Facultad

de

Ciencias

Profundidad Estadística por Reflexiones

(Statistical Depth by Reflections)

Trabajo de Fin de Grado para acceder al

Grado en Matemáticas.

Autor: Francisco Javier Soto Sánchez.
Directora: Alicia Nieto Reyes.

Junio de 2021

Abstract

We define a two-dimensional statistical depth function and study two important aspects of it: its robustness and computability. We begin by formally proving that the function is, indeed, a statistical depth function. To achieve this, we introduce a new notion of symmetry for distributions in $\mathbb{R}^p$. We study the robustness through the concept of breakdown point. In terms of computability, we provide an implementable algorithm to calculate the depth contours with temporal complexity $\Theta(n^2)$ and spatial complexity $\Theta(n)$, where n is the size of the data set. As an application of the proposed depth function, we provide a hypothesis test for the independence of two absolutely continuous variables.

Key words : statistical depth, reflections, median, robustness, breakdown point, band depth, algorithm, computational geometry, hyphotesis test.

Resumen

Definimos una función de profundidad estadística bi-dimensional y estudiamos dos aspectos importantes de la misma: la robustez y la computabilidad. Comenzamos probando formalmente que la función es, de hecho, una función de profundidad estadística. Para conseguirlo , introducimos una nueva noción de simetría para distribuciones en $\mathbb{R}^p$. Estudiamos la robustez a través del concepto de breakdown point. En cuanto a la computabilidad, proporcionamos un algoritmo implementable para calcular los contornos de profundidad con complejidad temporal $\Theta(n^2)$ y complejidad espacial $\Theta(n)$, siendo n el tamaño del conjunto de datos. Como aplicación de la función de profundidad propuesta, proporcionamos un contraste de hipótesis para la independencia de dos variables absolutamente continuas.

Palabras clave : profundidad estadística, reflexiones, mediana, robustez, breakdown point, profundidad de la banda, algoritmo, geometría computacional, test de hipótesis.

Contents

List of Figures

List of Tables

Reflexiones

En esta apartado me gustaría sintetizar las observaciones que tengo del trabajo realizado. En primer lugar considero que este TFG me ha servido para darme cuenta de que las matemáticas son complejas. Me explico. En mi época de instituto pensaba que las matemáticas eran básicamente manejar una notación concreta bajo unas reglas. Me llamaba la atención y disfrutaba haciendo ejercicios. Me parecía todo muy mecánico, y eso era cómodo mentalmente. Había seguridad, si practicabas mucho tenías el éxito asegurado. Un primer aviso, de que la cosa no iba bien, fue cuando me animaron a participar en las olimpiadas de matemáticas, aquí, en la Facultad de Ciencias de Cantabria, en mis años de bachillerato. Por primera vez en mi vida no sabía qué es lo que tenía que hacer en aquellos problemas. Lejos de desanimarme, pensé que con el tiempo y durante los años de universidad descubriría lo sucedido. Durante el grado he aprendido muchas matemáticas, desde luego nada que ver que con las del instituto, mucho menos interesantes, pero en el fondo notaba otro problema oculto. Aquí, en la universidad, notaba que más que hacer ejercicios y más ejercicios lo importante era entender y estudiar la teoría. Sin lugar a dudas fue un gran salto, pero no toda la verdad. Sin embargo, es en este trabajo donde me he dado cuenta que lo importante en matemáticas es sobretodo pensar, ser creativo y ser honesto con uno mismo, como cuando dices, ahora sí esta bien esta demostración. Puedo decir con seguridad que en este trabajo he dedicado tiempo a concentrarme y pensar durante varias horas seguidas. Estoy seguro que este entrenamiento, que ha resultado ser la realización de este TFG, me ayudará en el futuro. Lo que más valoro de este trabajo, por encima de los detalles técnicos que he aprendido, es la sensación de que tras su finalización, he mejorado aunque sea un poco, mi nivel en matemáticas. Finalmente considero que he disfrutado el trabajo y he sacado cosas positivas de él.

Agradecimientos

En primer lugar quisiera agradecer a mi familia que siempre me haya animado en cada una de las cosas que he hecho, en particular estudiar Matemáticas. En segundo lugar agradecer el buen trato recibido por parte de todos lo profesores que he tenido en esta facultad. Me he sentido muy cómodo con todos ellos durante estos años. También me gustaría recordar aquí a mis profesores de etapas académicas anteriores, los profesores del Colegio Puente III de El Astillero y los profesores del IES El Astillero. Todos ellos junto con mi familia hicieron que entrara en la universidad creyendo en mi mismo y con confianza. Por último agradecer a mi directora Alicia todo lo que me ha ayudado en este trabajo y toda la paciencia que ha tenido conmigo.

