



UNIVERSIDAD DE BUENOS AIRES  
Facultad de Ciencias Exactas y Naturales  
Departamento de Matemática

## **Tiempo de salida de difusiones y función de profundidad**

Tesis presentada para optar al título de Doctor de la Universidad de Buenos Aires en el área  
Ciencias Matemáticas

**Lic. Guido Giussani**

Directores de tesis: Pablo Ferrari y Ricardo Fraiman.

Consejero de estudios: Pablo Ferrari

Lugar de trabajo: Universidad de San Andrés

Buenos Aires, 2021



## Tiempo de salida de difusiones y función de profundidad

### Resumen

Una de las herramientas más utilizadas para el análisis e inferencia no paramétrica multivariada es la *función de profundidad*. Su principal uso es el de dar un ordenamiento de los puntos en el soporte de la distribución, “del centro hacia afuera”.

La propuesta es utilizar difusiones de Itô para definir una nueva función de profundidad. Dada una distribución  $F$  con soporte  $D$ , la profundidad en un punto  $x$  es la transformada de Laplace del tiempo de salida de  $D$  de una difusión de Itô cuyos coeficientes dependen de la densidad.

Esta *profundidad de salida* ( $\mathcal{PS}$ ), es la solución de un problema de frontera de tipo Dirichlet 0 para un operador diferencial de segundo orden uniformemente elíptico. Las herramientas empleadas para mostrar sus propiedades son usuales en ecuaciones en derivadas parciales.

Así como la profundidad de Tukey [58], entre otras,  $\mathcal{PS}$  está inspirada en su versión empírica; basada en el tiempo de salida de un paseo aleatorio en un grafo aleatorio, un objeto clásico en probabilidad. El grafo utilizado en la versión empírica tiene como vértices a los puntos de una muestra y aristas definidas intrínsecamente, dependiendo de las distancias entre los puntos; el paseo aleatorio se mueve entre vecinos del grafo. La consistencia de la profundidad empírica es una consecuencia de un principio de invarianza, es decir, de la convergencia del paseo aleatorio reescalado a la difusión, cuando el número de puntos de la muestra crece a infinito.

La profundidad  $\mathcal{PS}$  se puede aplicar a distribuciones con soportes no convexos. La versión empírica está bien definida en grafos genéricos, no asociados a una distribución y en espacios no Euclidianos, por ejemplo, sin un sistema de coordenadas de referencia.

Para mostrar una de las posibles aplicaciones, se realiza un estudio de clasificación para datos simulados y reales obteniéndose resultados superlativos en el caso de que las muestras aleatorias provengan de distribuciones con soporte no convexo.

*Palabras Claves:* Función de profundidad, Difusiones de Itô, Paseos Aleatorios, Tiempo de salida, Datos Multivariados.

## Exit time of diffusions and depth function

**Abstract**

One of the most used tools for multivariate nonparametric analysis and inference is the *depth function*. Its main use is to give an ordering of the points in the support of the distribution, “ from the center outwards ”.

The proposal is to use Itô diffusions to define a new depth function. Given a distribution  $F$  with support  $D$ , the depth at a point  $x$  is the Laplace transform of the exit time from  $D$  of a Itô diffusion whose coefficients depend on the density.

This *exit depth* ( $\mathcal{PS}$ ), is the solution to a 0 Dirichlet boundary problem for an uniformly elliptical second-order differential operator. The tools used to show their properties are common in partial derivative equations.

As well as the Tukey depth [58], among others,  $\mathcal{PS}$  is inspired by its empirical version; based on the exit time of a random walk on a random graph, a classic object in probability. The graph used in the empirical version has as vertices the points of a sample and intrinsically defined edges, depending on the distances between the points; the random walk moves between neighbors of the graph. The consistency of the empirical depth is a consequence of an invariance principle, that is, of the convergence of the rescaled random walk to the diffusion, when the number of points in the sample grows to infinity.

The depth  $\mathcal{PS}$  can be applied to distributions with non-convex supports. The empirical version is well defined in generic graphs, not associated with a distribution and in non-Euclidean spaces, for example, without a reference coordinate system.

To show one of the possible applications, a classification study is carried out for simulated and real data, obtaining superlative results in the case that the random samples come from distributions with non-convex support.

*Key Words:* Depth functions, Itô Diffusions, Random Walks, Exit time, Multivariate Data.

*A mis viejos*

## **Agradecimientos**

A mis directores Ricardo y Pablo, por darme la oportunidad y guiarme en este proceso.

A Marcela, por toda su ayuda, imposible resumirla en una frase. Sin su apoyo constante y consejos esta tesis no sería realidad.

A Lucas, Claudio, Marisol y todos los compañeros de UdeSA por enseñarme tantas cosas y acompañarme todos estos años.

A mis amigos, los que siempre están, los que ansiaban que esta tesis este terminada tanto como yo.

A mi familia, sostén ineludible, su apoyo incondicional es el motor que me da energía.

A Vicka, por la S.R.A.M.T.D. y no hacerle caso a su amiga Anto ese día.

A Deborah, auxiliar en la subcomisión de doctorado y parte esencial de toda tesis, sin su predisposición y ayuda con los trámites todos estos años nunca habría terminado.

Al grupo de proba por todo lo que me enseñaron y tantas horas de discusiones, especialmente a Patu y Facundo quienes ayudaron a pensar los problemas y encontrar un camino a seguir.

A Jaime San Martín, por su ayuda para obtener el resultado de consistencia largamente esquivo.

A la Agencia I + D + i y CONICET por las becas y a UdeSA por brindarme un lugar de trabajo en el que siempre me sentí cómodo.

# Índice general

<b>Introducción</b>	<b>9</b>
<b>1. Funciones de Profundidad</b>	<b>17</b>
1.1. Simetrías y Profundidades clásicas . . . . .	17
1.2. Aplicaciones y extensiones . . . . .	21
<b>2. Profundidad de salida, difusiones en <math>\mathbb{R}^d</math></b>	<b>23</b>
2.1. Difusiones . . . . .	23
2.2. Definición de $\mathcal{PS}$ . . . . .	25
2.2.1. Propiedades de $\mathcal{PS}$ . . . . .	28
2.2.2. Extensión a distribuciones con soporte no acotado . . . . .	35
<b>3. Profundidad de salida en Grafos</b>	<b>37</b>
3.1. Definición de $\mathcal{PS}$ en grafo $\mathcal{G}$ . . . . .	37
3.1.1. $\mathcal{PS}_{\mathcal{G}}$ y detección de comunidades en grafos . . . . .	39
3.2. Grafos inducidos por puntos en $\mathbb{R}^d$ . . . . .	40
3.2.1. Grafos inducidos por procesos puntuales . . . . .	40
3.2.2. Marcado de frontera . . . . .	42
3.2.3. Cálculo de $\mathcal{PS}_{\mathcal{G}_n}$ para una muestra aleatoria y extensión a $\mathbb{R}^d$ . . . . .	45
<b>4. Resultados asintóticos y <math>\mathcal{PS}</math> para <math>d = 1</math></b>	<b>47</b>
4.1. Consistencia . . . . .	47
4.1.1. Escalamiento temporal y coeficientes de $Y_t$ . . . . .	47
4.1.2. Consistencia para grafos de vecindad generalizada . . . . .	51
4.1.3. Consistencia para Delaunay en Poisson . . . . .	54
4.2. $\mathcal{PS}$ en dimensión 1 . . . . .	55

<b>5. Simulaciones y aplicaciones</b>	<b>63</b>
5.1. Comparación de conjuntos de nivel . . . . .	64
5.2. DD-plots y clasificación . . . . .	68
5.2.1. Datos simulados . . . . .	69
5.2.2. Datos reales . . . . .	74
5.3. Comentarios y futuras líneas de investigación . . . . .	76
<b>Bibliografía</b>	<b>79</b>

# Introducción

El espíritu de esta tesis es el de dar una nueva mirada a las funciones de profundidad. A través de la aplicación de técnicas, procedimientos e ideas de otras ramas de la matemática y tratando de explotar la interrelación que existe entre las diferentes áreas, se espera propiciar un nuevo punto de vista de este concepto tan importante dentro de la estadística.

A modo de introducción comenzaremos haciendo un repaso de la definición clásica de profundidad junto con los conceptos que permiten comprender el problema que vamos a tratar; presentaremos la nueva idea de manera informal a partir de la motivación empírica que tuvo lugar al principio de esta tesis y algunos resultados que se relacionan con los obtenidos.

## Profundidades.

Los conceptos mediana y percentiles permiten ordenar una muestra aleatoria univariada y conocer características de la distribución de la cuál provienen, por ejemplo, a través del uso de test no paramétricos basados en rangos. La noción de función de profundidad es una herramienta que generaliza estas ideas a varias dimensiones. El primer limitante obvio es que en más de una dimensión no tiene sentido un orden “de menor a mayor”, sino que se debe comenzar a hablar de un orden “del centro hacia afuera”.

Para poder hablar de *centro* y de *orden* en una distribución multivariada y de las herramientas que permiten detectarlos y definirlos es preciso pensar primero en simetrías en vectores aleatorios. Comenzaremos presentando los conceptos de una manera informal que nos permita avanzar en la comprensión del tema, dejando para el primer capítulo las definiciones formales. Tomando como base Serfling [51], en creciente orden de generalidad, es posible decir que un vector multivariado  $\mathbf{X}$  tiene simetría respecto de un punto  $\boldsymbol{\theta} \in \mathbb{R}^d$  del tipo *esférica* si toda rotación de  $\mathbf{X}$  respecto de  $\boldsymbol{\theta}$  no altera su distribución; simetría del tipo *elíptica* si  $\mathbf{X}$  tiene igual distribución que una transformación afín de un vector aleatorio con simetría esférica  $\mathbf{Y}$ ; simetría del tipo *central* si la distribución de  $\mathbf{X}$  es simétrica respecto de una reflexión en  $\boldsymbol{\theta}$  y simetría del tipo *angular* si la variable  $(\mathbf{X} - \boldsymbol{\theta})/\|\mathbf{X} - \boldsymbol{\theta}\|$  es centralmente simétrica respecto del origen.

En todos estos casos,  $\boldsymbol{\theta}$  se considera el *centro* de la distribución y a partir de éste es que se piensa en un *orden*. En Zuo y Serfling [67] se establece a manera de definición las propie-

dades fundamentales a cumplir por una función de profundidad. Denotamos  $\mathcal{F}$  el conjunto de distribuciones en los Borelianos de  $\mathbb{R}^d$  y  $F_{\mathbf{X}}$  la distribución del vector aleatorio  $\mathbf{X}$ .

**Definición de Profundidad** Sea  $D(\cdot; \cdot) : \mathbb{R}^d \times \mathcal{F} \rightarrow \mathbb{R}$  acotada, no negativa, con las siguientes propiedades:

- P1** Invarianza afín:  $D(\mathbf{A}x + b; F_{\mathbf{A}\mathbf{X}+b}) = D(x; F_{\mathbf{X}})$  vale para todo vector aleatorio  $\mathbf{X} \in \mathbb{R}^d$ , toda matriz  $d \times d$  no singular  $\mathbf{A}$  y todo vector  $d$ -dimensional  $b$ .
- P2** Máximo en el centro:  $D(\boldsymbol{\theta}; F) = \sup_{x \in \mathbb{R}^d} D(x; F)$  vale para toda  $F \in \mathcal{F}$  con centro  $\boldsymbol{\theta}$ .
- P3** Monotonía desde el máximo: para  $F \in \mathcal{F}$  con punto más profundo  $\boldsymbol{\theta}$ ,  $\forall \lambda \in [0, 1]$ , se tiene que  $D(x; F) \leq D(\boldsymbol{\theta} + \lambda(x - \boldsymbol{\theta}); F)$ .
- P4** Límite 0:  $D(x; F) \rightarrow 0$  si  $\|x\| \rightarrow \infty$  para  $F \in \mathcal{F}$ .

Entonces  $D(\cdot; \cdot)$  es una función de profundidad.

Otras propiedades deseables son:

- P5** Continuidad en  $x$ . Sea  $F \in \mathcal{F}$ ,  $D(\cdot; F) : \mathbb{R}^d \rightarrow \mathbb{R}$  es continua.
- P6** Continuidad en  $F$ . Sea  $x \in \mathbb{R}^d$ ,  $D(x; \cdot) : \mathcal{F} \rightarrow \mathbb{R}$  es continua.

Algunas profundidades cumplen una versión un poco más laxa de **P1**:

- P1'** Invarianza por transformaciones ortogonales y traslaciones:  $D(\mathbf{A}x+b; F_{\mathbf{A}\mathbf{X}+b}) = D(x; F_{\mathbf{X}})$  vale para todo vector aleatorio  $\mathbf{X} \in \mathbb{R}^d$ , toda matriz  $d \times d$  ortogonal  $\mathbf{A}$  y todo vector  $d$ -dimensional  $b$ .

Generalmente, una versión de la función de profundidad  $D(\cdot, F)$  para aplicar a una muestra aleatoria se obtiene reemplazando  $F$  por una medida empírica  $F_n$  y así por el método “plug in” queda definida la versión empírica  $D_n(x) = D(x, F_n)$ . En el caso de la nueva definición de profundidad que daremos en esta tesis veremos que la versión poblacional dependerá de un tiempo de parada de una difusión, mientras que la versión empírica será calculada a partir de un tiempo de parada de paseos aleatorios definidos en grafos; estos paseos aleatorios pueden llegar a pensarse informalmente como una “muestra empírica” de las trayectorias de las difusiones que a través de un escalamiento temporal convergen a las trayectorias continuas de la difusión.

La existencia de una versión empírica y una versión poblacional para una función de profundidad plantea la necesidad de que sea posible obtener también **consistencia** entre las versiones, es decir, que para  $F \in \mathcal{F}$  se cumpla la convergencia

$$\sup_x |D_n(x) - D(x; F)| \rightarrow 0, \quad n \rightarrow \infty, \quad \text{con probabilidad 1.}$$

Esta propiedad permite por ejemplo que en He y Wang [28] se demuestre la convergencia de los contornos (líneas de nivel) empíricos  $\{x \in \mathbb{R}^d : D_n(x) = c\}$  a los poblacionales.

## Grafos, laplacianos, cluster, LTHT y resultados relacionados.

Antes de presentar la nueva función de profundidad de manera informal, realizamos un recorrido por los resultados que se relacionan con los obtenidos aquí y que sirven de base o punto de comparación para los aportes de este trabajo.

Para un grafo  $G = (V, E)$  donde  $V = \{v_1, \dots, v_n\}$  es un conjunto de vértices y  $E = \{(v_i, v_j) : v_i, v_j \in V\}$  es el conjunto de aristas, existe una matriz  $A = \{a_{ij}\}$  que permite su representación y contiene toda la información generalmente denominada matriz de adyacencia (o incidencia), donde  $a_{ij} = 1$  si existe arista  $(v_i, v_j) \in E$  o  $a_{ij} = 0$  en caso contrario. En la teoría espectral de grafos, una de las matrices asociadas a grafos más estudiada es la denominada laplaciano del grafo  $L = D - A$ , donde  $D$  es la matriz diagonal  $d_{ii} = \sum_j a_{ij}$ . A la matriz  $L$  se le suele denominar *laplaciano no normalizado* o *laplaciano combinatorio*, mientras que existen dos maneras de definir el *laplaciano normalizado*:

$$\tilde{L} = D^{-1/2} L D^{-1/2} = I - D^{-1/2} A D^{-1/2}, \quad \hat{L} = D^{-1} L = I - D^{-1} A,$$

donde  $I$  es la matriz identidad y todo puede ser generalizado a grafos con pesos en las aristas asignando a la matriz  $A$  valores no negativos.

Dada una muestra aleatoria en el espacio euclídeo  $\mathbb{R}^d$ , es posible construir grafos de vecindad que pueden ser vistos como aproximación del soporte de la distribución. El laplaciano de estos grafos es usado en aplicaciones de estadística no paramétrica y machine learning como aprendizaje supervisado, reducción de dimensiones y clustering. Los métodos utilizados se basan en la idea de que existe convergencia del laplaciano del grafo (i.e. discreto) al laplaciano continuo cuando el tamaño de muestra va a infinito y la vecindad del grafo se achica. En Hein, Audibert y Von Luxburg [29], [30] prueban rigurosamente estos resultados (para distribuciones con soporte en la dimensión del espacio o menor), basados en grafos generados por kernels suaves; resultados de consistencia para algoritmos de clusterización basados en el espectro del laplaciano del grafo es posible encontrar en Von Luxburg, Nekin y Bousquet [59]; mientras que en Ting, Huang y Jordan [57] los resultados de convergencia se extienden a kernels menos restrictivos (entre ellos el que genera el grafo de vecinos más cercanos Knn) e introducen un marco más general para analizar los grafos que pueden utilizarse comenzando a emplear el punto de vista de la caracterización de difusiones como procesos estocásticos a través de la deriva y el término difusivo para poder mostrar consistencia.

Por otro lado, el concepto de paseo aleatorio tiene una vasta historia en el área de Probabilidad y también es posible encontrar usos en otras áreas como redes eléctricas y para resolver ecuaciones en derivadas parciales en análisis numérico después de discretizar el espacio. Zhou [64] define una distancia entre vértices basada en el concepto de un Movimiento Browniano (paseo aleatorio en grafo) que permite abordar el problema de encontrar la cantidad de clusters que forman una red biológica o social. Esta distancia es utilizada para definir vértices “atractores” de importancia local y global lo que induce comunidades dentro de la red global. Basado en esta distancia, en Zhou [63] se genera un índice de diferenciación entre

vértices vecinos y un algoritmo para partir un grafo en comunidades que están jerárquicamente organizadas.

En Smith et al. [53] desarrollan otra idea basada en aplicar paseos aleatorios en grafos para detectar pequeñas redes encubiertas dentro de grafos mucho mayores, en este caso a través de una aproximación Bayesiana teniendo en cuenta que  $P = D^{-1}A$  puede ser vista como la matriz de salto de un paseo aleatorio, interpretando  $p_{ij}$  como la probabilidad de que en el siguiente paso el paseo aleatorio en estado  $i$  salte al estado  $j$ ; además establecen una relación de sus resultados con la clusterización espectral basada en el laplaciano del grafo. En el Capítulo 3 veremos que la nueva definición de profundidad, que puede ser aplicada en grafos, tiene una estrecha relación con este procedimiento de detección de comunidades en grafos.

En Hashimoto, Sun y Jaakkola [26] introducen una técnica para analizar paseos aleatorios en grafos utilizando cálculo estocástico, estableciendo una correspondencia entre los tiempos de parada de un Movimiento Browniano y sus análogos en los grafos. A través del concepto de Transformada de Laplace del “hitting time” (LTHT) los autores definen una nueva métrica que demuestran es consistente con respecto a la métrica subyacente utilizada para crear el grafo y conserva la tendencia de agrupamiento de los puntos; muestran además que sus resultados pueden ser utilizados para realizar análisis de cluster, probando también robustez frente a la introducción de aristas añadidas que no tienen que ver con las características del grafo. Destacan que la misma métrica, con variantes y sin la correspondiente relación con el Movimiento Browniano, es utilizada como métrica de similaridad en redes colaborativas en Yazdani [61], búsqueda de distancia robusta mínima entre vértices en Yen et al. [62] y es la aplicada en detección de subgrafos en Smith et al. [53].

## Contribuciones de la tesis.

Hasta aquí hemos hablado de funciones de profundidad (buscan el centro de distribuciones y un orden de datos) y de que grafos (y matrices asociadas a estos) junto con el concepto de paseo aleatorio en grafos han sido utilizados principalmente en métodos de clusterización. Presentaremos de manera informal, a partir de la idea que inspiró esta tesis, cómo vamos a conseguir una función de profundidad a través de paseos aleatorios en grafos. Usaremos la *Teselación de Voronoi* y *Triangulación de Delaunay*, definidas en la Sección 3.2.1.

Comenzamos fijando una distribución  $F$ , por ejemplo la uniforme en el círculo unitario en  $\mathbb{R}^2$ . Sea  $\mathcal{X}_n = \{X_1, \dots, X_n\}$  muestra aleatoria de puntos,  $X_i \sim F$ . Queremos marcar algunos puntos como los “menos profundos” a partir de su relación con el resto de la muestra. Una manera de hacerlo es elegir los puntos para los cuáles su celda de Voronoi es no acotada en una ventana suficientemente grande que contiene a toda la muestra aleatoria (Figura1). Estos puntos serán considerados la “frontera” del grafo con vértices  $\mathcal{X}_n$ .

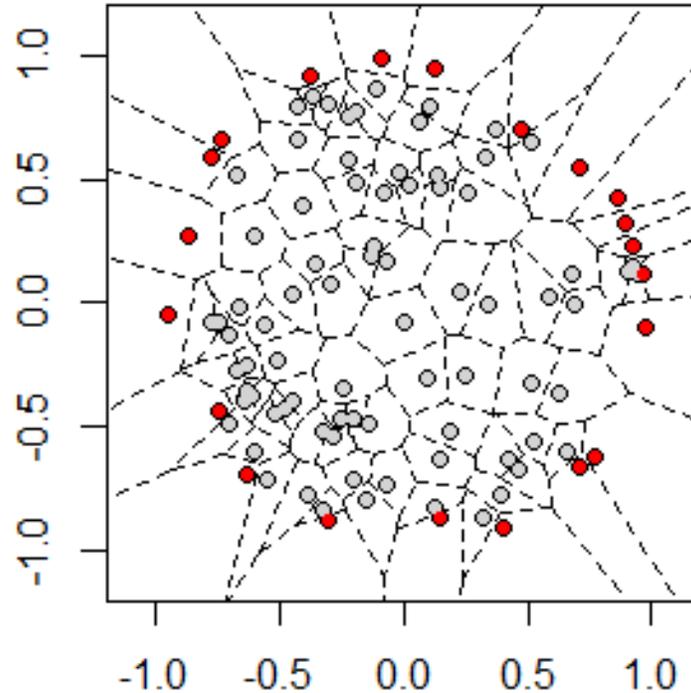


Figura 1: Los puntos en rojo son los marcados como “frontera” de la muestra  $\mathcal{X}_n$ .

Luego, es posible utilizar el grafo dual a la Teselación de Voronoi, que se obtiene uniendo con una arista los puntos cuyas celdas comparten una cara para obtener la Triangulación de Delaunay de  $\mathcal{X}_n$  con un subconjunto de vértices marcados como “frontera” (Figura 2).

Ahora podemos pensar que tenemos un espacio de estados (los puntos de la muestra) en el que realizamos un paseo aleatorio simple (es decir, que se elige el nuevo vértice a visitar uniformemente entre los adyacentes en el grafo) comenzando en cada uno de los puntos, a tiempo discreto o continuo. La idea principal entonces para ver qué puntos son “mas centrales” es que si se calcula el tiempo esperado de arribo a cualquiera de los vértices marcados como frontera (“expected hitting time”), es factible suponer que puntos más alejados de la frontera tengan una esperanza más grande, mientras que puntos más cerca de la frontera tengan esperanza menor (con esperanza 0 para los puntos marcados).

Una vez que tenemos un procedimiento empírico para ordenar una muestra de puntos, nos preguntamos cuál va a ser el límite de esta función cuando el número de puntos vaya a infinito. Resultados clásicos de probabilidad indican que el paseo aleatorio en puntos distribuidos uniformemente en el espacio, después de un escalamiento espacio-temporal, se aproxima a un Movimiento Browniano, mientras que la convergencia de la distribución empírica a la distribución uniforme en el círculo indica que los puntos marcados como frontera aproximen la circunferencia  $\{x \in \mathbb{R}^2 : \|x\| = 1\}$ . Así, deberíamos pensar a la definición poblacional de

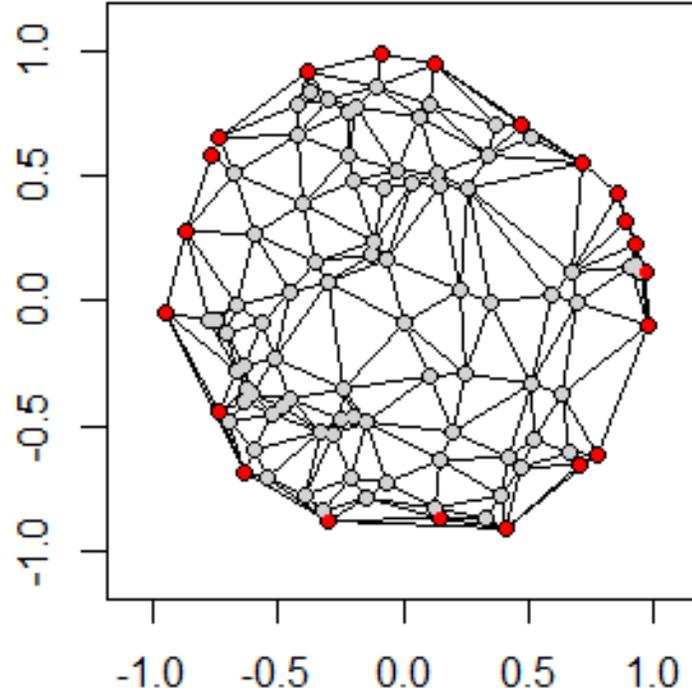


Figura 2: Triangulación de Delaunay de los puntos  $\mathcal{X}_n$  de la Figura 1 con los vértices de la frontera marcados en rojo.

la función de profundidad como la esperanza del tiempo de salida del círculo unitario de un Movimiento Browniano que comienza en cada punto  $x$ .

Generalizando, sería posible trabajar calculando la esperanza de cualquier función monótona  $\varphi$  del tiempo de salida y en lugar de utilizar el Movimiento Browniano trabajar con difusiones cuyos coeficientes tengan en cuenta la distribución de la cual provienen los puntos.

Ahora sí, dada una distribución  $F$  en  $\mathbb{R}^d$  con soporte  $D$  acotado y densidad  $f$ , podemos plantear los objetivos de esta tesis de forma general de la siguiente manera:

1. *empírico*: dada una muestra aleatoria  $\mathcal{X}_n$ , queremos utilizar procedimientos para fijar algún subconjunto  $C \subset \mathcal{X}_n$  como frontera (típicamente uno pensaría en los extremos de la cápsula convexa de  $\mathcal{X}_n$  pero queremos explorar otras opciones que nos permitan trabajar con soportes no convexos); luego definir grafos  $\mathcal{G}_n = (\mathcal{X}_n, E)$  que ayuden a identificar un orden de centralidad a través de alguna función  $\varphi$  aplicada al tiempo de llegada del paseo aleatorio al conjunto de vértices marcado.
2. *poblacional*: definir una difusión de Itô  $d$ -dimensional,  $dY_t = b(Y_t) dt + \sigma(Y_t) dB_t$  donde los parámetros  $b$  y  $\sigma$  sean funciones de  $f$  y tal que si  $\tau = \inf\{t > 0 : Y_t \notin D\}$  sea posible definir una función de profundidad con las propiedades descritas anteriormente

aplicando  $\mathbb{E}(\varphi(\tau))$ .

3. *consistencia*: encontrar condiciones para los grafos utilizados, los coeficientes de la difusión  $Y_t$  y la función  $\varphi$  que impliquen que la versión empírica resulte consistente con la versión poblacional.

En resumen, en esta tesis se define una nueva función de profundidad para datos multivariados a través del tiempo de parada de una difusión. Una de las propiedades que explotaremos del concepto de LTHT para difusiones comentado anteriormente es que puede ser presentada en forma cerrada a partir del uso de la fórmula de Feynman-Kac como solución de una ecuación en derivadas parciales con frontera del tipo Dirichlet como veremos en la Sección 2.2.

Esta nueva definición de profundidad difiere en gran medida de las clásicas de la literatura tanto en su concepción probabilística como en las herramientas utilizadas y conceptos relacionados. No obstante, mantiene las mismas propiedades descritas anteriormente en el caso en que la distribución cumple ciertos criterios de simetría; además, es posible utilizarla tanto para distribuciones con soporte no convexo como en grafos con vértices que no provengan de una distribución o que no tengan un sistema de coordenadas “latente” y se establece la relación con un procedimiento de detección de comunidades en grafos. En el caso empírico es posible calcularla tanto a partir de simulaciones a través de montecarlo simulando caminos aleatorios, como a partir de la inversión de una matriz como veremos en la Sección 3.1.

En términos asintóticos, se demuestra consistencia para un amplio espectro de grafos basados en muestras aleatorias en varias dimensiones y se prueba que en dimensión 1 recobra los mismos conceptos clásicos de todas las profundidades. Para la consistencia usaremos un principio de invarianza presentado en Hashimoto et al. [26], donde se demuestra la convergencia en el espacio de Skorokhod (conjunto de funciones continuas a izquierda con límite a derecha) de los paseos aleatorios en grafos a difusiones de Itô. Este resultado implica que funciones acotadas de toda la trayectoria del paseo aleatorio en el grafo, apropiadamente escalado, convergen a sus versiones poblacionales.

Además, se obtienen buenos resultados en un estudio de simulación destacándose en casos donde la distribución tiene soporte no convexo y se aplica la nueva definición a clasificadores basados en profundidades para mostrar que mejora el clasificador en algunos casos.

La tesis está organizada de la siguiente manera:

En el Capítulo 1 ampliamos el análisis de simetrías en distribuciones, presentamos las definiciones de las funciones de profundidad clásicas de la literatura en su versión poblacional y empírica y vemos algunas de las aplicaciones y extensiones del concepto de profundidad.

En el Capítulo 2 comenzamos haciendo un repaso de difusiones de Itô, presentamos formalmente la nueva profundidad para distribuciones en  $\mathbb{R}^d$  y probamos sus propiedades.

En el Capítulo 3 vemos la definición de profundidad para grafos generales y hacemos un análisis de aplicación a grafos con vértices que provienen de una muestra aleatoria de una

distribución multivariada y un resumen de las posibles técnicas de marcado de frontera.

En el Capítulo 4 se encuentran los resultados asintóticos donde mostramos consistencia entre las definiciones empírica y poblacional y se desarrolla el caso de dimensión 1.

En el Capítulo 5 hacemos un análisis de simulación comparando los conjuntos de nivel con otras profundidades y aplicamos a clasificadores con datos simulados y reales para mostrar la performance de esta nueva función de profundidad. Finalmente, presentamos a modo de conclusión comentarios y posibles líneas de investigación futuras que surgieron a lo largo de la realización de este trabajo y escapan al objetivo principal de la tesis.

# Capítulo 1

## Funciones de Profundidad

### 1.1. Simetrías y Profundidades clásicas

Comenzaremos presentando las definiciones formales de las nociones de simetría comentadas en la Introducción, a partir de la más restrictiva.

