} respectivamente.

2.5.1.1 Propiedades de la covarianza y variograma espacio-temporal

Una condición necesaria y suficiente para que la función de valores reales $C[(s_i, t_i), (s_j, t_j)]$ definido en $\mathbb{R}^d \times \mathbb{R}$ sea una función de covarianza es que sea simétrica y definida positiva, es decir

$$C[(s_i, t_i), (s_j, t_j)] = C[(s_j, t_j), (s_i, t_i)] \quad (2.137)$$

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j C[(s_i, t_i), (s_j, t_j)] \geq 0 \quad (2.138)$$

para cualquier $n \in \mathbb{N}$ y para cualquier $(s_i, t_i) \in \mathbb{R}^d \times \mathbb{R}$ y $a_i \in \mathbb{R}$, $i = 1, \dots, n$.

De manera similar, una condición necesaria y suficiente para una función no negativa de valores reales $\gamma[(s_i, t_i), (s_j, t_j)]$ definidas en $\mathbb{R}^d \times \mathbb{R}$ para que una función de variograma es que sea simétrica y condicionalmente definida negativa, es decir

$$\gamma[(s_i, t_i), (s_j, t_j)] = \gamma[(s_j, t_j), (s_i, t_i)] \quad (2.139)$$

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma[(s_i, t_i), (s_j, t_j)] \leq 0 \quad (2.140)$$

para cualquier $n \in \mathbb{N}$ y para cualquier $(s_i, t_i) \in \mathbb{R}^d \times \mathbb{R}$ y $a_i \in \mathbb{R}$, $i = 1, \dots, n$ con $\sum_{i=1}^n a_i = 0$.

En el caso de estacionariedad, se puede deducir que la función de valores reales $C(\mathbf{h}, u)$ definido en $\mathbb{R}^d \times \mathbb{R}$ es una función de covarianza, si y solo si, es una función par y definida positiva, es decir

$$C(\mathbf{h}, u) = C(-\mathbf{h}, -u) \quad (2.141)$$

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j C(\mathbf{h}_i, u_i) \geq 0 \quad (2.142)$$

para cualquier $n \in \mathbb{R}$ y para cualquier $(\mathbf{h}_i, u_i) \in \mathbb{R}^d \times \mathbb{R}$, $a_i \in \mathbb{R}$, $i = 1, \dots, n$.

En este sentido, una función no negativa de valores reales $\gamma(\mathbf{h}, u)$ definida en $\mathbb{R}^d \times \mathbb{R}$ es una función de variograma intrínsecamente estacionaria, si y solo si, es una función para y condicionalmente definida negativa, es decir

$$\gamma(\mathbf{h}, u) = \gamma(-\mathbf{h}, -u) \quad (2.143)$$

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma(\mathbf{h}_i, u_i) \leq 0 \quad (2.144)$$

para cualquier $n \in \mathbb{R}$ y para cualquier $(\mathbf{h}_i, u_i) \in \mathbb{R}^d \times \mathbb{R}$, $a_i \in \mathbb{R}$, $i = 1, \dots, n$ con $\sum_{i=1}^n a_i = 0$

Algunas de las propiedades más importantes son las siguientes

- Si $C_1(\mathbf{h}, u)$ es una función de covarianza definida en $\mathbb{R}^d \times \mathbb{R}$ y $b > 0$, entonces

$$C(\mathbf{h}, u) = bC_1(\mathbf{h}, u) \quad (2.145)$$

- Sean $C_1(\mathbf{h}, u)$ y $C_2(\mathbf{h}, u)$ dos funciones de covarianza definidas en $\mathbb{R}^d \times \mathbb{R}$, entonces

$$C(\mathbf{h}, u) = C_1(\mathbf{h}, u) + C_2(\mathbf{h}, u) \quad (2.146)$$

es también una función de covarianza definida en $\mathbb{R}^d \times \mathbb{R}$.

- Dadas dos funciones de covarianza $C_1(\mathbf{h}, u)$ y $C_2(\mathbf{h}, u)$ definidas en $\mathbb{R}^d \times \mathbb{R}$, entonces

$$C(\mathbf{h}, u) = C_1(\mathbf{h}, u)C_2(\mathbf{h}, u) \quad (2.147)$$

es también una función de covarianza definida en $\mathbb{R}^d \times \mathbb{R}$.

- Si $\{C_n(\mathbf{h}, u), n = 1, \dots\}$ es una secuencia de funciones de covarianza definidas en $\mathbb{R}^d \times \mathbb{R}$ convergen a $C(\mathbf{h}, u), \forall (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}$, entonces

$$C(\mathbf{h}, u) = \lim_{n \rightarrow \infty} C_n(\mathbf{h}, u) \quad (2.148)$$

es una función de covarianza en $\mathbb{R}^d \times \mathbb{R}$.

MODELOS NO SEPARABLES: sherman 127, g_est_temp_casal: 157 (modelos)

Modelos separables: montero 198, g_esp_temp 52

Estimacion de varis: montero 183

montero 179 nuestro 11 sherman 127 wikle para ejemplos de R pg 46
g_esp_tiempo pg 49 y apoyo c01 pg 12 apoyo 2 martha pg69

2.5.2 Modelos separables

2.5.3 Modelos no separables

2.5.4 Estimación del variograma espacio-temporal (Ejemplos)

Chapter 3

Introducción al análisis de datos funcionales (FDA)

Los objetivos fundamentales del FDA son los mismos que la estadística convencional e incluyen: (a) Representar y transformar los datos de manera que faciliten el análisis posterior. (b) Mostrar los datos de manera que se destaquen las variables en estudio. (c) Estudiar fuentes relevantes de patrones y variación de los datos. (d) Explicar la variación de una variable resultante o dependiente utilizando la información de la variable independiente (?).

En la estadística elemental las observaciones son escalares, pero en la estadística multivariante las observaciones son vectores que pertenecen al espacio Euclídeo \mathbb{R}^d , donde d es la dimensión de los vectores observados. Sin embargo, en el FDA las observaciones son funciones o curvas y aunque estas pueden ser medidas en puntos discretos se las pueden pensar como funciones en espacios infinitos dimensionales (?).

Los datos funcionales por tanto son una generalización natural de datos multivariantes, pasando de una dimensión finita a una dimensión infinita (?). La alta dimensionalidad de los datos, representa un reto para la teoría y para el cálculo computacional, donde estos retos varían dependiendo de como se tomaron las muestras. Sin embargo, la alta o infinita dimensionalidad de los datos, son una fuente rica en información, lo que conlleva a poder realizar investigaciones y análisis de datos mas robustos (?).

La extensa variedad de aplicaciones y herramientas hacen que definir de forma precisa el FDA sea de cierta manera difícil. El FDA surge cuando una variable de interés, en un conjunto de datos, puede ser naturalmente vista como una curva o función suave; entonces el FDA puede pensarse como un análisis de alto nivel estadístico sobre muestras de curvas (?). En la práctica, los datos funcionales son obtenidos por medio de observaciones sobre el tiempo, ESPACIO

u otros dominios continuos; los datos funcionales resultantes pueden ser curvas, superficies u objetos complejos (?).

INSERTAR IMÁGENES DATO FUNCIONAL CON DOMINIOS

Las funciones o curvas pueden ser vistas como realizaciones de un proceso estocástico uno dimensional; usualmente se asume que estas realizaciones pertenecen al espacio de *Hilbert*.

3.1 Espacios de Hilbert

El espacio de *Hilbert* generaliza la noción de espacio euclidiano, extendiendo las propiedades de dimensión finita para dimensión infinita.

3.1.1 El Espacio L^2

El espacio $L^2 = L^2[0, 1]$ es el conjunto de funciones reales medibles de Lebesgue f definidas en $[0, 1]$, tales que satisfacen (?):

$$\int_0^1 f^2(t)dt < \infty \quad (3.1)$$

Este espacio se le conoce como el *Espacio de Funciones Cuadrado Integrables*.

Si $f, g \in L^2$, la igualdad $f = g$, significa que (?):

$$\int [f(t) - g(t)]dt = 0 \quad (3.2)$$

Las funciones cuadrado integrables forman un espacio vectorial; esto significa que, si $f, g \in L^2$, entonces se cumple que, para cualquier escalar a, b :

$$af + bg \in L^2 \quad (3.3)$$

$$(af + bg)(t) = af(t) + bg(t), \quad t \in [0, 1] \quad (3.4)$$

Se debe tener en cuenta que los elementos de L^2 y operaciones sobre estos no necesitan estar definidos para todo punto t , sino que están definidos para casi todo punto t (?).

Para funciones $f, g \in L^2$, se define el producto interno de dos funciones como:

$$\langle f, g \rangle = \int f(t)g(t)dt \quad (3.5)$$

El producto interno permite entender la noción de ortogonalidad, lo cual hace que la estructura geométrica del espacio L^2 sea intuitivamente muy similar al espacio euclidiano finito dimensional.

Se dice que dos funciones f, g son ortogonales si:

$$\langle f, g \rangle = 0 \quad (3.6)$$

El producto interno permite definir la noción de distancia entre funciones a través de la norma:

$$\|f\| = \sqrt{\langle f, f \rangle} = \left\{ \int f^2(t)dt \right\}^{\frac{1}{2}} \quad (3.7)$$

Por otro lado, la distancia entre dos funciones en el espacio L^2 es la norma de su diferencia; es decir:

$$d(f, g) = \|f - g\| \quad (3.8)$$

Un conjunto de funciones $\{e_1, e_2, e_3, \dots\}$ es una base en L^2 si toda función $f \in L^2$ admite una única expansión de la forma:

$$f(t) = \sum_{j=1}^{\infty} a_j e_j(t) \quad (3.9)$$

y se dice que $\{e_1, e_2, e_3, \dots\}$ es una base ortonormal, si:

$$\langle e_j, e_{j'} \rangle = 0, \quad j \neq j' \quad \text{y} \quad \|e_j\| = 1 \quad (3.10)$$

Para una base ortonormal, como por ejemplo la base de Fourier, se tiene que:

$$a_j = \langle f, e_j \rangle \quad (3.11)$$

y tenemos la igualdad de *Parseval*:

$$\int f^2(t)dt = \|f\|^2 = \sum_{j=1}^{\infty} \langle f, e_j \rangle^2 = \sum_{j=1}^{\infty} \left\{ \int f(t)e_j(t)dt \right\}^2 \quad (3.12)$$

3.1.2 Funciones Aleatorias Cuadrado Integrables

Dado que Z denota una función aleatoria en \mathbb{R}^d . Así, como una variable aleatoria, una función aleatoria es definida en un espacio de probabilidad, digamos Ω ; para cada $\omega \in \Omega$, $Z(\omega)$ es una función determinística. Se asume que cada una de las realizaciones $Z(\omega), \omega \in \Omega$, son elementos del espacio L^2 de funciones cuadrado integrables, lo que significa que para cada $\omega \in \Omega$, se verifica que:

$$\|Z(\omega)\|^2 = \int \{Z(\omega)(t)\}^2 < \infty \quad (3.13)$$

La función $\|Z(\omega)\|$ es por tanto una variable aleatoria; si esta variable aleatoria tiene momento de segundo orden finito; es decir, $E\|Z\|^2 < \infty$, por consiguiente Z es una función aleatoria cuadrado integrable.