Chapter 1

Introduction

The study of statistical depth functions is an active research field, both for the multidimensional case [17] and for the functional one [11]. A statistical depth function informally is a mathematical object that allows ordering data that are not simple real numbers, where there is a natural order. In addition, it allows to generalize the idea of *quantile* and *median* to dimensions higher than one. The strategy is to study complicated spaces through $\mathbb{R}$, as if this object were an abstraction of some tool to measure like the usual *tape measure* that everyone has at home. These tools allow us to obtain information about the environment and a better understanding of it. In mathematics as characteristic examples of this strategy we have the *inner product* that abstracts the notion of *angle* and the *norm* that abstracts the notion of *length*.

The band depth was introduced in [8] as a functional depth. The proves of the results that appear there are in [9]. However, its properties when applied to multivariate spaces are the ones studied there and [1] proves it suffers from degeneracy for some standard probability model in functional spaces. Thus, the initial objective of this work is to study the band depth in the multidimensional case, specifically $\mathbb{R}^2$.

We denote by $\mathcal{BP}_p$ the family of distributions on the Borel sets of $\mathbb{R}^p$. Let $(\Omega, \sigma, \mathbb{P})$ and $(\mathbb{R}^2, \mathcal{A}(\mathbb{R}^2), P)$ be two probabilistic spaces where $\mathcal{A}(\mathbb{R}^2)$ is the borel sigma-algebra for $\mathbb{R}^2$. Let us take the independent and identically distributed random vectors $X, Y : \Omega \longrightarrow \mathbb{R}^2$. The following function $BD : \mathbb{R}^2 \times \mathcal{BP}_2 \longrightarrow \mathbb{R}$ is called *bivariate band depth*:

$BD((x_1, x_2), P) = \mathbb{P}[\{\omega \in \Omega : \min(X_i(\omega), Y_i(\omega)) \leq x_i \leq \max(X_i(\omega), Y_i(\omega)),$
$i = 1, 2\}]$.

Among the properties studied in [9] it is proved the following statement.

Proposition 1.0.1 If P is an absolutely continuous distribution such that its marginals are symmetric with respect to 0 and the density function of P is positive in a neighborhood of $(0, 0) \in \mathbb{R}^2$ then $BD(\cdot, P)$ is uniquely maximized at $(0, 0)$.

However, this statement is not correct. In effect, let us take the function:

$$f_P(\mathbf{x}) = \frac{1}{10}\left(\mathbf{I}_{[-1,0]\times[0,1]}(\mathbf{x}) + \mathbf{I}_{[0,1]\times[-1,0]}(\mathbf{x})\right) + \tag{1.1}$$

$$+\frac{1}{1000}\left(\mathbf{I}_{(-1,0)^2}(\mathbf{x}) + \mathbf{I}_{(0,1)^2}(\mathbf{x})\right) + \frac{399}{1000}\left(\mathbf{I}_{[-2,-1]^2}(\mathbf{x}) + \mathbf{I}_{[1,2]^2}(\mathbf{x})\right),$$

where

$$\mathbf{I}_A(\mathbf{x}) := \begin{cases} 1 & if \quad \mathbf{x} \in A \\ 0 & if \quad \mathbf{x} \notin A \end{cases}.$$

For a visual idea of this function see Figure 1.1 where the regions of the plane where the function takes positive values are shaded and the number that appears is the value of the function in that region. It is easy to see that f_P is a density function because it is non-negative and its integral along the plane is 1. On the other hand, its marginals are symmetric because f_P is an antipodal function, $f_P(-\mathbf{x}) = f_P(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^2$, and it is also positive in a neighborhood of (0,0).

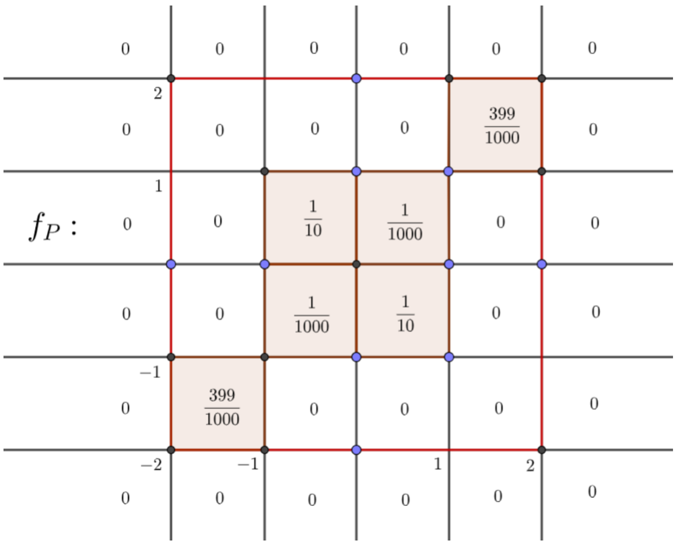


Figure 1.1: Display of the function provided in the Equation (1.1).