Sea  $\theta \in \mathbb{R}^d$ , un vector aleatorio  $\mathbf{X}$  tiene distribución *esféricamente simétrica* respecto de  $\theta$  si para toda matriz ortogonal  $\mathbf{A}$

$$\mathbf{X} - \theta \stackrel{d}{=} \mathbf{A}(\theta - \mathbf{X}),$$

es decir, si una rotación de  $\mathbf{X}$  sobre  $\theta$  no cambia su distribución. En este caso, la densidad  $f$  del vector  $\mathbf{X}$  puede escribirse  $f(x) = g((x - \theta)^T(x - \theta))$  para alguna función escalar no negativa  $g$ , lo que permite caracterizar los contornos de igual densidad con círculos para este tipo de distribuciones.

Generalizando,  $\mathbf{X}$  es *elípticamente simétrico* respecto de  $\theta$  si

$$\mathbf{X} \stackrel{d}{=} \mathbf{A}\mathbf{Y},$$

donde  $\mathbf{Y}$  es un vector aleatorio esféricamente simétrico respecto de  $\theta$  y  $\mathbf{A} \in \mathcal{M}^{k \times d}$  es una matriz que satisface  $\mathbf{A}^T \mathbf{A} = \Sigma$  tal que  $\text{rango}(\Sigma) \leq d$ . De manera análoga al caso anterior, es posible caracterizar a la densidad  $f$  del vector  $\mathbf{X}$ ,  $f(x) = |\Sigma|^{-1/2} g((x - \theta)^T \Sigma^{-1} (x - \theta))$  para alguna función escalar no negativa  $g$ ; en este caso los contornos de igual densidad son elipses.

Continuando con la generalización, se dice que  $\mathbf{X}$  tiene distribución *centralmente simétrica* respecto de  $\theta$  si

$$\mathbf{X} - \theta \stackrel{d}{=} \theta - \mathbf{X},$$

es decir, si una reflexión de  $\mathbf{X}$  sobre  $\theta$  no cambia su distribución; esto implica que su densidad satisface  $f(x - \theta) = f(\theta - x)$ . Esta noción de simetría es considerada la extensión no paramétrica más directa de la simetría univariada.

Llegamos así al último paso en la generalización del concepto dado por Liu [36]

**Definición 1.1** (Simetría angular). *El vector aleatorio  $X$  tiene simetría angular respecto de  $\theta$  si*

$$\frac{X - \theta}{\|X - \theta\|} \stackrel{d}{=} \frac{\theta - X}{\|X - \theta\|},$$

es decir, la variable  $(X - \theta)/\|X - \theta\|$  es centralmente simétrica respecto del origen.

Debemos tener en cuenta que en Zuo y Serfling [68] dan la definición “más general” para el concepto de simetría en la literatura que es la siguiente.

**Definición 1.2** (Simetría de semiespacio). *El vector aleatorio  $X$  tiene simetría de semiespacio respecto de  $\theta$  si*

$$\mathbb{P}(X \in H) \geq 1/2 \text{ para todo semiespacio cerrado } H \text{ con } \theta \text{ en el borde.}$$

Sin embargo, en esa misma publicación los autores muestran en la Observación 2.1(i) y Teorema 2.6 que si la distribución es absolutamente continua, el cuál va a ser nuestro caso de estudio, entonces la simetría de semiespacio es equivalente a la simetría angular.

Para una distribución que cumple alguno de los tipos de simetría, la noción de función de profundidad generaliza entonces el estudio de medianas, estadísticos de orden y rangos a varias dimensiones.

Existen varias definiciones de funciones de profundidad, siendo las más destacadas en la literatura las que daremos a continuación. Para un punto  $x \in \mathbb{R}^d$  respecto de una medida  $\mathbb{P} \in \mathbb{R}^d$  y una muestra aleatoria  $\{X_1, \dots, X_n\}$  se definen:

- la *profundidad del semiespacio*, “*halfspace depth*”, *HS*, Tukey [58]:

$$HS(x, \mathbb{P}) = \inf\{\mathbb{P}(H) : H \text{ semiespacio cerrado, } x \in H\},$$

es decir, la profundidad de un punto es definida como la mínima probabilidad que puede tener un subespacio que contenga a  $x$ .

La versión empírica es

$$HS_n(x) = \min_{v \in \mathbb{R}^d} \frac{\#\{i : \langle X_i, v \rangle \geq \langle x, v \rangle\}}{n}$$

donde  $\langle \cdot, \cdot \rangle$  es el producto interno usual.

Algoritmos para calcular eficientemente esta profundidad son conocidos para dimensiones 2 o 3 (ver Rousseeuw y Struyf [47]), mientras que en dimensiones mayores el algoritmo de Cuesta-Albertos y Nietos-Reyes [11] es computacionalmente eficiente en todas las dimensiones.

- la *profundidad de simplices*, “*simplicial depth*”,  $SD$ , Liu [36]:

$$SD(x, \mathbb{P}) = \mathbb{P}(x \in S[\tilde{X}_1, \dots, \tilde{X}_{d+1}]),$$

donde  $\tilde{X}_1, \dots, \tilde{X}_{d+1}$  son i.i.d  $\mathbb{P}$  y  $S[\tilde{X}_1, \dots, \tilde{X}_{d+1}]$  es un simplex de vértices  $\tilde{X}_1, \dots, \tilde{X}_{d+1}$ .

La profundidad de un punto en este caso es la probabilidad de que pertenezca a un simplex aleatorio cuyos vértices son i.i.d  $\mathbb{P}$ . La versión empírica es

$$SD_n(x) = \binom{n}{d+1}^{-1} \sum_{1 \leq i_1 < i_2 < \dots < i_{d+1} \leq n} \mathbb{1}\{x \in S[X_{i_1}, \dots, X_{i_{d+1}}]\}.$$

- la *profundidad de zonoides*, “*Zonoid depth*”,  $ZD$ , Dyckerhoff, Mosler y Koshevoy [21]:

$$ZD(x, \mathbb{P}) = \sup_{\alpha} \{x \in D_{\alpha}(\mathbb{P})\},$$

donde  $D_{\alpha}(\mathbb{P})$  es la región zonoide  $\alpha$  podada de  $\mathbb{P}$ ,

$$D_{\alpha}(\mathbb{P}) = \left\{ \int_{\mathbb{R}^d} x g(x) d\mathbb{P}(x) : g : \mathbb{R}^d \rightarrow \left[0, \frac{1}{\alpha}\right] \text{ es medible y } \int_{\mathbb{R}^d} g(x) d\mathbb{P}(x) = 1 \right\}.$$

Mientras que la versión empírica es

$$ZD_n(x) = \sup_{\alpha} \{x \in D_{\alpha}(X_1, \dots, X_n)\},$$

con

$$D_{\alpha}(X_1, \dots, X_n) = \left\{ \sum_{i=1}^n \lambda_i X_i : \sum_{i=1}^n \lambda_i = 1, \alpha \lambda_i \leq \frac{1}{n} \forall i \right\}.$$

- la *profundidad de proyecciones*, “*Projection depth*”,  $PD$ , Stahel [54], Donoho [17]:

$$PD(x, \mathbb{P}) = 1/[1 + O(x, \mathbb{P})], \quad (1.1)$$

basada en la función,

$$O(x, \mathbb{P}) = \sup_{\|u\|=1} \frac{|\langle u, x \rangle - \text{Med}(F_u)|}{MAD(F_u)}, \quad (1.2)$$

donde  $F_u$  es la distribución de  $\langle u, \mathbf{X} \rangle$ ,  $\text{Med}(F_u)$  es la mediana de  $F_u$  y  $MAD(F_u) = \text{Med}\{|F_u - \text{Med}(F_u)|\}$ .

Para calcular la versión empírica se debe reemplazar la mediana y  $MAD$  por sus versiones muestrales.

La función (1.2) mide el grado de lejanía de un punto respecto a la distribución, por lo que es un concepto equivalente pero inverso al de profundidad, i.e., puntos con mayor valor de la función están más alejados del centro.

En Zuo [66] se amplía la definición (1.2) a toda medida univariada de posición  $\mu$  y de escala  $\sigma$  definiendo las *funciones de outlyingness*, generalmente denotadas  $O(x, \mathbb{P})$  de la siguiente manera

$$O(x, \mathbb{P}) = \sup_{\|u\|=1} \frac{|u'x - \mu(F_u)|}{\sigma(F_u)}.$$

Si la función  $O$  tiene rango  $[0, +\infty)$ , es posible definir una función de profundidad  $D$  como en (1.1), mientras que si el rango es acotado, es posible definir

$$D(x, \mathbb{P}) = 1 - \frac{O(x, \mathbb{P})}{\sup_{x \in \mathbb{R}^d} (O(x, \mathbb{P}))}.$$

Ya fue demostrado que las profundidades listadas cumplen con todas las propiedades descriptas **P1** a **P6** y sus versiones empíricas son consistentes. (Ver Zuo y Serfling [67] para HS y PD, Liu [36] para SD y Dyckerhoff et al. [21] para ZD).

También es posible encontrar casos en los que la definición original de una profundidad no cumple con alguna de las propiedades listadas anteriormente, pero con pequeños ajustes es posible que si lo hagan; por ejemplo, la profundidad  $L^p$  cumple **P1'**, mientras que cambiando la norma a utilizar, la misma profundidad cumple **P1** como fue mostrado en Zuo y Serfling [67].

Existen otras definiciones de profundidad basadas en una construcción empírica como las desarrolladas a partir del concepto “convex hull peeling” y la profundidad de Delaunay [3] las cuales no son tenidas en cuenta en este trabajo por carecer de una definición poblacional.

Es necesario destacar que, por su definición, algunas profundidades en su versión empírica tienen como soporte  $\mathbb{R}^d$ , mientras que para otras el soporte resulta ser la cápsula convexa de la muestra aleatoria. Además, también se sabe que algunas (como HS) siempre tendrán contornos convexos por lo que cuando la distribución tenga un soporte no convexo “no van a poder distinguirlo”.

Una manera de generalizar la idea de ordenamiento del centro hacia afuera que permite trabajar con distribuciones con soportes no convexos y está comenzando a tomar impulso es basar la definición de percentiles y rangos a partir de la utilización de funciones de transporte óptimo.

En Chernozhukov et al. [7] definen la profundidad “Monge - Kantorovich” (MK) cuya idea principal es la siguiente: para calcular la profundidad respecto de la distribución  $F_1$ , se debe elegir una distribución de “referencia”  $F_2$  para la cual se conocen los contornos de profundidad (típicamente  $F_2$  es elegida como la distribución uniforme en el círculo unitario y los contornos de profundidad son las circunferencias de centro 0 y radio  $r \in (0, 1)$ ) y a través de un mapeo “bien elegido” (en términos de medidas el mapeo elegido es el gradiente de  $\psi$

donde  $\nabla\psi\#F_2 = F_1$ ) se transforman los contornos de profundidad respecto de  $F_2$  en los contornos de profundidad respecto de  $F_1$ . En el caso de que ambas distribuciones tengan segundo momento finito, el mapeo utilizado es la función de transporte de Monge - Kantorovich de  $F_2$  a  $F_1$  para la pérdida cuadrática. El caso empírico para una muestra  $\mathcal{X}_n$  consiste en generar una grilla que refleje de la mejor manera la distribución empírica de  $F_2$  con la cantidad de puntos que se tienen en la muestra original y a partir de un algoritmo de asignamiento óptimo se mapea la profundidad de los puntos de la grilla de referencia a la muestra  $\mathcal{X}_n$ .

En una reciente publicación de Hallin et al. [25] se presentan ideas muy interesantes extendiendo y generalizando el enfoque de MK.

Otro punto a tener en cuenta es que para una misma distribución (o muestra aleatoria) utilizar distintas profundidades puede dar como resultados distintos “centros”, los cuales suelen ser parecidos entre sí mientras la distribución cumpla alguna de las simetrías descritas anteriormente, mientras que pueden llegar a diferir mucho en caso de trabajar con distribuciones asimétricas (por más que las funciones de profundidad cumplan **P1-P6**).

Por esto, es entendible que no exista una profundidad que sea “mejor” que todas. A veces una puede ajustar mejor a los datos con ciertas características en su distribución, mientras que en otros casos la facilidad de cálculo o implementación es lo que se busca. Además es necesario tener en cuenta la aplicación que se va a hacer de la profundidad, las cuales pueden ser muy variadas como veremos a continuación.

## 1.2. Aplicaciones y extensiones

La principal aplicación de las funciones de profundidad se da dentro de procedimientos de inferencia multivariada. Por ejemplo, los contornos de profundidades (conjunto de puntos en  $\mathbb{R}^d$  con igual valor de profundidad), cumplen el rol de los percentiles univariados y ayudan geoméricamente a entender la forma o estructura de una distribución y pueden ser utilizados en estadísticos robustos. En Liu, Parelius y Singh [37] se muestra cómo a través del orden dado por una profundidad es posible encontrar diferencias entre dos poblaciones en posición, escala, skewness y kurtosis. También, a partir del ordenamiento dado por una profundidad en Liu y Singh [39] desarrollan un test de rangos para testear diferencia de escalas para datos multivariados.

Una idea de aplicación de funciones de profundidad introducida en Liu et al. [37] es utilizar DD-plot (Depth vs Depth plot). Esta idea se basa en un plot bidimensional en el que cada punto representa la profundidad del dato respecto de una y de otra distribución. La ventaja de los DD es que sin importar la dimensión en que están los datos, se trabaja con un plot en dos dimensiones y, más precisamente, en  $[0, 1] \times [0, 1]$  pues sin pérdida de generalidad puede estandarizarse la función de profundidad que uno desee usar. Además, muestran que distintos patrones en el gráfico pueden ser asociados a diferencias en algunos estadísticos de ambas distribuciones. Es por esto que, por ejemplo, en Li y Liu [35] describen varios test no

paramétricos de posición y de escala a partir de la utilización de DD-plot. Otra aplicación de los DD-plot muy interesantes es la clasificación (ver Sección 5.2).

Las profundidades también se han utilizado para mejorar la visualización de datos multivariados, una de las herramientas desarrolladas es la generalización del boxplot a 2 dimensiones a través del bagplot definido por Rousseuw, Ruts y Tukey [46] y la utilización de estos para detección de outliers.

Cuando los datos están en  $\mathbb{R}^d$  con  $d$  grande, es muy común que tengan estructura en dimensión menor lo que hace que sea necesario también extender el concepto de profundidad a distribuciones en variedades, tema que comentaremos hacia el final de la tesis. Además, mientras la capacidad de cómputo comenzó a dejar atrás el problema de trabajar con muestras de gran tamaño, existen todavía limitaciones en términos computacionales para datos en dimensiones muy altas. Para que una profundidad definida sea verdaderamente utilizable su cálculo tiene que ser posible y eficiente. El nuevo paso en este sentido se ve en una colaboración entre estadísticos y computadores que ha generado algunos resultados principalmente en dimensión 2 y que seguramente se amplíe a más dimensiones.

Todas las funciones de profundidad descritas en la sección anterior trabajan sobre la estimación no paramétrica de posición y centro de distribuciones lo cual tiene sentido en funciones unimodales. Típicamente, las características multimodales de la distribución no son captadas por las profundidades, por lo que de ser necesario hacerlo es posible utilizar las *profundidades locales* (ver Agostinelli y Romanazzi [4], Paindaveine y Van Bever [43]). Por otro lado, si se desea trabajar con datos funcionales el concepto de profundidad fue extendido para poder utilizarlo en espacios infinito dimensionales (ver López-Pintado y Romo [40]).

Además, fuera del concepto de posición se han hecho extensiones de la idea de profundidad por ejemplo a regresiones, Rousseuw y Hubert [45] definieron la “regression depth”, donde RD mide qué tan bien un hiperplano ajusta a los datos. El trabajo de Zuo [65] trata de extender el conjunto de criterios dados en la definición de profundidad (**P1-P6**) de Zuo y Serfling [67] a este tipo de profundidades.

Las funciones de profundidad continúan siendo uno de los tópicos más estudiados en la estadística. En términos de inferencia no paramétrica multivariada una publicación que resume los posibles usos y la bibliografía es Serfling [50]. Algunas otras aplicaciones ya conocidas y estudiadas son los gráficos de control (control charts), la exploración de datos, clustering, medición de riesgo (multivariate risk measurement), procesamiento de señales, econometría y aplicaciones a estudios sociales. En Zuo [65] se encuentran listadas estas y otras aplicaciones junto con la bibliografía correspondiente.

# Capítulo 2

## Profundidad de salida, difusiones en $\mathbb{R}^d$

### 2.1. Difusiones

Previamente a definir la nueva función de profundidad, necesitamos repasar el concepto de difusión de Itô y algunas de sus propiedades como la conexión con las ecuaciones en derivadas parciales.

Coloquialmente una difusión es la solución de una ecuación diferencial estocástica y podría pensarse que describe el movimiento de una partícula en un fluido con una deriva (o drift)  $b$  y un coeficiente difusivo  $\sigma$ .

La definición clásica es la siguiente:

**Definición 2.1** (Difusión de Itô. Oksendal [42], Definición 7.1.1). *Una difusión de Itô (homogénea en el tiempo) en  $\mathbb{R}^d$  es un proceso  $Y_t(\omega) = Y(t, \omega) : [0, \infty) \times \Omega \rightarrow \mathbb{R}^d$  definido en un espacio de probabilidad  $(\Omega, \Sigma, \mathbb{P})$  que satisface la ecuación diferencial estocástica*

$$dY_t = b(Y_t) dt + \sigma(Y_t) dB_t, \quad (2.1)$$

donde  $B_t$  es un movimiento Browniano  $m$ -dimensional y los coeficientes  $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$  y  $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$  satisfacen la condición

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leq D|x - y|; \quad x, y \in \mathbb{R}^d, \quad (2.2)$$

para alguna constante  $D > 0$  y  $|\sigma|^2 = \sum |\sigma_{ij}|^2$ .

La condición (2.2), asegura que la solución  $Y_t$  de (2.1) es única. Esto significa que si  $Y_1(t, \omega)$  e  $Y_2(t, \omega)$  son dos procesos  $t$ -continuos satisfaciendo (2.1), entonces  $Y_1(t, \omega) = Y_2(t, \omega), \forall t$  con probabilidad 1. Además, que  $b(\cdot)$  y  $\sigma(\cdot)$  sean Lipschitz continuas puede ser relajado cuando interesa trabajar con difusiones en conjuntos acotados como lo demuestra el siguiente lema.

**Lema 2.1** (Existencia. Ikeda y Watanabe [31], Teorema 4.3.1). *Supongamos  $b(x)$  y  $\sigma(x)$  continuas, localmente Lipschitz, es decir,  $\forall r > 0$ , existe  $K_r \in \mathbb{R}$  tal que*

$$|b(x) - b(y)|^2 + |\sigma(x) - \sigma(y)|^2 \leq K_r |x - y|^2, \quad x, y \in B(0, r).$$

*Entonces, existe una única solución fuerte de (2.1)  $\forall t \leq \tau_r$ ,  $\tau_r = \inf\{t : |Y_t| \geq r\}$ .*

A continuación definimos el generador de una difusión y su forma explícita en función de los coeficientes de la difusión.

**Definición 2.2** (Generador. Oksendal [42], Definición 7.3.1). *Sea  $Y_t$  una difusión de la forma (2.1). El generador  $\mathcal{A}$  de  $Y_t$ , aplicado a funciones  $g : \mathbb{R}^d \rightarrow \mathbb{R}$ , se define por*

$$\mathcal{A}g(x) = \lim_{t \rightarrow 0^+} \frac{\mathbb{E}_x(g(Y_t)) - g(x)}{t}, \quad x \in \mathbb{R}^d.$$

*Denotamos  $\mathcal{D}_{\mathcal{A}}$  al conjunto de funciones  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  para el cual el límite existe  $\forall x \in \mathbb{R}^d$ .*

**Lema 2.2** (Generador. Oksendal [42], Teorema 7.3.3). *Sea  $Y_t$  una difusión de la forma (2.1), con generador  $\mathcal{A}$ . Si  $g \in C_0^2(\mathbb{R}^d)$ , entonces  $g \in \mathcal{D}_{\mathcal{A}}$  y*

$$\mathcal{A}g(x) = \frac{1}{2} \sum_{i,j} (\sigma \sigma^T)_{i,j}(x) \frac{\partial^2 g}{\partial x_i \partial x_j} + \sum_i b_i(x) \frac{\partial g}{\partial x_i}.$$

Las difusiones de Itô son procesos continuos con la propiedad fuerte de Markov. Otra característica, que será utilizada para probar propiedades de la nueva profundidad, está dada por el siguiente lema.

**Lema 2.3** (Continuidad. Oksendal [42], Lema 8.1.4). *Sea  $Y_t$  de la forma (2.1) con generador  $\mathcal{A}$ . Sea  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  acotada inferiormente y definimos, para  $t \geq 0$  fijo*

$$v(x) = \mathbb{E}_x[g(Y_t)].$$

- a) *Si  $g$  es semi-continua inferior, entonces  $v$  es semi-continua inferior.*
- b) *Si  $g$  es acotada y continua, entonces  $v$  es continua.*

**Observación 2.1** (Feller-continuidad). *En Oksendal [42], todo proceso estocástico que satisface b) en el lema anterior es denominado Feller-continuo. Es decir, el resultado dice que toda difusión de Itô  $Y_t$  es Feller-continua.*

Llegamos así a poder plantear un resultado clásico, la *fórmula de Feynman-Kac*, que relaciona al generador de la difusión (i.e. el operador asociado a la difusión) con una ecuación en derivadas parciales y será herramienta a utilizar para definir la nueva función de profundidad en la Sección 2.2.

**Lema 2.4** (Feynmann-Kac. Oksendal [42], Teorema 8.2.1). *Sea  $Y_t$  de la forma (2.1) con generador  $\mathcal{A}$ . Sean  $g \in C_0^2(\mathbb{R}^d)$  y  $q \in C(\mathbb{R}^d)$  acotada inferiormente.*

a) Si

$$v(t, x) = \mathbb{E}_x \left[ \exp \left( - \int_0^t q(Y_s) ds \right) g(Y_t) \right], \quad (2.3)$$

entonces

$$\frac{\partial v}{\partial t} = \mathcal{A}v - qv; \quad t > 0, x \in \mathbb{R}^d \quad (2.4)$$

$$v(0, x) = g(x); \quad x \in \mathbb{R}^d \quad (2.5)$$

b) *Es más, si  $w(t, x) \in C^{1,2}(\mathbb{R} \times \mathbb{R}^d)$  es acotada en  $K \times \mathbb{R}^d$  para cada  $K \subset \mathbb{R}$  y  $w$  es solución de (2.4), (2.5), entonces  $w(t, x) = v(t, x)$ , dada por (2.3).*

## 2.2. Definición de $\mathcal{PS}$

En esta sección vamos a definir formalmente la nueva función de profundidad de manera general. Exceptuando que se indique lo contrario,  $\mathbf{X}$  es un vector aleatorio con distribución  $F$  y densidad  $f$  para la cual asumimos

- 1)  $\mathfrak{D} = \{x : f(x) > 0\} \subset \mathbb{R}^d$  es compacto y conexo con borde  $\partial\mathfrak{D}$  suave.
- H1** 2) Existen  $f_{min}, f_{max} \in \mathbb{R}^+$  tal que  $0 < f_{min} \leq f(x) \leq f_{max} < \infty, \forall x \in \mathfrak{D}$ .
- 3)  $f(x)$  es diferenciable y  $\nabla(\ln(f(x)))$  es acotada y localmente Lipschitz,  $\forall x \in \mathfrak{D}$ .

Definimos la difusión de Itô  $Y_t$  que toma valores en el espacio de funciones continuas  $C([0, \infty), \mathfrak{D})$

$$dY_t = \frac{\nabla \ln(f(Y_t))}{3(f(Y_t))^{2/d}} dt + \frac{1}{\sqrt{3}(f(Y_t))^{1/d}} dB_t, \quad (2.6)$$

que bajo las condiciones **H1**, por el Lema 2.1, tiene solución única.

Coloquialmente podemos interpretar que el término difusivo dice que  $Y_t$  “varía” más en áreas de poca densidad que en áreas de mayor densidad; mientras que para comprender el término de deriva, vemos que apunta hacia las zonas de mayor densidad y, al igual que el término difusivo, en áreas de poca densidad será más fuerte el movimiento que en áreas de densidad alta. Es posible entender a  $Y_t$ , entonces, como una difusión que “trata de ir hacia áreas densas”, pasando más tiempo (menos difusiva y con menos drift) cerca de la/s moda/s de la distribución.

En la Sección 4.1.1, más precisamente en (4.1) y (4.3), veremos que los coeficientes de (2.6) surgen naturalmente como la esperanza y varianza de un salto en el paseo aleatorio en el grafo utilizado y son los necesarios para alcanzar el objetivo **3**. de consistencia.

Conceptualmente entonces tenemos que el tiempo de salida del dominio partiendo de puntos ubicados en áreas densas debería ser mayor que el de puntos en áreas ralas. En ese sentido el tiempo esperado de salida (o cualquier función monótona de este tiempo) surgen naturalmente como medidas de profundidad.

Es necesario destacar que, como estamos interesados en el tiempo de parada

$$\tau = \inf\{t \geq 0 : Y_t \in \overline{\mathcal{D}^c}\}, \quad (2.7)$$

es posible pensar  $Y_t$  con condiciones de rebote en la frontera  $\partial\mathcal{D}$  o condiciones de Dirichlet (absorbentes en la frontera) de manera indistinta.

De manera análoga utilizaremos también como definición para  $x \in \mathcal{D}$

$$\tau = \inf\{t \geq 0 : Y_t \in \partial\mathcal{D}\}.$$

Ahora bien, el uso de  $\mathbb{E}(\tau)$  como función de profundidad plantea una dificultad extra: la variable aleatoria  $\tau$  no está acotada, por lo que las posibles aproximaciones empíricas a la difusión  $Y_t$ , a utilizar eventualmente, deberían converger de manera “más fuerte” que en distribución para poder realizar inferencia y obtener consistencia. Por otro lado, el uso de cualquier función  $\varphi(\tau)$  acotada, continua y monótona establecería un orden “del centro hacia afuera” similar (exactamente el mismo orden en caso de que  $\tau_x \leq_{st} \tau_y$ , donde  $\leq_{st}$  significa dominación estocástica) necesitando las mínimas condiciones de convergencia en los procesos empíricos a  $Y_t$ .

Como hemos comentado en la Introducción, en Hashimoto et al. [26] se muestra que la Transformada de Laplace de  $\tau$  conserva la tendencia de agrupamiento de los puntos y es posible utilizarla en métodos de clusterización. Aquí será usada para definir formalmente una nueva función de profundidad.

**Definición 2.3** (Profundidad de salida). *Sea  $Y_t$  una difusión dada por (2.6) y sea  $\tau$  el tiempo de parada dado en (2.7). Definimos la función profundidad de salida  $\mathcal{PS} : \mathbb{R}^d \rightarrow \mathbb{R}$ ,*

$$\mathcal{PS}(x) = -\ln(\mathbb{E}_x(e^{-\tau})). \quad (2.8)$$

Es decir, para  $Y_0 = x$ ,  $\mathcal{PS}(x)$  es el logaritmo de la transformada de Laplace del tiempo de salida de la difusión del dominio  $\mathcal{D}$ . Notar que  $\tau = 0$  con probabilidad 1 para  $x \notin \mathcal{D}$  (y consecuentemente  $\mathcal{PS}(x) = 0$  para  $x \notin \mathcal{D}$ ).

En la Observación 2.5 comentamos el posible uso de otra función monótona de  $\tau$  como definición de profundidad.

Lo primero que haremos va a ser asegurarnos que  $\mathcal{PS}$  está bien definida, para esto queremos utilizar el siguiente resultado que es una aplicación de la fórmula Feynman-Kac (Lema 2.4) para problemas de frontera del tipo Dirichlet.

**Lema 2.5** (Sistema de ecuaciones. Oksendal[42]. Ejercicio 9.12.). *Sea  $Y_t$  una difusión de Itô en  $\mathbb{R}^d$  con generador  $\mathcal{A}$  que coincide con el operador  $\mathcal{L}$  en  $C_0^2(\mathbb{R}^d)$ . Sean  $\mathfrak{D} \subset \mathbb{R}^d$  con borde suave,  $\phi \in C(\partial\mathfrak{D})$  acotada,  $q(x) \geq 0$  función continua en  $\mathbb{R}^d$  y sea  $g \in C(\mathfrak{D})$  tal que*

$$\mathbb{E}_x \left[ \int_0^{\tau_{\mathfrak{D}}} |g(Y_t)| dt \right] < \infty, \quad \forall x \in \mathfrak{D},$$

donde  $\tau_{\mathfrak{D}} = \inf\{t > 0 : Y_t \notin \mathfrak{D}\}$ . Si existe solución acotada  $h(x) \in C^2(\mathfrak{D}) \cap C(\bar{\mathfrak{D}})$  del sistema

$$\begin{cases} \mathcal{L}h(x) - q(x)h(x) = -g(x), & x \in \mathfrak{D} \\ \lim_{x \rightarrow y} h(x) = \phi(y), & y \in \partial\mathfrak{D}, \end{cases}$$

entonces

$$h(x) = \mathbb{E}_x \left[ \int_0^{\tau_{\mathfrak{D}}} e^{-\int_0^t q(Y_s) ds} g(Y_t) dt + e^{-\int_0^{\tau_{\mathfrak{D}}} q(Y_s) ds} \phi(Y_{\tau_{\mathfrak{D}}}) \right].$$

**Proposición 2.1** (Existencia de  $\mathcal{PS}$ ). *Bajo las condiciones dadas en **H1** para  $\mathcal{PS}$  definida en (2.8),  $\mathcal{PS} \in C^2(\mathfrak{D}) \cap C(\bar{\mathfrak{D}})$  es la única solución del problema de frontera de tipo Dirichlet:*

$$\Delta H(x) - \|\nabla H(x)\|^2 + 2\langle \nabla \ln(f(x)), \nabla H(x) \rangle + 6f^{2/d}(x) = 0, \quad x \in \mathfrak{D} \quad (2.9)$$

$$H(x) = 0, \quad x \in \partial\mathfrak{D}. \quad (2.10)$$

*Demostración.* Si tomamos  $q(x) = 1$ ,  $g(x) = 0$  y  $\phi(y) = 1$ , todas las hipótesis del Lema 2.5 se cumplen y obtenemos que si existe  $h(x) \in C^2(\mathfrak{D}) \cap C(\bar{\mathfrak{D}})$  solución acotada de

$$\begin{cases} \mathcal{L}h(x) - h(x) = 0, & x \in \mathfrak{D} \\ \lim_{x \rightarrow y} h(x) = 1, & y \in \partial\mathfrak{D}. \end{cases} \quad (2.11)$$

entonces

$$h(x) = \mathbb{E}_x [e^{-\tau_{\mathfrak{D}}}] .$$

Según el Lema 2.2, el generador de la difusión  $Y_t$  dada en (2.6) aplicado a  $\varphi \in C_0(\mathbb{R}^d)$  es

$$\mathcal{A}(\varphi)(x) = 3^{-1} f^{-2/d}(x) \left( \frac{1}{2} \Delta \varphi(x) + \langle \nabla \ln(f(x)), \nabla \varphi(x) \rangle \right).$$

Luego, el sistema (2.11) en este caso es

$$\begin{cases} \Delta h(x) + 2\langle \nabla \ln(f(x)), \nabla h(x) \rangle - 6f^{2/d}(x)h(x) = 0, & x \in \mathfrak{D} \\ h(x) = 1, & x \in \partial\mathfrak{D}. \end{cases} \quad (2.12)$$

Para poder ver que existe solución del sistema (2.12), utilizamos el siguiente resultado.