Es importante notar la diferencia entre una función cuadrado integrable determinística, donde la integración se define en el intervalo $[0, 1]$, y las funciones aleatorias cuadrado integrables, donde la integración está definida en el espacio de probabilidad Ω (?).

3.1.3 Estimación de la función de media y covarianza

En la práctica, se observa una muestra que consiste de N curvas Z_1, Z_2, \dots, Z_N ; se puede ver cada curva como una realización de una función aleatoria Z , o como un elemento aleatorio de L^2 con la misma distribución de Z . Comúnmente se asume que los Z_i son independientes, en particular si las curvas surgen de mediciones sobre sujetos seleccionados aleatoriamente de una población grande.

Ahora bien, suponiendo que Z_1, Z_2, \dots, Z_N son iid en L^2 , y que tienen la misma distribución de Z , la cual se asume de cuadrado integrable. De esta forma se define lo siguiente:

- Función de media:

$$\mu(t) = E[Z(t)] \quad (3.14)$$

- Función de media estimada:

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N Z_i(t) \quad (3.15)$$

- Función de covarianza:

$$c(t, s) = E[(Z(t) - \mu(t))(Z(s) - \mu(s))] \quad (3.16)$$

- Función de covarianza estimada:

$$\hat{c}(t, s) = \frac{1}{N} \sum_{i=1}^N (Z_i(t) - \hat{\mu}(t))(Z_i(s) - \hat{\mu}(s)) \quad (3.17)$$

- Operador de covarianza:

$$C = E[\langle (Z - \mu), \cdot \rangle (Z - \mu)] \quad (3.18)$$

- Estimación del operador de covarianza:

$$\hat{C}(z) = \frac{1}{N} \sum_{i=1}^N \langle Z_i - \hat{\mu}, z \rangle (Z_i - \hat{\mu}), \quad z \in L^2 \quad (3.19)$$

Nótese que \hat{C} proyecta el espacio L^2 en un subespacio finito dimensional generado por Z_1, Z_2, \dots, Z_N , mostrando las limitaciones de la inferencia estadística para observaciones funcionales; una muestra finita puede cubrir un objeto infinito dimensional solo con reducida exactitud (?).

INSERTAR GRAFICOS, CODIGO, EJEMPLOS

3.2 Campos aleatorios funcionales espaciales

3.3 Campos aleatorios funcionales espaciales multivariantes

3.4 Reducción de Dimensionalidad

Como se mencionó, un dato funcional es intrínsecamente infinito dimensional, incluso si un gran número de datos son medidos de manera discreta sobre un conjunto finito de cualquier intervalo la dimensionalidad es alta, produciendo dificultades en el análisis de datos (?), como la ralentización de algoritmos

estadísticos convencionales o incluso produciendo que algunos de estos sean inviables. En algoritmos de clasificación, el costo computacional y la precisión de clasificación pueden ser mejorados de manera significativa utilizando un subconjunto de mediciones representativas; una razón es que instantes en el tiempo consecutivos usualmente están altamente correlacionados, incluyendo redundancias que pueden incrementar el nivel de ruido y ocultar información útil acerca de la estructura de los datos (?).

Para tratar el problema de alta dimensionalidad, se hace uso de funciones base, las que constituyen una herramienta para almacenar información acerca de las funciones y da la flexibilidad para combinar el poder computacional para ajustar cientos y miles de observaciones. Además, permite expresar los cálculos requeridos mediante matrices (?).

El conjunto de datos más simple, en el contexto de FDA, es una muestra de la forma:

$$z_n(t_{j,n}) \in \mathbb{R}, t_{j,n} \in [T_1, T_2], n = 1, \dots, N, j = 1, \dots, J_n \quad (3.20)$$

es decir, tenemos N curvas que son observadas en un intervalo común $[T_1, T_2]$. Los valores de las curvas nunca son conocidos en todos los puntos $t \in [T_1, T_2]$, pues solo están disponibles en puntos específicos $t_{j,n}$, los cuales pueden ser diferentes para curvas z_n diferentes. Algunas aplicaciones relevantes del FDA, tratan con situaciones donde el número de puntos $\{t_{j,n}\}$, por curva son pequeños. La idea fundamental del FDA es que los objetos en estudio sean curvas suaves:

$$\{z_n(t) : t \in [T_1, T_2], n = 1, \dots, N\} \quad (3.21)$$

para los cuáles los valores $z_n(t)$ existen en todo punto t , pero solo son observados en puntos específicos t_j .

El enfoque principal en el FDA es la forma de las funciones observadas o de las funciones que resumen las propiedades de los datos en un sentido específico (?).

Bases de expansión

En la práctica se trabaja con funciones con características que pueden ser impredecibles y complicadas, por lo que se requiere de estrategias para construir funciones que trabajen con parámetros que sean fáciles de estimar y que puedan ser lo más cercano posible a cualquier característica de la curva.

Se utiliza un conjunto de bloques funcionales $\phi_k, k = 1, \dots, K$ llamados *funciones base*, los cuales son combinados linealmente; es decir, una función $z(t)$ definida en este sentido es expresada de la siguiente forma:

$$z(t) = \sum_{k=1}^K c_k \phi_k(t) = c' \phi(t) \quad (3.22)$$

conocido como *expansión en funciones base*. Los parámetros c_1, \dots, c_K son coeficientes de la expansión. La expresión matricial usa c para referirse al vector de K coeficientes y ϕ para denotar al vector de tamaño K que contiene las funciones suaves de base que comparten las mismas propiedades. (?).

Usualmente se considera una muestra de funciones de tamaño N :

$$z_i(t) = \sum_{k=1}^K c_{ik} \phi_k(t), \quad i = 1, \dots, N \quad (3.23)$$

en este caso, la notación matricial está dada por:

$$\mathbf{z}(\mathbf{t}) = C \phi(t) \quad (3.24)$$

donde $\mathbf{z}(\mathbf{t})$ es un vector de tamaño N conteniendo las funciones $z_i(t)$, y la matriz de coeficientes C tiene N filas y K columnas (?).

Lo ideal, es que las funciones base tengan características que coincidan con las características de las funciones que se están estimando, lo que permite la obtención de una aproximación adecuada utilizando un número más pequeño que K de funciones; cuánto menor sea K y mejor reflejen las funciones base ciertas características de los datos: se tendrá más grados de libertad para realizar pruebas de hipótesis y calcular intervalos de confianza más precisos, menos cálculo computacional es requerido y es factible que los mismos coeficientes logren describir los datos (?).

El concepto de sistemas de bases no es nuevo; por ejemplo, un polinomio de cualquier orden, no es más que una combinación lineal de las funciones base monomial $\{1, t, t^2, \dots, t^n\}$; no obstante, los polinomios no son muy útiles cuando se trabajan con formas funcionales complejas, por lo que, sistemas de base de tipo splines y Fourier son ampliamente utilizados en la práctica (?).

Bases de Fourier

Probablemente la expansión de base más conocida está dada por las series de Fourier. Esta serie es especialmente útil para funciones extremadamente estables; es decir, funciones en las que no hay características locales fuertes y

donde la curva tiende a ser del mismo orden en todas partes. Estas series suelen dar lugar a expansiones uniformes, sin embargo son inadecuadas en cierto grado para los datos que se sospecha que tienen discontinuidades en la función o en las derivadas de bajo orden(?). En algunas ocasiones, es necesario que las funciones se puedan repetir sobre un cierto período de tiempo T ; por ejemplo, pueden ser requeridos para expresar comportamiento estacional en series de tiempo extensas. La serie de Fourier, está dada por (?):

$$\phi_1(t) = 1 \quad (3.25)$$

$$\phi_2(t) = \sin(\omega t) \quad (3.26)$$

$$\phi_3(t) = \cos(\omega t) \quad (3.27)$$

$$\phi_4(t) = \sin(2\omega t) \quad (3.28)$$

$$\phi_5(t) = \cos(2\omega t) \quad (3.29)$$

$$\vdots \quad (3.30)$$

donde la constante ω está relacionado con el período T por la relación:

$$\omega = \frac{2\pi}{T} \quad (3.31)$$

Se observa que, después de la primera función base constante, las funciones base de Fourier son ordenadas de forma sucesiva en pares de seno/coseno, los dos con argumentos en cualquier par que se multiplique por uno de los enteros $1, 2, \dots$ hasta un límite superior m . Si las series contienen ambos elementos de cada par, como es usual, el número de funciones base es $K = 1 + 2m$, puesto que, de la forma en que se definió ω , cada función base se repite por si misma después de T unidades de tiempo transcurrido.

Para definir una base de tipo Fourier, solo son necesario dos parámetros: el número de funciones base K y el período T ; no obstante, este último valor usualmente puede ser por defecto el rango de los valores t en los cuales están definidos los datos (?).

Base de Splines

Las funciones spline son utilizadas usualmente para aproximaciones de datos funcionales no periódicos o perimétricos y estas son parecidas a los polinomios. Esta base ofrece una mayor flexibilidad, sin embargo es más compleja que la series finitas de Fourier. Los splines combinan el cálculo eficiente de los polinomios con una sustancial flexibilidad, usualmente conseguidos con un moderado número de funciones base. El principal camino para ganar flexibilidad en

un spline es incrementar el número de puntos de división del intervalo en el cual se define la función (?).

Funciones Splines y Grados de Libertad

Para definir un spline, el primer paso es dividir el intervalo sobre el cual una función va a ser aproximada en L subintervalos separados por valores $\tau_l, l = 1, \dots, L$ que se conocen como *puntos de ruptura* o *nodos*.