Finally, if we take a glimpse to the next chapter, Proposition 2.3.5, we have that:

$BD((0,0),P) = 2((\frac{1}{1000} + \frac{399}{1000})^2 + \frac{1}{10^2}) = \frac{34}{100} = .34$

$BD((1,1),P) = 2(\frac{2}{1000} + \frac{399}{1000} + \frac{2}{10})\frac{399}{1000} = \frac{239799}{500000} = .479598.$

Furthermore, [9] shows that the function $BD(\cdot)$ is not *affine invariant*, that is, it depends on affine transformations of the data. The affine invariance is a property that is requested for a multivariate depth function. What, was mentioned earlier is an objective of this work, both, the previous statement and this fact are examined in this work.

On the other hand, it is common to study together with the statistical depth functions the notions of robustness and computational complexity. See for example [13], where some well-known multivariate depth functions appear. A statistical estimator is considered *robust* if it supports *noise* well in the data. That is, for an estimate to be affected by a bad data set, the noise must necessarily be a significantly large proportion of the data set it intends to study. We study robustness with the breakdown point concept [2]. Today, where massive information and large amounts of data are everywhere, it is becoming increasingly important that computing and statistics work together. For this, the tools provided by statistics must be efficient from the point of view of computational complexity. For this reason, in this work we are concerned with the study of the efficiency of algorithms related to the world of multivariate statistical depth functions. The study of algorithms for calculating depth contours is mainly interesting, see [10], in particular the calculation of the generalized median, as the point with the greatest depth, and the calculation of the depth of a given arbitrary point [13]. The mathematical contributions of this TFG are the following:

(i). We give an alternative definition of statistical depth function inspired by the one that appears in [17]. It simply consists of adding *movement* by considering for each fixed distribution P a family of parameterized functions instead of a single *static* function. We also generalize the bivariate band depth. We refer to this generalization as the statistical depth by reflections, and prove that it satisfies the definition of statistical depth function. For this purpose, we introduce a new notion of symmetry, which we name the reflective symmetry.

(ii). We demonstrate some theoretical results on the breakdown point of the statistical depth by reflections. In particular we prove that in favorable situations it has an acceptable robustness.

(iii). We build an implementable algorithm that calculates the α-depth-contours with temporal complexity $\Theta(n^2)$ and spatial complexity $\Theta(n)$.

(iv). Finally we give as an application a new hypothesis test for the independence of two absolutely continuous random variables. However in this case we do not give the proof of the fundamental result, we leave it as a conjecture.

In this work we have made small programs in the languages MATLAB, Java and R. According to the final degree project regulations, we have included them in a separate .zip file that includes several folders. We explain how they are organized.

The folder whose name is **JavaTFG** contains the Java files that implement the algorithms developed in Chapter 4.

The folder whose name is **exampleChapter4** has the files to produce the plots of the Figure 4.2.

The folder whose name is **independenceTest** contains the files that allow to build the Tabla 5.1 and the histogram of Figure 5.1.

We warn the reader that some programs can take a long time to run, as is the case of **distribucion.m** inside **independenceTest**, which on a laptod, Honor MagicBook 15, took us more than 24 hours. The reason why this program takes so long is that its mission is to build the K_n distribution table that is discussed in Chapter 5.

Chapter 2

Statistical depth

First of all, we give the definition of statistical depth function. We give a definition similar to the definition [17]. The definition that we use in this work, although it may be similar to the one that appears in [17], differs in the following aspects. In the first place, the *affine invariance* property that appears there seems too demanding and we think that for certain potential depth functions, the modification required to satisfy it may imply a high computational cost. The explanation given for this property in [17] is that the depth of a point should not depend on the coordinate system, in particular the scale. At this point we simply do not agree with this necessity because for example classical statistical parameters such as the mean or the variance would be considered "bad" from this point of view and we do not see it that way. Instead our philosophy is as follows. A depth function intuitively is a tool for studying a multidimensional data set. Let us think about another type of tool, more traditional and from the "real" world, the *wrenches*. The standard definition of the previous reference could informally say that it is a *monkey wrench* while the definition that we give below would be a whole *set of wrenches* with their different sizes (*parameters*). For this reason, our definition depends on a family of parameters that allows "correcting" the depth function so that it is "affine invariant". The substitution of that property is Condition (i) of the Definition **2.1.1**.