**Lema 2.6** (Existencia y unicidad de solución. Gilbarg y Trudinger [24], Teorema 6.13). *Sea  $\mathcal{D} \subset \mathbb{R}^d$  un dominio acotado tal que  $\partial\mathcal{D}$  es  $C^2$ . Sea  $\mathcal{L}$  el operador*

$$\mathcal{L}h(x) = \sum_{i,j} a_{ij}(x) \frac{\partial^2 h}{\partial x_i \partial x_j} + \sum_i b_i(x) \frac{\partial h}{\partial x_i} + c(x)h(x), \quad a_{ij} = a_{ji},$$

*estrictamente elíptico, es decir, existe  $\lambda > 0$  tal que*

$$\sum_{i,j} a_{ij} \xi_i \xi_j \geq \lambda |\xi|^2, \quad \forall x \in \mathcal{D}, \xi \in \mathbb{R}^d.$$

*Si  $c(x) \leq 0$  y los coeficientes  $a_{ij}$ ,  $b_i$  y  $g(x)$  son  $C^\alpha$ ,  $\forall i, j$  y  $\phi$  continua en  $\partial\mathcal{D}$ , entonces el problema de frontera Dirichlet*

$$\begin{aligned} \mathcal{L}h(x) &= g(x), & x \in \mathcal{D} \\ h(x) &= \phi(x), & x \in \partial\mathcal{D}. \end{aligned}$$

*tiene una única solución  $h(x) \in C^{2,\alpha}(\mathcal{D}) \cap C(\bar{\mathcal{D}})$ .*

Luego, las condiciones dadas en **H1** aseguran que existe una solución única y acotada  $h(x)$  de (2.12) y por el Lema 2.5 resulta  $h(x) = \mathbb{E}_x(e^{-\tau})$ .

Definimos  $H(x) = -\ln(h(x))$ , luego

$$\begin{cases} \nabla H(x) = -\frac{1}{h(x)} \nabla h(x) \\ \Delta H(x) = \frac{1}{h^2(x)} \|\nabla h(x)\|^2 - \frac{1}{h(x)} \Delta h(x) \end{cases} \implies \begin{cases} \nabla h(x) = -h(x) \nabla H(x) \\ \Delta h(x) = -h(x) \Delta H(x) + \frac{1}{h(x)} \|\nabla h(x)\|^2 \end{cases}$$

Sustituyendo en (2.12) y simplificando,  $H(x)$  es la única solución del sistema dado en (2.9-2.10) y podemos concluir entonces que  $\mathcal{PS} \in C^2(\mathcal{D}) \cap C(\bar{\mathcal{D}})$ .  $\square$

**Observación 2.2** (Dominios no conexos). *En **H1** asumimos que el soporte de la distribución  $f$  es conexo, para facilitar la exposición. Si  $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$ , con  $\mathcal{D}_1, \mathcal{D}_2$  conexos disjuntos,  $\mathcal{PS}$  va a estar definida en cada componente conexa, veremos en el Capítulo 5 que esto permitirá a  $\mathcal{PS}$  identificar cuándo los datos provienen de poblaciones distintas y tener una buena performance en clasificación de datos con muestras contaminadas.*

### 2.2.1. Propiedades de $\mathcal{PS}$

Ahora que tenemos la nueva definición de función de profundidad, queremos ver que  $\mathcal{PS}$  cumple las propiedades descritas en la Introducción.

Comenzaremos por las propiedades que se cumplen por definición de  $\mathcal{PS}$ . En primer lugar, que la profundidad tienda a 0 cuando  $\|x\| \rightarrow \infty$  (**P4**) se cumple trivialmente al tratarse de distribuciones con soporte acotado. En la Sección 3.2.3 extenderemos la definición de  $\mathcal{PS}$  a distribuciones no acotadas.

Como hemos visto, la continuidad de  $\mathcal{PS}$  como función de  $x$  (**P5**) fue probada en la Proposición 2.1. Desde el punto de vista probabilístico, esta propiedad podía ser demostrada a partir de la definición, utilizando que las difusiones de Itô son *Feller-continuas*, como vimos en el Lema 2.3.

A partir de (2.9) uno puede intuir que la función  $\mathcal{PS}$  no va a tener la propiedad de invarianza afín; a continuación veremos que  $\mathcal{PS}$  cumple **P1'** y además es invariante por cambios de escala. Para que la exposición sea más clara denotaremos  $f_W$  a la densidad de la variable  $W$  y  $\mathcal{PS}_W$  a la solución de (2.9)-(2.10) con densidad  $f_W$ .

Definimos el operador  $\mathcal{L}_W : C^2(\mathbb{R}^d) \rightarrow \mathbb{R}$ ,

$$\mathcal{L}_W(\cdot)(w) = \Delta_w(\cdot) - \|\nabla_w(\cdot)\|^2 + \frac{2}{f_W(w)} \langle \nabla_w f_W(w), \nabla_w(\cdot) \rangle.$$

Luego  $\mathcal{PS}_X(x)$  es la única solución de

$$\begin{aligned} \mathcal{L}_X(\varphi)(x) &= -6f_X^{2/d}(x), & x \in \mathfrak{D} \\ \varphi(x) &= 0, & x \in \mathfrak{D}^c. \end{aligned}$$

**Lema 2.7** ( $\mathcal{PS}$  cumple **P1'** y es invariante por cambio de escala). Sean  $b \in \mathbb{R}^d$ ,  $\mathbf{A}$  matriz ortogonal y  $c > 0$  entonces

$$\mathcal{PS}(c\mathbf{A}x + b; F_{c\mathbf{A}x+b}) = \mathcal{PS}(x; F_X)$$

vale para todo vector aleatorio  $\mathbf{X} \in \mathbb{R}^d$  con distribución  $F_X$ .

*Demostración.* Sea  $\mathbf{Y} = c\mathbf{A}\mathbf{X} + b$ . Para  $y \in \mathbb{R}^d$ , tomamos  $x = c^{-1}\mathbf{A}^{-1}(y - b)$  y definimos  $\mathfrak{ps}(y) = \mathcal{PS}_X(c^{-1}\mathbf{A}^{-1}(y - b)) = \mathcal{PS}_X(x)$ . Queremos ver que  $\mathfrak{ps}(y) = \mathcal{PS}_Y(y)$ .

- Si  $y \notin c\mathbf{A}\mathfrak{D} + b$ , resulta  $c^{-1}\mathbf{A}^{-1}(y - b) \notin \mathfrak{D}$  y por (2.10)  $\mathfrak{ps}(y) = 0$ .
- Si  $y \in c\mathbf{A}\mathfrak{D} + b$ , resulta  $c^{-1}\mathbf{A}^{-1}(y - b) \in \mathfrak{D}$  y además

$$\begin{aligned} f_Y(y) &= |\det(c\mathbf{A})|^{-1} f_X(c^{-1}\mathbf{A}^{-1}(y - b)) = c^{-d} f_X(c^{-1}\mathbf{A}^{-1}(y - b)) = c^{-d} f_X(x), \\ \nabla_y f_Y(y) &= \nabla_y (c^{-d} f_X(c^{-1}\mathbf{A}^{-1}(y - b))) = c^{-d-1} (\mathbf{A}^{-1})^T (\nabla_x f_X)(x), \\ \nabla_y (\mathfrak{ps}(y)) &= \nabla_y (\mathcal{PS}_X(c^{-1}\mathbf{A}^{-1}(y - b))) = c^{-1} (\mathbf{A}^{-1})^T (\nabla_x \mathcal{PS}_X)(x), \\ \Delta_y (\mathfrak{ps}(y)) &= \Delta_y (\mathcal{PS}_X(c^{-1}\mathbf{A}^{-1}(y - b))) = c^{-2} (\Delta_x \mathcal{PS}_X)(x), \\ \langle \nabla_y f_Y(y), \nabla_y (\mathfrak{ps}(y)) \rangle &= c^{-d-2} \langle (\nabla_x f_X)(x), (\nabla_x \mathcal{PS}_X)(x) \rangle. \end{aligned}$$

Por lo tanto,

$$\begin{aligned}
\mathcal{L}_Y(\text{ps})(y) &= \Delta_y(\text{ps}(y)) - \|\nabla_y(\text{ps}(y))\|^2 + \frac{2}{f_Y(y)} \langle \nabla_y f_Y(y), \nabla_y(\text{ps}(y)) \rangle \\
&= c^{-2} (\Delta_x \mathcal{P}S_X)(x) - \|c^{-1}(\mathbf{A}^{-1})^T (\nabla_x \mathcal{P}S_X)(x)\|^2 + \frac{2c^{-d-2} \langle (\nabla_x f_X)(x), (\nabla_x \mathcal{P}S_X)(x) \rangle}{c^{-d} f_X(x)} \\
&= c^{-2} (\Delta_x \mathcal{P}S_X)(x) - c^{-2} \|(\nabla_x \mathcal{P}S_X)(x)\|^2 + \frac{2c^{-2}}{f_X(x)} \langle (\nabla_x f_X)(x), (\nabla_x \mathcal{P}S_X)(x) \rangle \\
&= c^{-2} \mathcal{L}_X(\mathcal{P}S_X)(x) \\
&= c^{-2} [-6f_X^{2/d}(x)] = c^{-2} [-6(c^d f_Y(y))^{2/d}] = -6f_Y^{2/d}(y).
\end{aligned}$$

Por unicidad de solución, tenemos que  $\mathcal{P}S_Y(y) = \text{ps}(y) = \mathcal{P}S_X(x)$ .  $\square$

Utilizando la misma idea que en el lema anterior es posible demostrar la siguiente propiedad de simetría.

**Lema 2.8** (Simetría de  $\mathcal{P}S$ ). *Si  $X$  tiene distribución centralmente simétrica en  $\theta$  entonces*

$$\mathcal{P}S(x - \theta) = \mathcal{P}S(\theta - x), \quad \forall x.$$

*Demostración.* Sea  $f$  la densidad del vector  $\mathbf{X}$ , sin pérdida de generalidad podemos suponer  $\theta = \mathbf{0}$  y por lo tanto  $f(x) = f(-x)$ .

Definiendo  $\text{ps}(x) = \mathcal{P}S(-x)$ , obtenemos

$$\Delta \text{ps}(x) = \Delta \mathcal{P}S(-x), \quad \nabla \text{ps}(x) = -\nabla \mathcal{P}S(-x) \quad \text{y además} \quad \nabla f(x) = -\nabla f(-x).$$

Resulta que entonces  $\text{ps}(x)$  también es solución de (2.9-2.10) y por unicidad de solución tenemos que  $\mathcal{P}S(x) = \text{ps}(x) = \mathcal{P}S(-x)$ .  $\square$

Para probar las propiedades **P2** y **P3** de  $\mathcal{P}S$  es necesario primero introducir los siguientes conceptos y resultados de Kennington [32].

**Definición 2.4** (Funciones armónicas cóncavas). *Si  $D \subset \mathbb{R}^d$  es un conjunto convexo, entonces  $b : D \rightarrow \mathbb{R}$  es armónica cóncava si para todo  $(y, z, \lambda) \in D \times D \times [0, 1]$ ,*

$$\begin{aligned}
b((1 - \lambda)y + \lambda z) &\geq b(y)b(z)((1 - \lambda)b(z) + \lambda b(y))^{-1}, & \text{si } (1 - \lambda)b(z) + \lambda b(y) > 0 \\
b((1 - \lambda)y + \lambda z) &\geq 0, & \text{si } b(y) = b(z) = 0.
\end{aligned}$$

**Observación 2.3** (Condiciones para concavidad). *Esta definición extiende el concepto de (-1)-concavidad de funciones positivas a funciones en  $\mathbb{R}$  en general. Además, una función cóncava (positiva o no) es armónica cóncava, mientras que una función positiva  $b$  es armónica cóncava si y sólo si  $1/b$  es convexa (es decir, (-1)-concava).*

El siguiente resultado se obtiene combinando los Teoremas 3.1, 3.2, 3.3 y la Observación 3.1.1 de Kennington [32], para nuestro caso en particular.

**Teorema 2.1** (Condiciones para concavidad). *Supongamos  $d \geq 2$  y  $\mathfrak{D} \subset \mathbb{R}^d$  convexo acotado.*

*Sea  $u \in C^2(\mathfrak{D}) \cap C(\bar{\mathfrak{D}})$ ,  $u|_{\partial\mathfrak{D}} = 0$ ,  $u|_{\mathfrak{D}} > 0$  que satisfice*

$$\Delta(u)(x) + b(x, u(x), \nabla u(x)) = 0, \quad \forall x \in \mathfrak{D},$$

donde  $b : \mathfrak{D} \times I \times \mathbb{R}^d \rightarrow \mathbb{R}$  con  $I \subset \mathbb{R}$ , la imagen de  $u$ . Si  $u$  y  $b$  satisfacen

- i) Para todo  $x \in \mathfrak{D}$  y  $\mathbf{p} \in \mathbb{R}^d$ ,  $b(x, \cdot, \mathbf{p})$  es estrictamente decreciente.
- ii) Para todo  $\mathbf{p} \in \mathbb{R}^d$ ,  $b(\cdot, \cdot, \mathbf{p}) : \mathfrak{D} \times I \rightarrow \mathbb{R}$  es armónica cóncava.
- iii) Para todo  $y \in \partial\mathfrak{D}$  y  $z \in \mathfrak{D}$ ,  $\limsup_{t \rightarrow 0^+} t^{-1}u(y + t(z - y)) > u(z)$ .

Entonces,  $u$  es cóncava en  $\bar{\mathfrak{D}}$ .

Estamos en condiciones de probar el siguiente resultado.

**Lema 2.9** ( $\mathcal{PS}$  cumple **P2** y **P3**). *Sea  $X$  vector aleatorio con distribución centralmente simétrica en  $\theta$  y sea  $f$  su densidad. Si  $f$  cumple **H1** y  $b(\cdot, \cdot, \mathbf{p}) : \mathfrak{D} \times [0, 1) \rightarrow \mathbb{R}$  dada por*

$$b(x, t, \mathbf{p}) = 2\langle \nabla \ln(f(x)), \mathbf{p} \rangle + 6f^{2/d}(x)(1 - t)$$

*es armónica cóncava para todo  $\mathbf{p} \in \mathbb{R}^d$ , entonces  $\mathcal{PS}$  se maximiza en  $\theta$  y decrece monótonamente en todo rayo que parte de  $\theta$ .*

*Demostración.* Suponemos, sin pérdida de generalidad, que el centro de simetría de  $\mathbf{X}$  es  $\mathbf{0}$ .

Sea  $\tilde{g}(x) = 1 - \mathbb{E}_x(e^{-\tau})$ , con una demostración análoga al Lema 2.8, resulta  $\tilde{g}(x) = \tilde{g}(-x)$ .

Si suponemos que  $\tilde{g}$  es cóncava, entonces se maximiza en  $\mathbf{0}$  pues

$$\tilde{g}(\mathbf{0}) = \tilde{g}((1/2)x + (1/2)(-x)) \geq (1/2)\tilde{g}(x) + (1/2)\tilde{g}(-x) = \tilde{g}(x), \quad \forall x \in \mathbb{R}^d.$$

Esto implica, para  $\lambda \in [0, 1]$  y  $x \in \mathbb{R}^d$ ,

$$\tilde{g}(\lambda x) = \tilde{g}((1 - \lambda)\mathbf{0} + \lambda x) \geq (1 - \lambda)\tilde{g}(\mathbf{0}) + \lambda\tilde{g}(x) \geq (1 - \lambda)\tilde{g}(x) + \lambda\tilde{g}(x) = \tilde{g}(x).$$

Por último,

$$\tilde{g}(\mathbf{0}) = \sup_{x \in \mathbb{R}^d} \tilde{g}(x) \implies \mathcal{PS}(\mathbf{0}) = \sup_{x \in \mathbb{R}^d} \mathcal{PS}(x)$$

y

$$\tilde{g}(x) \leq \tilde{g}(\lambda x) \quad \forall x \in \mathbb{R}^d \quad \forall \lambda \in [0, 1] \implies \mathcal{PS}(x) \leq \mathcal{PS}(\lambda x) \quad \forall x \in \mathbb{R}^d \quad \forall \lambda \in [0, 1].$$

Por lo tanto, utilizaremos el Teorema 2.1 para ver que la función  $\tilde{g}$  es cóncava y así obtener los resultados del lema.

Recordemos que  $\mathbb{E}_x(e^{-\tau})$  es solución de (2.12), entonces  $\tilde{g}(x) \in C^2(\mathfrak{D}) \cap C(\bar{\mathfrak{D}})$  es solución de

$$\begin{cases} \Delta(\tilde{g})(x) + b(x, \tilde{g}(x), \nabla\tilde{g}(x)) = 0, & x \in \mathfrak{D} \\ \tilde{g}(x) = 0, & x \in \partial\mathfrak{D}, \end{cases}$$

con

$$b(x, \tilde{g}(x), \nabla\tilde{g}(x)) = 2\langle \nabla \ln(f(x)), \nabla\tilde{g}(x) \rangle + 6f^{2/d}(x)(1 - \tilde{g}(x)).$$

Ahora bien, la condición i) del teorema siempre se satisface pues

$$\frac{\partial}{\partial \tilde{g}} b(x, \tilde{g}(x), \nabla\tilde{g}(x)) = -6f^{2/d}(x) < 0$$

independientemente de la distribución del vector aleatorio  $\mathbf{X}$ .

La condición ii) del teorema se cumple por hipótesis, mientras que la condición iii) se cumple para todo dominio  $\mathfrak{D}$  que tenga la propiedad de que una esfera ruede en el interior de  $\partial\mathfrak{D}$  (en particular el dominio  $\mathfrak{D}$  con borde suave como en nuestro caso) pues el Lema de Hopf (del punto en la frontera) asegura que la derivada normal (interior) para  $\tilde{g}$  es estrictamente positiva en todo  $\partial\mathfrak{D}$  y por lo tanto el límite es  $+\infty$ . En conclusión,  $\tilde{g}$  es cóncava, lo cual implica que se  $\mathcal{PS}$  cumple **P2** y **P3**.  $\square$

**Observación 2.4** (Caso uniforme). *Para el caso en el que  $\mathbf{X}$  es uniforme en  $\mathfrak{D}$  convexo simétrico basta notar que*

$$b(x, t, \mathbf{p}) = 6|\mathfrak{D}|^{-2/d}(1 - t)$$

*es constante en  $x$  y lineal positiva en  $t \in [0, 1)$ , por lo que resulta armónica cóncava.*

**Observación 2.5** (Otras profundidades). *A partir de la demostración del lema anterior es posible ver que la función  $\tilde{g}(x) = 1 - \mathbb{E}_x(e^{-\tau})$  podría ser tomada como otra función monótona de  $\tau$  que define una profundidad con las mismas propiedades que  $\mathcal{PS}$ ; el uso del logaritmo en la definición de  $\mathcal{PS}$  tiene como objetivo el destacar dentro del conjunto de puntos más centrales, los que tardan “un poquito más”, teniendo en cuenta que para  $\tau$  grande,  $\mathbb{E}_x(e^{-\tau})$  puede ser un número positivo muy pequeño.*

Ahora nos enfocaremos en probar **P6**, la continuidad de  $\mathcal{PS}$  como función  $F$ .

Sea  $\{\mathbf{X}_n\}_{n \geq 1}$  una sucesión de vectores aleatorios que convergen en distribución a  $\mathbf{X}_0$ . Para que se pueda definir  $\mathcal{PS}(x, F_n)$ , donde  $F_n$  es la distribución del vector  $\mathbf{X}_n$ , es necesario que la densidad de cada variable  $\mathbf{X}_n$ ,  $f_n(x)$ , así como las densidad límite,  $f_0(x)$  cumplan **H1**. Si denotamos  $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ , entonces debemos asumir:

$\forall n \in \mathbb{N}_0 : 1) \mathfrak{D}_n = \{x : f_n(x) > 0\} \subset \mathbb{R}^d$  es compacto y conexo con borde  $\partial\mathfrak{D}_n$  suave.

**H1n** 2) Existen  $f_{min}^n, f_{max}^n \in \mathbb{R}^+$  tal que  $0 < f_{min}^n \leq f_n(x) \leq f_{max}^n < \infty, \forall x \in \mathfrak{D}_n$ .

3)  $f_n(x)$  es diferenciable y  $\nabla(\ln(f_n(x)))$  es acotada y localmente Lipschitz,  $\forall x \in \mathfrak{D}_n$ .

Denotamos  $Y_t^0 = Y_t$  dada en (2.6) y

$$dY_t^n = \frac{\nabla \ln(f_n(Y_t^n))}{3(f_n(Y_t^n))^{2/d}} dt + \frac{1}{\sqrt{3}(f_n(Y_t^n))^{1/d}} dB_t, \quad n \in \mathbb{N}.$$

Para  $n \in \mathbb{N}_0$  tomamos

$$\tau_x^n = \inf\{t \geq 0 : Y_t^n \in \overline{\mathfrak{D}_n^c}, Y_0^n = x\}, \quad (2.13)$$

donde  $\mathfrak{D}_0 = \mathfrak{D}$ .

Queremos probar que

$$\mathcal{PS}(x, F_n) = -\ln(\mathbb{E}_x(e^{-\tau_x^n})) = -\ln(\mathbb{E}(e^{-\tau_x^n})) \xrightarrow{n \rightarrow \infty} -\ln(\mathbb{E}(e^{-\tau_x^0})) = -\ln(\mathbb{E}_x(e^{-\tau_x^0})) = \mathcal{PS}(x, F_0) \quad \forall x.$$

Utilizaremos que  $\tau_x^n \xrightarrow[n \rightarrow \infty]{P} \tau_x^0 \quad \forall x$  a partir del siguiente resultado de Shevchenko [52] que presentamos reescrito para nuestro caso particular.

**Teorema 2.2** (Convergencia de tiempos de parada. Shevchenko [52], Teorema 5.2.). *Sean*

$$d\tilde{Y}_t^n = b^n(\tilde{Y}_t^n) dt + \sigma^n(\tilde{Y}_t^n) dB_t,$$

donde  $B_t$  es un movimiento Browniano  $d$ -dimensional y los coeficientes  $b^n : \mathbb{R}^d \rightarrow \mathbb{R}^d$  y  $\sigma^n : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$  satisfacen las condiciones

$$\mathbf{c1)} \quad |b^n(x)|^2 + |\sigma^n(x)|^2 \leq D(1 + |x|^2); \quad x \in \mathbb{R}^d,$$

$$\mathbf{c2)} \quad |b^n(x) - b^n(y)|^2 + |\sigma^n(x) - \sigma^n(y)|^2 \leq D_r |x - y|^2; \quad x, y \in B(0, r) \subset \mathbb{R}^d,$$

para constante  $D, D_r > 0$ , para todo  $n \in \mathbb{N}_0$  y tales que

$$\mathbf{c3)} \quad b^n(x) \xrightarrow[n \rightarrow \infty]{} b^0(x), \quad \sigma^n(x) \xrightarrow[n \rightarrow \infty]{} \sigma^0(x), \quad x \in \mathbb{R}^d.$$

Sean  $\tau_x^n$  tiempos de parada dados por

$$\tau_x^n = \inf\{t \geq 0 : \varphi^n(\tilde{Y}_t^n) \geq 0, \tilde{Y}_0^n = x\}, \quad n \in \mathbb{N}_0, \quad (2.14)$$

donde

$$\mathbf{c4)} \quad \tau^0 < \infty \text{ casi seguramente.}$$

$$\mathbf{c5)} \quad \varphi^0 \in C^1(\mathbb{R}^d) \text{ y } \nabla \varphi^0 \text{ es localmente Lipschitz continua.}$$

$$\mathbf{c6)} \quad \text{Para todo } x \in \mathbb{R}^d, |\nabla \varphi^0(x)| > 0.$$

$$\mathbf{c7)} \quad \text{Para todo } r > 0,$$

$$\sup_{x \in B(0, r)} \{|\varphi^n(x) - \varphi^0(x)|\} \xrightarrow[n \rightarrow \infty]{} 0.$$

Entonces

$$\tau_x^n \xrightarrow[n \rightarrow \infty]{P} \tau_x^0.$$

Estamos en condiciones, ahora sí, de enunciar el siguiente lema.

**Lema 2.10** ( $\mathcal{PS}$  cumple **P6**). *Sea  $\{X_n\}$  una sucesión de vectores aleatorios tales que convergen en distribución a  $X_0$ . Si  $\{X_n\}_{n \geq 0}$  cumple **H1n** y  $\nabla f_n(x) \rightarrow \nabla f_0(x) \forall x$ , entonces*

$$\mathcal{PS}(x, F_n) \xrightarrow[n \rightarrow \infty]{} \mathcal{PS}(x, F_0), \quad \forall x.$$

*Demostración.* Las hipótesis en **H1n**, sumado a la convergencia de los gradientes implican que las condiciones **c1**) - **c3**) se satisfagan.

Para ver que se cumple **c4**), si tomamos  $q(x) = 0$ ,  $g(x) = 1$  y  $\phi(y) = 0$ , todas las hipótesis del Lema 2.5 se verifican y el Lema 2.6 asegura que existe solución única acotada  $h(x) \in C^2(\mathfrak{D}) \cap C(\bar{\mathfrak{D}})$  para el sistema

$$\begin{aligned} \frac{1}{6f_0^{2/d}(x)} \Delta h(x) + \frac{1}{3f_0^{2/d}(x)} \langle \nabla \ln(f_0(x)), \nabla h(x) \rangle &= -1, & x \in \mathfrak{D} \\ h(x) &= 0, & x \in \partial \mathfrak{D}. \end{aligned}$$

por lo que concluimos que  $\mathbb{E}(\tau_x^0) < \infty$  y por lo tanto  $\mathbb{P}(\tau_x^0 < \infty) = 1$ .

Para  $n \in \mathbb{N}_0$ ,  $x \in \mathbb{R}^d$  la distancia euclídea a  $\partial \mathfrak{D}_n$  es

$$d(x, \partial \mathfrak{D}_n) = \inf_{y \in \partial \mathfrak{D}_n} \|x - y\|.$$

Sea  $\psi^n$  la distancia signada a  $\partial \mathfrak{D}_n$ , definida de la siguiente manera

$$\psi^n(x) = \begin{cases} -d(x, \partial \mathfrak{D}_n), & x \in \mathfrak{D}_n \\ 0, & x \in \partial \mathfrak{D}_n \\ d(x, \partial \mathfrak{D}_n), & x \notin \mathfrak{D}_n. \end{cases}$$

Esta sucesión de funciones surge naturalmente a partir de la definición de  $\tau_x^n$  en (2.14) y permite que las definiciones de  $\tau_x^n$  en (2.13) y (2.14) coincidan.

Como  $\partial \mathfrak{D}_n$  es suave, existe  $\varepsilon_n > 0$  tal que  $\psi^n$  es suave en  $\Gamma_n = \{x \in \mathbb{R}^d : |\psi^n(x)| < \varepsilon_n\}$  (ver Lema 14.16 de Gilbarg y Trudinger [24]). Además  $\psi^0$  es solución de la ecuación Eikonal y vale que  $|\nabla \psi^0| = 1$  en  $\Gamma_0$ . Según la Observación 5.4. de Shevchenko [52], sólo es necesario que **c6**) valga en el conjunto  $\{x \in \mathbb{R}^d : \psi = 0\}$ .

La convergencia de  $f_n(x)$  a  $f_0(x)$  y la continuidad de las densidades implica que los soportes  $\mathfrak{D}_n$  convergen a  $\mathfrak{D}$  y por lo tanto  $\psi^n(x) \rightarrow \psi^0(x)$  y **c7**) vale.

Luego la sucesión de funciones  $\{\psi^n\}_{n \in \mathbb{N}_0}$  cumple casi con todas las propiedades necesarias, pero  $\nabla \psi^0$  puede no ser Lipschitz continua. Finalmente, solo basta definir las funciones  $\varphi^n$

como  $\varphi^n(x) = \psi^n(x)$  si  $x \in \Gamma_n$  para mantener todas las propiedades ya vistas y utilizar una convolución en la región restante para que  $\varphi^n(x)$  resulte  $C^1$  y  $\nabla\varphi^n$  sea localmente Lipschitz continua, para todo  $n \in \mathbb{N}_0$ .

A partir del Teorema 2.2 concluimos que  $\tau_x^n \xrightarrow[n \rightarrow \infty]{P} \tau_x^0 \forall x$ , lo cual implica  $\tau_x^n \xrightarrow[n \rightarrow \infty]{d} \tau_x^0 \forall x$ , que por el Lema de Portmanteau es equivalente a

$$\mathbb{E}_x(e^{-\tau^n}) = \mathbb{E}(e^{-\tau_x^n}) \xrightarrow[n \rightarrow \infty]{} \mathbb{E}(e^{-\tau_x^0}) = \mathbb{E}_x(e^{-\tau^0}) \forall x,$$

lo que finalmente conlleva a que

$$\mathcal{PS}(x, F_n) \xrightarrow[n \rightarrow \infty]{} \mathcal{PS}(x, F_0) \forall x$$

y el lema queda probado.  $\square$

### 2.2.2. Extensión a distribuciones con soporte no acotado

Para extender la Definición 2.3 de  $\mathcal{PS}$  a distribuciones con soporte no acotado, aplicamos la idea de “cortar” el dominio con bolas centradas en el origen que van creciendo.

Para  $\lambda > 0$ , sea  $B(\mathbf{0}, \lambda)$  la bola  $d$ -dimensional de centro  $\mathbf{0}$  y radio  $\lambda$ , en este caso las hipótesis se pueden resumir de la siguiente manera

$\forall \lambda > 0$  : 1)  $\mathcal{D}_\lambda := \mathcal{D} \cap B(\mathbf{0}, \lambda) \subset \mathbb{R}^d$  es compacto y conexo con borde  $\partial\mathcal{D}_\lambda$  suave.