El término *puntos de ruptura*, se refiere al número único de valores de los *nodos* mientras el que término *nodos* se refiere a la secuencia de valores en los *puntos de ruptura*, donde algunos *puntos de ruptura* pueden estar asociados con múltiples *nodos*. Todos los *nodos* son distintos en varias aplicaciones, y por tanto *puntos de ruptura* y *nodos* son la misma cosa, sin embargo, se encuentran sistemas de datos de entrada y salida donde las entradas son variadas en un sentido discreto, y estos requieren nodos coincidentes para modelizar estos cambios (?).

Sobre cada intervalo, un spline es un polinomio de orden específico m . El orden de un polinomio es el número de constantes requeridos para definirlo, y a medida que mayor es su grado, más alta es su potencia. El número total de grados de libertad en un ajuste es igual al orden de los polinomios más el número de *puntos de ruptura* en el interior $m + L - 1$. *puntos de ruptura* interior, el spline se convierte en un polinomio sencillo. Un spline gana flexibilidad al incrementar el número de *puntos de ruptura*; generalmente se requiere más de estos puntos sobre regiones donde la función presenta la variación más compleja, y menos donde la función es levemente no lineal (?).

Para splines de grado mayor que uno, los valores de la función están restringidos a ser iguales en su unión, puesto que los polinomios adyacentes se unen suavemente en el punto de ruptura que los separa, incluso las derivadas de orden $m - 2$ también deben coincidir en esta unión (?).

Base B-Spline

Este sistema de funciones base $\phi_k(t)$ tiene la siguientes propiedades:

- Cada función base $\phi_k(t)$ por si misma es un función spline y está definido por un orden m y una secuencia de nodos τ .
- La combinación lineal de funciones base spline, también es una función spline.
- Cualquier función spline definida por m y τ puede expresarse como combinación lineal de estas funciones base.

La propiedad de que una base de funciones B-spline de orden m sea positiva sobre no más de m intervalos, y que estos se junten, es denominada la **propiedad de soporte compacto**, y es de suma importancia para un cálculo computacional eficiente (?).

La notación $B_k(t, \tau)$ se utiliza para indicar el valor en t de la base de funciones B-spline definido por la secuencia de puntos de ruptura τ ; k hace referencia al número del nodo más grande en o inmediatamente a la izquierda de t , teniendo en cuenta que cuando los nodos son discretos, se requieren $m + L - 1$ funciones base. Una función spline se puede escribir como:

$$S(t) = \sum_{k=1}^{m+L-1} c_k B_k(t, \tau) \quad (3.32)$$

En cuanto a las posiciones de los nodos τ_l , varias aplicaciones por defecto tienen la misma separación siempre y cuando los datos estén también separados de manera regular. Si no lo están, es recomendable ubicar un nodo en cada punto j de los datos, donde j es un número fijado previamente (?).

Estimación de curvas utilizando sistemas de bases por mínimos cuadrados

Se tiene por objetivo ajustar observaciones discretas $y_j, j = 1, \dots, n$ utilizando el modelo $y_j = x(t_j) + \epsilon_j$ y se utiliza una base de funciones para la expansión de $x(t)$ de la forma:

$$x(t) = \sum_{k=1}^K c_k \phi_k(t) = \mathbf{c}' \phi \quad (3.33)$$

donde el vector \mathbf{c} de tamaño K contiene los coeficientes c_k y ϕ la matriz de orden $n \times K$ que contiene los valores $\phi_k(t_j)$.

El suavizamiento básico se obtiene si se determinan los coeficientes de la expansión c_k minimizando el criterio de mínimos cuadrados:

$$SMSSE(y|c) = \sum_{j=1}^n \left[y_j - \sum_{k=1}^K c_k \phi_k(t_j) \right]^2 \quad (3.34)$$

en términos matriciales la expresión anterior está dada de la siguiente manera:

$$SMSSE(y|c) = (y - \Phi c)'(y - \Phi c) = \|y - \Phi c\|^2 \quad (3.35)$$

luego, tomando el criterio de derivada de $SMSSE(y|c)$ con respecto a \mathbf{c} tenemos:

$$2\Phi\Phi'c - 2\Phi'y = 0 \quad (3.36)$$

resolviendo para \mathbf{c} tenemos la estimación $\hat{\mathbf{c}}$ que minimiza la solución de mínimos cuadrados:

$$\hat{\mathbf{c}} = (\Phi'\Phi)^{-1}\Phi'y \quad (3.37)$$

así, el vector $\hat{\mathbf{y}}$ de valores estimados es:

$$\hat{\mathbf{y}} = \Phi\hat{\mathbf{c}} = \Phi(\Phi'\Phi)^{-1}\Phi'y \quad (3.38)$$

Este método básico de aproximación es adecuado cuando se asume que los residuos ϵ_j sobre la curva verdadera son independientes e idénticamente distribuidas con media cero y varianza σ^2 constante.

Ajuste por mínimos cuadrados ponderados

En la práctica se trabaja con errores no estacionarios y/o autocorrelacionados, por lo que se usa el método de mínimos cuadrados ponderados como una extensión del método anterior de la forma:

$$SMSSE(y|c) = (y - \Phi c)'W(y - \Phi c) \quad (3.39)$$

donde W es una matriz simétrica definida positiva que permite la ponderación de cuadrados y producto de los residuos.

Si la matriz de varianza-covarianza Σ_e para los residuos ϵ_j se conoce, entonces:

$$W = \Sigma_e^{-1} \quad (3.40)$$

en aplicaciones donde una estimación completa de Σ_e no es factible, los valores de la covarianza de los errores son usualmente asumidos iguales a cero, entonces W es diagonal con los recíprocos de la varianza de los errores asociados con los y_j .

La estimación $\hat{\mathbf{c}}$ de los coeficientes del vector \mathbf{c} está dada por:

$$\hat{\mathbf{c}} = (\Phi'W\Phi)^{-1}\Phi'W'y \quad (3.41)$$

Suavizamiento de datos funcionales mediante el uso de bases de Fourier

Una función f continua a trozos en el intervalo de $[-\pi, \pi]$, puede ser representada en un sistema de tipo Fourier de la siguiente manera:

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx) \quad (3.42)$$

donde, para todo k se tiene que:

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx dx \quad (3.43)$$

$$b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin kx dx \quad (3.44)$$

esta representación de f se conoce como la expansión en series de Fourier en el intervalo $[-\pi, \pi]$ y donde a_k, b_k son conocidos como los coeficientes de Euler-Fourier de f (?) y este tipo de expansión se utiliza para aproximar datos periódicos.

El suavizamiento de datos, utilizando una base de Fourier, siempre implica que se tenga más puntos de los necesarios para modelizar la suavidad de la función en estudio. En la práctica, la frecuencia de la serie no sobrepasa una frecuencia de corte; es decir, que a partir de cierto término, digamos m , todos los coeficientes, a_k, b_k , de Fourier serán prácticamente cero; por lo tanto, la ecuación anterior, puede escribirse de la siguiente manera:

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^m (a_k \cos kx + b_k \sin kx) \quad (3.45)$$

esta ecuación se conoce como la expansión de Fourier truncada y es la función suavizada de nuestros datos.

Usualmente, el intervalo donde una función está definida no es $[-\pi, \pi]$ o $[0, \pi]$; también podría ser un intervalo $[a, b]$ arbitrario, por lo cual la expansión en serie de Fourier queda de la siguiente forma:

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k \cos \frac{2k\pi x}{b-a} + b_k \sin \frac{2k\pi x}{b-a} \right) \quad (3.46)$$

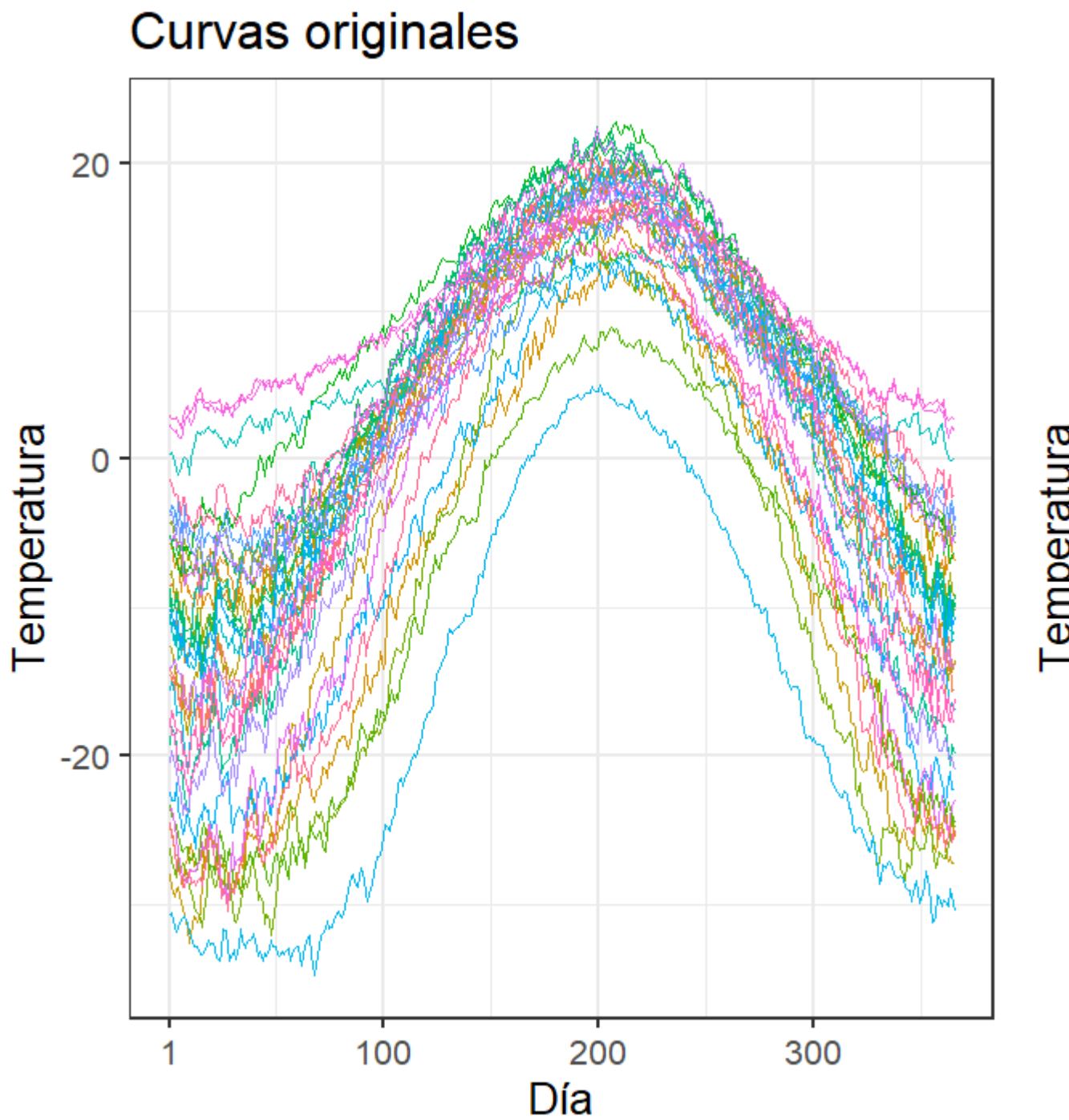


Figure 3.1: Suavización de curvas.