On the other hand, our definition differs quite a bit in how the idea of symmetry of a distribution is used. This affects the way we consider the *maximality at center* property that appears in [17]. We think that talking about symmetry of a distribution is ambiguous and is not even rigorous to be able to formalize and prove results. Instead we prefer to define a family of distributions and that the properties that are requested are closed for that family. In particular, it is possible to define this family of distributions as the distributions that satisfy a certain property that can be interpreted as symmetry. In the classical definition [17] the maximality at center is expressed for any distribution having an unique center with respect to some notion of symmetry. If we wanted to formally demonstrate this property we would have to start by saying: Let P be a distribution with an unique center with respect to some notion of symmetry [...]. However, since this is not rigorously defined in our opinion, in order to prove the results that appear there, they end up saying that since the *$H-$ symmetry* or *halfspace symmetry* [18] is the broadest known notion of symmetry, in the sense of inclusion, it is the one they should fix. Thus, in Theorem 2.1 of [17] it is stated that the *halfspace depth function* [16] is a statistical depth function and to prove the maximality at center they start by writing: Suppose that P is $H-$symmetric about an unique point $\theta \in \mathbb{R}^p$. We do not like these types of situations and that is why the adaptation of this property is the Condition (ii) of the Definition 2.1.1. The Condition (iii) of the Definition 2.1.1 is also an adaptation to our point of view of the third property that appears in [17]. We think that the definition we give still preserves the intuition of what a statistical depth function should be. In this chapter we will prove that there is a mathematical object that satisfies this definition but not the standard one. We will call it depth by reflections.

2.1 Formal definition

Definition 2.1.1 Let $\mathcal{J}$ be a nonempty set and $\mathcal{P} \subseteq \mathcal{BP}_p$. The bounded and non-negative mapping $D(\cdot,\cdot,\cdot) : \mathbb{R}^p \times \mathcal{P} \times \mathcal{J} \longrightarrow \mathbb{R}$ is called a **$\mathcal{J}$-statistical depth function over** $\mathcal{P}$ if it satisfies the following conditions:

(i). For all nonsingular real matrix A and $b \in \mathbb{R}^p$, there exists a *scale-function* $\mathbb{A} : \mathcal{J} \longrightarrow \mathcal{J}$ such that: $D(\mathbf{x}, P, \zeta) = D\left(A\mathbf{x} + b, P_{AX+b}, \mathbb{A}(\zeta)\right)$, where X is a random vector with distribution $P \in \mathcal{P}$.

For all $P \in \mathcal{P}$, there exists $\zeta \in \mathcal{J}$ such that:
(ii). Exists and is unique: $Me(P, \zeta) := \arg\max_{\mathbf{x} \in \mathbb{R}^p} D(\mathbf{x}, P, \zeta)$.
(iii). Let $t \in [0, 1]$ and $\mu := Me(P, \zeta)$, $D(\mathbf{x}, P, \zeta) \leq D(\mu + t(\mathbf{x} - \mu), P, \zeta)$, for all $\mathbf{x} \in \mathbb{R}^p$
(iv). $D(\mathbf{x}, P, \zeta) \longrightarrow 0$ as $||\mathbf{x}||$[1] $\longrightarrow \infty$.

From the previous definition it follows that the statistical depth functions that we have defined have two fundamental attributes, $\mathcal{J}$ and $\mathcal{P}$. It is clear that we are interested in the set $\mathcal{P}$ being as large as possible in the sense of inclusion. As an observation we have that if we take a singleton set $\mathcal{J}_0 = \{\zeta_0\}$, we obtain the affine invariance [17]. The objective of this chapter is to prove that there exists, $\mathcal{J}$ and $\mathcal{P}$ such that the depth by reflections (Definition 2.3.1) satisfies the previous definition. In addition, in Theorem 2.3.1 we correct the Statement 1.0.1 we talked about in the introduction and we gave a counterexample.

[1] Let us remember that in $\mathbb{R}^p$ all the norms are equivalent from the topological point of view.