**H1 $\lambda$**  2) Existen  $f_{min}^\lambda, f_{max}^\lambda \in \mathbb{R}^+$  tal que  $0 < f_{min}^\lambda \leq f(x) \leq f_{max}^\lambda < \infty, \forall x \in \mathcal{D}_\lambda$ .

3)  $f(x)$  es diferenciable y  $\nabla(\ln(f(x)))$  es acotada y localmente Lipschitz,  $\forall x \in \mathcal{D}_\lambda$ .

**Definición 2.5** (Profundidad de salida para soportes no acotados). *Sea  $f$  densidad con soporte  $\mathcal{D}$ , tal que se cumplen las hipótesis **H1 $\lambda$** . Para la difusión ya definida en (2.6)*

$$dY_t = \frac{\nabla \ln(f(Y_t))}{3(f(Y_t))^{2/d}} dt + \frac{1}{\sqrt{3}(f(Y_t))^{1/d}} dB_t,$$

tomamos  $\tau_\lambda = \inf\{t \geq 0 : Y_t \notin \mathcal{D}_\lambda\}$  y definimos

$$\mathcal{PS}_\lambda(x) = -\ln(\mathbb{E}_x(e^{-\tau_\lambda})), \quad \mathcal{PS}(x) = \lim_{\lambda \rightarrow \infty} \mathcal{PS}_\lambda(x).$$

**Observación 2.6.** *Claramente es una extensión de la definición anterior y para distribuciones con soporte acotado la definición de  $\mathcal{PS}$  es la misma que antes pues existe  $\lambda_0$  tal que  $\mathcal{D}_\lambda = \mathcal{D}$  para todo  $\lambda > \lambda_0$ ,*

En cuanto a las **propiedades**, las funciones  $\mathcal{PS}_\lambda$  son cuasi-cóncavas bajo las hipótesis vistas anteriormente. Como el límite de funciones cuasi-cóncavas también es cuasi-cóncava, todas las propiedades demostradas para distribuciones con soporte acotado valen para el caso en que la distribución tenga soporte no acotado.



# Capítulo 3

## Profundidad de salida en Grafos

### 3.1. Definición de $\mathcal{PS}$ en grafo $\mathcal{G}$

En esta sección daremos la versión de la profundidad de salida aplicada a grafos. La naturaleza discreta de estos objetos motiva el uso de paseos aleatorios en lugar de las difusiones utilizadas en el contexto continuo.

Sea  $V$  un conjunto finito y  $\mathcal{G} = (V, E)$  un grafo conexo sin bucles ni aristas múltiples. Denotamos  $u, v \in V$  a los vértices y  $u \rightarrow v$  si  $(u, v) \in E$ . Suponemos marcado un subconjunto de vértices  $\partial\mathcal{G} \subset V$  como “frontera” de  $\mathcal{G}$ , en la próxima sección veremos herramientas que permitan identificar este subconjunto de puntos cuando el grafo sea inducido por puntos de  $\mathbb{R}^d$ .

Consideramos el paseo aleatorio simple  $(X_i)_{i \in \mathbb{N}}$  con espacio de estados  $V \cup \Delta$  que da saltos a tiempo discreto según la siguiente dinámica:

- $p(u, v) = |u|^{-1} \mathbb{1}\{(u, v) \in E\} \mathbb{1}\{u \in V \setminus \partial\mathcal{G}\}$ , donde  $|u|$  es el grado del vértice  $u$ .
- $p(u, \Delta) = \mathbb{1}\{u \in \partial\mathcal{G}\}$ .

Es decir, estamos pensando al estado  $\Delta$  como *el afuera* del grafo (análogo a  $\overline{\mathcal{D}^c}$  en el capítulo anterior) y que el paseo  $(X_i)_{i \in \mathbb{N}}$  es un paseo en  $\mathcal{G}$  a tiempo discreto que salta a los vértices adyacentes eligiendo uniformemente hasta que llega a la frontera y una vez en  $\partial\mathcal{G}$  “sale” con probabilidad 1.

**Observación 3.1** (Estado  $\Delta$ ). *Desde el punto de vista probabilístico, la inclusión de  $\Delta$  al espacio de estados no es muy relevante para la dinámica, pero veremos en el Capítulo 5 los beneficios de esa incorporación a partir de las aplicaciones de  $\mathcal{PS}$  (ver Observación 5.1).*

**Definición 3.1** (Profundidad de salida en grafos). *Para el paseo  $(X_i)_{i \in \mathbb{N}}$  definido anteriormente, sea  $t = \hat{c}t$  un escalamiento temporal y  $T = \min\{t \in \mathbb{N} \cup \{0\} : X_t \in \Delta\}$ , tomamos*

$$\gamma_{\mathcal{G}}(u) = \gamma(u, \mathcal{G}) := \mathbb{E}(e^{-cT} | X_0 = u) = \mathbb{E}_u(e^{-cT}). \quad (3.1)$$

Definimos la función  $\mathcal{PS}_{\mathcal{G}} : V \rightarrow \mathbb{R}$ , profundidad de salida para el grafo  $\mathcal{G}$ ,

$$\mathcal{PS}_{\mathcal{G}}(u) = -\ln(\gamma(u)).$$

Por definición, si  $u \in \partial\mathcal{G}$ ,  $\gamma_{\mathcal{G}}(u) = e^{-c}$  pues en ese caso  $T = 1$  con probabilidad 1. Para calcular  $\gamma_{\mathcal{G}}(u)$  si  $u \in V \setminus \partial\mathcal{G}$  hacemos un análisis condicionando a un paso

$$\mathbb{E}_u(e^{-cT}) = \mathbb{E}_u(\mathbb{E}(e^{-c(T+1)}|X_1)) = e^{-c} \sum_v p(u, v) \mathbb{E}_v(e^{-cT}) = \frac{e^{-c}}{|u|} \sum_{u \rightarrow v} \gamma_{\mathcal{G}}(v). \quad (3.2)$$

Por lo tanto el problema puede reescribirse de la siguiente manera

$$\begin{cases} \gamma_{\mathcal{G}}(u) - e^{-c}|u|^{-1} \sum_{u \rightarrow v} \gamma_{\mathcal{G}}(v) & = 0, & u \in V \setminus \partial\mathcal{G} \\ \gamma_{\mathcal{G}}(u) & = e^{-c}, & u \in \partial\mathcal{G}. \end{cases} \quad (3.3)$$

Denotamos  $\mathbf{A}$  a la matriz de adyacencias de  $\mathcal{G}$ ,  $\mathbf{I}$  la matriz identidad  $|V| \times |V|$  y  $\mathbf{D}$  la matriz diagonal  $e^c|u|$ ; si definimos  $\mathbf{W} = (\mathbf{I} - \mathbf{D}^{-1}\mathbf{A})$  y reordenamos filas y columnas ( $i$  interiores,  $b$  frontera), obtenemos

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_{ii} & \mathbf{W}_{ib} \\ \mathbf{W}_{bi} & \mathbf{W}_{bb} \end{pmatrix}$$

Esto permite escribir (3.3) de manera matricial

$$\begin{pmatrix} \mathbf{W}_{ii} & \mathbf{W}_{ib} \end{pmatrix} \begin{pmatrix} \gamma_{\mathcal{G}}(x_i) \\ \gamma_{\mathcal{G}}(x_b) \end{pmatrix} = 0 \quad (3.4)$$

donde  $\gamma_{\mathcal{G}}(x_b)$  es un vector con valores  $e^{-c}$  de longitud  $|\partial\mathcal{G}|$  y vemos que  $\gamma_{\mathcal{G}}$  para el vector  $x_i$  de vértices interiores es solución de

$$\gamma_{\mathcal{G}}(x_i) = -\mathbf{W}_{ii}^{-1} (\mathbf{W}_{ib} \gamma_{\mathcal{G}}(x_b)). \quad (3.5)$$

Además se puede pensar a  $\gamma_{\mathcal{G}}$  como solución de un problema de punto fijo tomando  $\gamma_{\mathcal{G}}^{(1)}(u) = e^{-c} \forall u$  e iterando

$$\gamma_{\mathcal{G}}^{(i+1)} = \mathbf{D}^{-1}\mathbf{A}\gamma_{\mathcal{G}}^{(i)}, \quad i \in \mathbb{N} \quad (3.6)$$

recordando  $\gamma_{\mathcal{G}}^{(i)}(u) = e^{-c} \forall u \in \partial\mathcal{G}, \forall i$ .

Luego, para calcular la profundidad de cada vértice,  $\mathcal{PS}_{\mathcal{G}}(u) = -\ln(\gamma_{\mathcal{G}}(u)) \forall u$ .

**Observación 3.2.** Si el grafo “no es tan grande” es posible el calcular  $\mathbf{W}_{ii}^{-1}$  y  $\mathcal{PS}_{\mathcal{G}}$  puede ser resuelta de manera matricial como en (3.5); mientras que si el número de vértices es grande, es preferible el cálculo de  $\mathcal{PS}_{\mathcal{G}}$  como solución de un problema de punto fijo (3.6). Notar que grafos con matriz de adyacencia rala reducen notablemente el tiempo de cálculo. En teoría, es posible utilizar Monte Carlo, simulando  $K$  paseos diferentes partiendo de cada vértice interior  $u$  ( $K$  grande) y calculando  $\gamma_{\mathcal{G}}(u)$  como el promedio de  $e^{-cT_k}$  para  $k \in \{1, \dots, K\}$ ; vale aclarar que esto es sólo una suposición y no utilizamos esta herramienta en las aplicaciones mostradas en el Capítulo 5.

Conexidad. En la Observación 2.2 explicamos que la condición de conexidad del dominio sólo aparece para facilitar la exposición. En este caso, si el grafo  $\mathcal{G}$  contiene más de una componente conexa, cada componente  $\mathcal{G}'$  tiene su propia frontera y profundidad  $\mathcal{PS}_{\mathcal{G}'}$ .

Aristas con pesos. Si  $\tilde{\mathcal{G}} = (V, \tilde{E})$  es un grafo con pesos  $\omega_{uv}$  en las aristas  $(u, v) \in E$ , es posible introducir esta información en el paseo aleatorio considerando  $(\tilde{X}_t)_{t \in \mathbb{N}}$  con el mismo espacio de estados que antes pero que da saltos a tiempo discreto según la siguiente dinámica modificada:

- $p(u, v) = (w_{uv} / \sum_v w_{uv}) \mathbb{1}\{u \in V \setminus \partial\tilde{\mathcal{G}}\}$ .
- $p(u, \Delta) = \mathbb{1}\{u \in \partial\tilde{\mathcal{G}}\}$ .

El paseo en  $\tilde{\mathcal{G}}$  ya no salta a los vértices adyacentes eligiendo uniformemente, sino que lo hace con probabilidades proporcionales a los pesos del grafo hasta que llega a la frontera y sale con probabilidad 1.

Por último, en caso de trabajar con un multigrafo, las aristas múltiples pueden ser consideradas como aristas con mayor peso, por lo que se puede trabajar del mismo modo que en la situación anterior.

### 3.1.1. $\mathcal{PS}_{\mathcal{G}}$ y detección de comunidades en grafos

Como habíamos comentado en la Introducción,  $\mathcal{PS}_{\mathcal{G}}$  tiene relación con un procedimiento de detección de comunidades en grafos dado en Smith et al. [53], el cual fue comentado en Hashimoto et al. [26] que puede ser visto como una aplicación de la métrica LTHT.

Comenzamos introduciendo el problema de detección y la solución hallada por los autores y comentaremos en qué caso los resultados son análogos a los aquí obtenidos.

Sea  $G = (V, E)$  un grafo y  $\Theta : V \rightarrow \{0, 1\}^{|V|}$  tal que  $\Theta(v) = 1$  implica que el vértice está *infectado* y  $\Theta(v) = 0$  que no. Supongamos que existe un conjunto de vértices  $z = \{v_1, \dots, v_C\} \subset V$  “observados” y una densidad (o función de probabilidad puntual)  $f$  tal que  $f(z(v)|\Theta)$  es la probabilidad de que el vértice observado  $v$  esté infectado.

El problema que se plantea es el de detectar el subconjunto de vértices  $V_1 \subset V$ ,  $V_1 = \{v \in V : \Theta(v) = 1\}$ . Este problema es abordado calculando  $\theta_v = \mathbb{P}(\Theta(v) = 1|z)$  para todos los vértices del grafo a partir del desarrollo de un método Bayesiano basado en paseos aleatorios en el grafo y que cada vértice tiene una probabilidad *a priori* de infectar a sus vecinos  $\psi_v$ .

En el caso en que las probabilidades de salto dadas *a priori* al paseo aleatorio sean uniformes entre los vecinos, los autores obtienen como solución

$$\theta_v = \frac{\psi_v}{|V|} \sum_{v \rightarrow u} \theta_u, \quad v \in V \setminus z, \quad (3.7)$$

mientras que  $\theta_v = f(z(v)|\Theta)$  para  $v \in z$ .

Para hacer una conexión con métodos de detección de clusters basados en el espectro del laplaciano de un grafo, los autores definen el *operador laplaciano generalizado*

$$L^\psi = I - \Psi D^{-1} A$$

donde  $\Psi$  es la matriz diagonal  $\psi_v$ ,  $D$  es la matriz diagonal  $|v|$  y  $A$  es la matriz de adyacencias de  $G$  y llaman *ecuación de propagación armónica* a

$$\begin{pmatrix} L_{ii}^\psi & L_{ib}^\psi \end{pmatrix} \begin{pmatrix} \theta_i \\ \theta_b \end{pmatrix} = 0 \quad (3.8)$$

donde las filas y columnas de  $L^\psi$  fueron permutadas del mismo modo que en (3.4).

Por lo tanto, es posible observar en (3.7) que tomando  $\psi_v = e^{-c}$  y  $f(z(v)|\Theta) = e^{-c} \forall v \in \mathbf{z}$ , resulta  $\theta_v = \mathcal{PS}_{\mathcal{G}}(v)$ , mientras que (3.8) tiene como solución al problema de detección de comunidades el mismo resultado que en (3.5) obtenemos para calcular la profundidad de los puntos interiores.

## 3.2. Grafos inducidos por puntos en $\mathbb{R}^d$

Queremos aplicar  $\mathcal{PS}_{\mathcal{G}}$  a un grafo cuyo vértices provengan de una muestra aleatoria de una distribución  $F$  en  $\mathbb{R}^d$ . Aquí presentamos los grafos utilizados en las aplicaciones del Capítulo 5; para completar la implementación de  $\mathcal{PS}_{\mathcal{G}}$  a grafos en  $\mathbb{R}^d$ , mostraremos un procedimiento para “marcar la frontera” de un grafo en general y daremos ideas de cómo extender la profundidad de la muestra aleatoria (vértices de  $\mathcal{PS}_{\mathcal{G}}$ ) a todo el espacio.

En los *grafos de vecindad*, que definiremos formalmente en el Capítulo 4, el grado de los vértices crece con  $n$  pero converge a 0 cuando es dividido por  $n$ ; mientras que en el grafo llamado Triangulación de Delaunay, el número de vecinos por vértice es una variable aleatoria que no crece con  $n$ .

Nos centramos en la aplicación a grafos donde la matriz de adyacencias puede ser obtenida a través de una función determinística de los puntos pues al utilizar grafos aleatorios es posible tener diferentes grafos para una misma muestra aleatoria, generando diferentes  $\mathcal{PS}_{\mathcal{G}_n}$  para una misma muestra. Sin embargo, se demostrará en el próximo capítulo que los resultados de consistencia entre la definición empírica y teórica de la profundidad se cumplen para un amplio espectro de grafos.

### 3.2.1. Grafos inducidos por procesos puntuales

#### Grafos de vecindad

Sea  $X_n$  un conjunto de puntos en  $\mathbb{R}^d$ , comenzamos introduciendo grafos clásicos de la literatura donde la matriz de adyacencias depende únicamente de la vecindad (distancia) entre los puntos.

El grafo  $r$  vecinos (que denotaremos  $\mathcal{R}_n$ ) es aquel en el cual cada punto  $x \in \mathcal{X}_n$  se conecta con todos los puntos de la muestra dentro de la bola cerrada  $B(x, r_n)$  (la cuál puede ser calculada con cualquier métrica, no necesariamente euclídea y utiliza el mismo radio  $r_n$  para todos los puntos). Este grafo posee una matriz de adyacencias simétrica pues  $y \in B(x, r_n)$  si y solo si  $x \in B(y, r_n)$ .

En el grafo de vecinos cercanos ( $\mathcal{K}_n$ ), fijado  $k_n \in \mathbb{N}$  (la cantidad de vecinos a utilizar), se conecta cada punto con sus  $k_n$  vecinos más cercanos; el grafo puede ser visto de manera similar al anterior, pero en este caso la bola utilizada  $B(x, r_n^x)$  tiene un radio variable  $r_n^x$  que depende del punto  $x \in \mathcal{X}_n$ . Este es un grafo dirigido y la matriz de adyacencias puede no ser simétrica pues es posible que  $y$  sea uno de los  $k_n$  vecinos más cercanos de  $x$ , pero que esto no suceda a la inversa.

En la literatura existen otros dos grafos (no dirigidos) similares a  $\mathcal{K}_n$  pero con matriz de adyacencia simétrica; uno es el que une con aristas a los puntos  $x$  e  $y$  si por lo menos uno de los dos puntos está dentro de los  $k_n$  vecinos más cercanos del otro; mientras que el otro grafo (generalmente denominado  $k_n$  vecinos mutuos) es el que une a los puntos si y solo si ambos puntos están dentro de los  $k_n$  vecinos más cercanos del otro.

Para poder asegurar que estos grafos sean conexos y obtener consistencia claramente es necesario imponer condiciones a los parámetros  $r_n$  y  $k_n$ , las cuales explicitaremos en el próximo capítulo.

### Triangulación de Delaunay $\mathcal{T}_n$

Hemos visto en la Sección 3.1 que en el caso en que el paseo se realiza en un grafo  $\mathcal{G}$ , la solución del problema depende de la matriz de adyacencias de  $\mathcal{G}$  y concluimos que utilizar grafos en los que los vértices tengan pocos vecinos hace que la solución se alcance más rápido computacionalmente hablando. Ahora presentaremos el grafo que inspiró esta Tesis y comentamos en la Introducción, la *Triangulación de Delaunay*.

Sea  $\mathcal{X}_n = \{X_1, \dots, X_n\}$  un conjunto de puntos en  $\mathbb{R}^d$  en posición general, es decir, tal que  $\text{span}(\mathcal{X}_n) = \mathbb{R}^d$  y ningún subconjunto de  $d + 2$  puntos en  $\mathcal{X}_n$  pertenece al borde de una bola que no contiene puntos de  $\mathcal{X}_n$  en su interior ( $\mathcal{X}_n$  es una muestra aleatoria de una variable con distribución continua en nuestro caso y con probabilidad 1 está en posición general).

Se define la *celda de Voronoi*  $\mathcal{V}_i$  del punto  $X_i$  como el conjunto de puntos del espacio que están más cerca a  $X_i$  que de cualquier otro punto de la muestra; más precisamente  $\mathcal{V}_i = \{x \in \mathbb{R}^d : |x - X_i| \leq |x - X_j|, j = 1, \dots, n, j \neq i\}$ . El conjunto  $\{\mathcal{V}_1, \dots, \mathcal{V}_n\}$  genera una partición de  $\mathbb{R}^d$  que se denomina *teselación de Voronoi*. Para generar la *Triangulación de Delaunay* de  $\mathcal{X}_n$ , que denotaremos  $\mathcal{T}_n$ , se une con una arista cada par de puntos  $X_i$  y  $X_j$  que compartan una cara  $(d - 1)$ -dimensional de la teselación; es por esto que se dice que la Triangulación de Delaunay es el grafo dual de la teselación de Voronoi.

El nombre Triangulación de Delaunay es comunmente utilizado para cualquier dimensión

$d$ , por más de que claramente sólo en dimensión 2 los simples generados por este grafo son triángulos. Una caracterización conocida que puede ser tomada como definición de  $\mathcal{T}_n$  es que para todo simplex de esta triangulación (“Simplex Delaunay”), la esfera circunscrita no contiene puntos de la muestra  $\mathcal{X}_n$  en su interior. La condición de que los puntos estén en posición general no sólo implica la existencia de la triangulación sino también su unicidad.

En Bern, Eppstein y Yao [5], para un Proceso Puntual de Poisson  $\mathcal{N}$  de intensidad 1 en  $\mathbb{R}^d$ ,  $\mathcal{G}_{\mathcal{N}}$  la Triangulación de Delaunay de  $\mathcal{N}$  y  $\mathcal{D} = [0, n^{1/d}]^d$ , muestran que el grado máximo esperado para los vértices de  $\mathcal{G}_{\mathcal{N}} \cap \mathcal{D}$  es  $\Theta(\log(n)/\log(\log(n)))$ , con una constante que depende sólo de la dimensión  $d$ .

Por otro lado en Dwyer [20], para una muestra aleatoria de  $n$  puntos distribuidos uniformemente en la bola unitaria  $d$ -dimensional, se prueba que la esperanza de los grados de un vértice en la Triangulación de Delaunay de los puntos es independiente de  $n$  y crece con  $d$ .

Es de esperar entonces (por lo menos en el caso uniforme) que el grado de los vértices en este grafo sea sustancialmente menor a la cantidad de vecinos generalmente utilizados en los grafos de vecindad presentados anteriormente.

Existen varios algoritmos para calcular el grafo entre los que se encuentra el de Bowyer - Watson (algoritmo incremental) y algoritmos de “dividir y conquistar” los cuales son computacionalmente rápidos.

La Triangulación de Delaunay puede ser utilizada para estimar densidades (Schaap [49]) y veremos en la próxima sección que existen métodos de inferencia de frontera del soporte de la distribución basados en  $\mathcal{T}_n$ .

### 3.2.2. Marcado de frontera

Por definición, las profundidades  $\mathcal{PS}$  y  $\mathcal{PS}_{\mathcal{G}}$  presentadas dependen de las fronteras tanto del soporte como del grafo. Como la única información de una muestra aleatoria es la posición de los puntos, es necesario definir alguna “frontera empírica” que converja en algún sentido a la frontera teórica.

Existen herramientas para estimar el soporte de una densidad y/o el borde de éste. Típicamente, si quisieramos aplicar el concepto sólo a distribuciones con soporte convexo, la cápsula convexa de las observaciones puede ser vista como el mejor estimador. Algunos resultados sobre este estimador se encuentran en Dümbgen y Walther [19], Efron [22] y Reitzner [44].

El uso de la cápsula convexa coincide con la técnica de marcar los puntos  $X_i$  que tienen una celda de Voronoi no acotada en la teselación de Voronoi. Para poder incluir soportes no convexos, se puede utilizar una proporción  $\alpha$  de puntos cuyas celdas de Voronoi sean las más grandes; aplicando esta idea se realizaron algunas simulaciones con resultados aceptables en casos puntuales, pero para usar este enfoque sistemáticamente habría que pensar en crosvalidar ese parámetro  $\alpha$  y obtener resultados de consistencia como los que comentamos más adelante en **H2**.

En el caso general, el estimador de soporte más estudiado es el propuesto por Devroye y Wise [14] (ver Chevalier [8] también), denominado estimador de Devroye-Wise:

$$\mathfrak{D}_r = \bigcup_i^n B(X_i, r),$$

donde  $B(x, r)$  es la bola cerrada centrada en  $x$  de radio  $r > 0$  y  $r = r_n$  debe ser optimizado.

Muchas variantes de este estimador se han propuesto, por ejemplo permitiendo que  $r$  varíe localmente, demostrando la versatilidad de esta herramienta. En Cuevas y Rodríguez-Casal [13] se muestra un resultado de consistencia universal y tasas de convergencia bajo condiciones amplias para este estimador.

En las simulaciones utilizaremos una generalización del estimador de Devroye-Wise, la versión de Aaron y Bodart [2]:

$$\mathfrak{D}_n = \bigcup_{X \in \mathcal{X}_n} \mathcal{H}(B(X, r_n) \cap \mathcal{X}_n), \quad (3.9)$$

donde  $\mathcal{H}(A)$  es la cápsula convexa del conjunto  $A$  y el radio  $r_n$  es el sugerido por los autores y está dado por

$$M_n(\mathcal{X}_n) = \max_{j \neq i} \min_i \{\|X_i - X_j\|\}, \quad r_n = M_n(\mathcal{X}_n)(n/\ln(n))^{1/(d(d+1))}. \quad (3.10)$$

Este estimador puede verse como una extensión de  $\mathcal{H}(\mathcal{X}_n)$  a soportes no convexos.

En la literatura también existen otros tipos de estimadores como el presentado en Aaron [1] donde se utilizan poliedros provenientes de la Triangulación de Delaunay y  $k$  vecinos más cercanos y Cuevas y Fraiman [12] donde se considera un método “plug-in” basado en estimadores de densidad.

El estimador (3.9), y muchos otros, no sólo cumplen con  $|\mathfrak{D} \Delta \mathfrak{D}_n| \rightarrow 0$  cuando la muestra crece, sino que también sus fronteras convergen en distancia Hausdorff a la verdadera frontera con probabilidad uno, es decir

$$\mathbf{H2} \quad d_H(\partial \mathfrak{D}_n, \partial \mathfrak{D}) = \max \left\{ \sup_{x \in \partial \mathfrak{D}_n} \inf_{y \in \partial \mathfrak{D}} \|x - y\|, \sup_{y \in \partial \mathfrak{D}} \inf_{x \in \partial \mathfrak{D}_n} \|x - y\| \right\} \xrightarrow{n \rightarrow \infty} 0, \quad \text{casi seguramente.}$$

Esta propiedad es fundamental para que  $\mathcal{PS}_{\mathcal{G}_n}$  tenga aplicaciones en soportes generales no sólo convexos y nos va a permitir mostrar consistencia en el próximo capítulo.

Para poder definir  $\partial \mathcal{G}_n$  nos basamos en el siguiente lema:

**Lema 3.1** (Existencia de la frontera. Aaron y Bordart[2]. Lema 5). *Sea  $\mathfrak{D}_n$  dado en (3.9) con  $r_n$  dado por (3.10). Existe eventualmente casi seguramente un número finito  $N_\delta$  de poliedros  $d - 1$  dimensionales  $\sigma_1, \dots, \sigma_{N_\delta}$  tales que:*

$$i) \partial \mathcal{D}_n = \bigcup_{i=1}^{N_\partial} \sigma_i$$

ii) Para todo  $i \in \{1, \dots, N_\partial\}$  existe un conjunto de  $d$  puntos  $\{X_{1,i}, \dots, X_{d,i}\} \subset \mathcal{X}_n$  y un único vector unitario  $v_i$ , normal a  $\sigma_i$  y apuntando hacia afuera de  $\mathcal{D}_n$ , tal que

$$ii-a) \sigma_i \subset \mathcal{H}(\{X_{1,i}, \dots, X_{d,i}\}),$$

$$ii-b) \text{ existe } i^* \text{ tal que } \{X_{1,i}, \dots, X_{d,i}\} \subset B(X_{i^*}, r_n),$$

$$ii-c) \text{ para todo } j \text{ tal que } \{X_{1,i}, \dots, X_{d,i}\} \subset B(X_j, r_n), \text{ para todo } k \text{ tal que } X_k \in B(X_j, r_n), \\ \text{ para todo } x \in \sigma_i, \text{ se tiene que } \langle x - X_k, v_i \rangle \geq 0.$$

Luego, tomamos

$$\partial \mathcal{G}_n = \bigcup_{i=1}^{N_\partial} \{x \in \mathcal{X}_n : x \in \sigma_i\}. \quad (3.11)$$

Un ejemplo del marcado de esta elección se puede ver en la Figura 3.1 para una muestra aleatoria de 500 puntos con distribución uniforme en el anillo  $B(0, 2) \setminus B(0, 1)$ .

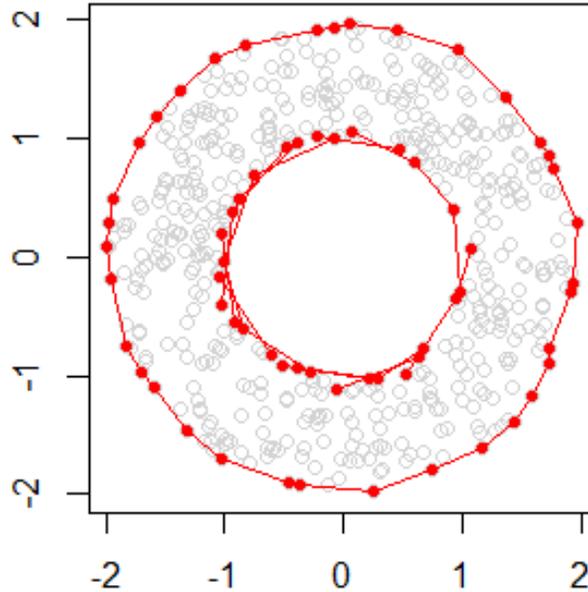


Figura 3.1: Puntos considerados  $\partial \mathcal{G}_n$  y frontera  $\partial \mathcal{D}_n$ .

### 3.2.3. Cálculo de $\mathcal{PS}_{\mathcal{G}_n}$ para una muestra aleatoria y extensión a $\mathbb{R}^d$

Sea  $\mathcal{X}_n = \{X_1, \dots, X_n\}$  una muestra aleatoria de  $F$  en  $\mathbb{R}^d$ , donde  $f$  cumple las hipótesis **H1** y sean  $\mathcal{D}_n$  y  $\partial\mathcal{D}_n$  estimadores de  $\mathcal{D}$  y  $\partial\mathcal{D}$  tal que se cumple **H2**.

Para  $\mathcal{G}_n = (\mathcal{X}_n, E_n)$  grafo inducido por un proceso puntual, tomamos  $\partial\mathcal{G}_n$  definida en (3.11) como frontera de  $\mathcal{G}_n$ ,  $\Delta = \mathcal{D}_n^c$  y calculamos  $\gamma_n(X_i) \forall i$  siguiendo (3.5) o (3.6) según se prefiera, para un escalamiento temporal  $c = c_n$  que asegura la consistencia del procedimiento y definiremos en el próximo capítulo.