donde, para todo k :

$$a_k = \frac{2}{b-a} \int_a^b f(x) \cos \frac{2k\pi x}{b-a} dx \quad (3.47)$$

$$b_k = \frac{2}{b-a} \int_a^b f(x) \sin \frac{2k\pi x}{b-a} dx \quad (3.48)$$

Suavizamiento de datos funcionales mediante el uso de bases B-splines

Se tiene que, en el caso “más general”, la estimación del vector de coeficientes c , \hat{c} es de la forma:

$$\hat{c} = (\Phi' W \Phi)^{-1} \Phi' W' y \quad (3.49)$$

donde Φ es una matriz de orden $n \times K$ que contiene los valores de las K funciones base en los n puntos de muestra, W es la matriz de ponderación que permite tener una estructura de covarianza en los residuos, y y es el vector de los datos discretos a ser suavizados; por tanto, la expresión para el vector de ajuste de los datos es:

$$\hat{y} = \Phi(\Phi' W \Phi)^{-1} \Phi' W' y = S_\phi y \quad (3.50)$$

donde S_ϕ es el operador proyección correspondiente al sistema de base ϕ :

$$S_\phi = \Phi(\Phi' W \Phi)^{-1} \Phi' W. \quad (3.51)$$

Dado que en el proceso de suavización se tiene un comportamiento de discontinuidades, calcular las rugosidades es de utilidad para compensar dicho comportamiento en los datos suavizados.

Para medir la rugosidad de una función, la integral del cuadrado de la segunda derivada $[D^2 x(t)]^2$ se utiliza:

$$PEN_2(x) = \int [D^2 x(s)]^2 ds \quad (3.52)$$

Para funciones con alta variabilidad, se espera un valor alto de $PEN_2(x)$, pues su segunda derivada es grande al menos en algún rango de interés. De forma

general, la penalización de la rugosidad permite una derivada $D^m x$ de orden arbitrario (?):

$$PEN_m(x) = \int [D^m x(s)]^2 ds \quad (3.53)$$

matricialmente, la expresión anterior está dada por:

$$PEN_m(x) = c' Rc \quad (3.54)$$

dónde

$$R = \int D^m \phi D^m \phi'. \quad (3.55)$$

Ahora bien, añadiendo esta penalización al ajuste por mínimos cuadrados y multiplicando por un parámetro λ de suavizamiento, se obtiene:

$$PENSSSE_m(y|c) = (y - \Phi c)' W (y - \Phi c) + \lambda c' Rc \quad (3.56)$$

y tomando la derivada con respecto al vector c , se obtiene:

$$-2\Phi' W y + \Phi' W \Phi c + \lambda Rc = 0 \quad (3.57)$$

de donde la expresión para el vector de coeficientes estimados \hat{c} es:

$$\hat{c} = (\Phi W \Phi + \lambda R)^{-1} \Phi' W y \quad (3.58)$$

3.4.1 Elección de número de bases

3.5 Detección de atípicos

El tratamiento de datos atípicos es un aspecto importante en cualquier análisis estadístico, pese a que los datos atípicos influyen de manera significativa en metodologías estadísticas en varios sentidos, su análisis en datos funcionales ha sido poco abordado (?).

De manera intuitiva, se podría pensar en la estadística multivariante para tratar casos de datos atípicos en muestras de datos funcionales. Desafortunadamente, hay varios motivos por los que la estadística multivariante falla en el caso funcional. La primera es que los datos funcionales son realizaciones de un proceso aleatorio suave medido en un conjunto discreto de tiempo, por lo cual la estructura de correlación temporal es ignorada cuando se utilizan métodos de estadística multivariante. La segunda es que por la naturaleza infinito-dimensional, los métodos de estadística multivariante son afectados por la alta dimensionalidad; es decir, los métodos no son capaces de manejar las situaciones en donde el número de mediciones de una o más variables es más grande que el número de individuos en la muestra. La tercera es que pocas asunciones sobre la distribución son impuestas sobre los conjuntos de datos funcionales, mientras que los métodos multivariantes están implícitamente sujetos a poblaciones Gaussianas (?).

Los datos atípicos en un conjunto de datos funcionales pueden surgir principalmente por dos razones. La primera es que las curvas pueden tener graves errores de medición, registro o de digitación; estos errores deben ser identificados y corregidos siempre que sea posible. La segunda es que los datos atípicos pueden ser curvas de datos que expresan la realidad, en el sentido de que no tienen errores graves; sin embargo, no se comportan como el resto de curvas. Por tanto, se considera que una curva es un dato atípico si ha sido generado por un proceso estocástico con diferente distribución que el resto de curvas, las cuales se asumen que son idénticamente distribuidas. Una curva puede ser un dato atípico si esta está significativamente alejada del proceso estocástico o si tiene diferente forma y/o comportamiento del resto de curvas (?).

Para identificar datos atípicos en un conjunto de datos funcionales, se hace uso de la *profundidad funcional*. Este concepto fue originalmente introducido en el análisis multivariante para medir la centralidad de un punto en relación con la nube de puntos; es decir, la profundidad da una forma de ordenar los puntos en un espacio Euclídeo desde el centro hacia el exterior; así, los puntos cercanos al centro tendrán una profundidad más grande (?). Por otro lado, para datos funcionales, si una curva es un dato atípico, esta curva tendrá una profundidad baja, por lo que para detectar la presencia de datos atípicos funcionales basta analizar las curvas con profundidades más bajas (?). Existen tres medidas principales para calcular la profundidad funcional, las cuales son: *Profundidad de Fraiman y Muniz*, *Profundidad H-modal*, y *Profundidad de proyección aleatoria*. Para mayor detalle revisar (?).

3.5.1 Medidas de profundidad

El objetivo de profundidad para datos funcionales es medir la centralidad de una curva dada z_i dentro de un conjunto de curvas z_1, \dots, z_n generadas por un proceso estocástico $Z(\cdot)$ con muestra de trayectorias en $C[a, b]$, el espacio de las funciones continuas definidas en el intervalo $[a, b] \subset \mathbb{R}$; es decir, la idea bajo la

profundidad funcional es medir cuánto tiempo una curva permanece en el centro de todo el grupo de trayectorias (?).

Profundidad de Fraiman y Muniz

Sea $F_{n,t}(z_i(t))$ la función de distribución acumulada empírica de los valores de las curvas $z_1(t), \dots, z_n(t)$ en un punto de tiempo $t \in [a, b]$ dado, dado por:

$$F_{n,t}(z_i(t)) = \frac{1}{n} \sum_{k=1}^n I(z_k(t) \leq z_i(t)) \quad (3.59)$$

donde $I(\cdot)$ es la función indicatriz.

La medida de profundidad funcional de *Fraiman y Muniz* de una curva z_i con respecto al conjunto de curvas z_1, \dots, z_n , está dada por:

$$FMD_n(z_i) = \int_a^b D_n(z_i(t)) dt \quad (3.60)$$

donde $D_n(z_i(t))$ es la profundidad univariante del punto $z_i(t)$ dada por:

$$D_n(z_i(t)) = 1 - \left| \frac{1}{2} - F_{n,t}(z_i(t)) \right| \quad (3.61)$$

Profundidad H-modal

Esta medida de profundidad se base en el concepto de moda funcional, definiendo la moda funcional como la curva más densamente rodeada por el resto del conjunto de curvas. Así, la profundidad *h-modal* de una curva z_i , con respecto al conjunto de curvas z_1, \dots, z_n , está dada por:

$$MD_n(z_i, h) = \sum_{k=1}^n K\left(\frac{\|z_i - z_k\|}{h}\right) \quad (3.62)$$

Profundidad de proyección aleatoria

La idea de esta medida es proyectar cada curva funcional y su primera derivada a través de una dirección aleatoria, definiendo un punto en \mathbb{R}^2 , luego, el valor de la profundidad en \mathbb{R}^2 , da un orden de puntos proyectados. Si se trabaja con

un gran número de proyecciones aleatorias, el valor medio de profundidad de los puntos proyectados define la profundidad para datos funcionales. Así, dado un conjunto de curvas z_1, \dots, z_n y una dirección ν que pertenece a un proceso independiente direccional $V(\cdot)$, así la proyección de z_i a través de la dirección ν , está dada por:

$$T_{i,\nu} = \langle \nu, z_i \rangle = \int_a^b \nu(t) z_i(t) dt \quad (3.63)$$

de manera similar, $T'_{i,\nu}$ es la proyección de la derivada de z_i a través de la dirección ν , obteniendo así el par $(T_{i,\nu}, T'_{i,\nu})$ que es un punto en \mathbb{R}^2 . Para P direcciones aleatorias independientes, la profundidad de proyección aleatoria de una curva z_i está definida por:

$$RPD_n(z - I) = \frac{1}{P} \sum_{p=1}^P D_n(\langle \nu_p, z_i \rangle, \langle \nu_p, z'_i \rangle, h) \quad (3.64)$$

donde D_n es cualquier profundidad definida en \mathbb{R}^2 de los puntos $(\langle \nu_p, z_i \rangle, \langle \nu_p, z'_i \rangle) \in \mathbb{R}^2$ (?). Para la visualización de datos atípicos se han desarrollado varias herramientas gráficas, principalmente propuestas para el caso univariante; sin embargo, para el caso multivariante han sido poco desarrolladas (?).

3.5.2 Gráfico de Magnitudes y Formas

Una de la herramientas gráficas usadas es el **MS-Plot** o **Gráfico de Magnitudes y Formas** propuesto por (?), el cual se basa en el marco de la periferia funcional direccional, que mide la centralidad de los datos funcionales considerando el nivel y la dirección de sus desviaciones de la región central.

Periferia Funcional Direccional

Teniendo en cuenta que la externalidad es la medida, grado o cualidad de ser atípico; a este concepto se añade el de dirección, ya que la dirección de externalidad es crucial para describir la centralidad de datos funcionales multivariantes.