2.2 Reflective symmetry

We introduce the definition of **reflective symmetry**, Definition 2.2.1 , that we will use to define a family of distributions $\mathcal{P}$ where to apply the Definition 2.1.1. Other notions of symmetry for multivariate distributions can be seen in [15] and the references therein. Although it is not the objective of this work to study the symmetry of the distributions, we can say that our definition is inspired by the spherical symmetry that appears in [15], that is, a distribution $P \in \mathcal{BP}_p$ is spherically symmetric about a point $\mu \in \mathbb{R}^p$ if for any random random vector X in $\mathbb{R}^p$ with distribution P and for any orthogonal matrix A, $X - \mu$ and $A(X - \mu)$ are identically distributed, which is obviously a particular case of our definition.

The reason for introducing this new definition is because the existing symmetry notions either assumed too strong hypothesis, as in the case of spherical symmetry, or they did not provide us with enough information to prove our results. That is why we have decided to give a "*tailor-made*" symmetry for the purposes of this work. In our opinion, mathematics is the art of drawing inconspicuous conclusions from the weakest possible hypotheses. This has been part of the motivation to introduce the following definition and reject spherical symmetry. We also emphasize that in the results that we prove under the hypothesis of being reflective symmetric, we have never proven that they are necessary conditions for those same results. Therefore, there might be a better notion of symmetry than the one we introduce below. However, for our work and the proposed objectives mentioned in the introduction, apart from what we have already said, the symmetry that we define is useful and convenient from the notational point of view and to introduce the concepts we need.

Let $\mu \in \mathbb{R}^p$ and $\beta := \{v_1, v_2, ..., v_p\}$ be a family of p linearly independent vectors of $\mathbb{R}^p$. This family is called a **base** of $\mathbb{R}^p$. We denote by $\mathcal{B}_p$ the set of all bases on $\mathbb{R}^p$. We define a **reference** to be $\mathcal{R} := \{\mu; \beta\}$ where μ is the center of the reference. Let us consider any subfamily of $p - 1$ vectors of β . We denote it β' and we take $\{v_0\} := \beta \backslash \beta'$. Some important functions in linear algebra are **reflections**. We denote by $\sigma^{\mu}_{\beta', v_0}(\cdot)$ the reflection with respect to the *hyperplane* $H := \mu + < \beta' >$ in the direction of v_0, where $< \beta' > := \{\alpha_1 v'_1 + \alpha_2 v'_2 + ... + \alpha_{p-1} v'_{p-1} : \alpha_i \in \mathbb{R}, v'_i \in \beta'\}$. So, for any $\mathbf{x} \in \mathbb{R}^p$, $\sigma^{\mu}_{\beta', v_0}(\mathbf{x}) := 2\mathbf{x}_H - \mathbf{x}$ where $\mathbf{x}_H := H \cap L$, $L := \mathbf{x} + < v_0 >$.

Definition 2.2.1 $P \in \mathcal{BP}_p$ is **reflective symmetric** about the point $\mu \in \mathbb{R}^p$ if it exists a base of $\mathbb{R}^p$, β, such that for any random vector X in $\mathbb{R}^p$ with distribution P and any subfamily of $p - 1$ vectors of β, β', we have that X and $\sigma^{\mu}_{\beta', v_0}(X)$ are identically distributed.

If P is reflective symmetric with respect to a reference $\mathcal{R}$, we emphasize it by making use of the notation $P_{\mathcal{R}}$. In general, this reference is not unique. On the other hand, although we have defined it this way, *intrinsically*, for many of the results that we are going to give, it is more convenient to call the points by their coordinates in a fixed reference. We will say $\mathbf{x} = (x_1, x_2, ..., x_p)$ in a reference $\mathcal{R} = \{\mu; \beta\}$ if $\mathbf{x} - \mu = x_1 v_1 + x_2 v_2 + ... + x_p v_p$. Furthermore, since reflections are affine transformations, it is often convenient to think about their matricial interpretation. An idea of this is given in the following example.

Example. The uniform distribution $\mathcal{U}$ in the square $[-a, a]^2$ with $a > 0$ is reflective summetric. We take the *canonical reference*, $\mathcal{R}_c := \{\mathbf{0}; \beta_c\}$ where $\mathbf{0} := (0, 0)$ and $\beta_c := \{e_1, e_2\}$ with $e_1 = (1, 0)$ and $e_2 = (0, 1)$. So, if $\mathbf{X} = (X, Y)$ in $\mathcal{R}_c$ we have that $\sigma^{\mu}_{e_1,e_2}(X, Y) = (X, -Y)$ and $\sigma^{\mu}_{e_2,e_1}(X, Y) = (-X, Y)$ and by definicion of the uniform distribution $\mathcal{U}$ it is clear that $X, \sigma^{\mu}_{e_1,e_2}(X), \sigma^{\mu}_{e_2,e_1}(X)$ are identically distributed.