Denotaremos  $\mathcal{PS}_n$  a  $\mathcal{PS}_{\mathcal{G}_n}$ . Para extender a  $x \notin \mathcal{X}_n$ , si  $x \notin \mathcal{D}_n$  por definición  $\gamma_n(x) = 1$  y  $\mathcal{PS}_n(x) = 0$ , mientras que si  $x \in \mathcal{D}_n$  siguiendo (3.2),

$$\gamma_n(x) = (e^c |v(x)|)^{-1} \sum_{x \rightarrow X_j} \gamma_n(X_j), \quad (3.12)$$

donde  $x \rightarrow X_j$  son los vecinos que tendría  $x$  dentro de la muestra  $\mathcal{X}_n$  y  $|v(x)|$  es la cantidad de vecinos que tendría. Luego calculamos  $\mathcal{PS}_n(x) = -\ln(\gamma_n(x))$ .

Esta forma de extender  $\mathcal{PS}_n$  se basa en que es conocido el grafo utilizado para calcularla (la función de vecindad utilizada) y puede ser reemplazada por una manera más simple, la cuál es que  $\mathcal{PS}_n(x)$  tome el valor de  $\mathcal{PS}_n(X_i)$  para el punto  $X_i \in \mathcal{X}_n$  más cercano a  $x$ , lo cuál sería extender  $\mathcal{PS}_n$  como constante en toda la celda de voronoi  $\mathcal{V}_i$  y dejando como definición

$$\mathcal{PS}_n(x) = \sum_{i=1}^n \mathcal{PS}_n(X_i) \mathbb{1}\{x \in \mathcal{V}_i\}. \quad (3.13)$$

Otra forma de extender para los puntos  $x \in \mathcal{D}_n$  si  $x \notin \mathcal{X}_n$ , es la de interpolación lineal de los valores  $\mathcal{PS}_n(X_{j_1}), \dots, \mathcal{PS}_n(X_{j_{d+1}})$ , para los  $X_{j_k}$  tales que  $x$  pertenece al “simplex Delaunay”  $X_{j_1}, \dots, X_{j_{d+1}}$  definido anteriormente.

En la Sección 4.2 aplicaremos estas técnicas a un ejemplo en dimensión 1 para poder apreciar las diferencias de extensión gráficamente.



# Capítulo 4

## Resultados asintóticos y $\mathcal{PS}$ para $d = 1$

### 4.1. Consistencia

Queremos mostrar convergencia de la profundidad de salida empírica a la poblacional para grafos donde la relación entre vértices adyacentes esté dada con la mayor generalidad posible; comenzamos presentando, a través del grafo  $\mathcal{K}_n$  para ganar intuición, cuáles son las condiciones a imponer a los grafos utilizados y el necesario escalamiento temporal para poder conseguir el resultado en los casos en que el grado de los vértices crece con  $n$  pero converge a 0 cuando es dividido por  $n$ . Además comentaremos el porqué de los coeficientes de la difusión  $Y_t$  definida en (2.6), para después mostrar consistencia tanto para grafos de vecindad generalizada como para la Triangulación de Delaunay de puntos provenientes de un proceso de Poisson homogéneo.

#### 4.1.1. Escalamiento temporal y coeficientes de $Y_t$

Para obtener consistencia entre las definiciones de  $\mathcal{PS}_{\mathcal{G}}$  y  $\mathcal{PS}$  es necesario utilizar un principio de invarianza que implique la convergencia de las trayectorias debidamente escaladas de los paseos aleatorios a las trayectorias de la difusión.

En el caso de principio de invarianza para cadenas de Markov convergiendo a difusiones con condiciones de frontera, el resultado clásico se debe a Stroock y Varadhan [55]; este resultado se basa en que si los momentos de un salto en el paseo (primer momento = esperanza = deriva, segundo momento = varianza = difusión, momentos de orden mayor = existencia de saltos “grandes”) son continuos y están “controlados” y además se cumplen condiciones que aseguran la reflexión en el borde del dominio, el paseo aleatorio escalado en el tiempo convergerá a una difusión de Itô.

Antes de mostrar consistencia queremos adentrarnos en la comprensión del escalamiento temporal que vamos a utilizar y el cómo surgen los coeficientes de  $Y_t$  en la definición de  $\mathcal{PS}$ ;

para esto introduciremos de manera informal los razonamientos probados en Hashimoto et al. [27] y utilizaremos un grafo de vecindad donde los vértices provengan de una muestra aleatoria.

Más precisamente, sea  $\mathcal{X}_n$  una muestra aleatoria de la distribución  $F$  en  $\mathbb{R}^d$ , donde la densidad  $f$  cumple las hipótesis **H1** y sea  $\mathcal{G}_n = (\mathcal{X}_n, E_n)$ , donde  $E_n = \{(x, y) : x, y \in \mathcal{X}_n, y \in B(x, \varepsilon_n(x))\}$ , con  $\varepsilon_n(x)$  una función de escala local. Consideramos el paseo aleatorio simple  $\{X_n(t)\}_{t \in \mathbb{N}}$  con incrementos de tiempo  $h_n$  (es decir, “salta” a tiempos  $h_n, 2h_n, 3h_n, \dots$ ) en  $\mathcal{G}_n$ ; queremos encontrar las condiciones a imponer a  $\varepsilon_n(x)$  y  $h_n$  para obtener consistencia.

Utilizaremos en esta sección la misma notación de Stroock y Varadhan [55] y Hashimoto et al. [27] para los coeficientes.

Si denotamos  $v_n(x)$  al conjunto de vértices adyacentes de  $x \in \mathcal{X}_n$ , para  $A \subset \mathbb{R}^d$  tenemos que el kernel de transición de  $\{X_n(t)\}$  es

$$\mathbb{P}(X_n(t+1) \in A | X_n(t) = x) = \frac{|v_n(x) \cap A|}{|v_n(x)|}$$

lo cual implica que la deriva de  $\{X_n(t)\}$  es

$$b_n^i(x) = \frac{1}{h_n} \sum_{y \in v_n(x)} \frac{y_i - x_i}{|v_n(x)|}, \quad i = 1, \dots, d,$$

mientras que los coeficientes de difusión vienen dados por

$$a_n^{ij}(x) = \frac{1}{h_n} \sum_{y \in v_n(x)} \frac{(y_i - x_i)(y_j - x_j)}{|v_n(x)|}, \quad 1 \leq i, j \leq d,$$

y los momentos de orden mayor pueden denotarse

$$\Delta_{n,\alpha}(x) = \frac{1}{h_n} \sum_{y \in v_n(x)} \frac{|y - x|^{2+\alpha}}{|v_n(x)|}, \quad \alpha > 0.$$

Luego, surge naturalmente la necesidad de que existan límites de los coeficientes  $b_n \rightarrow b$ ,  $a_n \rightarrow a$  y que estos sean uniformes en subconjuntos compactos y con probabilidad 1; además, para converger a una difusión con trayectorias continuas es necesario que los saltos “grandes” desaparezcan, lo cuál es consecuencia de que los momentos de orden mayor vayan a 0 y puede resumirse en la necesidad de que se cumpla que

$$\sup_{x \in \mathcal{X}_n} \Delta_{n,1}(x) \rightarrow 0, \quad \text{con probabilidad 1.}$$

Para mejorar la exposición, denotamos

$$H_n(x) = \frac{1}{h_n} \sum_{y \in v_n(x)} \frac{h(y-x)}{|v_n(x)|}.$$

Notar que para  $h(y - x) = y_i - x_i$  resulta  $H_n(x) = b_n^i(x)$ ; mientras que si  $h(y - x) = (y_i - x_i)(y_j - x_j)$ ,  $H_n(x) = a_n^{ij}(x)$  y si  $h(y - x) = |y - x|^{2+\alpha}$ ,  $H_n(x) = \Delta_{n,\alpha}(x)$ .

Para obtener convergencia, al utilizar la desigualdad de Hoeffding

$$\mathbb{P}(|H_n(x) - \mathbb{E}(H_n(x))| \geq t) \leq \exp\left(-\frac{2h_n^2|v_n(x)|^2t^2}{|v_n(x)|(h(y-x))^2}\right),$$

resulta que las condiciones necesarias para utilizar Borel-Cantelli (y que la convergencia sea casi segura), pueden resumirse de la siguiente manera:

- a)  $h_n \rightarrow 0$ , lo cual es claro teniendo en cuenta que es el escalamiento temporal.
- b)  $|v_n(x)| \in \omega(n^{2/(d+2)} \ln(x)^{d/(d+2)})$ .
- c) existe función determinística continua positiva  $\bar{\varepsilon}(x)$  tal que

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathcal{X}_n} |h_n^{-1/2} \varepsilon_n(x) - \bar{\varepsilon}(x)| \rightarrow 0, \quad \text{con probabilidad 1.}$$

En caso de que se satisfagan **a)**, **b)** y **c)**, los límites de los coeficientes son:

$$b_n^i(x) \rightarrow \frac{f_i(x)}{3f(x)}(\bar{\varepsilon}(x))^2, \quad a_n^{ii}(x) \rightarrow \frac{1}{3}(\bar{\varepsilon}(x))^2 \quad \text{y} \quad a_n^{ij}(x) \rightarrow 0, \quad (4.1)$$

donde  $f_i(x)$  es la derivada parcial en la  $i$ -ésima coordenada de la densidad  $f$ .

Si pensamos en el grafo  $\mathcal{K}_n$ ,  $|v_n(x)| = k_n$ , por lo que la sucesión de números positivos  $k_n$  según **b)** debe satisfacer

$$\forall C > 0, \exists N_C > 0, \text{ tal que si } n \geq N_C \implies k_n \geq C n^{\frac{2}{d+2}} \ln(n)^{\frac{d}{d+2}}. \quad (4.2)$$

Por otro lado,  $\varepsilon_n(x)$  es la distancia al  $k_n$ -ésimo vecino más cercano de  $x$ . Recordamos que el estimador de densidad de vecinos más cercanos dado en Devroye y Wagner [15] cumple

$$\hat{f}_n(x) \stackrel{\text{def}}{=} \frac{k_n}{n V_n^x} \rightarrow f(x), \text{ uniformemente, con probabilidad 1,}$$

donde  $V_n^x$  es el volumen de la bola centrada en  $x$  que contiene los  $k_n$  puntos de la muestra más cercanos a  $x$ . Reescribiendo  $V_n^x = (\varepsilon_n(x))^d \omega_d$ , con  $\omega_d$  el volumen de la bola unitaria  $d$ -dimensional, obtenemos que  $h_n = (k_n/(n\omega_d))^{2/d}$  es el cuadrado de la velocidad con la que  $\varepsilon_n$  va a 0 que verifica **a)** y para la condición **c)** resulta  $\bar{\varepsilon}(x) = (f(x))^{-1/d}$ .

Reemplazando en (4.1), obtenemos que

$$b_n(x) \rightarrow b(x) = \frac{\nabla f(x)}{3f(x)}(f(x))^{-2/d} = \frac{\nabla(\ln(f(x)))}{3(f(x))^{2/d}}, \quad a_n(x) \rightarrow a(x) = \frac{1}{3(f(x))^{2/d}} \mathbf{I}, \quad (4.3)$$

donde  $\mathbf{I}$  es la matriz identidad  $d$ -dimensional.

Estamos en condiciones, ahora sí, de definir los grafos a utilizar.

**Definición 4.1** (Grafos de vecindad). Sea  $\varepsilon_n : \mathcal{X}_n \rightarrow \mathbb{R}^+$  una función de escala local y sean  $c_n$  constantes positivas tales que

$$c_n \rightarrow 0, \quad c_n n^{1/(d+2)} \ln(n)^{-1/(d+2)} \rightarrow \infty \quad (4.4)$$

y

$$\sup_{x \in \mathcal{X}_n} \{|c_n^{-1} \varepsilon_n(x) - (f(x))^{-1/d}\}| \xrightarrow{n \rightarrow \infty} 0, \quad \text{con probabilidad 1.} \quad (4.5)$$

El Grafo de vecindad  $\mathcal{G}_n$  correspondiente a  $\varepsilon_n$  es el grafo con vértices  $\mathcal{X}_n$  y aristas dirigidas de  $X_i$  a  $X_j$  si y solo si

$$|X_i - X_j| \leq \varepsilon_n(X_i). \quad (4.6)$$

En este caso  $c_n = h_n^{1/2}$  en el razonamiento anterior, las condiciones (4.4) implican **a)** y **b)**, mientras que (4.5) es **c)** forzando  $\bar{\varepsilon}(x) = (f(x))^{-1/d}$ . Por último, la matriz de adyacencia (4.6) es una función determinística de los puntos de la muestra aleatoria.

**Observación 4.1** (Comportamiento de  $Y_t$ ). Cuando observamos los grafos de vecindad y analizamos la dinámica del paseo aleatorio (salto con distribución uniforme entre los vértices adyacentes) notamos que predominan los saltos hacia las zonas de mayor densidad y tienen distancia más corta en esas zonas que en las de menor densidad. Esto mismo se puede observar en los coeficientes en (4.3) al forzar  $\bar{\varepsilon}(x) = (f(x))^{-1/d}$ , pues la deriva tiene dirección  $\nabla \ln(f(x))$  y ambos coeficientes son menores en áreas densas.

*Grafos de vecindad generalizada.* Sean  $\varepsilon_n$  y  $c_n$  tales que satisfacen las condiciones (4.4) y (4.5). Sea  $h : \mathbb{R}_0^+ \rightarrow [0, 1]$  una función continua a trozos con  $h(x) = 0$  para  $x > 1$ ,  $h(1) > 0$ , y  $h$  continua por izquierda en 1. El grafo de vecindad generalizada  $\mathcal{G}_n$  correspondiente a  $\varepsilon_n$  y  $h$  es el grafo aleatorio con vértices  $\mathcal{X}_n$  y aristas dirigidas de  $X_i$  a  $X_j$  con probabilidad  $p_{ij} = h(|X_i - X_j|/\varepsilon_n(X_i))$ .

En el caso del grafo  $\mathcal{R}_n$ , tenemos que  $\varepsilon_n(x) = r_n$  para todo  $x \in \mathcal{X}_n$  y debemos tomar  $c_n = [(\#\{y \in \mathcal{X}_n : y \in B(x, r_n)\})/n]^{1/d}$ ; es necesario imponer las condiciones  $r_n \rightarrow 0$  y  $nr_n^d/\ln(n) \rightarrow \infty$  a la sucesión de números positivos  $r_n$  para satisfacer las condiciones (4.4). La convergencia (4.5) se deduce como caso particular del Teorema 4.9 de Devroye [16], el cuál es un resultado más general para estimadores de densidad que provienen de un kernel que sea acotado, con soporte compacto y cuyo conjunto de puntos de discontinuidad tiene clausura con medida nula; este es un resultado que puede ser utilizado para probar la convergencia (4.5) para un amplio espectro de grafos que provienen del uso de kernels y que pueden considerarse como de *vecindad generalizada*.

En caso de kernels como uno Gaussiano, estos pueden ser aproximados utilizando  $h(x) = \exp(-x^2/\sigma^2)\mathbb{1}\{x \in [0, 1]\}$  y la “ventana” a utilizar puede ser variable para cada vértice a través de  $\varepsilon_n$ .

Es importante resaltar nuevamente que una de las cosas a tener en cuenta al elegir el grafo a utilizar es que la matriz de adyacencias sea lo más rala posible, por lo que, si bien un

kernel con estructura Gaussiana puede mejorar la performance de  $\mathcal{PS}_{\mathcal{G}_n}$  para distribuciones elípticas, el cálculo para una muestra grande puede ser más dificultoso. Además, si la función  $h$  elegida no es una función indicadora (grafo de vecindad generalizada), es posible tener diferentes grafos para una misma muestra aleatoria, generando diferentes  $\mathcal{PS}_{\mathcal{G}_n}$  para una misma muestra.

#### 4.1.2. Consistencia para grafos de vecindad generalizada

Comenzamos enunciando un teorema dado en Hashimoto et al. [26] que nos permitirá demostrar consistencia de  $\mathcal{PS}$ . Este resultado es una aplicación del Teorema 11.2.3 de Stroock y Varadhan [56], utilizando las verificaciones de las hipótesis dadas en Hashimoto et al. [27]. Lo presentamos para nuestro caso en particular.

Recordamos la métrica utilizada en los espacios de Skorokhod. Sea  $T > 0$  fijo o  $T = +\infty$ , sea  $\Lambda$  el conjunto de biyecciones estrictamente crecientes de  $[0, T)$  en  $[0, T)$ , para  $\nu, \omega$  en  $D([0, T), \mathcal{D})$  o  $D([0, T), \mathbb{R}_0^+)$ , la distancia  $\sigma$  se define

$$\sigma(\nu, \omega) = \inf_{\lambda \in \Lambda} \max\{\|\lambda - I\|, \|\nu - \omega \circ \lambda\|\}, \quad (4.7)$$

donde  $I : [0, T) \rightarrow [0, T)$  es la identidad y  $\|\omega\| = \sup\{|\omega(t)|, 0 \leq t < T\}$ .

Sea  $\mathcal{X}_n$  una muestra aleatoria de una distribución  $F$  y densidad  $f$  que cumple **H1**. Sea  $\mathcal{G}_n$  un grafo de vecindad generalizada y sea  $(X_i^n)_{i \in \mathbb{N}}$  un paseo aleatorio simple a tiempo discreto en  $\mathcal{G}_n$ .

**Teorema 4.1** (Convergencia de paseo aleatorio a difusión. Teorema 2.2 Hashimoto, Sun y Jaakkala [26]). *El paseo aleatorio  $(X_i^n)_{i \in \mathbb{N}}$  converge uniformemente en el espacio de Skorokhod  $D([0, T), \overline{\mathcal{D}})$  para todo  $T$  finito, luego de un escalamiento temporal  $t = c_n^2 \hat{t}$  a una difusión de Itô  $Y_t$  que toma valores en el espacio de funciones continuas  $C([0, T), \overline{\mathcal{D}})$  con condiciones de rebote en la frontera  $\partial \mathcal{D}$  definida por*

$$dY_t = \frac{\nabla \ln(f(Y_t))}{3(f(Y_t))^{2/d}} dt + \frac{1}{\sqrt{3}(f(Y_t))^{1/d}} dB_t,$$

**Observación 4.2** (Convergencia uniforme de trayectorias). *La convergencia en la métrica de Skorokhod (4.7) en el espacio de funciones continuas  $C([0, T), \overline{\mathcal{D}})$  coincide con la convergencia uniforme de funciones, es decir en  $\|\cdot\|_\infty$  (ver Billingsley [6] página 124). Por lo tanto la convergencia de las trayectorias aleatorias del paseo a la trayectoria aleatoria de la difusión puede pensarse en sentido uniforme pues las hipótesis **H1** aseguran que la trayectoria de  $Y_t$  es continua.*

Para poder utilizar el principio de invarianza, es necesario el próximo resultado.

**Proposición 4.1** (Tiempo de salida es función continua). *Sea  $D([0, \infty), \mathfrak{D})$  el conjunto de trayectorias  $\xi = \{\xi(t) : t \geq 0\}$  y sea*

$$\tau(\xi) = \inf\{t \geq 0 : \xi(t) \in \partial\mathfrak{D}\}$$

*el tiempo de salida de una trayectoria  $\xi$ . Luego la función*

$$g : D([0, \infty), \mathfrak{D}) \rightarrow \mathbb{R}_0^+, \quad g(\xi) := \exp(-\tau(\xi)) \quad (4.8)$$

*es una función acotada y continua en un conjunto de probabilidad uno.*

*Demostración.* Como  $\tau(\xi) \geq 0$ ,  $g$  es una función acotada; para poder verificar que es continua, mostraremos que  $\tau$  es función continua de las trayectorias.

Denotamos  $d_\partial : \mathfrak{D} \rightarrow \mathbb{R}_0^+$  a la función distancia al borde, i.e.,  $d_\partial(x) = \inf_{y \in \partial\mathfrak{D}} \|x - y\|$ ,  $x \in \mathfrak{D}$  y  $dd_\partial : D([0, \infty), \mathfrak{D}) \rightarrow D([0, \infty), \mathbb{R}_0^+)$  a la función que a cada trayectoria  $\xi$  le asigna la “trayectoria distancia”  $\Xi$  definida por:

$$\Xi(t) = dd_\partial(\xi) = d_\partial \circ \xi = \{d_\partial(\xi(t)), t \geq 0\}.$$

Afirmamos que  $dd_\partial$  es continua. En efecto, sea  $\epsilon > 0$ , como  $d_\partial$  es uniformemente continua, existe  $0 < \delta < \epsilon$  tal que

$$\forall x, y \in \mathfrak{D}, \quad \|x - y\| < \delta \implies |d_\partial(x) - d_\partial(y)| < \epsilon. \quad (4.9)$$

Sean  $\psi, \xi \in D([0, \infty), \mathfrak{D})$  tal que  $\sigma(\psi, \xi) < \delta$ , entonces por (4.7) existe  $\lambda^* \in \Lambda$  para el cual  $\|\lambda^* - I\| < \delta$  y  $\|\psi - \xi \circ \lambda^*\| < \delta$  lo cual por (4.9) implica que  $\|d_\partial \circ \psi - d_\partial \circ \xi \circ \lambda^*\| < \epsilon$ .

Si  $\Psi = d_\partial \circ \psi$  y  $\Xi = d_\partial \circ \xi$  entonces

$$\sigma(\Psi, \Xi) = \inf_{\lambda \in \Lambda} \max\{\|\lambda - I\|, \|\Psi - \Xi \circ \lambda\|\} \leq \max\{\|\lambda^* - I\|, \|\Psi - \Xi \circ \lambda^*\|\} < \epsilon,$$

lo que demuestra que  $dd_\partial$  es continua.

Sea  $\tau_0 : D([0, \infty), \mathbb{R}_0^+) \rightarrow \mathbb{R}$ ,  $\tau_0(\Xi) = \inf\{t \geq 0 : \Xi(t) \leq 0\}$  el primer momento en el que  $\Xi$  es 0; es decir, el primer momento en que la trayectoria  $\xi$  toca la frontera. El Teorema 7.1. de Whitt [60] demuestra que  $\tau_0$  es una función continua en  $C([0, \infty), \mathbb{R}_0^+)$ .

Por lo tanto, el tiempo de salida  $\tau = \tau_0 \circ dd_\partial$  es composición de funciones continuas por lo que resulta continuo; luego  $g(\xi) = \exp(-\tau(\xi))$  también es continua y el resultados queda probado.  $\square$

Estamos en condiciones, ahora sí, de enunciar el siguiente teorema.

**Teorema 4.2** (Consistencia). *Bajo **H1** y **H2**, para el grafo de vecindad generalizada  $\mathcal{G}_n$  resulta*

$$\sup_{x \in \mathbb{R}^d} \{\|\mathcal{P}\mathcal{S}_n(x) - \mathcal{P}\mathcal{S}(x)\|\} \xrightarrow{n \rightarrow \infty} 0, \text{ con probabilidad 1.}$$

*Demostración.* Recordemos que las hipótesis **H1** son necesarias para la existencia de solución fuerte  $Y_t$  y buena definición de  $\mathcal{PS}$ .

Teniendo en cuenta la continuidad del tiempo de salida dada por la Proposición 4.1, para  $T > 0$  fijo definimos la función  $g$  truncada en  $T$

$$g^T(\xi) := \exp(-(\tau(\xi) \wedge T)).$$

El Teorema 4.1 implica que

$$\lim_n \mathbb{E}_x(g^T(X_t^n)) = \mathbb{E}_x(g^T(Y_t)), \quad (4.10)$$

donde  $\mathbb{E}_x$  se refiere a la esperanza de las trayectorias con punto inicial en  $x$  y  $(X_t^n)$  es el paseo  $(X_t^n)$  escalado en el tiempo  $t = c_n^2 \hat{t}$ .

Es necesario ahora extender el resultado a convergencia en el intervalo  $[0, \infty)$  pues  $\tau(\xi)$  no esta acotada en el tiempo.

Para mejorar la exposición denotamos  $T_n = \tau(X_t^n)$  y  $T = \tau(Y_t)$  a los tiempos de salida del paseo aleatorio y de la difusión respectivamente. Siguiendo la definición (4.8),  $g(X_t^n) = e^{-T_n}$ ,  $g(Y_t) = e^{-T}$  y para cualquier  $t > 0$  tenemos

$$\mathbb{E}_x(\mathbb{1}\{T_n < t\} e^{-T_n}) \leq \mathbb{E}_x(e^{-T_n}) \leq \mathbb{E}_x(\mathbb{1}\{T_n < t\} e^{-T_n}) + \mathbb{E}_x(\mathbb{1}\{T_n \geq t\} e^{-T_n}).$$

Tomando límite en  $n$ , aplicando (4.10) obtenemos

$$\mathbb{E}_x(\mathbb{1}\{T < t\} e^{-T}) \leq \lim_n \mathbb{E}_x(e^{-T_n}) \leq \mathbb{E}_x(\mathbb{1}\{T < t\} e^{-T}) + e^{-t}.$$

Si ahora tomamos límite cuando  $t \rightarrow \infty$  llegamos a que

$$\lim_n \mathbb{E}_x(e^{-T_n}) = \mathbb{E}_x(e^{-T}). \quad (4.11)$$

Luego, para probar la convergencia de  $\mathcal{PS}_n$ , volviendo a la notación anterior, (4.11) implica que si fijamos  $\epsilon > 0$ , existe  $N_1 \in \mathbb{N}$  tal que

$$|\mathbb{E}_x(g(X_t^n)) - \mathbb{E}_x(g(Y_t))| < \epsilon/2 \quad \forall x, \text{ si } n > N_1.$$

Para  $n \in \mathbb{N}$  grande es posible definir  $\tau^n$ ,  $d_\theta^n$ ,  $dd_\theta^n$  y  $g_n$  reemplazando  $\partial\mathcal{D}_n$  en lugar de  $\partial\mathcal{D}$  en las definiciones de la Proposición 4.1. La hipótesis **H2** asegura que  $d_\theta^n$  converge uniformemente a  $d_\theta$  con probabilidad 1, por lo que combinado con la continuidad de  $g_n$ , existe  $N_2 \in \mathbb{N}$  tal que

$$|\mathbb{E}_x(g_n(X_t^n)) - \mathbb{E}_x(g(X_t^n))| < \epsilon/2 \quad \forall x, \text{ si } n > N_2.$$

Tomando  $n \geq \max\{N_1, N_2\}$ , por desigualdad triangular tenemos que

$$|\mathbb{E}_x(g_n(X_t^n)) - \mathbb{E}_x(g(Y_t))| < \epsilon \quad \forall x.$$

Por último, notando que  $\mathcal{PS}_n(x) = -\ln(\mathbb{E}_x(g_n(X_t^n)))$  y  $\mathcal{PS}(x) = -\ln(\mathbb{E}_x(g(Y_t)))$ , el resultado queda probado.  $\square$

### 4.1.3. Consistencia para Delaunay en Poisson

En el caso de los grafos de vecindad comenzamos pensando en un paseo al azar simple en el grafo  $\mathcal{G}_n$ , a través de la ley de los grandes números obtuvimos los coeficientes de la difusión límite y utilizando un resultado de Hashimoto et al. [26] obtuvimos la consistencia de  $\mathcal{PS}_n$  para grafos de vecindad generalizada.

Presentamos ahora un resultado análogo pero para paseos en la Triangulación de Delaunay de un proceso de Poisson homogéneo que permite pensar de la misma manera la consistencia en este caso. La difusión “límite”  $Y_t$  en esta sección sera  $B_t$ , un Movimiento Browniano  $d$ -dimensional, en concordancia con lo visto anteriormente para la distribución uniforme.

Sea  $\mathcal{N} \subset \mathbb{R}^d$  un Proceso de Poisson en  $\mathbb{R}^d$  de parámetro 1 y  $\mathcal{G}(\mathcal{N})$  la Triangulación de Delaunay de  $\mathcal{N}$ . Sea  $S[\mathcal{N}] = (S[\mathcal{N}, t])_{t \geq 0}$  un paseo aleatorio simple en  $\mathcal{G}(\mathcal{N})$  a tiempo continuo: cuando el paseo está en  $x \in \mathcal{N}$ , a tasa  $(1/\#\text{vecinos de } x)$  salta a cada uno de los vecinos de  $x$ .

Sea

$$S^k[\mathcal{N}] = (S_t^k[\mathcal{N}])_{t \geq 0}, \quad S_t^k := \frac{1}{k} S[\mathcal{N}, k^2 t],$$

el paseo empírico escalado de manera espacial y temporalmente. Sea  $x^k$  el punto de  $\frac{1}{k}\mathcal{N}$  más cercano a  $x$ . El siguiente teorema nos da la convergencia débil de las trayectorias empíricas del paseo aleatorio al movimiento Browniano.

**Teorema 4.3** (Teorema 1 Rousselle [48]). *Para  $x \in \mathbb{R}^d$ , el paseo escalado  $S^k[\mathcal{N}]$  con condición inicial  $S_0^k = x^k$  verifica*

$$S^k[\mathcal{N}] \xrightarrow[k \rightarrow \infty]{d} B \quad \mathcal{N} \text{ c.s.}, \quad (4.12)$$

donde  $B = (B_t)_{t \geq 0}$  es un Browniano con covarianza  $\sigma I$  y condición inicial  $B_0 = x$ ;  $\sigma$  no depende de  $\mathcal{N}$ .

Es decir, para casi toda realización del proceso puntual, la distribución del paseo aleatorio escalado empezando en  $x$  converge a la distribución del Browniano comenzando en  $x$ .

Sea  $\mathcal{D} \subset \mathbb{R}^d$  compacto con frontera  $\partial\mathcal{D}$  suave, recordamos

$$\tau(\xi) = \inf\{t \geq 0 : \xi(t) \in \partial\mathcal{D}\}$$

es el tiempo de salida de una trayectoria  $\xi$  y en la Proposición 4.1 se demostró que la función

$$g : D([0, \infty), \mathcal{D}) \rightarrow \mathbb{R}_0^+, \quad g(\xi) := \exp(-\tau(\xi))$$

es una función acotada y continua en un conjunto de probabilidad uno.

En este caso

$$\mathcal{PS}(x) = -\ln \mathbb{E}_x(g(B_t))$$

y el estimador es

$$\mathcal{PS}_k(x) = -\ln \mathbb{E}_x(g(S^k[\mathcal{N}])).$$

La convergencia (4.12) es equivalente a

$$\mathbb{E}_x F(S^k[\mathcal{N}]) \xrightarrow[k \rightarrow \infty]{} \mathbb{E}_x F(B_t) \quad \mathcal{N} \text{ c.s.},$$

para funciones  $F$  continuas, acotadas, que dependan de las trayectorias en un intervalo finito de tiempo  $[0, T]$ . Por lo tanto, este resultado cumple el mismo rol que el Teorema 4.1 en el caso de grafos de vecindad generalizada, por lo que es posible aplicar un razonamiento análogo al aplicado para deducir (4.11) y poder obtener el límite

$$\mathbb{E}_x(g(S^k[\mathcal{N}])) \xrightarrow[k \rightarrow \infty]{} \mathbb{E}_x(g(B_t)) \quad \mathcal{N} \text{ c.s.},$$

Luego, la consistencia de  $\mathcal{PS}_k$  se obtiene de manera análoga en este caso recordando que es necesaria la hipótesis **H2** que asegura que la frontera empírica marcada converge uniformemente a la verdadera frontera con probabilidad 1.