De manera formal, la externalidad direccional de un dato escalar está definida por:

$$O(Y, F_Y) = \left\{ \frac{1}{d(Y, F_Y) - 1} \right\} v, \quad d(Y, F_Y) > 0 \quad (3.65)$$

donde F_Y es la distribución de una variable aleatoria, d es una medida convencional de profundidad, y v es el vector unitario que va desde la mediana de F_Y a Y . Asumiendo que Z es la única mediana de F_Y para la medida de profundidad d , v puede expresarse como:

$$v = \frac{Y - Z}{\|Y - Z\|} \quad (3.66)$$

donde $\|\cdot\|$ es la norma L_2 .

Ahora bien, si X es una función p -dimensional definida en un dominio \mathcal{J} , a partir de la distribución de los datos funcionales F_X , para cada punto fijado $t \in \mathcal{J}$, $F_{X(t)}$ es la función de distribución de $X(t)$ con dimensión p . De esta forma, se definen tres medidas de externalidad direccional para datos funcionales:

- 1) Media de externalidad direccional:

$$MO(X, F_X) = \int_{\mathcal{J}} O(X(t), F_{X(t)}) w(t) dt \quad (3.67)$$

- 2) Variación de externalidad direccional:

$$VO(X, F_X) = \int_{\mathcal{J}} \|O(X(t), F_{X(t)}) - MO(X, F_X)\|^2 w(t) dt \quad (3.68)$$

- 3) Externalidad direccional funcional:

$$FO(X, F_X) = \int_{\mathcal{J}} \|O(X(t), F_{X(t)})\|^2 w(t) dt \quad (3.69)$$

donde $w(t)$ es una función de peso definida en \mathcal{J} , la cual puede ser constante o proporcional a la variación local en cada punto de diseño; en particular, se usa $w(t) = \{\lambda(\mathcal{J})\}^{-1}$, donde $\lambda(\cdot)$ es la medida de Lebesgue.

Las tres medidas anteriores de externalidad pueden escribirse en una sola ecuación de la siguiente manera:

$$FO = \|MO\|^2 + VO \quad (3.70)$$

La ecuación anterior realiza una descomposición total de la externalidad funcional (**FO**) en dos términos: la magnitud de externalidad ($\|MO\|$) y la cantidad de variación de externalidad direccional (**VO**). Esta descomposición permite una gran flexibilidad para describir la centralidad de un conjunto de datos funcionales y para identificar curvas potencialmente atípicas (?).

MS-Plot

El **MS-Plot** es un gráfico de dispersión de puntos, $(MO', VO)'$, para un grupo de datos funcionales. Cuando la dimensión es más grande que dos, se usa $(\|MO\|, VO)'$, la cual representa la magnitud general de externalidad y la forma de externalidad sin información de la dirección (?).

De las definiciones de MO Y VO, se esperaría que:

- Las curvas centrales se asignen a la región central baja del MS-Plot; es decir, $\|MO\|$ pequeño y VO pequeño.
- Valores atípicos desplazados se sitúan en la región inferior; es decir, $\|MO\|$ grande y VO pequeño, y las diferentes direcciones de MO indican las diferentes direcciones de sus desplazamientos.
- Los valores atípicos aislados, que están en una pequeña parte del espacio de soporte, se asignan a la región superior central; es decir, $\|MO\|$ pequeño y VO grande.
- Los puntos en la región superior exterior corresponden a las curvas que son sustancialmente atípicas tanto en magnitud como en forma; es decir, $\|MO\|$ grande y VO grande.

3.6 ANOVA Para datos funcionales

En la práctica se vuelve un problema decidir si existen o no diferencias en el proceso de interés cuando varían condiciones que pueden afectarlo. Este problema se suele abordar mediante un modelo que supone la existencia de una función base que describe la evolución típica del proceso estudiado, suponiendo que los datos con los que se trabaja se han obtenido agregando variaciones aleatorias a esta función. Entonces, este problema se convierte en un tipo de problema de ANOVA funcional (?).

Existen varios procedimientos para tratar problemas de tipo ANOVA. En el presente trabajo se tratará el problema de *one-way* ANOVA para datos funcionales univariantes, conocido también como el problema de la *prueba de k-muestras*.

3.6.1 One-Way ANOVA

Este problema puede ser formulado de la siguiente manera:

- Sea $Z_{i1}(t), Z_{i2}(t), \dots, Z_{in_i}(t)$, $i = 1, \dots, k$, donde k hace referencia a los k grupos de funciones aleatorias definidas sobre un intervalo finito $T = [a, b]$ dado.
- Sea $SP(m, \gamma)$ un proceso estocástico con función de media $m(t)$, $t \in T$ y función de covarianza $\gamma(s, t)$, $s, t \in T$.
- Asumiendo que $Z_{i1}(t), Z_{i2}(t), \dots, Z_{in_i}(t)$ son $SP(m_i, \gamma)$, $i = 1, \dots, k$ i.i.d.

Es interesante comprobar la igualdad de las k funciones de media; es decir:

$$H_0 : m_1(t) = \dots = m_k(t), \quad t \in T \quad (3.71)$$

$$H_a : m_1(t) \neq \dots \neq m_k(t), \quad t \in T \quad (3.72)$$

A continuación se presentan varias pruebas para este tipo de ANOVA.

CS-Test

El siguiente estadístico fue propuesto en (?):

$$V_n = \sum_{1 \leq i < j \leq k} n_i \int_T (\bar{Z}_i(t) - \bar{Z}_j(t))^2 dt \quad (3.73)$$

Bajo la hipótesis nula y la asunción de que :

$$n_i, n \rightarrow \infty \quad \text{tal que} \quad \frac{n_i}{n} \rightarrow p_i > 0, \quad i = 1, \dots, k \quad (3.74)$$

se demuestra que la distribución aproximada de V_n es la del estadístico:

$$V = \sum_{1 \leq i < j \leq k} n_i \int_T \left(\bar{Y}_i(t) - \sqrt{\frac{p_i}{p_j}} \bar{Y}_j(t) \right)^2 dt \quad (3.75)$$

donde $Y_1(t), \dots, Y_k(t)$ son procesos Gaussianos independientes de media cero y función de covarianza $\gamma(s, t)$. El valor crítico empírico se calcula mediante remuestreo sobre $Y_i(t)$, $i = 1, \dots, k$ de un proceso Gaussiano con media cero y función de covarianza $\hat{\gamma}$, definida por:

$$\hat{\gamma}(s, t) = \frac{1}{n-k} \sum_{i=1}^k \sum_{j=1}^{n_i} (Z_{ij}(s) - \bar{Z}_i(s))(Z_{ij}(t) - \bar{Z}_i(t)) \quad (3.76)$$

Para el caso heterocedástico y mediante el cálculo del valor crítico empírico, el valor p de V_n puede calcularse como en el caso anterior, mediante procesos Gaussianos independientes con funciones de covarianza dadas por:

$$\hat{\gamma}_i(s, t) = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Z_{ij}(s) - \bar{Z}_i(s))(Z_{ij}(t) - \bar{Z}_i(t)) \quad (3.77)$$

Prueba basadas en la norma \mathcal{L}^2

Este test usa el siguiente estadístico:

$$S_n = \int_T SSR_n(t) dt \quad (3.78)$$

donde:

$$SSR_n(t) = \sum_{i=1}^k n_i (\bar{Z}_i(t) - \bar{Z}(t))^2 \quad (3.79)$$

$$\bar{Z}(t) = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} Z_{ij}(t) \quad (3.80)$$

$$\bar{Z}_i(t) = \frac{1}{n_i} \sum_{j=1}^{n_i} Z_{ij}(t) \quad (3.81)$$

Bajo la hipótesis nula, se tiene que $S_n \sim \beta \chi_d^2$ aproximadamente, donde:

$$\beta = \frac{tr(\gamma^{\otimes 2})}{tr(\gamma)}, \quad (3.82)$$

$$d = (k-1)\kappa, \quad (3.83)$$

$$\kappa = \frac{tr^2(\gamma)}{tr(\gamma^{\otimes 2})}, \quad (3.84)$$

$$\gamma^{\otimes 2}(s, t) = \int_T \gamma(s, u) \gamma(u, t) du. \quad (3.85)$$

Esta distribución aproximada es usada para calcular el p-valor de S_n ; es decir, $P(\chi_d^2 \geq S_n/\beta)$ o su valor crítico $\beta\chi_d^2(1-\alpha)$. Los parámetros β y κ son estimados con base en los datos funcionales por el método de *naïve* o el método de *reducción de sesgo*.

Utilizando el método naïve, y con el estimador $\hat{\gamma}(s, t)$, se tiene que:

$$\hat{\beta} = \frac{\text{tr}(\hat{\gamma}^{\otimes 2})}{\text{tr}(\hat{\gamma})}, \quad (3.86)$$

$$\hat{d} = (k-1)\hat{\kappa}, \quad (3.87)$$

$$\hat{\kappa} = \frac{\text{tr}^2(\hat{\gamma})}{\text{tr}(\hat{\gamma}^{\otimes 2})}. \quad (3.88)$$

Utilizando el método de reducción de sesgo, se tiene que:

$$\hat{\beta} = \frac{\widehat{\text{tr}(\gamma^{\otimes 2})}}{\widehat{\text{tr}(\gamma)}}, \quad (3.89)$$

$$\hat{d} = (k-1)\hat{\kappa}, \quad (3.90)$$

$$\hat{\kappa} = \frac{\widehat{\text{tr}^2(\gamma)}}{\widehat{\text{tr}(\gamma^{\otimes 2})}}, \quad (3.91)$$

$$\widehat{\text{tr}(\gamma^{\otimes 2})} = \frac{(n-k)^2}{(n-k-1)(n-k+2)} \left(\text{tr}(\hat{\gamma}^{\otimes 2}) - \frac{\text{tr}^2(\hat{\gamma})}{n-k} \right), \quad (3.92)$$

$$\widehat{\text{tr}^2(\gamma)} = \frac{(n-k)(n-k+1)}{(n-k-1)(n-k+2)} \left(\text{tr}^2(\hat{\gamma}) - \frac{2\text{tr}(\hat{\gamma}^{\otimes 2})}{n-k+1} \right). \quad (3.93)$$

Prueba F

Este test usa el estadístico:

$$F_n = \frac{\int_T SSR_n(t)dt/(k-1)}{\int_T SSE_n(t)dt/(n-k)} \quad (3.94)$$

$$SSE_n(t) = \sum_{i=1}^k \sum_{j=1}^{n_i} (Z_{ij}(t) - \bar{Z}_i(t))^2 \quad (3.95)$$

Bajo la hipótesis nula, se tiene que $F_n \sim F_{d_1, d_2}$ aproximadamente, donde $d_1 = (k-1)\kappa$ y $d_2 = (n-k)\kappa$. De manera similar, la distribución aproximada de la prueba F puede ser usada para calcular el p-valor de F_n o su valor crítico; y el

parámetro κ puede ser estimado por el método naive o el método de reducción de sesgo.