In what follows we work on $\mathbb{R}^2$ unless otherwise stated because it is easier to do the proofs and also the notation is simpler. Conceptually it is equivalent to what we would obtain in $\mathbb{R}^p$. Moreover, we are going to concentrate on absolutely continuous distributions. On the other hand, as it is usual in the literature, we characterize absolutely continuous distributions by their density function, because it is more useful to make calculations and check that a bivariate distribution is reflective symmetric since a function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is a more manageable mathematical object than its corresponding distribution.

Proposition 2.2.1 *Let $P \in \mathcal{BP}_2$ be an absolutely continuous distribution with density function f. Then, $P = P_{\mathcal{R}}$ for a reference $\mathcal{R} = \{\mu; v_1, v_2\}$ if and only if $f \circ \sigma^{\mu}_{v_2,v_1} = f \circ \sigma^{\mu}_{v_1,v_2} = f$.*

Proof. Let $P = P_{\mathcal{R}}$, $\mathbf{x} \in \mathbb{R}^2$ and X a random vector with distribution P. In the reference $\mathcal{R}$ we have that $\mathbf{x} = (x_1, x_2)$ and $X = (X_1, X_2)$ where x_1, x_2, X_1, X_2 are their coordinates respectively. Let us start with the implication in which P is reflective symmetric with respect to $\mathcal{R}$. So, if we take the reflection $\sigma^{\mu}_{v_2,v_1}(\cdot)$ we have that X and $\sigma^{\mu}_{v_2,v_1}(X)$ are identically distributed, i.e, (X_1, X_2) and $(-X_1, X_2)$ are identically distributed because $\sigma^{\mu}_{v_2,v_1}(\cdot) = \left(\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ in the reference $\mathcal{R} = \{\mu; v_1, v_2\}$. Consequently, $P[(-\infty, x_1] \times (-\infty, x_2]] = P[[-x_1, \infty) \times (-\infty, x_2]]$. Thus,

$$\int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(t_1, t_2) dt_1 dt_2 = \int_{-\infty}^{x_2} \int_{-x_1}^{\infty} f(t_1, t_2) dt_1 dt_2.$$

Deriving with respect to x_2 we obtain,

$$\frac{\partial}{\partial x_2}\left(\int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(t_1, t_2) dt_1 dt_2\right) = \frac{\partial}{\partial x_2}\left(\int_{-\infty}^{x_2} \int_{-x_1}^{\infty} f(t_1, t_2) dt_1 dt_2\right),$$

then $\int_{-\infty}^{x_1} f(t_1, t_2) dt_1 = \int_{-x_1}^{\infty} f(t_1, t_2) dt_1 = \int_{-\infty}^{\infty} f(t_1, t_2) dt_1 - \int_{-\infty}^{-x_1} f(t_1, t_2) dt_1$.

We note that $f_{X_2}(x_2) = \int_{-\infty}^{\infty} f(t_1,t_2)dt_1$. In addition, if we derivate with respect to x_1, $\frac{\partial}{\partial x_1}(\int_{-\infty}^{x_1} f(t_1,t_2)dt_1) = \frac{\partial}{\partial x_1}(f_{X_2}(x_2)) - \frac{\partial}{\partial x_1}(\int_{-\infty}^{-x_1} f(t_1,t_2)dt_1)$ and then, $f(x_1,x_2) = 0 - (-f(-x_1,x_2))$ so $f(x_1,x_2) = f(-x_1,x_2)$. So, $f(\mathbf{x}) = f(\sigma^{\mu}_{v_2,v_1}(\mathbf{x}))$. Taking the other reflection $\sigma^{\mu}_{v_1,v_2}(\cdot)$ leads to $f(\mathbf{x}) = f(\sigma^{\mu}_{v_1,v_2}(\mathbf{x}))$.

For the other implication let $\Gamma \in \mathcal{A}(\mathbb{R}^2)$. We prove that $\mathbb{P}[\sigma^{\mu}_{v_2,v_1}(X) \in \Gamma] = \mathbb{P}[X \in \Gamma]$. We know that $\sigma^{\mu}_{v_2,v_1}(\cdot)$ is a bijective function and the following equality $\sigma^{\mu}_{v_2,v_1} \circ \sigma^{\mu}_{v_2,v_1} = Id$ because is a reflection, in consecuense $\mathbb{P}[\sigma^{\mu}_{v_2,v_1}(X) \in \Gamma] = \mathbb{P}[X \in \sigma^{\mu}_{v_2,v_1}(\Gamma)]$.