En el Capítulo 5 veremos que la Triangulación de Delaunay puede utilizarse para datos con distintos tipos de distribuciones obteniendo buenos resultados en las aplicaciones, mientras que la demostración de consistencia en el caso general no ha sido obtenida.

## 4.2. $\mathcal{PS}$ en dimensión 1

En dimensión 1 ninguno de los resultados usados anteriormente aplican, pero la idea es utilizar fuertemente que la muestra aleatoria puede ser ordenada y que la “frontera” claramente en este caso debe ser el dato más chico y el más grande de la muestra. Además un resultado clásico de teoría de probabilidad es que un paseo aleatorio en  $\mathbb{Z}$  con escalamiento temporal el cuadrado del escalamiento espacial converge a un movimiento Browniano, por lo que nos basaremos en eso para encontrar cuál es la difusión necesaria para la definición teórica de la función de profundidad.

Empezamos entonces con la versión empírica de la profundidad como para después “encontrar” la difusión límite. Sea  $\mathbf{X}$  una variable aleatoria con densidad  $f$  que cumple **H1**,  $F$  su distribución acumulada y  $F^{-1}$  su inversa.

Para  $\mathcal{X}_n = \{X_1, \dots, X_n\}$ , una muestra aleatoria de la distribución  $F$  que *sin pérdida de generalidad podemos suponer ordenada* en esta sección, consideramos el grafo  $\mathcal{G}_n = (\mathcal{X}_n, E_n)$  donde  $E_n = \{(X_i, X_j) : |i - j| = 1\}$ , es decir que consiste en unir cada punto de la muestra con el anterior y siguiente. Definimos la frontera  $\partial\mathcal{G}_n = \{X_1, X_n\}$ , pues el primer y último dato deberían ser los menos profundos y  $\mathfrak{D}_n = [X_1, X_n]$  como estimador del soporte  $\mathfrak{D}$ .

En este caso pensamos  $(X_i^n)_{i \in \mathbb{N}}$  paseo aleatorio simple con espacio de estados  $\mathcal{X}_n \cup \Delta$ , con  $\Delta = \mathfrak{D}_n^c$ , que salta con probabilidad  $1/2$  hacia adelante o hacia atrás a tiempo discreto y “sale” según la dinámica:

- $p(X_i, X_j) = 1/2 \mathbb{1}\{|i - j| = 1\}$ ,  $i = 2, 3, \dots, n - 1$ .
- $p(X_i, \Delta) = 1$ , si  $i = 1, n$ .

Considerando el escalamiento temporal conocido para paseos aleatorios en una dimensión, definimos  $c_n = 1/n^2$  y tomamos  $t = c_n \hat{t}$  y  $T_n = \min\{t \in \mathbb{N} \cup \{0\} : X_t^n \in \Delta\}$ .

Recordamos la definición (3.1)

$$\gamma_n(i) = \gamma(X_i, \mathcal{G}_n) := \mathbb{E}(e^{-c_n T_n} | X_0^n = X_i) = \mathbb{E}_i(e^{-c_n T_n}).$$

En este caso, haciendo el análisis condicionando a un paso como en (3.2) para calcular  $\gamma_n(i)$  obtenemos

$$\gamma_n(i) = \frac{1}{2e^{1/n^2}} [\gamma_n(i-1) + \gamma_n(i+1)], \quad i = 2, \dots, n-1,$$

con las condiciones de borde  $\gamma_n(1) = \gamma_n(n) = e^{-1/n^2}$ .

Podemos resolver de manera cerrada para obtener la expresión para todo  $i$

$$\gamma_n(i) = e^{-1/n^2} \cosh \left[ \sqrt{2} \left( \frac{i}{n} A_n - \frac{1}{2} - B_n \right) \right] \left[ \cosh \left( \sqrt{2} \left( A_n - \frac{1}{2} - B_n \right) \right) \right]^{-1}$$

donde  $A_n = \frac{n}{\sqrt{2}} \operatorname{arcosh}(e^{1/n^2})$  y  $B_n = ((n+1)A_n - n)/2n$ .

Luego, para todo  $i$ ,  $\mathcal{PS}_n(i) = \mathcal{PS}_n(X_i) = -\ln(\gamma_n(i))$  es

$$\mathcal{PS}_n(i) = \frac{1}{n^2} + \ln \left[ \cosh \left( \sqrt{2} \left( A_n - \frac{1}{2} - B_n \right) \right) \right] - \ln \left[ \cosh \left( \sqrt{2} \left( \frac{i}{n} A_n - \frac{1}{2} - B_n \right) \right) \right]. \quad (4.13)$$

## Extensión a $\mathbb{R}$

Para extender la función  $\mathcal{PS}_n$  a  $\mathbb{R}$ , análogamente a lo referido en la Sección 3.2.3, si  $x < X_1$  o  $x > X_n$ , es decir  $x \notin \mathcal{D}_n$ , resulta  $\gamma_n(x) = 1$  y por lo tanto  $\mathcal{PS}_n(x) = 0$ .

Por otro lado, para  $x \in \mathcal{D}_n$ ,  $x \notin \mathcal{X}_n$ , existen diferentes maneras para extender las funciones, análogas a las dadas para  $d \geq 2$ .

Utilizando que existe único  $i \in \{1, \dots, n-1\}$  tal que  $X_i < x < X_{i+1}$ , una manera de extender  $\gamma_n$  es a partir su definición, es decir,

$$\gamma_n(x) = \frac{1}{2e^{1/n^2}} [\gamma_n(i) + \gamma_n(i+1)].$$

Como segunda opción, emulando la extensión a celdas de Voronoi en dimensión mayor, es posible tomar  $\gamma_n(x) = \gamma_n(i)$  si  $X_i$  es el punto más cercano de la muestra a  $x$ .

En ambos casos una vez obtenida  $\gamma_n$ , calculamos  $\mathcal{PS}_n(x) = -\ln(\gamma_n(x))$ , la versión empírica de la profundidad  $\mathcal{PS}$  para la muestra  $X_n$  extendida a  $\mathbb{R}$ .

Por último, la tercer opción es la interpolación lineal de  $\mathcal{PS}_n$ , es decir,

$$\mathcal{PS}_n(x) = \frac{x - X_i}{X_{i+1} - X_i} \mathcal{PS}_n(i) + \frac{X_{i+1} - x}{X_{i+1} - X_i} \mathcal{PS}_n(i + 1).$$

En la Figura 4.1 se muestra el cálculo de  $\mathcal{PS}_n$  para una muestra de 21 datos con distribución normal estándar y las tres opciones de extensión propuestas.

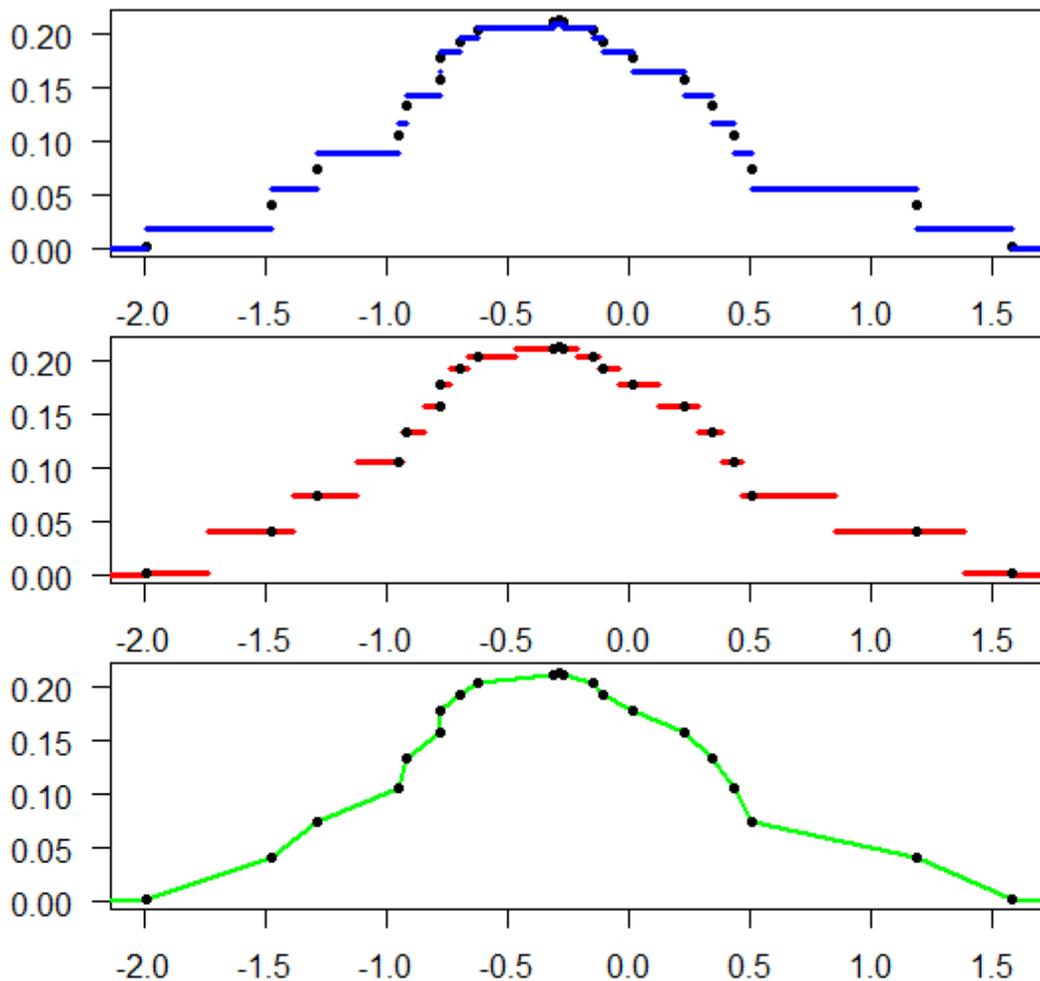


Figura 4.1: Extensiones (de arriba hacia abajo): por definición de  $\gamma_n$ , Voronoi e interpolación.

### Versión Teórica en dimensión 1

Veremos ahora entonces por qué el escalamiento temporal  $1/n^2$  resulta necesario para la convergencia del paseo aleatorio a una difusión límite a medida que la cantidad de puntos en la muestra  $\mathcal{X}_n$  crece y obtendremos cuál es la difusión para dar la definición de  $\mathcal{PS}$  en este caso.

Haremos la deducción de cuál debe ser la difusión límite sin preocuparnos por probar de manera estricta la convergencia en distribución del paseo aleatorio, pues en este caso tanto  $\mathcal{PS}_n$  como  $\mathcal{PS}$  tienen una forma cerrada y la consistencia va a ser demostrada tomando límite en  $n$  para  $\mathcal{PS}_n$ .

Dada la muestra aleatoria  $\mathcal{X}_{n+1}$  (que recordemos suponemos ordenada) y el paseo aleatorio a tiempo discreto  $\{X_i^n\}_{i \in \mathbb{N}}$  en  $\mathcal{G}_{n+1}$ , sea  $F_{n+1}(x) = \#\{y \in \mathcal{X}_{n+1} : y \leq x\}/(n+1)$  la distribución acumulada empírica de  $\mathcal{X}_{n+1}$ ; definimos

$$h_n(x) \stackrel{\text{def}}{=} \frac{\#\{y \in \mathcal{X}_{n+1} : y \leq x\} - 1}{n} = \frac{(n+1)F_{n+1}(x) - 1}{n} = \frac{n+1}{n}F_{n+1}(x) - \frac{1}{n}.$$

Sabemos que  $h_n(x)$  es una biyección entre  $\{X_1, \dots, X_{n+1}\}$  y  $\{0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\}$ , por lo tanto  $\{\tilde{X}_i^n\}_{i \in \mathbb{N}} \stackrel{\text{def}}{=} \{h_n(X_i^n)\}_{i \in \mathbb{N}}$  es un paseo aleatorio a tiempo discreto con espacios de estados  $\{0, \frac{1}{n}, \dots, \frac{n-1}{n}, 1\}$  y tasas  $q(i, j) = \frac{1}{2n^2} \mathbb{1}\{|i - j| = n^{-1}\}$  (Ver Figura 4.2).

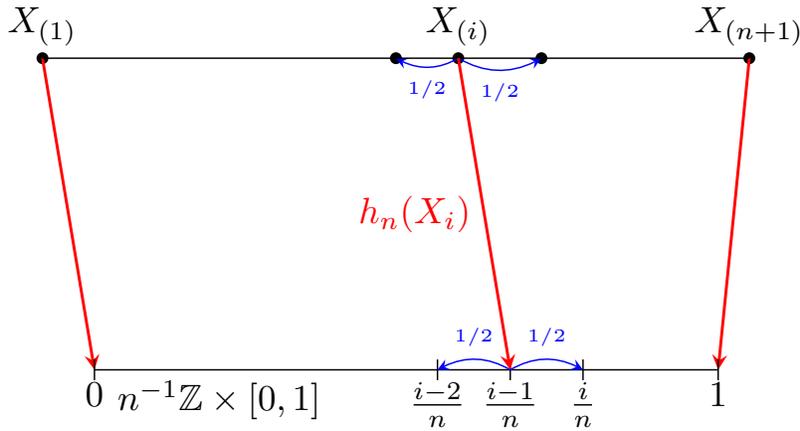


Figura 4.2: paseo aleatorio en  $\mathcal{X}_{n+1}$  transformado en paseo aleatorio en  $n^{-1}\mathbb{Z} \times [0, 1]$

Además

$$\lim_{n \rightarrow \infty} h_n(x) = F(x), \quad \text{casi seguramente y uniformemente en } x.$$

Sea  $\{B_t, t \geq 0\}$  un Movimiento Browniano estandar, por el teorema de invarianza de Donsker sabemos que un paseo aleatorio simple en  $\mathbb{Z}$  con escalamiento temporal el cuadrado del escalamiento espacial converge en distribución a un Movimiento Browniano.

Por lo tanto, si consideramos  $\tilde{S}_t$  a la interpolación lineal de  $\{\tilde{X}_t^n\}_{t \in \mathbb{N}}$ , sabemos que

$$\tilde{S}_t \xrightarrow[n \rightarrow \infty]{d} B_t, \quad t \geq 0.$$

Así es que deberíamos utilizar tasas  $1/n^2$  para obtener convergencia en distribución.

Para encontrar cuál es la difusión límite, como  $F^{-1}$  es una función continua, tenemos

$$F^{-1}(\tilde{S}_t) \xrightarrow[n \rightarrow \infty]{d} F^{-1}(B_t), \quad t \geq 0.$$

Por Borel Cantelli

$$\sup_{0 < p < 1} |F_{n+1}^{-1}(p) - F^{-1}(p)| \xrightarrow[n \rightarrow \infty]{} 0 \quad (4.14)$$

con probabilidad 1, entonces,

$$F_{n+1}^{-1}(\tilde{S}_t) \xrightarrow[n \rightarrow \infty]{d} F^{-1}(B_t), \quad t \geq 0.$$

De la definición de  $h_n$  vemos que  $h_n^{-1}(x) = F_{n+1}^{-1}\left(\frac{nx+1}{n+1}\right)$ . Tomando  $\tilde{p} = \frac{np+1}{n+1}$ , para  $p \in [0, 1]$  obtenemos

$$\sup_{0 < p < 1} |h_n^{-1}(p) - F_{n+1}^{-1}(p)| = \sup_{0 < p < 1} |F_{n+1}^{-1}(\tilde{p}) - F_{n+1}^{-1}(p)| \xrightarrow[n \rightarrow \infty]{} 0,$$

pues  $|p - \tilde{p}| \leq (n+1)^{-1}$  y por (4.14)  $F_{n+1}^{-1}$  converge uniformemente a  $F^{-1}$  que es uniformemente continua.

Finalmente,

$$h_n^{-1}(\tilde{S}_t) \xrightarrow[n \rightarrow \infty]{d} F^{-1}(B_t), \quad t \geq 0,$$

pero  $h_n^{-1}(\tilde{S}_t) = S_t$ , la interpolación lineal del paseo en  $\{X_t^n\}_{t \in \mathbb{N}}$ , por lo tanto

$$S_t \xrightarrow[n \rightarrow \infty]{d} F^{-1}(B_t), \quad t \geq 0.$$

Queremos definir entonces  $Y_t = F^{-1}(B_t)$ , por esto necesitamos el siguiente lema.

**Lema 4.1** (Fórmula de Itô en dimensión 1. Oksendal[42], Teorema 4.1.2 y Remark). *Sea  $Y_t$  una difusión de la forma*

$$dY_t = b(Y_t) dt + \sigma(Y_t) dB_t.$$

*Si  $g(t, x)$  es una función determinística real definida para todo  $t \in [0, \infty)$ ,  $x \in \mathbb{R}$ , con derivadas continuas  $\frac{\partial g}{\partial t}$ ,  $\frac{\partial g}{\partial x}$  y  $\frac{\partial^2 g}{\partial x^2}$ , entonces*

$$dg(t, Y_t) = \tilde{b}(t, Y_t) dt + \tilde{\sigma}(t, Y_t) dB_t,$$

donde  $\tilde{b}(t, x) = \left[ \frac{\partial g}{\partial t} + b \frac{\partial g}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 g}{\partial x^2} \right](t, x)$  y  $\tilde{\sigma}(t, x) = \left[ \sigma \frac{\partial g}{\partial x} \right](t, x)$ .

Utilizando la fórmula de Itô,  $Y_t = F^{-1}(B_t)$  es una difusión y

$$dY_t = -\frac{f'(Y_t)}{2f^3(Y_t)}dt + \frac{1}{f(Y_t)}dB_t \quad (4.15)$$

**Observación 4.3.** *Notar la similitud en los coeficientes con la difusión (2.6). El término difusivo dice esencialmente que  $Y_t$  “varía” más en áreas de poca densidad que en áreas de mayor densidad. El término de drift también en áreas de poca densidad será más fuerte que en áreas de densidad alta, pero a diferencia de (2.6) la difusión trata de irse (por el signo  $-$ ) de las áreas con más densidad.*

*En conclusión podemos entender a  $Y_t$  como una difusión que “trata de salir”, pasando más tiempo (menos difusiva y con menos drift) cerca del centro de la distribución.*

Estamos en condiciones entonces de dar la definición de  $\mathcal{PS}$  en dimensión 1.

**Definición 4.2.** *Sea  $X$  variable aleatoria con densidad  $f$  que cumple **H1**, para  $x \in \mathbb{R}$  e  $Y_t$  dada por (4.15), definimos la función profundidad de salida  $\mathcal{PS}$  de la siguiente manera:*

*Si  $\mathcal{D}$  es acotado,  $\tau = \inf\{t \geq 0 : Y_t \in \overline{\mathcal{D}^c}\}$*

$$\mathcal{PS}(x) = -\ln(\mathbb{E}_x(e^{-\tau})).$$

*Si  $\mathcal{D}$  no es acotado, para  $\lambda > 0$  tomamos  $\tau_\lambda = \inf\{t \geq 0 : Y_t \notin [-\lambda, \lambda] \cap \mathcal{D}\}$*

$$\mathcal{PS}_\lambda(x) = -\ln(\mathbb{E}_x(e^{-\tau_\lambda})), \quad \mathcal{PS}(x) = \lim_{\lambda \rightarrow \infty} \mathcal{PS}_\lambda(x).$$

Planteamos ahora los resultados obtenidos en dimensión 1.

**Teorema 4.4.** *Bajo las condiciones de la Definición 4.2,*

$$\mathcal{PS}(x) = \ln\left(\cosh\left(\sqrt{2}/2\right)\right) - \ln\left(\cosh\left(\sqrt{2}(F(x) - 1/2)\right)\right), \quad (4.16)$$

*donde  $F(x)$  es la función de distribución acumulada de  $X$ .*

*Demostración.* Según el Lema 2.2, si  $g \in C^2$  el generador de la difusión en (4.15) es

$$\mathcal{A}g(x) = \frac{1}{2f^2(x)}g''(x) - \frac{f'(x)}{2f^3(x)}g'(x).$$

Supongamos primero  $f$  con soporte acotado  $[a, b]$ , utilizando los Lemas 2.5 y 2.6 sabemos que  $h(x) = \mathbb{E}_x(e^{-\tau})$  es la única solución de

$$\begin{aligned} h''(x) - \frac{f'(x)}{f(x)}h'(x) - 2f^2(x)h(x) &= 0, & x \in [a, b] \\ h(x) &= 1, & x \notin [a, b]. \end{aligned}$$

Luego  $H(x) = -\ln(h(x))$  es la única solución de

$$\begin{cases} H''(x) - [H'(x)]^2 - \frac{f'(x)}{f(x)}H'(x) + 2f^2(x) = 0, & x \in [a, b] \\ H(x) = 0, & x \notin [a, b]. \end{cases} \quad (4.17)$$

Para  $\mathcal{PS}$  dada por (4.16) derivando y reemplazando verificamos que es la solución.

En el caso de  $f$  con soporte no acotado en  $\mathbb{R}$ , para todo  $\lambda > 0$ , análogamente a (4.17) tenemos que  $H_\lambda(x) = -\ln(\mathbb{E}_x(e^{-\tau_\lambda}))$  es la única solución de

$$\begin{cases} H''(x) - [H'(x)]^2 - \frac{f'(x)}{f(x)}H'(x) + 2f^2(x) = 0, & x \in [-\lambda, \lambda] \cap \mathfrak{D} \\ H(x) = 0, & x \notin [-\lambda, \lambda] \cap \mathfrak{D}. \end{cases}$$

y derivando y reemplazando verificamos que

$$\mathcal{PS}_\lambda(x) = \ln \left[ \cosh \left( \frac{\sqrt{2}}{2} (F(\lambda) - F(-\lambda)) \right) \right] - \ln \left[ \cosh \left( \frac{\sqrt{2}}{2} (2F(x) - F(\lambda) - F(-\lambda)) \right) \right]$$

es la solución.

Tomando límite para  $\lambda \rightarrow \infty$  se verifica (4.16) para el caso en que el soporte es no acotado.  $\square$

Ahora que tenemos a  $\mathcal{PS}$  de forma explícita, es posible demostrar que cumple todas las propiedades descriptas en la Introducción.

**Corolario 4.1.** *Bajo las condiciones de la Definición 4.2,  $\mathcal{PS}$  dada por (4.16) es una función de profundidad que cumple **P1-P6** y la versión empírica  $\mathcal{PS}_n(x)$  dada por (4.13) converge a  $\mathcal{PS}(x)$  uniformemente.*

*Demostración.* Sean  $c, d \in \mathbb{R}$ ,  $\mathbf{Y} = c\mathbf{X} + d$ . Si denotamos  $F_X(x)$ ,  $F_Y(y)$  a las distribuciones acumuladas de  $\mathbf{X}$  e  $\mathbf{Y}$  respectivamente, para  $x \in \mathbb{R}$ ,  $y = cx + d$ , si  $c > 0$  se verifica

$$F_Y(y) = F_X \left( \frac{y-d}{c} \right) = F_X(x),$$

mientras que si  $c < 0$  tenemos

$$F_Y(y) = 1 - F_X \left( \frac{y-d}{c} \right) = 1 - F_X(x).$$

Luego, si  $\mathcal{PS}_X(x)$  y  $\mathcal{PS}_Y(y)$  son la función dada en (4.16) para las variables aleatorias  $\mathbf{X}$  e  $\mathbf{Y}$  respectivamente, resulta  $\mathcal{PS}_Y(y) = \mathcal{PS}_X(x)$  y por lo tanto  $\mathcal{PS}$  es invariante afín (**P1**).

Por otro lado, si  $\theta$  es la mediana de  $\mathbf{X}$ , es decir, es el único punto tal que  $F_X(\theta) = 1/2$ , entonces  $\mathcal{PS}(\theta) = \ln(\cosh(1/\sqrt{2}))$  que es el máximo de  $\mathcal{PS}$  y por lo tanto se cumple (**P2**).

Como

$$\frac{d \mathcal{PS}(x)}{d x} = -\sqrt{2} f_X(x) \tanh\left(\sqrt{2}\left(F_X(x) - 1/2\right)\right),$$

tenemos que la derivada de  $\mathcal{PS}$  es positiva si  $x < \theta$  y negativa si  $x > \theta$ , por lo que  $\mathcal{PS}$  decae monótonamente desde el centro  $\theta$  (**P3**).

Además por definición de distribución acumulada,

$$\lim_{x \rightarrow -\infty} F_X(x) = 0, \quad \lim_{x \rightarrow \infty} F_X(x) = 1,$$

por lo que la simetría de  $\cosh(x)$  implica  $\lim_{\|x\| \rightarrow \infty} \mathcal{PS}(x) = 0$  (**P4**).

La continuidad de  $\mathcal{PS}$  como función de  $x$  (**P5**) se verifica por ser composición de funciones continuas, mientras que la continuidad como función de  $F$  (**P6**) se cumple por definición de convergencia en distribución. En efecto, si  $\mathbf{X}_1, \mathbf{X}_2, \dots$  es una sucesión de variables aleatorias tales que  $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ , entonces  $F_{X_n}(x) \rightarrow F_X(x) \forall x \in \mathbb{R}$  y por lo tanto  $\mathcal{PS}(x, F_{X_n}) \rightarrow \mathcal{PS}(x, F_X)$ .

Finalmente, la versión empírica  $\mathcal{PS}_n$  en (4.13) puede reescribirse como

$$\mathcal{PS}_n(X_i) = \frac{1}{n^2} + \ln \left[ \cosh \left( \sqrt{2} \left( A_n - \frac{1}{2} - B_n \right) \right) \right] - \ln \left[ \cosh \left( \sqrt{2} \left( F_n(X_i) A_n - \frac{1}{2} - B_n \right) \right) \right] \quad \forall i,$$

donde  $F_n(X_i) = i/n$  es la acumulada empírica de la muestra  $X_n$ .

Los límites  $A_n \rightarrow 1$ ,  $B_n \rightarrow 0$  y  $F_n(x) \rightarrow F(x)$  cuando  $n \rightarrow \infty$  uniformemente, implican que la convergencia de  $\mathcal{PS}_n$  a  $\mathcal{PS}$  es uniforme en  $x$ .  $\square$

Podemos concluir que la profundidad  $\mathcal{PS}$  básicamente coincide con todas las funciones de profundidad clásicas en dimensión uno. En la Figura 4.3 se muestra  $\mathcal{PS}$  para la distribución uniforme en  $[0, 1]$  y la normal estándar.

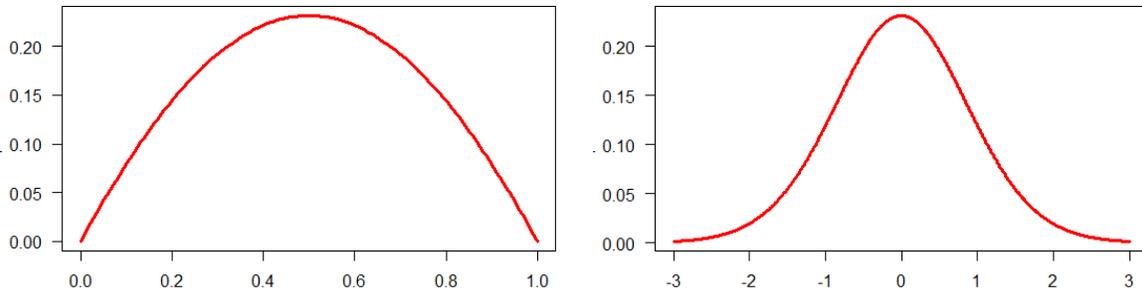


Figura 4.3:  $\mathcal{PS}$  para la distribución  $U[0, 1]$  (izquierda) y  $N(0, 1)$  (derecha).

# Capítulo 5

## Simulaciones y aplicaciones

En este capítulo mostraremos el desempeño de  $\mathcal{PS}$  a través de diferentes herramientas; haremos un análisis comparando los conjuntos de nivel que generan para distintas distribuciones profundidades clásicas y  $\mathcal{PS}$  y aplicaremos clasificadores con datos simulados y reales. Hacia el final, presentaremos algunos comentarios y posibles futuras líneas de investigación.

Para calcular  $\mathcal{PS}_n$  se usaran los siguientes grafos con la especificaciones dadas:

- $\mathcal{K}_n$ : grafo de  $k$  de vecinos más cercano tomando  $C = \omega_d$  en (4.2),

$$k_n = \lceil d^{-1} \omega_d n^{\frac{2}{d+2}} \ln(n)^{\frac{d}{d+2}} \rceil \quad \text{y} \quad c_n = d^{-1/d} (\ln(n)/n)^{1/(d+2)}.$$

- $\mathcal{R}_n$ : grafo de radio único  $r_n$  utilizando distancia de Mahalanobis para  $B(x, r_n)$ , con

$$r_n = 2^{d+1} \left( \frac{\ln^{2/d}(n)}{n^{(4+d)/2d}} \right)^{1/(d+2)} \quad \text{y} \quad c_n = \left( 2^{d+1} \sqrt{n} \ln^{2/d}(n) \right)^{-1/(d+2)}.$$

- $\mathcal{T}_n$ : Triangulación de Delaunay con tasa  $(v_i/n)^2/d$ .

Denotaremos PSK, PSR y PST a  $\mathcal{PS}_n$  calculada para los grafos  $\mathcal{K}_n, \mathcal{R}_n$  y  $\mathcal{T}_n$  respectivamente.

Para unificar y mejorar comparaciones siempre se va a utilizar para estimar el soporte de la distribución al estimador  $\mathfrak{D}_n$  dado en (3.9) con las especificaciones dadas en (3.10) y su frontera  $\partial\mathfrak{D}_n$  como estimador de  $\partial\mathfrak{D}$ ; mientras que los puntos marcados como frontera de  $\partial\mathcal{G}_n$  para todos los grafos utilizados es el conjunto dado en (3.11).

Para obtener la profundidad en puntos que no pertenecen a la muestra la extensión va a estar calculada a partir de (3.12) para los casos PSK y PSR, mientras que para PST se utilizará (3.13).