Cuando las muestras nos son Gaussianas o son pequeñas, estas pruebas no son utilizadas; en este caso, las versiones bootstrap de estas pruebas son usadas para calcular el p-valor de S_n y F_n (?).

Prueba basada en representación de funciones base

Se puede utilizar una función de representación de base para obtener una prueba más robusta.

Asumiendo que se trabaja con k grupos de un proceso estocástico $Z_{ij} \in L_2(T), i = 1, \dots, k, j = 1, \dots, n_i$ y sea $\{\phi_l\}$ una base ortonormal de $L_2(T)$.

Esta prueba usa el siguiente estadístico:

$$F = \frac{\frac{1}{k-1} \sum_{i=1}^k n_i \|\bar{Z}_i - \bar{Z}\|^2}{\frac{1}{n-k} \sum_{i=1}^k \sum_{j=1}^{n_i} \|Z_{ij} - \bar{Z}_i\|^2} \quad (3.96)$$

donde el proceso estocástico, la media funcional muestral y media funcional grupal, escrita en notación matricial a partir de la representación en la base ϕ , están definidos de la siguiente manera:

$$Z_{ij}(t) = c'_{ij} \phi(t), \quad (3.97)$$

$$\bar{Z}(t) = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} c'_{ij} \phi(t), \quad \forall t \in T \quad (3.98)$$

$$\bar{Z}_i(t) = \frac{1}{n_i} \sum_{j=1}^{n_i} c'_{ij} \phi(t) \quad (3.99)$$

para todo $t \in T$.

El numerador del estadístico F , mide la variabilidad externa entre las diferentes muestras, mientras que el denominador mide la variabilidad interna dentro de las muestras (?).

Chapter 4

Introducción a la estadística esepacial funcional

- 4.1 Proceso Espacial Funcional
- 4.2 Datos funcionales espaciales
- 4.3 Variable Regionalizada Funcional
- 4.4 Función de Distribución Conjunta Funcional
- 4.5 Función de Media Funcional
- 4.6 Función de varianza Funcional
- 4.7 Función de Covarianza Funcional
- 4.8 Función de correlación o Correlograma Funcional
- 4.9 Función de semi-variograma Funcional
- 4.10 Estacionariedad de Funciones Aleatorias
- 4.11 No estacionariedad de Funciones Aleatorias
- 4.12 Correlación Espacial Funcional
- 4.13 Índices de correlación espacial

Chapter 5

Agrupación

De manera empírica, agrupación significa encontrar grupos en un conjunto de datos. La primera agrupación en la historia fue de tipo jerárquico, cuando Aristóteles clasificó a los seres vivos. El análisis de agrupación ha sido desarrollado en diferentes campos y con diversas aplicaciones (?).

En el análisis de agrupación el objetivo es examinar datos que no contienen etiquetas, para luego encontrar agrupaciones de los datos; se debe tener en cuenta que se desconoce el número de grupos. Esta metodología es una forma de aprendizaje no supervisado; es decir, no se utilizan datos de entrenamiento. (?).

Formalmente, si se tiene un conjunto de datos, $D = \{z_1, z_2, \dots, z_n\}$, el cual contiene n datos, la tarea de los métodos de agrupación es asignarlos a K subconjuntos disjuntos de D , denotados por C_1, C_2, \dots, C_K (?). Estos métodos con frecuencia se utilizan como un paso preliminar en la exploración de datos para identificar patrones que tengan una interpretación útil para el usuario (?).

5.1 Algoritmos de agrupación

Algoritmo 1 Algoritmo 2 Algoritmo 3 Algoritmo 4

5.2 Agrupación: Datos Multivariantes (apéndice)

5.3 Agrupación: Datos Funcionales

Como se mencionó, el objetivo principal del análisis de agrupación es construir grupos homogéneos de observaciones que representen realizaciones de alguna

variable aleatoria Z , y se busca que las observaciones asignadas a los grupos sean similares entre ellas y lo más diferente posible de las observaciones asignadas a otros grupos (?). En el espacio finito dimensional, Z es un vector aleatorio con valores en \mathbb{R}^p , $Z = (Z_1, \dots, Z_p)$, $p \geq 1$. Un caso particular es cuando las variables aleatorias toman valores en un espacio infinito dimensional, usualmente un espacio de funciones definidas en algún conjunto continuo \mathcal{T} ; de esta forma los datos son representados por medio de curvas y la variable aleatoria que corresponde a los datos es un proceso estocástico $Z = \{Z(t), t \in \mathcal{T}\}$. En la actualidad este tipo de datos son más fáciles de observar y se han desarrollado herramientas para poderlos almacenar y procesar (?).

Se conocen cuatro metodologías para la agrupación de datos funcionales, las cuales son:

- Métodos de datos en bruto: Estos métodos consisten en agrupar directamente las curvas sobre la base de sus puntos de evaluación.
- Métodos de filtrado: Estos métodos realizan la aproximación de las curvas en alguna base de funciones y luego realizan la agrupación utilizando los coeficientes de la expansión en la base.
- Métodos adaptativos: Estos métodos consideran que la representación de un dato funcional depende del grupo, y se realiza simultáneamente la reducción de dimensión y la agrupación.
- Métodos con base en distancias: Estos métodos usan algoritmos de agrupación con base en distancias específicas para datos funcionales.

En el presente trabajo se hará uso de los métodos de filtrado y los métodos basados en distancias.

5.3.1 Métodos de datos en bruto

Este primer enfoque, uno de los más sencillos, consiste en utilizar directamente la discretización de las funciones en algunos puntos de tiempo; en este caso no es necesario reconstruir la forma funcional de los datos, debido al gran tamaño de la discretización, por lo cual se requiere utilizar métodos de agrupación para datos vectoriales de alta dimensión.

Existen métodos que realizan agrupación extrayendo las observaciones a partir de datos con ruido, por lo cual se espera pérdidas de información; incluso curvas observadas en diferentes puntos de evaluación pueden no ser tomadas en cuenta por este método. Por otro lado, existen métodos que realizan la agrupación con base en la distribución de probabilidad de las observaciones discretas de las curvas en cada intervalo de alguna partición de algún conjunto continuo \mathcal{T} . En general, estos métodos tienen la ventaja de que no dependen de ninguna asunción sobre la naturaleza de los datos, puesto que la expansión de base no se utiliza (?).

5.3.2 Métodos de filtrado

Este método consiste de dos etapas para la agrupación de datos funcionales. La primera hace referencia al método de filtrado, el cual reduce la dimensión de los datos, pues consiste generalmente en aproximar las curvas utilizando una base finita de funciones; esta base puede ser de tipo B-splines, Fourier o Wavelets; otra técnica de reducción de dimensión es el análisis de componentes principales funcionales (FACP). La segunda etapa consiste en que a partir de los coeficientes de la base de expansión o de los *scores* de los componentes principales, se aplican algoritmos habituales de agrupación para determinar los grupos (?).

Se debe tener en cuenta que el método de filtrado tiene algunos inconvenientes; uno de ellos es que si las curvas son medidas en diferentes puntos de tiempo, la varianza de la estimación de los coeficientes de base es distinto para cada individuo. Para conjuntos de datos dispersos, varios coeficientes de base podrían tener varianza infinita, haciendo imposible obtener estimaciones razonables (?).

5.3.3 Métodos adaptativos

Contrario al método de filtrado en donde las curvas son aproximadas en una base finita de funciones y sus curvas son identificadas mediante los coeficientes de expansión de la base, en este método en lugar de tratar los coeficientes como parámetros, estos son considerados variables aleatorias con una distribución de probabilidad de un grupo específico. Estos métodos se dividen en dos grupos, los que modelan los scores del FPCA y los que modelizan directamente los coeficientes de expansión de una base finita de funciones (?).

5.3.4 Clasificación basados en modelos

En el marco multivariante, este método utiliza la técnica de agrupación basada en modelos; es decir, se asume que las observaciones z_1, \dots, z_n son generadas de acuerdo a una distribución mixta con G componentes. Sea $f_k(z, \theta_k)$ la densidad correspondiente al k -ésimo grupo con parámetro θ_k , y sea $y_i = (y_{i1}, \dots, y_{iG})$ un vector que indica la pertenencia a un grupo para la i -ésima observación donde $y_{ik} = 1$ si z_i es un miembro del k -ésimo grupo y 0 caso contrario. Los y_i son desconocidos y generalmente son tratados con uno de los dos siguientes enfoques: *clasificación por verosimilitud* o *verosimilitud mixta*.

En el primer enfoque, los y_i son vistos como parámetros y el modelo es ajustado para maximizar la verosimilitud:

$$L_c(\theta_1, \dots, \theta_G; y_1, \dots, y_n | z_1, \dots, z_n) = \prod_{i=1}^n f_{y_i}(z_i | \theta_{y_i})$$

donde $f_k(z|\theta_k)$ es una normal multivariante con matriz identidad como covarianza; este enfoque produce una solución por *k-medias* (?).

Por otro lado, el segundo enfoque también considera $f_k(z, \theta_k)$ como una normal multivariante y con matriz identidad como covarianza; sin embargo, puede ocurrir que los miembros de algún grupo pueden ser tratados como datos faltantes, donde y_i es una multinomial con parámetros (π_1, \dots, π_G) y π_k es la probabilidad de que una observación pertenezca al *k-ésimo* grupo; entonces, los parámetros son estimados maximizando lo siguiente:

$$L_M(\theta_1, \dots, \theta_G; \pi_1, \dots, \pi_n | z_1, \dots, z_n) = \prod_{i=1}^n \sum_{k=1}^G \pi_k f_{y_i}(z_i | \theta_{y_k})$$

La principal diferencia entre estos dos enfoques es que en el primero, cada punto se asigna a un único grupo, mientras que en el segundo a cada punto se le asigna a una probabilidad originada en cada grupo y así influye en la estimación de todos los parámetros (?).