On the other hand, $\mathbb{P}[X \in \sigma^{\mu}_{v_2,v_1}(\Gamma)] = \int_{\sigma^{\mu}_{v_2,v_1}(\Gamma)} f$. By the *change of variables theorem* we have that $\int_{\sigma^{\mu}_{v_2,v_1}(\Gamma)} f = \int_{\Gamma} f(\sigma^{\mu}_{v_2,v_1}(\mathbf{x})) \cdot |\det Jac(\sigma^{\mu}_{v_2,v_1}(\mathbf{x}))| d\mathbf{x}$. In addition, $\sigma^{\mu}_{v_2,v_1}(\cdot)$ is an affine transformation, so it is of the form $A\mathbf{x} + b$ where $A^2 = Id$ (because the implicit linear application is also a reflection). Then, $Jac(\sigma^{\mu}_{v_2,v_1}(\mathbf{x})) = A$ and $det(A) = \pm 1$, so, $|\det Jac(\sigma^{\mu}_{v_2,v_1}(\mathbf{x}))| = 1$ and we have that $\int_{\Gamma} f(\sigma^{\mu}_{v_2,v_1}(\mathbf{x})) \cdot |\det Jac(\sigma^{\mu}_{v_2,v_1}(\mathbf{x}))| d\mathbf{x} = \int_{\Gamma} f(\sigma^{\mu}_{v_2,v_1}(\mathbf{x})) \cdot 1 d\mathbf{x} = \int_{\Gamma} f(\mathbf{x}) d\mathbf{x}$. In conclusion, $\mathbb{P}[\sigma^{\mu}_{v_2,v_1}(X) \in \Gamma] = \mathbb{P}[X \in \Gamma]$. The other equality of probabilities, for the case $\sigma^{\mu}_{v_1,v_2}$, is equivalent. □

Let an absolutely continuous distribution $P \in \mathcal{BP}_2$ with density function f such that the marginals P_1, P_2 are independent. Then, $f(x,y) = f_1(x) f_1(y)$ here f_1, f_2 are the density function of P_1, P_2 respectively. If f_1, f_2 are even functions, i.e; $f_1(x) = f_1(-x)$, $f_2(y) = f_2(-y)$, we have that: $f(-x,y) = f_1(-x) f_2(y) = f_1(x) f_2(y) = f(x,y)$ and $f(x,-y) = f_1(x) f_1(-y) = f_1(x) f_2(y) = f(x,y)$. Then, by the Proposition 2.2.1 P is reflective symmetric. In conclusion, it is easy to build symmetric reflective distributions from symmetric real distributions.

2.3 Statistical depth by reflections

We introduce an generalization of the definition of the bivariate band depth [8] following the style and the geometric intuition of the simplicial depth [6]. The idea is to give a definition without fixing some coordinates, like those of the canonical reference, which is how it is done in [8]. We remind the reader that part of our motivation is to solve the problems that the bivariate band depth has relative to affine invariance and maximality at center (originally thinking of the standard definition) since the hypotheses that appear in [8] to prove this property are not sufficient as we show in the introduction. Let us first introduce some notation.

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ and a base $\beta = \{v_1, v_2\}$ of $\mathbb{R}^2$. We denote the **parallelogram** determined by the points $\mathbf{x}, \mathbf{y}$ in the directions v_1, v_2 by:

$$L_{v_1,v_2}(\mathbf{x},\mathbf{y}) := \begin{cases} (1-\alpha)\mathbf{x} + \alpha\mathbf{y} : \alpha \in [0,1], & if \quad \mathbf{x} - \mathbf{y} \in < v_1 > \cup < v_2 > \\ \\ Q + \alpha(\mathbf{x} - Q) + \gamma(\mathbf{y} - Q) : \alpha, \gamma \in [0,1], & otherwise, \end{cases} \tag{2.1}$$

where $Q := r \cap s$ with $r := \mathbf{x} + < v_1 >$ and $s := \mathbf{y} + < v_2 >$.

Due to the following trivial proposition we can use the notation $L_\beta(\mathbf{x}, \mathbf{y})$ to refer to the set (parallelogram) in **(2.1)** .