## 5.1. Comparación de conjuntos de nivel

Comenzamos realizando comparaciones de los conjuntos de nivel que generan diferentes profundidades para distintas distribuciones calculados de la siguiente manera: primero generamos una muestra aleatoria de tamaño 500 en cada uno de los casos dados y calculamos las profundidades para ese conjunto de puntos; luego, generamos una grilla de  $50 \times 50$  puntos equiespaciados en  $[x_{min} - e, x_{max} + e] \times [y_{min} - e, y_{max} + e]$ , donde  $e = \max\{x_{max} - x_{min}, y_{max} - y_{min}\}/50$ ; por último para estos puntos calculamos las profundidades respecto de la muestra aleatoria y graficamos los contornos de percentiles fijados, más precisamente, la curvas de nivel de 5 %, 25 %, 50 %, 75 % y 95 % del conjunto de puntos cuya profundidad no sea 0.

El análisis lo realizaremos tanto para distribuciones elípticas como para algunas con soporte no convexo y otras “contaminadas”, por lo que utilizaremos las profundidades “Half Space” (HS), “Simplicial” (SD) y “Monge - Kantorovich” (MK). Las profundidades HS y SD fueron calculadas con las funciones *mdepth.HS* y *mdepth.SD* del paquete *fda.usc*, mientras que para MK se utilizó una aproximación discreta de la distribución uniforme en el círculo unitario como distribución de referencia, con el emparejamiento dado por la función *assignment* del paquete *adagio* y la extensión dada por el paquete *alphahull*, con  $\alpha = 0,5$  en todos los casos. Los puntos resaltados en los gráficos de PSK, PSR y PST son los marcados como frontera del grafo utilizado.

### Soportes no convexos

- Caso 1: puntos en un soporte “banana”:  $(X + R \cos(\phi), X^2 + R \sin(\phi))$  donde  $X \sim U[-1, 1]$ ,  $\phi \sim U[0, 2\pi]$ ,  $Z \sim U[0, 1]$ ,  $X, \phi, Z$  independientes y  $R = 0,2 Z(1 + (1 - |X|)/2)$ .
- Caso 2: puntos uniformes en el disco  $B(0, 2) \setminus B(0, 1)$ .

Como puede apreciarse en la Figura 5.1, las profundidades HS y SD no son capaces de registrar la forma del soporte en el Caso 1. Por el contrario PST funciona muy bien tanto en la identificación del soporte como encontrando “centralidad”, mientras que MK tiene un rendimiento similar pero depende fuertemente del parámetro  $\alpha$  para que la estimación del soporte sea la adecuada.

En la Figura 5.2 se puede ver que en el Caso 2 las profundidades HS y SD nuevamente tienen problemas para registrar la forma del soporte y en particular que el centro del círculo no es parte de la distribución, dándole la mayor profundidad. Por otro lado, si bien MK funciona bien en la identificación del soporte, sitúa a los puntos más centrales muy cerca de la frontera interior del anillo. En cambio, PST no solo identifica bien el soporte sino que le da más profundidad a los puntos que tienen norma cercana a 1,5, es decir, el punto medio entre las fronteras interior y exterior del anillo.

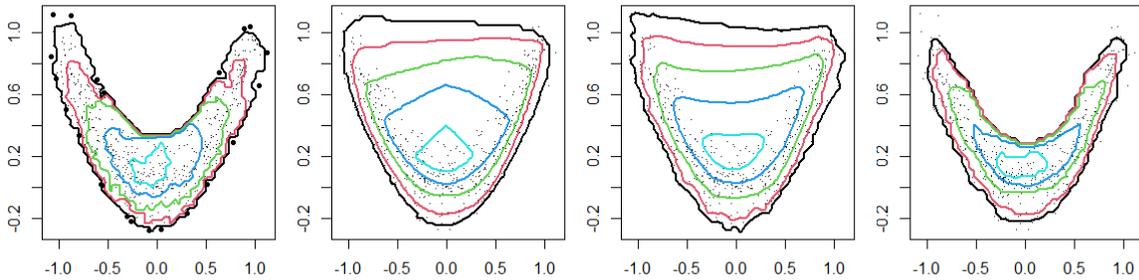


Figura 5.1: Caso 1: de izquierda a derecha, PST, HS, SD, MK.

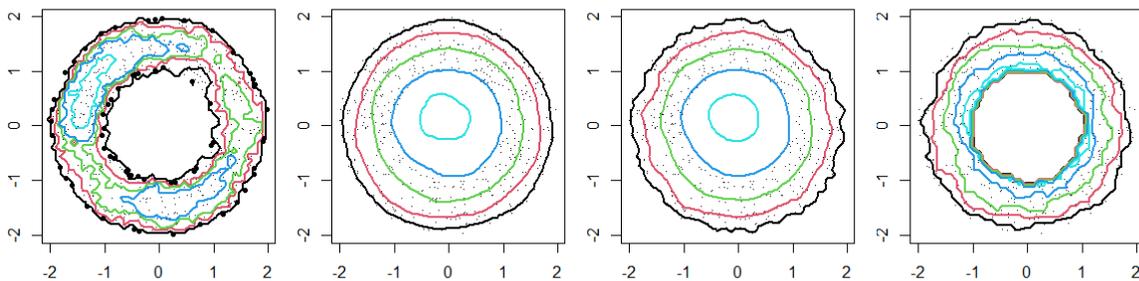


Figura 5.2: Caso 2: de izquierda a derecha, PST, HS, SD, MK.

En soportes como estos, cuando además la muestra es chica, el efecto de la frontera en el cálculo de  $\mathcal{PS}_n$  puede ser grande y el resultado depende fuertemente de la cantidad de puntos marcados como frontera del grafo y de su distribución a lo largo de  $\partial\mathcal{D}_n$ . Es por esto que en estos primeros dos casos se utilizó PST, pues es preciso hacer que la frontera incida lo menos posible en los puntos interiores, lo cual se consigue teniendo la menor cantidad de vecinos por vértice.

En los casos de distribuciones elípticas a continuación, utilizaremos PSR para poder aprovechar el uso de la métrica de Mahalanobis para la obtención de los vértices adyacentes.

### Distribuciones elípticas

- Caso 3: puntos con distribución  $N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}\right)$ .
- Caso 4: combinación convexa 50 %  $N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}\right)$  y 50 %  $N\left(\begin{bmatrix} -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}\right)$ .

En el Caso 3, Figura 5.3, donde la muestra proviene de un normal bivariada vemos que las profundidades HS, SD y PSR funcionan muy bien; por otro lado MK identifica eficazmente los puntos más centrales, pero la sensibilidad en la elección del  $\alpha$  en la extensión a puntos fuera de la muestra queda en evidencia en los contornos de los percentiles más bajos. Es necesario destacar que si bien el rendimiento de PST en este caso sería muy similar al de PSR, los contornos generados por PSK no serían óptimos por más que el soporte de la distribución

y la frontera esté bien identificados pues los *vecinos cercanos* calculados utilizando distancia euclídea generan que los contornos de PSK sean más circulares que elípticos.

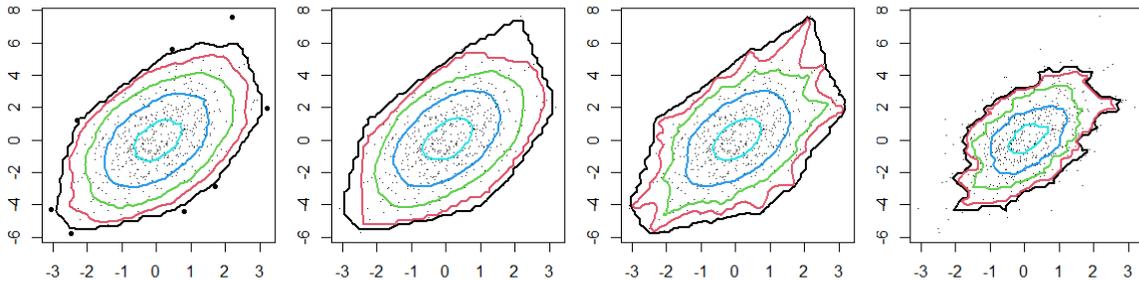


Figura 5.3: Caso 3: de izquierda a derecha, PSR, HS, SD, MK.

En el Caso 4 (Figura 5.4), las conclusiones son idénticas al caso anterior; las tres profundidades PSR, HS y SD funcionan de manera muy similar, mientras que sería posible mejorar la performance de MK a partir de una mejor elección del parámetro  $\alpha$  por ejemplo crossvalidando el valor dentro de un rango amplio, pero esto indicaría que el uso de la profundidad estaría supeditada a ese parámetro para cada caso en particular lo cuál no es el objetivo que perseguimos en este trabajo.

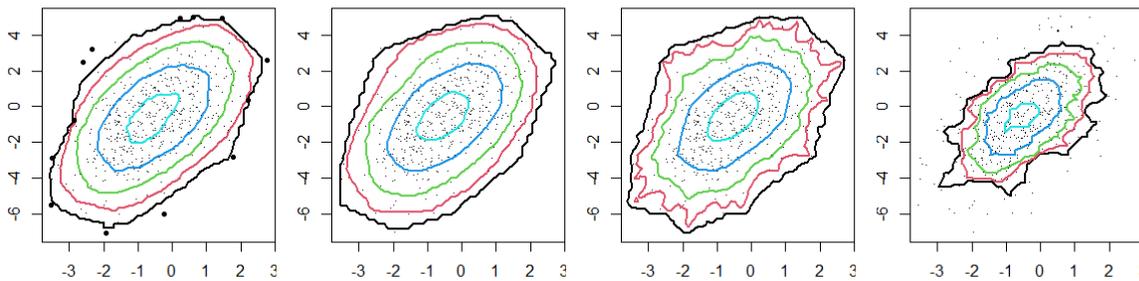


Figura 5.4: Caso 4: de izquierda a derecha, PSR, HS, SD, MK.

### Distribuciones contaminadas

Para mostrar la versatilidad de la profundidad  $\mathcal{PS}$ , continuamos el análisis con situaciones en donde las distribuciones son las mismas a las estudiadas hasta ahora, pero con *contaminación*, que en todos los casos es de un 10 % de los puntos.

- Caso 5: 90 % Caso 1 y 10 % puntos uniformes en  $[2, 3] \times [1, 2]$ .
- Caso 6: 90 % Caso 2 y 10 % puntos uniformes en  $[2, 3] \times [2, 3]$ .
- Caso 7: 90 %  $N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}\right)$  y 10 %  $N\left(\begin{bmatrix} 10 \\ 10 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}\right)$ .

- Caso 8: 45 %  $N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}\right)$  y 45 %  $N\left(\begin{bmatrix} -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}\right)$  y 10 %  $N\left(\begin{bmatrix} 10 \\ 10 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}\right)$ .

Los resultados pueden verse en las Figuras 5.5 a 5.8. Vemos que HS y SD en todos los casos tienen problemas para mostrar que existe contaminación en la muestra e identificar el soporte hasta en el caso de distribuciones elípticas; y no sólo eso, sino que corren los percentiles centrales estimados de las muestras hacia el espacio entre los clusters de puntos.

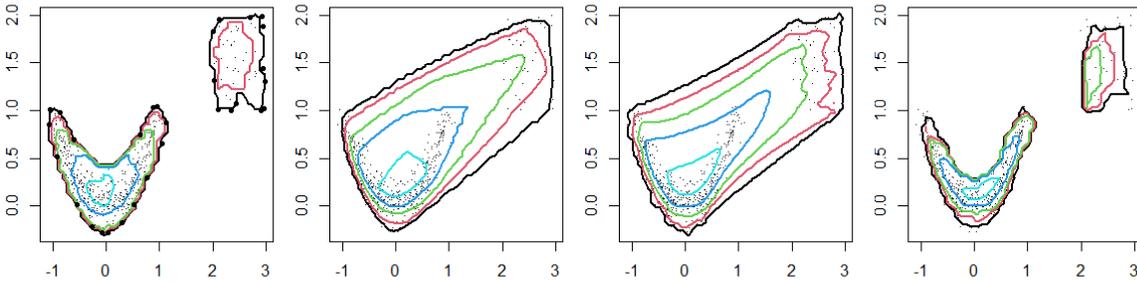


Figura 5.5: Caso 5: de izquierda a derecha, PST, HS, SD, MK.

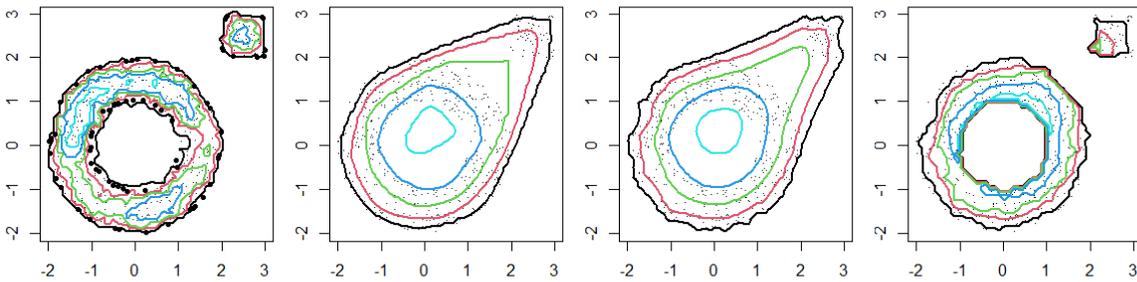


Figura 5.6: Caso 6: de izquierda a derecha, PST, HS, SD, MK.

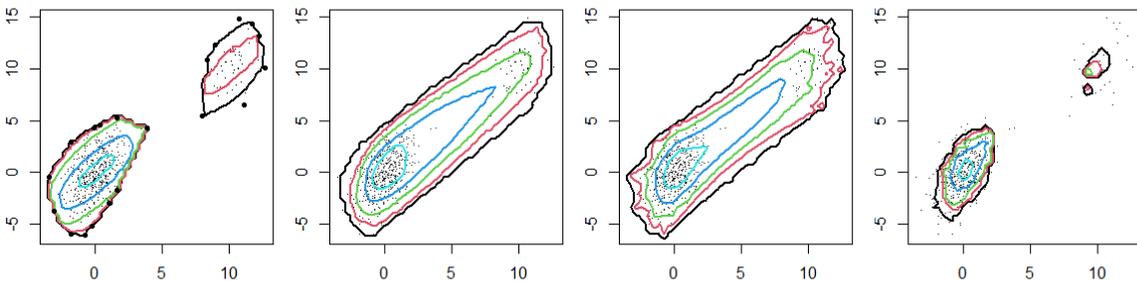


Figura 5.7: Caso 7: de izquierda a derecha, PSR, HS, SD, MK.

Por otro lado, tanto MK como PSR y PST funcionan de manera similar a los casos en los que no había contaminación. En todas las situaciones identifican al 10 % de outliers y consiguen generar un orden dentro de cada componente, que en el caso de MK sigue dependiendo fuertemente del parámetro  $\alpha$  utilizado.

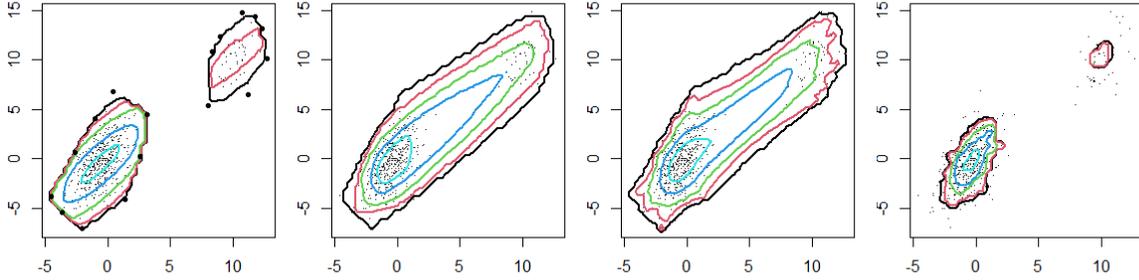


Figura 5.8: Caso 8: de izquierda a derecha, PSR, HS, SD, MK.

## 5.2. DD-plots y clasificación

Hemos comentado en la Sección 1.2 que una de las aplicaciones más utilizadas de las funciones de profundidad es el uso de éstas para generar el DD-plot (Depth vs Depth plot), un gráfico bidimensional en el que cada punto representa la profundidad de un punto del espacio  $\mathbb{R}^d$  respecto de una y de otra distribución, más precisamente:

**Definición 5.1** (DD-plot). Sean  $F_X$  y  $F_Y$  dos distribuciones en  $\mathbb{R}^d$  y sea  $D(\cdot, \cdot)$  una función de profundidad. El DD-plot se define como

$$DD(F_X, F_Y) = \{(D(z, F_X), D(z, F_Y)), z \in \mathbb{R}^d\}.$$

Para la versión empírica, sean  $\mathcal{X}_n = \{X_1, \dots, X_n\}$  e  $\mathcal{Y}_m = \{Y_1, \dots, Y_m\}$  dos muestras aleatorias de las distribuciones  $F_X$  y  $F_Y$  respectivamente. Si denotamos  $F_X^n$  y  $F_Y^m$  a las medidas empíricas se define

$$DD(F_X^n, F_Y^m) = \{(D(z, F_X^n), D(z, F_Y^m)), z \in \mathcal{X}_n \cup \mathcal{Y}_m\}.$$

Esto permite trabajar con un plot en dos dimensiones por mas que la dimensión en que se encuentren los datos sea alta. El uso de DD-plot tiene muchas aplicaciones en estadística no paramétrica como hemos comentado, pero en este caso nos interesa que permite realizar tareas de clasificación sin la necesidad de estimar parámetros como la media y varianza o algún otro parámetro de escala, lo que es requerido a veces en los clasificadores.

La idea de utilizar funciones de profundidad para clasificación fue comentada por Liu [36] y presentada por Ghosh y Chaudhuri [23] clasificando al punto dependiendo de qué distribución daba mayor profundidad, es decir, usando la línea que pasa por origen de pendiente 1 en el gráfico para decidir a qué población asignar el punto. Esta idea fue generalizada en Li, Cuesta-Albertos y Liu [34] permitiendo polinomios como curvas que separan las distribuciones y donde además se muestra que la mejor curva que separa dos poblaciones con distribución elíptica en  $\mathbb{R}^d$  es equivalente a una curva que separe a las dos muestras aleatorias en el DD-plot.

En Lange, Mosler y Mozharovskyi [33] proponen otro algoritmo de clasificación basado en DD-plots llamado procedimiento  $DD\alpha$  y generalizan a clasificación de más de 2 poblaciones; luego en Cuesta-Albertos, Febrero-Bande y de la Fuente [10] se extienden las ideas

a la aplicación de otros clasificadores en el DD-plot (kNN, LDA, QDA, etc) y la utilización para datos funcionales.

Una idea que es necesaria tener en cuenta es que un clasificador basado en el DD-plot puede alcanzar distintas propiedades según las profundidades utilizadas y que a veces la elección de la profundidad puede mejorar la performance del clasificador. Este concepto comentado en Li et al. [34] hace que nuestro objetivo aquí no sea ver cuál es el mejor método de clasificación basado en DD-plots, sino que nos proponemos mostrar que el uso de  $\mathcal{PS}$  puede ser beneficioso en los rendimientos de clasificadores ya establecidos; para esto utilizaremos datos simulados y reales.

### 5.2.1. Datos simulados

Comenzamos realizando un estudio de simulación que consta de 8 testeos diferentes que consisten en clasificar entre “muestra 1” y “muestra 2”; en cada testeo se realizan 100 experimentos y se registra el porcentaje de error de clasificación. Más precisamente, cada experimento tiene dos fases; en la fase *entrenamiento* 400 observaciones (200 de cada distribución) son generadas, calculado el DD-plot y entrenado el clasificador; en la fase *evaluación*, 1000 observaciones (500 de cada distribución) son generadas y una vez calculadas las profundidades de estas muestras respecto de las de entrenamiento, se genera el DD-plot y registra el porcentaje de error de clasificación dado por el clasificador.

Sin pérdida de generalidad, para mejorar la visualización y comparación, todas las profundidades son estandarizadas para que los DD-plot sean graficados en  $[0, 1] \times [0, 1]$  y en todos los casos el clasificador utilizado es Random Forest calculado con la función `randomForest` del paquete *randomForest*; recordando que nuestro objetivo es verificar que el uso de  $\mathcal{PS}$  en ciertos escenarios mejora la performance de los clasificadores basados en DD-plots, optamos por utilizar un clasificador genérico que no requiere de parámetros a fijar.

En todos los testeos reportaremos los resultados de PSK, PSR, PST, HS y SD calculadas con las especificaciones dadas anteriormente; para el caso de *MK* utilizamos los parámetros  $\alpha \in \{0,3, 0,5, 0,9\}$  pero sólo reportamos el que resulte con menor error de clasificación como MK.3, MK.5 o MK.9 respectivamente. Sumaremos las profundidades “Zonoid” y “Random Projection” (RP) calculadas con las funciones `depth.zonoid` y `depth.projection` del paquete *ddalpha*. En la práctica, puede ser computacionalmente muy costoso el cálculo de la *profundidad de proyecciones* (PD), especialmente en dimensiones altas; es por esto que trabajaremos con la *profundidad de proyecciones aleatorias*, “*Random projection depth*”, que consiste en aproximar PD al reemplazar  $\{u : \|u\| = 1\}$  en (1.2) por un conjunto finito de direcciones aleatorias. Claramente, mayor sea la cantidad de direcciones utilizadas, mejor resulta la aproximación, en todos los casos reportados se utilizaron 1000 direcciones aleatorias.

Para la “muestra 1”, en los testeos *I* a *IV* las muestras de entrenamiento y evaluación tienen las distribuciones de los casos 1 a 4 respectivamente, mientras que en los testeos *V* a *VIII* las muestras de entrenamiento tienen las distribuciones de los casos 5 a 8 y las muestras

de evaluación tienen las distribuciones de los caso 1 a 4 respectivamente, es decir, en los testeos *V* a *VIII* se entrena con muestra contaminada y testea con muestra sin contaminar.

En todos los testeos la “muestra 2” de entrenamiento y evaluación proviene de la misma distribución, las cuales son según el testeo:

- Test *I* y *V*:  $(1 + Y + T \cos(\varphi), 1,5 - Y^2 + T \sin(\varphi))$  donde  $Y \sim U[-1, 1]$ ,  $\varphi \sim U[0, 2\pi]$ ,  $W \sim U[0, 1]$ ,  $Y, \varphi, W$  independientes y  $T = 0,2 W(1 + (1 - |Y|)/2)$ .
- Test *II* y *VI*:  $N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}\right)$ .
- Test *III* y *VII*:  $N\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}\right)$ .
- Test *IV* y *VIII*: 50 %  $N\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}\right)$  y 50 %  $N\left(\begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}\right)$ .

En palabras, lo que se testea es la clasificación de dos poblaciones con las siguientes características: en *I* los soportes son disjuntos (pero no convexos); en *II* los soportes no son disjuntos pero uno es un anillo; en *III* normales bivariadas con cambio de media al igual que en *IV*; los testeos *V* a *VIII* tienen el mismo espíritu pero suman contaminación a las muestras de entrenamiento de la población 1.

En la Figura 5.9 se puede apreciar que en los testeos *I* y *V* todas las versiones de  $\mathcal{PS}$  tienen una performance muy buena con aproximadamente el 95 % de buena clasificación, la cual no se ve afectada por la contaminación en la muestra de entrenamiento; por otro lado el resto de las profundidades tiene un error de clasificación del orden del 10 % y algunas sí se ven afectadas por la aparición de datos atípicos. Los resultados pueden ser comprendidos a partir de ver los DD-plot que generan algunas de las profundidades, por ejemplo para el testeo *I* (Figura 5.10). El porcentaje de mala clasificación de MK puede resultar llamativo pues el DD-plot correspondiente es muy similar al de PST, así como eran muy similares los plots de conjuntos de nivel, pero el resultado se explica por la cantidad de puntos en el  $(0, 0)$  que el DD-plot de MK tiene. Con la profundidad  $\mathcal{PS}$  no se tiene ese problema por la inclusión del estado  $\Delta$  dentro del espacio de estados del paseo aleatorio, en la Observación 5.1 se ilustra y comenta esta idea para el testeo *VI*. Todos los resultados tienen una clara relación y son una consecuencia de la situación mostrada en términos de los conjuntos de nivel de las profundidades para los casos 1 y 5 (Figuras 5.1 y 5.5).

En los testeos *II* y *VI*, al solaparse casi por completo los soportes de las distribuciones, como es de esperarse todos los porcentajes de error de clasificación son superiores a los anteriores. Sin embargo, en la Figura 5.11 puede verse que cuando las muestras no están contaminadas casi todas las profundidades mantienen un error de clasificación similar; por otro lado, mientras que todas las variantes de  $\mathcal{PS}$  mantienen una performance semejante en el testeo *VI* respecto del testeo *II*, las restantes profundidades se ven afectadas por la contaminación de la muestra de entrenamiento de la población 1. Puede apreciarse en la Figura 5.12 que PST al tener una buena estimación del soporte, consigue diferenciar todos los puntos fuera del anillo y del bloque extra de puntos uniformes como de la población

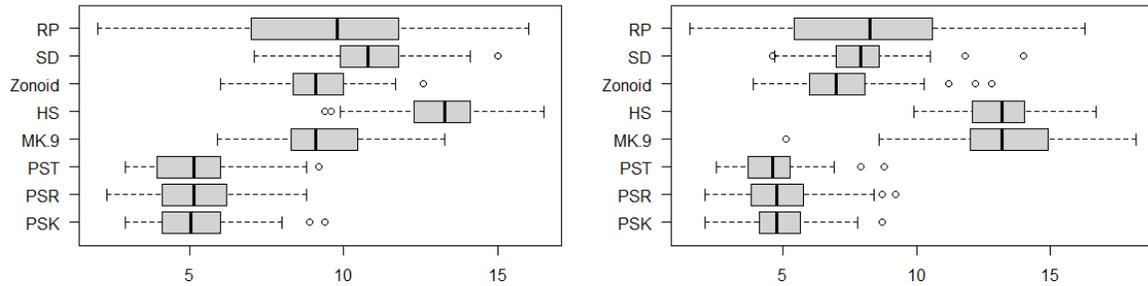


Figura 5.9: % de error de clasificación testeo *I* (izquierda) y *V* (derecha).

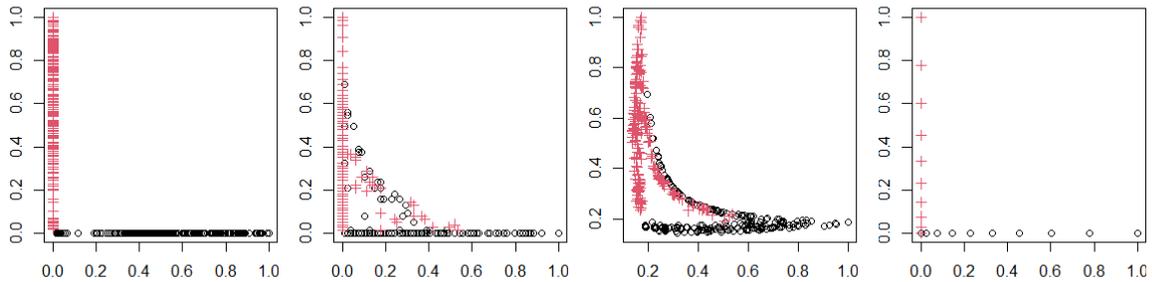


Figura 5.10: Testeo *I*, DD-plot entrenamiento de PST, HS, RP y MK.9.

2 (puntos '+' sobre el eje *Y*), mientras que justamente a la “contaminación” la identifica como con poca profundidad para ambas poblaciones (puntos 'o' sobre el eje *X*), haciendo que no influya al momento de testear con poblaciones no contaminadas. Debemos destacar que el formato que adquiere el DD-plot de MK se debe a que para su cálculo se usó como distribución de referencia a la distribución uniforme en el círculo unitario que fue aproximada con 10 circunferencias concéntricas de 20 puntos equiespaciados.

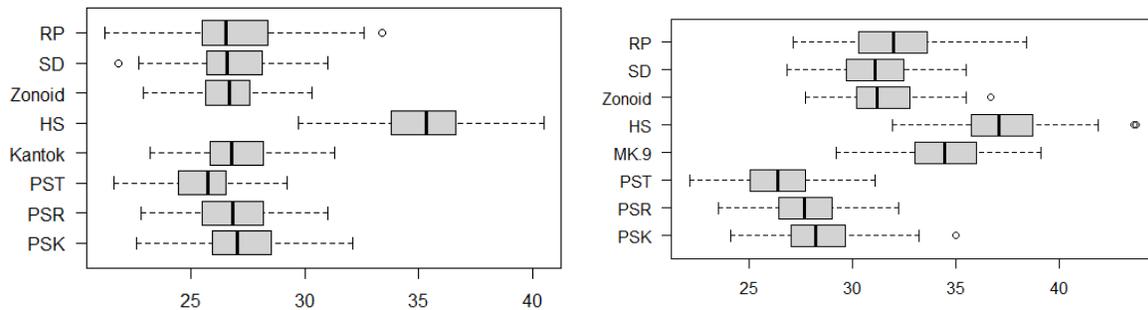


Figura 5.11: % de error de clasificación testeo *II* (izquierda) y *VI* (derecha).

**Observación 5.1.** Como notamos en la Obsevación 3.1 la inclusión del estado  $\Delta$  al paseo aleatorio en el grafo no influye desde el punto de vista probabilístico. El DD-plot de PST en

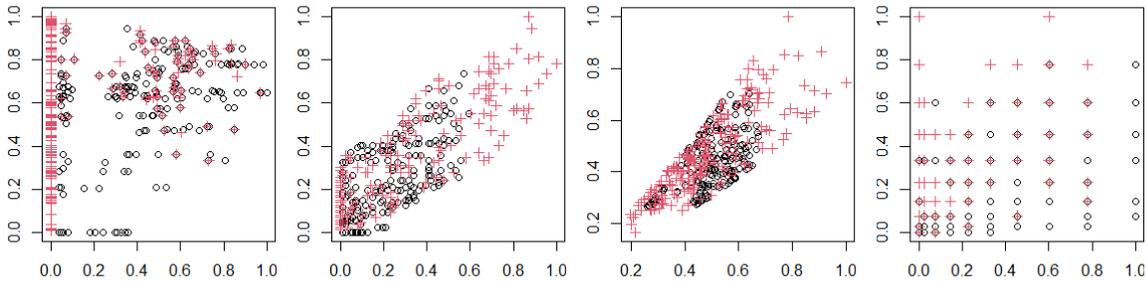


Figura 5.12: Testeo VI, DD-plot entrenamiento de PST, HS, RP y MK.9.

la Figura 5.12 es una muestra de cómo si influye desde el punto de vista estadístico. Ayuda a diferenciar los puntos de la población 2 fuera del soporte de la población 1 (puntos '+' sobre el eje Y) de los puntos en el límite del soporte de la población 1 (puntos negros cercanos al eje Y).