5.3.5 Método adaptativo de agrupación usando coeficientes de bases de expansión

El algoritmo **fclust** es el primer algoritmo de agrupación de este tipo y fue propuesto por (?), donde consideran que los coeficientes de la base de expansión tipo spline de las curvas se distribuyen de acuerdo a una mezcla de distribuciones Gaussianas con media específica para cada grupo μ_k y una varianza común Σ ; es decir:

$$\alpha_i \sim \mathcal{N}(\mu_k, \Sigma)$$

A diferencia de los métodos de filtrado, en el cual los coeficientes de la base de expansión son considerados fijos, aquí son considerados como variables aleatorias, que además permite proceder eficientemente con una muestra de curvas dispersa (?).

5.3.6 Método adaptativo de agrupación usando FPCA

Utilizando la noción de densidad de probabilidad para variables aleatorias funcionales desarrollada en (?) y asumiendo una distribución Gaussiana de los componentes principales y definiendo las técnicas de agrupación basadas en el modelo combinado dador por:

$$f_Z^{(q)}(z; \theta) = \sum_{k=1}^{q_k} \pi_k \prod_{j=1}^{q_k} f_{C_{j|Y_k=1}}(c_{jk}(z); \lambda_{jk})$$

donde $\theta = (\pi_k, \lambda_{1k}, \dots, \lambda_{q_k k})_{1 \leq k \leq K}$ son los parámetros del modelo referentes a las proporciones de los grupos y las varianzas de los componentes principales y q_k es el orden de la densidad de probabilidad específica para el grupo k (?). En (?) se desarrolló el modelo denominado *funclust* y los scores $c_{jk}(z)$ de Z son calculados en cada grupo (?). Si el FPCA no fuera calculado para cada grupo, tales métodos corresponden a una agrupación con base en modelos con probabilidades Gaussianas aplicados a los scores del FPCA y podrían ser considerados como métodos de filtrado.

Estos métodos son importantes puesto que asumiendo una distribución de probabilidad de los coeficientes de la base de expansión permiten tomar en cuenta la incertidumbre gracias a la reconstrucción de los datos funcionales (?).

5.3.7 Métodos basados en distancia

Estos métodos también son conocidos como métodos no paramétricos y se dividen en dos categorías, la primera agrupa los métodos que hacen uso de distancias o disimilitudes específicas, mientras que la segunda hace uso de heurísticas.

Métodos que usan distancias

Estos métodos aplican técnicas de agrupación no paramétrica como el método k-medias o el método jerárquico, los cuales consideran distancias específicas o disimilitudes entre curvas. Para este propósito se utiliza una métrica basada en derivadas que mide la proximidad en las curvas z_i y z'_i mediante la siguiente expresión:

$$d_q(z_i, z'_i)^2 = \int \left(z_i^{(q)}(t) - z_{i'}^{(q)}(t) \right)^2 dt$$

donde $z^{(q)}$ denota la q -ésima derivada de z y $d_0(z, 0)$ es la norma \mathcal{L}^2 (?). Si el cálculo de d_0 se lo realiza utilizando las observaciones discretas de la curva, los métodos no paramétricos son equivalentes a los métodos de agrupación de datos brutos. Por otro lado, si el cálculo de d_0 se lo realiza mediante la representación de las curvas en una base finita, estos métodos son equivalentes a los métodos de filtrado (?).

Métodos Heurísticos

Estos métodos proponen usar heurísticas para la agrupación de datos funcionales. En (?) se desarrollaron dos algoritmos de programación dinámica que realizan la agrupación y la estimación de los centros de cada grupo simultáneamente. Por otro lado, en (?) se desarrolló un procedimiento para identificar simultáneamente grupos óptimos de funciones y subespacios óptimos para la agrupación; para esto se define una función objetivo como la suma de las distancias entre las observaciones y sus proyecciones más las distancias entre las proyecciones y las medias de los grupos (?).

5.4 Agrupación: Datos Espaciales

El incremento de los datos espaciales y el uso extendido de bases de este tipo de datos ponen en evidencia la necesidad de generar procesos que permitan la extracción de información espacial; en este sentido se ha desarrollado la minería de datos espaciales, debido a que esta permite descubrir patrones interesantes y previamente desconocidos pero potencialmente útiles (?). A medida que se dispone de más y más datos sobre un área espacial es deseable identificar las diferentes funciones y papeles que desempeñan las distintas partes de esta área; en particular, un objetivo bastante deseable es identificar regiones homogéneas y descubrir sus características, creando resúmenes de alto nivel para los datos en estudio y obteniendo información valiosa para planificadores, científicos, entre otros (?).

Los datos espaciales poseen dos atributos distintos: uno espacial y otro no espacial. Los atributos espaciales de este tipo de datos incluyen información relacionada a la ubicación espacial como longitud, latitud, elevación, forma, entre otros. Por otro lado, los atributos no espaciales se utilizan para describir características como nombre, población, tasa de desempleo, entre otros. Se debe tener en cuenta que la información espacial referente a las relaciones entre los datos, tales como la influencia de los datos con los de sus vecinos, usualmente está implícita; por lo cual, para capturar esta información se hace uso de técnicas o instrumentos que la incorporen (?). Para este propósito se hace uso de las siguientes técnicas:

- Generalizaciones basadas en descubrimiento de conocimiento.
- Métodos de agrupación.
- Medidas de proximidad agregada.
- Reglas de asociación espacial.

El presente trabajo se enfocará en los métodos de agrupación. Este clase de métodos han sido desarrollados por la estadística a lo largo del tiempo, en

primera instancia se trabajó en una dimensión, y se centraron en la búsqueda de medidas de tendencia, variabilidad o dispersión. Luego se extendió para dos dimensiones, que tratan atributos con media o propiedades modales y variabilidad o dispersión, estos atributos son bidimensionales; en el caso de datos espaciales estas cantidades tienen interpretación como la ubicación del centro del grupo y la dispersión del grupo (?).

Los métodos clásicos de agrupación utilizados para particionar un conjunto de datos espaciales dan como resultado grupos que espacialmente están mezclados o carecen de sentido (?); para dar solución a este problema, se han desarrollado métodos que toman en cuenta la información espacial de los datos; la aplicación de estos incluyen la identificación de áreas con características o factores similares y el objeto de estudio depende del campo de aplicación.

Una solución que ha sido utilizada a lo largo de los años es ponderar las disimilitudes entre las muestras mediante la función de variograma; es decir:

$$d_{ij}^* = d_{ij}\gamma(h)$$

donde $\gamma(h)$ es el variograma ajustado al variograma empírico $\hat{\gamma}(h)$; d_{ij} es la disimilitud entre las muestras i y j (?).

Por otro lado, en el caso multivariante, en cuanto a similaridad, se tiene que la ponderación está realizada mediante la función de autocovarianza multivariada; es decir:

$$S_{ij}^{*2} = S_{ij}^2 K(h)$$

Esta ecuación es más robusta en el sentido de que favorece a la formación de grupos que son espacialmente homogéneos.

Por otro lado, en cuanto a disimilitud y utilizando la función de variograma multivariado se tiene que:

$$d_{ij}^{*2} = d_{ij}^2 \Gamma(h)$$

Se debe tener en cuenta que esta última ecuación es un poco más general ya que permite variogramas multivariados sin silla. Ambas ecuaciones se desempeñan adecuadamente cuando el efecto pepita está presente (?).

Una vez que los cálculos son realizados para todos los i y j el resultado es una matriz de disimilitud modificada D^* con elementos d_{ij}^* . Esta matriz se utiliza por una amplia variedad de estrategias de clasificación; por ejemplo, el método jerárquico opera directamente en esta matriz, agrupando primero

aquellos individuos para los cuales d_{ij}^* es la menor y luego la menos similar según las reglas del algoritmo en uso. Por otro lado, la agrupación dinámica es más apropiada para agrupar individuos de poblaciones que no presentan estructura jerárquica (?).

A partir de esta ponderación se pueden aplicar algunos métodos de agrupación que se clasifican de la siguiente forma:

- **Métodos jerárquicos**

Este método agrupa los datos en forma de árboles y generalmente se clasifican en dos tipos de enfoque, uno aglomerativo y otro divisivo.

- **El enfoque aglomerativo:** Utiliza una estrategia ascendente para agrupar los objetos; se fusionan los grupos más pequeños en grupos cada vez más grandes hasta que todos los objetos hayan sido agrupados en uno solo. Los métodos más utilizados son AGNES y DIANA.
- **El enfoque divisivo:** Utiliza una estrategia descendente para agrupar los objetos, en este caso los grupos más grandes se dividen en grupos más pequeños hasta que cada objeto forme un grupo por sí mismo. Los métodos más utilizados son: CURE, BIRC Y CHAMALEON (?).
- **Métodos de particionamiento.**

Los algoritmos de partición organizan los objetos en grupos tales que la dispersión total de cada objeto con respecto a su centro de grupo se minimice. Inicialmente, cada objeto se clasifica como un único grupo; en los siguientes pasos, todos los puntos de datos se reasignan iterativamente a cada grupo hasta que se cumpla un cierto criterio de parada. Bajo este criterio existen los siguientes métodos (?): algoritmo del vecino más cercano, algoritmo K-medioes, algoritmo K-medias.

- **Métodos con base en grillas.**

Los algoritmos de agrupación con base en grillas primero cuantifican el espacio de agrupación en un número limitado de celdas y luego realizan las operaciones necesarias en la estructura de la grilla, las celdas que contienen más de un cierto número de puntos se tratan como densas, la principal ventaja de este enfoque es su rápido tiempo de procesamiento ya que el tiempo es independiente del número de objetos de datos pero depende del número de celdas. Los métodos más utilizados de este tipo son: STRING, Wave Grupo, Cliqué (?).

- **Métodos con base en densidades.**

El método considera los grupos como regiones densas de objetos separadas por regiones de baja densidad, que representan el ruido; a diferencia de los métodos de partición se pueden descubrir grupos de manera arbitraria. Los métodos con base en densidades pueden utilizarse para filtrar el ruido y los valores atípicos. Los métodos más utilizados de este tipo son: DBSCAN y OPTICS (?).

5.5 Agrupación: Datos funcionales con correlación espacial

Como se ha visto a lo largo de este capítulo, los métodos de agrupación se han ido adaptando a las necesidades para tratar los problemas del mundo real. Puesto que se ha ido generalizando esta técnica para trabajar con datos puntuales, luego datos espaciales y datos funcionales; ahora la nueva problemática es extender aun más estos métodos de agrupación para datos funcionales con dependencia espacial. Este es el caso cuando las muestras son funciones observadas en diferentes sitios de una región o cuando estas funciones son observadas sobre un conjunto discreto de tiempo. Si los datos funcionales son espacialmente dependientes, los métodos básicos de agrupación de FDA fallan, ya que la estructura espacial es ignorada dando como resultado grupos en donde las curvas no son similares en forma o comportamiento. Es así que, además de considerar las características individuales en cada una de las curvas, se debe considerar la ubicación espacial de la curva y así agrupar aquellas curvas que son homogéneas no solo con respecto a su forma sino también con respecto a su ubicación espacial (?).