En la Figura 5.13 puede verse que en el testeo III, clasificar normales bivariadas con cambio de media, todas las profundidades clásicas tienen un rendimiento similar al igual que PST, PSR y MK; el caso que se destaca por un rendimiento menor es el de PSK, debiéndose a que se utiliza distancia euclídea en la búsqueda de los  $k$  vecinos. Por otro lado, en el testeo VII, cuando se utilizan muestras de entrenamiento contaminadas, la performance de todas las  $\mathcal{P}\mathcal{S}$  vuelven a estar dentro de las mejores, destacándose RP dentro de las clásicas. Puede verse en la Figura 5.14 que en este caso, la contaminación no influye en RP (puntos 'o' cerca del (0,0)) pues existen muchas direcciones en las que las proyecciones de esos puntos quedan muy alejadas de la población general, por lo que en ambos casos sus profundidades son muy bajas, mientras que PST consigue diferenciar a la contaminación pues respecto de la población 1 le da baja profundidad mientras que respecto de la población 2 le da profundidad 0 (puntos 'o' sobre el eje X).

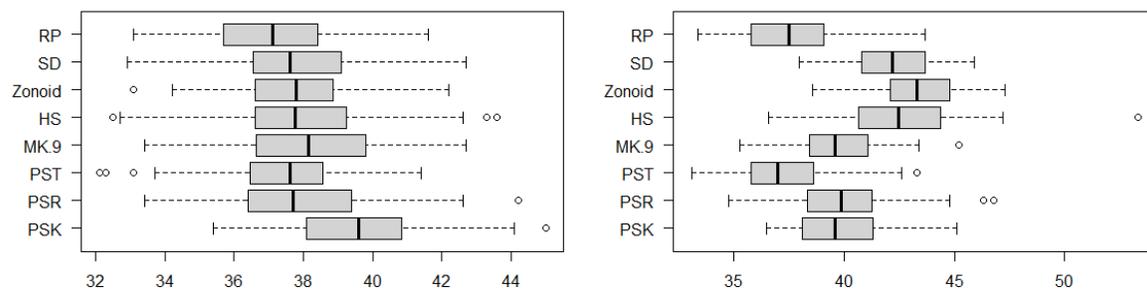


Figura 5.13: % de error de clasificación testeos III (izquierda) y VII (derecha).

Finalmente, en la Figura 5.15 se encuentran los resultados de los testeos IV y VIII, donde se puede apreciar una situación similar a la anterior, con PSK teniendo una performance no tan buena con respecto al resto de las profundidades en el testeo IV, pero teniendo un

rendimiento superior todas las  $\mathcal{P}S$  en el caso de muestra de entrenamiento contaminada, siendo RP nuevamente la más destacada en ambos casos.

Nuevamente puede verse en la Figura 5.16 que el rendimiento de MK se ve disminuido por el hecho de que la profundidad puede tomar una cantidad fija de valores distintos igual a la cantidad de niveles fijados de la distribución de referencia.

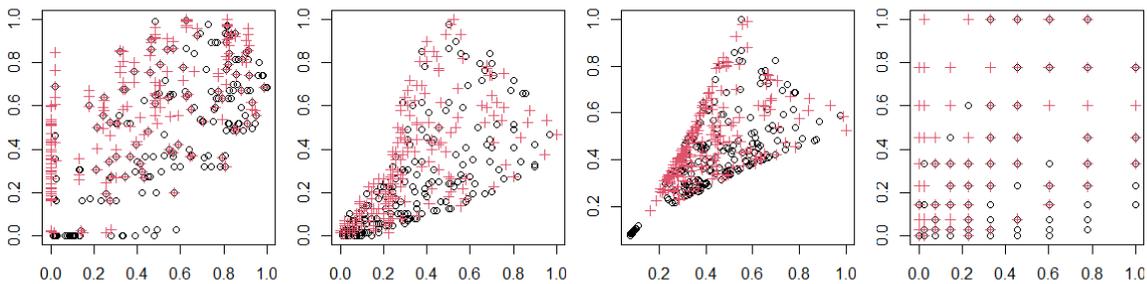


Figura 5.14: Testeo VII, DD-plot entrenamiento de PST, HS, RP y MK.9.

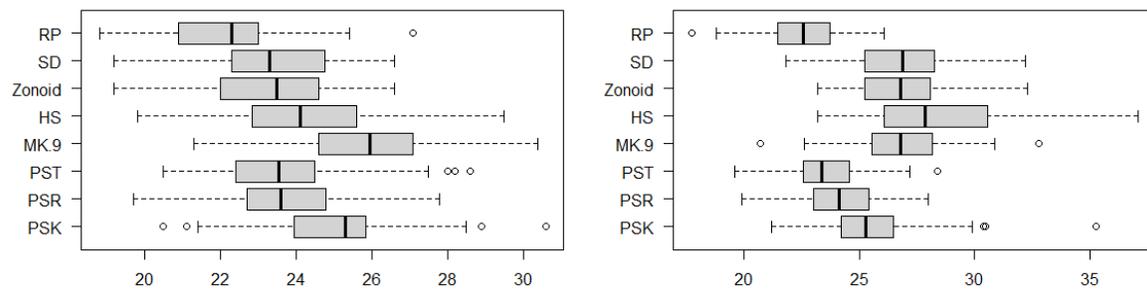


Figura 5.15: % de error de clasificación testeos IV (izquierda) y VIII (derecha).

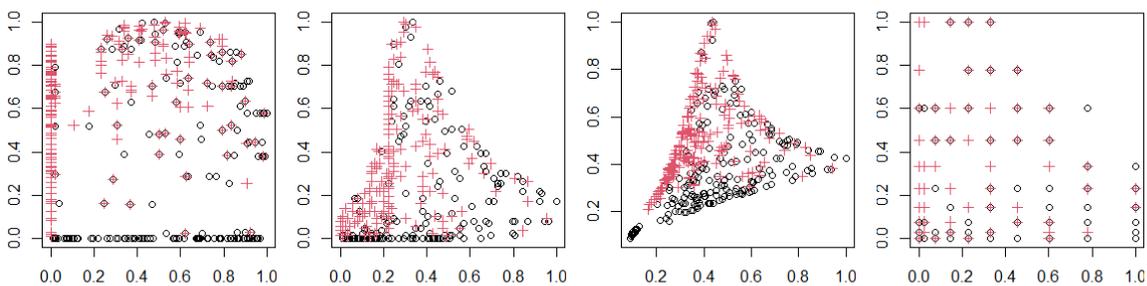


Figura 5.16: Testeo VIII, DD-plot entrenamiento de PST, HS, RP y MK.9.

### 5.2.2. Datos reales

Continuamos ahora con un estudio de clasificación aplicado a datos reales en dimensiones más altas para testear el uso de  $\mathcal{PS}$  en estos casos. En esta oportunidad no sólo se utilizará un clasificador basado en DD-plot para varias profundidades sino que se comparará también con la performance de los clásicos clasificadores LDA y QDA, calculadas con las funciones *lda* y *qda* de la librería *MASS*, kNN de la librería *class* y Random Forest (RF), en este caso no como clasificador en el DD-plot. Para determinar el  $k$  óptimo en el clasificador kNN en cada clasificación se realizó una crosvalidación ‘leave-one-out’ para un rango amplio de  $k$ .

En este caso el clasificador basado en DD-plot que vamos a utilizar es el  $DD\alpha$ , calculado con la función *ddalpha.classify* del paquete *ddalpha*; este procedimiento permite que si en la fase de clasificación un punto a clasificar no pertenece a ninguna de las cápsulas convexas de por lo menos una de las poblaciones de entrenamiento (punto denominado ‘outsider’), sea clasificado según otro procedimiento para el cuál elegimos kNN con  $k$  crosvalidado.

En el análisis, reemplazamos la profundidad SD por la profundidad “Mahalanobis”, “Mahalanobis depth” (MhD) basada en la distancia de Mahalanobis [41], definida en Liu y Singh [38]:

$$MhD(x, \mathbb{P}) = \left[ 1 + (x - \mu_{\mathbb{P}})^T \Sigma_{\mathbb{P}}^{-1} (x - \mu_{\mathbb{P}}) \right]^{-1},$$

donde  $\mu_{\mathbb{P}}$  es la esperanza del vector aleatorio y  $\Sigma_{\mathbb{P}}$  la matriz de covarianzas. En este caso la versión empírica se obtiene reemplazando  $\mu_{\mathbb{P}}$  y  $\Sigma_{\mathbb{P}}$  por la media y la matriz de covarianzas muestrales respectivamente.

Vale aclarar que en dimensión mayor a 3, la profundidad HS se calcula con proyecciones aleatorias y las profundidades HS, Zonoid y MhD son calculadas en esta ocasión con las funciones internas del paquete *ddalpha*.

Para cada una de las 3 bases de datos utilizadas presentamos primero una mínima descripción y comentamos brevemente el procedimiento de clasificación llevado a cabo junto con el resultado obtenido. En todos los casos se realizaron 100 repeticiones de los experimentos descriptos y reportamos en la Tabla 5.1 el porcentaje de error de clasificación junto con el desvío estándar de cada uno de los clasificadores utilizados.

#### Biomedical data

Esta base de datos puede encontrarse en <http://lib.stat.cmu.edu/datasets/>, fue utilizada por primera vez por Cox [9]; consiste inicialmente de dos grupos de pacientes (“normals” y “carriers”) con un total de 209 observaciones y 8 variables. Se descartan 4 variables (número de observación, id, edad y fecha) y 15 observaciones por datos faltantes (8 “carriers” y 7 “normals”) para quedar con una base de datos en dimensión 4 que consta de 127 y 67 mediciones de la sangre de personas denominadas normales y portadores de un desorden genético poco conocido respectivamente.

Se eligieron al azar 100 observaciones del grupo de normales y 50 de portadores para formar el grupo de entrenamiento de los clasificadores y el resto de las observaciones fueron al grupo test. Se puede observar que prácticamente todas las profundidades obtienen el mismo resultado, destacándose mínimamente HS, mientras que sólo la performance de RF es similar al clasificador basado en DD-plot.

### Vertebral column data set

Base de datos creada por el doctor Henrique da Mota, puede encontrarse en el repositorio UCI [18]; consta de 310 datos de pacientes, cada uno representado por seis atributos biomecánicos referidos a la forma y orientación de la pelvis y la columna lumbar: incidencia pélvica, inclinación pélvica, ángulo de lordosis lumbar, pendiente sacra, radio pélvico y grado de espondilolistesis. Los datos fueron organizados de dos maneras distintas para realizar clasificación; la base *Vert3* contiene 3 categorías de pacientes: 60 con hernia de disco (DH), 150 con espondilolistesis (SL) y 100 normales (NO); la base *Vert2* contiene 2 categorías, los pacientes DH y SL juntos, denominados “anormlaes” (AB) y los pacientes normales.

Para la clasificación con 3 categorías, en la fase de entrenamiento se eligieron al azar 50 datos de DH, 120 de SL y 80 de NO, mientras que el resto fueron al grupo test; en el caso de la clasificación en 2 categorías, se eligieron para la fase de entrenamiento 160 datos de AB y 80 de NO y se testeó con el resto.

Cuando la clasificación es entre 3 categorías puede verse un buen rendimiento del clasificador basado en DD-plot, siendo muy similar el resultado para todas las profundidades, mientras que al momento de clasificar entre normales y anormales, la utilización de PSK para el clasificador da el mejor resultado, siendo este mejor a la utilización de cualquier otro clasificador.

### Image Segmentation

Esta es otra base de datos obtenida del repositorio UCI [18]. Cada observación contiene información sobre una región de 3x3 píxeles de una imagen. Las imágenes fueron extraídas de una base de datos de imágenes de exterior; cada observación consta de 19 atributos y pertenece a una de las 7 clasificaciones diferentes (de las cuales seleccionamos 3: Brickface, Cement y Window).

Luego de remover las variables redundantes y dependientes, trabajamos con las variables:

- *rcc,rcw*: miden la columna y fila central de píxeles de la región.
- *vm,vs,hm,hs*: media y desvío de una medida del contraste vertical y horizontal de los píxeles de la región.
- *B, G*: promedio de luz azul y verde de la región.

Con los 330 datos de Brickface y Cement generamos la base *ImgBC* para realizar clasificación entre 2 categorías y junto con los 330 datos de Window generamos la base *Img3* para clasificar entre 3 categorías. Para cada una de estas bases, se eligieron al azar 100 observaciones de cada grupo para formar el grupo de entrenamiento de los clasificadores y el resto de las observaciones fueron al grupo test.

En este caso, las observaciones se encuentran en dimensión 8, por lo que los clasificadores basados en vecinos cercanos (recordar que RF puede ser visto como un clasificador basado en vecinos, los cuáles en este caso son los puntos que se encuentran en una misma “hoja” en alguno de los “árboles” de entrenamiento), tienen una performance destacada sobre el resto de los clasificadores. Con respecto al clasificador basado en los DD-plot, todas las versiones de  $\mathcal{PS}$  obtienen el mejor resultado en el caso que se clasifica entre 2 poblaciones; mientras que en el caso de 3 poblaciones se destaca el uso de HS.

Data	PSK	PSR	PST	Zonoid	HS	MhD	RF	Knn	LDA	QDA
Biom	12.89 (4.99)	12.52 (4.79)	12.73 (5.00)	12.34 (5.09)	11.75 (4.20)	12.64 (4.13)	12.61 (4.54)	13.82 (4.34)	16.77 (4.54)	13.52 (4.14)
Vert3	15.70 (3.89)	15.72 (3.78)	15.40 (3.72)	15.27 (4.17)	15.88 (3.97)	16.18 (3.66)	14.27 (3.69)	23.42 (4.85)	18.38 (5.12)	15.42 (3.42)
Vert2	14.80 (3.39)	15.66 (4.16)	15.31 (3.72)	15.69 (3.97)	15.76 (4.08)	15.27 (3.96)	15.84 (4.65)	20.21 (4.64)	15.50 (3.87)	17.77 (4.21)
ImgBC	4.12 (1.09)	4.20 (1.28)	4.15 (1.18)	4.21 (1.15)	5.67 (1.57)	4.27 (0.94)	2.86 (1.03)	2.95 (0.79)	6.35 (1.06)	4.07 (0.81)
Img3	18.05 (1.68)	18.07 (1.58)	18.06 (1.72)	18.12 (1.74)	15.47 (1.60)	18.98 (2.22)	7.12 (1.05)	11.44 (2.07)	23.91 (1.29)	25.89 (1.68)

Cuadro 5.1: Porcentaje de errores de los clasificadores junto con los desvíos standard.

### 5.3. Comentarios y futuras líneas de investigación

Llegamos al final de la tesis presentando algunas posibles futuras líneas de investigación. Las ideas listadas a continuación no fueron profundamente desarrolladas, son conjeturas y comentarios resultado de la intuición generada al resolver los diversos problemas en esta tesis abordados.

- La implementación de la versión empírica permitió comprobar que  $\mathcal{PS}_n$  distingue contaminación en las muestras. Se realizaron algunas simulaciones no mostradas en la tesis; estas sugieren que sería posible obtener una versión “local” de la profundidad  $\mathcal{PS}$ , que debería depender de un tiempo de parada diferente, por ejemplo, el de salida de áreas cuya densidad pasa un determinado umbral.

- Como vimos en la Observación 2.5, es posible trabajar con otras funciones  $\varphi(\tau)$  para definir la profundidad. No encontramos cuál es la función que daría mejores propiedades a  $\mathcal{PS}$ . Relacionado al comentario anterior, sería posible pensar  $\varphi$  como una función no monótona, con el objetivo de detectar las modas de la distribución.
- La aplicación de la Triangulación de Delaunay en todos los casos de simulación, dejó en evidencia que, de existir resultados más fuertes para la estimación de densidades usando este grafo, sería recomendable utilizarlo preferentemente para el cálculo de  $\mathcal{PS}_n$  pues tiene mucha más información intrínsecamente relacionada a la distribución estudiada que cualquiera de los otros grafos vistos. Existen diferentes ciencias que utilizan  $\mathcal{T}_n$  y sería de mucha utilidad que se obtengan propiedades y resultados de estimadores de densidad basados en  $\mathcal{T}_n$ .
- La Triangulación de Delaunay es comúnmente utilizada al generar grillas para discretizar el espacio en el método de elementos finitos o el método de volumen finito, por las características únicas de la triangulación. No fue posible probar consistencia del estimador  $\mathcal{PS}_n$  usando resultados de análisis numérico, pues no se tiene “control” de la grilla de puntos como es habitual en ese área. Sería interesante el desarrollo de herramientas que ayuden a encontrar aproximaciones de soluciones de PDE a partir de un enfoque probabilístico, que no parta de la idea de utilizar una grilla determinística sino aleatoria como en este trabajo.
- La idea usada en dimensión 1 de transformar una grilla no equiespaciada (aleatoria) en la típica grilla en  $[0, 1]$ , podría ser considerada para otras aplicaciones en probabilidad (por ejemplo, ser una herramienta para pensar límites de paseos aleatorios en grafos en general) y para aplicaciones como la del punto anterior.
- En  $d = 1$ , se obtiene que la inversa de la distribución acumulada de  $\mathbf{X}$  es el mapeo que relaciona al Movimiento Browniano con la difusión  $Y_t$ . Resulta que  $F^{-1}$  es el mapeo de transporte óptimo para la pérdida cuadrática de la distribución  $U[0, 1]$  a la de  $\mathbf{X}$ . Sería interesante ver si  $Y_t$  en el caso  $d \geq 2$  es el resultado de aplicar a un Movimiento Browniano  $d$ -dimensional una transformación, por ejemplo, la “pushforward” que mapea la distribución uniforme en el círculo unitario de  $\mathbb{R}^d$  en la de  $\mathbf{X}$ .
- Otra posible interpretación de  $Y_t$  es que se trata de una proyección en  $\mathbb{R}^d$  de un Movimiento Browniano que se mueve en la superficie  $(x, f(x)) \in \mathbb{R}^{d+1}$ , donde  $f(x)$  es la densidad de  $\mathbf{X}$ . Esto es claro en el caso uniforme pues la densidad constante hace que no haya cambios en la superficie y por lo tanto las proyecciones muestren un Movimiento Browniano en  $\mathbb{R}^d$ .
- Relacionado al punto anterior, se podría llegar a extender la definición de  $\mathcal{PS}$  a variables aleatorias en variedades, teniendo en cuenta, por ejemplo, que los resultados de Ting, Huang y Jordan [57] sobre convergencia de los laplacianos de grafos aleatorios

(a generadores de difusiones) son válidos para distribuciones en general. Éste sería un gran aporte, teniendo en cuenta la necesidad de existencia de funciones de profundidad definidas para variables aleatorias en superficies, concepto no muy estudiado.

# Bibliografía

- [1] C. Aaron, “Using the k- nearest neighbor restricted Delaunay polyhedron to estimate the density support and its topological properties”, 2012.
- [2] C. Aaron y O. Bodart, “Local convex hull support and boundary estimation”, *Journal of Multivariate Analysis*, vol. 147, págs. 82-101, 2016.
- [3] M. Abellanas, M. Claverol y F. Hurtado, “Point set stratification and Delaunay depth”, *Computational statistics & data analysis*, vol. 51, n.º 5, págs. 2513-2530, 2007.
- [4] C. Agostinelli y M. Romanazzi, “Local depth”, *Journal of Statistical Planning and Inference*, vol. 141, n.º 2, págs. 817-830, 2011.
- [5] M. Bern, D. Eppstein y F. Yao, “The expected extremes in a Delaunay triangulation”, *International Journal of Computational Geometry & Applications*, vol. 1, n.º 01, págs. 79-91, 1991.
- [6] P. Billingsley, *Convergence of probability measures*. John Wiley & Sons, 2013.
- [7] V. Chernozhukov, A. Galichon, M. Hallin, M. Henry y col., “Monge–Kantorovich depth, quantiles, ranks and signs”, *The Annals of Statistics*, vol. 45, n.º 1, págs. 223-256, 2017.
- [8] J. Chevalier, “Estimation du support et du contour du support d’une loi de probabilité”, en *Annales de l’IHP Probabilités et statistiques*, vol. 12, 1976, págs. 339-364.
- [9] L. Cox, “Exposition of statistical graphics technology”, *ASA Proceedings of the Statistical Computation Section, 1982*, págs. 55-56, 1982.
- [10] J. A. Cuesta-Albertos, M. Febrero-Bande y M. O. de la Fuente, “The DD-G-classifier in the functional setting”, *Test*, vol. 26, n.º 1, págs. 119-142, 2017.
- [11] J. A. Cuesta-Albertos y A. Nieto-Reyes, “The random Tukey depth”, *Computational Statistics & Data Analysis*, vol. 52, n.º 11, págs. 4979-4988, 2008.
- [12] A. Cuevas, R. Fraiman y col., “A plug-in approach to support estimation”, *The Annals of Statistics*, vol. 25, n.º 6, págs. 2300-2312, 1997.
- [13] A. Cuevas y A. Rodríguez-Casal, “On boundary estimation”, *Advances in Applied Probability*, págs. 340-354, 2004.

- [14] L. Devroye y G. L. Wise, “Detection of abnormal behavior via nonparametric estimation of the support”, *SIAM Journal on Applied Mathematics*, vol. 38, n.º 3, págs. 480-488, 1980.
- [15] L. P. Devroye y T. J. Wagner, “The strong uniform consistency of nearest neighbor density estimates”, *The Annals of Statistics*, págs. 536-540, 1977.
- [16] L. P. Devroye, “Nonparametric Discrimination and Density Estimation.”, TEXAS UNIV AT AUSTIN DEPT OF ELECTRICAL ENGINEERING, inf. téc., 1976.
- [17] D. L. Donoho, “Breakdown properties of multivariate location estimators”, Technical report, Harvard University, Boston. URL [http://www-stat.stanford . . .](http://www-stat.stanford.edu/~donoho/), inf. téc., 1982.
- [18] D. Dua y C. Graff, *UCI Machine Learning Repository*, 2017. dirección: <http://archive.ics.uci.edu/ml>.
- [19] L. Dümbgen y G. Walther, “Rates of convergence for random approximations of convex sets”, *Advances in applied probability*, págs. 384-393, 1996.
- [20] R. A. Dwyer, “Higher-dimensional Voronoi diagrams in linear expected time”, *Discrete & Computational Geometry*, vol. 6, n.º 3, págs. 343-367, 1991.
- [21] R. Dyckerhoff, K. Mosler y G. Koshevoy, “Zonoid data depth: Theory and computation”, en *COMPSTAT*, Springer, 1996, págs. 235-240.
- [22] B. Efron, “The convex hull of a random set of points”, *Biometrika*, vol. 52, n.º 3-4, págs. 331-343, 1965.
- [23] A. K. Ghosh y P. Chaudhuri, “On maximum depth and related classifiers”, *Scandinavian Journal of Statistics*, vol. 32, n.º 2, págs. 327-350, 2005.
- [24] D. Gilbarg y N. S. Trudinger, *Elliptic partial differential equations of second order*. springer, 2015.
- [25] M. Hallin, E. Del Barrio, J. Cuesta-Albertos y C. Matrán, “Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach”, *The Annals of Statistics*, vol. 49, n.º 2, págs. 1139-1165, 2021.
- [26] T. Hashimoto, Y. Sun y T. Jaakkola, “From random walks to distances on unweighted graphs”, en *Advances in Neural Information Processing Systems*, 2015, págs. 3429-3437.
- [27] ———, “Metric recovery from directed unweighted graphs”, en *Artificial Intelligence and Statistics*, PMLR, 2015, págs. 342-350.
- [28] X. He y G. Wang, “Convergence of depth contours for multivariate datasets”, *The Annals of Statistics*, págs. 495-504, 1997.
- [29] M. Hein, J.-Y. Audibert y U. v. Luxburg, “Graph laplacians and their convergence on random neighborhood graphs.”, *Journal of Machine Learning Research*, vol. 8, n.º 6, 2007.

- [30] M. Hein, J.-Y. Audibert y U. Von Luxburg, “From graphs to manifolds—weak and strong pointwise consistency of graph Laplacians”, en *International Conference on Computational Learning Theory*, Springer, 2005, págs. 470-485.
- [31] N. Ikeda y S. Watanabe, *Stochastic differential equations and diffusion processes*. Elsevier, 2014.
- [32] A. U. Kennington, “Power concavity and boundary value problems”, *Indiana University Mathematics Journal*, vol. 34, n.º 3, págs. 687-704, 1985.
- [33] T. Lange, K. Mosler y P. Mozharovskyi, “Fast nonparametric classification based on data depth”, *Statistical Papers*, vol. 55, n.º 1, págs. 49-69, 2014.
- [34] J. Li, J. A. Cuesta-Albertos y R. Y. Liu, “DD-classifier: Nonparametric classification procedure based on DD-plot”, *Journal of the American Statistical Association*, vol. 107, n.º 498, págs. 737-753, 2012.
- [35] J. Li y R. Y. Liu, “New nonparametric tests of multivariate locations and scales using data depth”, *Statistical Science*, págs. 686-696, 2004.
- [36] R. Y. Liu, “On a notion of data depth based on random simplices”, *The Annals of Statistics*, págs. 405-414, 1990.
- [37] R. Y. Liu, J. M. Parelus, K. Singh y col., “Multivariate analysis by data depth: descriptive statistics, graphics and inference,(with discussion and a rejoinder by Liu and Singh)”, *The annals of statistics*, vol. 27, n.º 3, págs. 783-858, 1999.
- [38] R. Y. Liu y K. Singh, “A quality index based on data depth and multivariate rank tests”, *Journal of the American Statistical Association*, vol. 88, n.º 421, págs. 252-260, 1993.
- [39] —, “Rank tests for multivariate scale difference based on data depth”, *DIMACS series in discrete mathematics and theoretical computer science*, vol. 72, pág. 17, 2006.
- [40] S. López-Pintado y J. Romo, “On the concept of depth for functional data”, *Journal of the American statistical Association*, vol. 104, n.º 486, págs. 718-734, 2009.
- [41] P. C. Mahalanobis, “On the generalized distance in statistics”, National Institute of Science of India, 1936.
- [42] B. Oksendal, *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.
- [43] D. Paindaveine y G. Van Bever, “From depth to local depth: a focus on centrality”, *Journal of the American Statistical Association*, vol. 108, n.º 503, págs. 1105-1119, 2013.
- [44] M. Reitzner y col., “Random polytopes and the Efron–Stein jackknife inequality”, *The Annals of Probability*, vol. 31, n.º 4, págs. 2136-2166, 2003.
- [45] P. J. Rousseeuw y M. Hubert, “Regression depth”, *Journal of the American Statistical Association*, vol. 94, n.º 446, págs. 388-402, 1999.

- [46] P. J. Rousseeuw, I. Ruts y J. W. Tukey, “The bagplot: a bivariate boxplot”, *The American Statistician*, vol. 53, n.º 4, págs. 382-387, 1999.
- [47] P. J. Rousseeuw y A. Struyf, “Computing location depth and regression depth in higher dimensions”, *Statistics and Computing*, vol. 8, n.º 3, págs. 193-203, 1998.
- [48] A. Rousselle y col., “Quenched invariance principle for random walks on Delaunay triangulations”, *Electronic Journal of Probability*, vol. 20, 2015.
- [49] W. E. Schaap, “The Delaunay tessellation field estimator”, *Ph. D. Thesis*, 2007.
- [50] R. Serfling, “Depth functions in nonparametric multivariate inference”, *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 72, pág. 1, 2006.
- [51] R. J. Serfling, “Multivariate symmetry and asymmetry”, *Wiley StatsRef: Statistics Reference Online*, 2014.
- [52] G. Shevchenko, “Convergence of hitting times for jump-diffusion processes”, *Modern Stochastics: Theory and Applications*, vol. 2, n.º 3, págs. 203-218, 2015.
- [53] S. T. Smith, E. K. Kao, K. D. Senne, G. Bernstein y S. Philips, “Bayesian discovery of threat networks”, *IEEE Transactions on Signal Processing*, vol. 62, n.º 20, págs. 5324-5338, 2014.
- [54] W. A. Stahel, “Robuste schätzungen: infinitesimale optimalität und schätzungen von kovarianzmatrizen”, Tesis doct., ETH Zurich, 1981.
- [55] D. W. Stroock y S. S. Varadhan, “Diffusion processes with boundary conditions”, *Communications on Pure and Applied Mathematics*, vol. 24, n.º 2, págs. 147-225, 1971.
- [56] ———, *Multidimensional diffusion processes*. Springer Science & Business Media, 1997, vol. 233.
- [57] D. Ting, L. Huang y M. Jordan, “An analysis of the convergence of graph Laplacians”, *arXiv preprint arXiv:1101.5435*, 2011.
- [58] J. W. Tukey, “Mathematics and the picturing of data”, en *Proceedings of the International Congress of Mathematicians, Vancouver, 1975*, vol. 2, 1975, págs. 523-531.
- [59] U. Von Luxburg, M. Belkin y O. Bousquet, “Consistency of spectral clustering”, *The Annals of Statistics*, págs. 555-586, 2008.
- [60] W. Whitt, “Some useful functions for functional limit theorems”, *Mathematics of operations research*, vol. 5, n.º 1, págs. 67-85, 1980.
- [61] M. Yazdani, “Similarity learning over large collaborative networks”, EPFL, inf. téc., 2013.
- [62] L. Yen, M. Saerens, A. Mantrach y M. Shimbo, “A family of dissimilarity measures between nodes generalizing both the shortest-path and the commute-time distances”, en *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2008, págs. 785-793.

- [63] H. Zhou, “Distance, dissimilarity index, and network community structure”, *Physical review e*, vol. 67, n.º 6, pág. 061 901, 2003.
- [64] —, “Network landscape from a Brownian particle’s perspective”, *Physical Review E*, vol. 67, n.º 4, pág. 041 908, 2003.
- [65] Y. Zuo, “On general notions of depth for regression”, *Statistical Science*, vol. 36, n.º 1, págs. 142-157, 2021.
- [66] Y. Zuo y col., “Projection-based depth functions and associated medians”, *The Annals of Statistics*, vol. 31, n.º 5, págs. 1460-1490, 2003.
- [67] Y. Zuo y R. Serfling, “General notions of statistical depth function”, *Annals of statistics*, págs. 461-482, 2000.
- [68] —, “On the performance of some robust nonparametric location measures relative to a general notion of multivariate symmetry”, *Journal of Statistical Planning and Inference*, vol. 84, n.º 1-2, págs. 55-79, 2000.