Para agrupar datos funcionales con correlación espacial se han desarrollado métodos como el dinámico y jerárquico, así como varias formas de capturar la dependencia espacial. En (?) se propuso el método jerárquico y en (?) el método dinámico; ambos métodos utilizan una medida de asociación espacial para resaltar y distinguir la dependencia espacial entre las curvas; esta medida está dada por la función trazo-variograma. Ambos métodos tienen en cuenta la dinámica funcional de las curvas en términos de tiempo y sus relaciones espaciales. El método jerárquico toma en cuenta la dinámica funcional a través de la distancia entre las curvas mediante la norma \mathcal{L}^2 , mientras que el método dinámico la considera a través de la distancia euclidiana al cuadrado entre las curvas calculando la función de variograma en cada grupo. No obstante, la dependencia espacial influye de manera diferente en el proceso de agrupación; en el caso jerárquico, la dependencia espacial es considerada por la función de trazo-variograma y en el caso dinámico, esta dependencia es considerada también por el trazo-variograma, pero calculado dentro de cada grupo (?). Por otro lado en (?) presentan un método para este tipo de datos tomando en cuenta la contribución de cada curva a la variabilidad espacial; de esta forma se define una función de dispersión espacial asociada a cada curva y llevan a cabo la agrupación utilizando el algoritmo k-medias; este algoritmo se basa en la

optimización de un criterio de ajuste entre las funciones de dispersión espacial asociadas a cada curva y el representante o centro de los grupos. En el trabajo realizado por (?) se extiende la estrategia de agrupación con base en modelos para datos funcionales con correlación espacial; esta estrategia se enfoca en clasificar curvas espacialmente dependientes y obtener un modelo funcional espacial base para cada grupo; el ajuste de estos modelos implica estimar la función trazo-variograma, por lo que proponen un estimador de variograma de kernel. No se ha definido lo de kernel

Ahora bien, dado que el objetivo de este trabajo es adaptar el método k-medias para datos funcionales espacialmente correlacionados, se hará uso de la metodología planteada en (?).

5.5.1 Método jerárquico para datos funcionales con correlación espacial

Suponiendo que $X_1(t), \dots, X_n(t)$ es una muestra de curvas definidas en $t \in T = [a, b] \subseteq \mathbb{R}$ y que además pertenecen al espacio de Hilbert de funciones cuadrado integrables definidas en $[a, b]$; es decir:

$$L_2(T) = \left\{ f : T \rightarrow \mathbb{R} : \int_T f(t)^2 < \infty \right\}.$$

De igual manera, se asume que las funciones son expandidas en términos de alguna base de funciones como sigue:

$$X_i(t) = \sum_{l=1}^K a_{il} B_l(t) = a_i^T B(t), \quad i = 1, \dots, n$$

El análisis de agrupación jerárquico funcional es desarrollado como en el enfoque clásico, pero considerando la distancia entre las curvas $X_i(t)$ y $X_j(t)$ a través de la norma \mathcal{L}^2 ; es decir:

$$d_{ij} = \sqrt{\int_{[a,b]} (X_i(t) - X_j(t))^2 dt}$$

de donde, utilizando la representación de la curva en la base de funciones, se tiene:

$$d_{ij} = \sqrt{\int_{[a,b]} ((a_i - a_j)^T B(t) B(t)^T (a_i - a_j)) dt}$$

$$d_{ij} = \sqrt{(a_i - a_j)^T W (a_i - a_j)}$$

donde:

$$W = \int_{[a,b]} B(t) B(t)^T dt$$

donde a_i y a_j son vectores de coeficientes de la base para las curvas i y j . Al utilizar bases ortonormales como la de Fourier, la matriz W es la identidad. Una vez calculada la matriz de disimilitud, el proceso estándar aglomerativo o divisivo es aplicado.

Cuando la estructura espacial es tomada en cuenta, y se considera el entorno geoestadístico, la agrupación permite encontrar grupos de sitios cercanos con características similares. Sea $\{Z(s) = (Z_1(s), \dots, Z_m(s)) : s \in D\}$ un proceso espacial m multivariante definido sobre un dominio $D \subseteq \mathcal{R}^d$. Cuando $m = 1$, el proceso de agrupación pondera las disimilitudes d_{ij} entre curvas por:

$$d_{ij}^w = d_{ij} \gamma(h)$$

donde $\gamma(h)$ es el variograma calculado para las distancias entre las ubicaciones i, j . Por otro lado, si $m > 1$ la ponderación es llevada a cabo por:

$$d_{ij}^w = d_{ij} \Gamma(h)$$

donde $\Gamma(h)$ es el variograma multivariado definido por:

$$\Gamma(h) = \frac{1}{2} E[Z(x) - Z(x+h)]^T M [Z(x) - Z(x+h)]$$

con M una matriz simétrica definida positiva usada como métrica. En particular si $M = I$, el variograma multivariado está dado por:

$$\Gamma(h) = \sum_{l=1}^m \frac{1}{2} E[Z_l(x) - Z_l(x+h)]^2$$

$$\Gamma(h) = \sum_l^m \gamma_l(h)$$

donde $\gamma_l(h)$ es el variograma de la l -ésima variable. Una alternativa a la matriz M es la inversa de la matriz de varianzas-covarianzas. En este caso el variograma multivariado es una suma ponderada de los variogramas y variogramas cruzados.

Ahora bien, en el contexto de datos funcionales con correlación espacial, se considera $\{\mathcal{Z}_s(t), s \in D \subseteq \mathbb{R}^d\}$ un proceso aleatorio funcional estacionario isotrópico y $\mathcal{Z}_1(t), \dots, \mathcal{Z}_n(t)$ una realización de este proceso aleatorio observado en n sitios con coordenadas s_1, \dots, s_n respectivamente. Es así que, para realizar el análisis de agrupación la estructura espacial se la considera a través del trazo-variograma, definido por γ , de la siguiente manera:

$$\gamma(h) = \frac{1}{2} E \left[\int_{[a,b]} (X_i(t) - X_j(t))^2 dt \right], \quad h = \|x_i - x_j\|$$

y la ponderación de las distancias entre las curvas está dada por:

$$d_{ij}^w = d_{ij} \gamma(h)$$

Utilizando el método de momentos la estimación de $\gamma(h)$ es la siguiente:

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{s_i, s_j \in N(h)} \int_{[a,b]} (\mathcal{Z}_{s_i}(t) - \mathcal{Z}_{s_j}(t))^2 dt$$

donde $N(h) = \{(s_i, s_j) : \|s_i - s_j\| = h\}$ es el número de elementos distintos en $N(h)$. Luego de estimar el trazo-variograma para una secuencia de P puntos h_p , se ajusta un modelo paramétrico $\gamma(h; \theta)$ a los puntos $(h_p, \hat{\gamma}(h_p))$, $p = 1, \dots, P$; este ajuste se lo hace usualmente mediante mínimos cuadrados ordinarios o ponderados. Este trazo-variograma, es un variograma válido debido a que tiene las mismas propiedades que las de un variograma paramétrico ajustado a partir de un conjunto de datos espaciales univariante.

Por otro lado, una segunda alternativa para la agrupación de este tipo de datos, es estimar los variogramas simples y variogramas cruzados a partir de los coeficientes de las funciones base obtenidos en la suavización.

5.5. AGRUPACIÓN: DATOS FUNCIONALES CON CORRELACIÓN ESPACIAL 111

Asumiendo que la curva para cada ubicación de muestra $i = 1, \dots, n$ se la ha expandido utilizando los coeficientes de las funciones base, se tiene la siguiente matriz de coeficientes:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nK} \end{pmatrix}_{(nxK)}$$

que forman una realización de un campo aleatorio de K-variables $\{A(s) = (A_1(s), \dots, A_K(s)) : s \in D \subseteq \mathbb{R}^d\}$ con $E(A_i(t)) = v_i$, y la matriz de variogramas y variogramas cruzados es de la siguiente forma:

$$\Upsilon(h) = \begin{pmatrix} \gamma_{11}(h) & \gamma_{12}(h) & \cdots & \gamma_{1K}(h) \\ \gamma_{21}(h) & \gamma_{22}(h) & \cdots & \gamma_{2K}(h) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{K1}(h) & \gamma_{K2}(h) & \cdots & \gamma_{KK}(h) \end{pmatrix}_{(KxK)}$$

donde $\gamma_{lq} = \frac{1}{2}E[A_l(s_l) - A_q(s_j)^2]$, $l, q = 1, \dots, K$, $h = ||s_i - s_j||$; para estimar la matriz anterior se hace uso del modelo lineal de coregionalización (LMC), este método permite modelizar los variogramas y variogramas cruzados de dos o más variables de modo que la varianza de cualquier combinación lineal posible de estas variables es siempre positiva. A partir del modelo estimado de LMC, se calcula el variograma multivariado utilizando los coeficientes de las funciones base.

$$\gamma_{lq}(h) = \sum_{k=1}^K b_{lq}^k g_k(h) \quad \forall l, q$$

donde los b_{lq}^k son las sillan o pendientes de los $g_k(h)$ y:

$$\frac{1}{2}E[A_{kl}(s) - A_{kl}(s+h)][A_{k'l'}(s) - A_{k'l'}(s+h)] = \begin{cases} g_l(h) & \text{Si } k = k', l = l' \\ 0 & \text{En otro caso} \end{cases}$$

y matricialmente se tiene que:

$$\Gamma(h) = \sum_{l=1}^K [b_{ij}^l] \gamma_l(h)$$

y finalmente se tiene que:

$$d_{ij}^w = d_{ij}\Gamma(h)$$

5.5.2 Método k-medias para datos funcionales con correlación espacial

Chapter 6

Aplicaciones

6.1 Aplicación 1

6.2 Aplicación 2

6.3 Aplicación 3

6.4 Aplicación 4

Chapter 7

Apéndice

7.1 Introducción a la simulación geoestadística condicional y no condicional

7.2 Estudio de simulación (referencia al artículo)

7.3 Índices

7.3.1 Selección de número de grupos

Índices basados en suma de cuadrados

Índices clásicos Índices funcionales

7.3.2 Validación de grupos

7.3.2.1 Correlación temporal

7.3.2.2 Índices de correlación espacial

Índice de Moran

Índice de Geary

Bibliography