Proposition 2.3.1 *The parallelogram does not depend on the order of the elements of the base* β*, i.e;* $L_{v_1,v_2}(\mathbf{x},\mathbf{y}) = L_{v_2,v_1}(\mathbf{x},\mathbf{y}) = L_\beta(\mathbf{x},\mathbf{y})$.

Proof. Let $Q = r \cap s$, where $r := \mathbf{x} + < v_1 >$ and $s := \mathbf{y} + < v_2 >$, and $Q' = r' \cap s'$ where $r' := \mathbf{x} + < v_2 >$ and $s' := \mathbf{y} + < v_1 >$. Firstly, we have that $Q = \mathbf{x} + t_1 v_1 = \mathbf{y} + t_2 v_2$ and $Q' = \mathbf{x} + t'_2 v_2 = \mathbf{y} + t'_1 v_1$. Then, $\mathbf{x} - \mathbf{y} = t_2 v_2 - t_1 v_1$ and also $\mathbf{x} - \mathbf{y} = t'_1 v_1 - t'_2 v_2$ so, $t_2 v_2 - t_1 v_1 = t'_1 v_1 - t'_2 v_2$, i.e, $(t_2 + t'_2) v_2 + (-t_1 - t'_1) v_1 = 0$. Then $t'_1 = -t_1$ and $t'_2 = -t_2$ because v_1, v_2 are linearly independent. Thus, as $\mathbf{x} + t_1 v_1 = \mathbf{y} + t_2 v_2$, adding the term $\alpha t_1 v_1 - \gamma t_2 v_2$ to both sides of the equality we obtain $\mathbf{x} + t_1 v_1 - \alpha t_1 v_1 - \gamma t_2 v_2 = \mathbf{y} + t_2 v_2 - \alpha t_1 v_1 - \gamma t_2 v_2$, which is equivalent to $\mathbf{x} + t_1 v_1 + \alpha(-t_1 v_1) + \gamma(-t_2 v_2) = \mathbf{y} - t_1 v_1 + t_1 v_1 - \alpha t_1 v_1 + t_2 v_2 - \gamma t_2 v_2$, because in the right side of the equality we add $-t_1 v_1 + t_1 v_1 = 0$. Rewriting the previous equality we have that $\mathbf{x} + t_1 v_1 + \alpha(-t_1 v_1) + \gamma(-t_2 v_2) = \mathbf{y} - t_1 v_1 + (1-\alpha) t_1 v_1 + (1-\gamma) t_2 v_2$ which is identical to the expression $Q + \alpha(\mathbf{x} - Q) + \gamma(\mathbf{y} - Q) = Q' + (1-\alpha)(\mathbf{y} - Q') + (1-\gamma)(\mathbf{x} - Q')$. ☐

Definition 2.3.1 Let $(\Omega, \delta, \mathbb{P})$ be a probabilistic space, $P \in \mathcal{BP}_2$, $\beta = \{v_1, v_2\}$ a base of $\mathbb{R}^2$ and $X, Y : \Omega \longrightarrow \mathbb{R}^2$ independent and identically distributed random vectors with distribution P. **The depth by reflections** of a point $\mathbf{x} \in \mathbb{R}^2$ with respect to P and β is:

$$RD(\mathbf{x}, P, \beta) := \mathbb{P}[\mathbf{x} \in L_\beta(X, Y)]. \tag{2.2}$$

We have enough terminology to prove Condition (i) of Definition **2.1.1** for the depth by reflections.

Proposition 2.3.2 *Let* $P \in \mathcal{BP}_2$ *and a base* $\beta = \{v_1, v_2\}$*. For all nonsingular real matrix* A *and* $b \in \mathbb{R}^p$*, if we take the scale-function* $\mathbb{A}(\beta) := \{Av_1, Av_2\}$*, then* $RD(\mathbf{x}, P, \beta) = RD\left(A\mathbf{x} + b, P_{AX+b}, \mathbb{A}(\beta)\right)$*, where* X *is a random vector with distribution* P*.*

Proof. Let $\mathbf{x} \in \mathbb{R}^2$, we have to prove the following equality, $\mathbb{P}[\mathbf{x} \in L_\beta(X,Y)] = \mathbb{P}[A\mathbf{x} + b \in L_{\mathbb{A}(\beta)}(AX + b, AY + b)]$. Firstly, $A\mathbf{x} + b$ is a bijective function so, $\mathbb{P}[\mathbf{x} \in L_\beta(X,Y)] = \mathbb{P}[A\mathbf{x} + b \in AL_\beta(X,Y) + b]$. Thus, it is enough to prove that $AL_\beta(X,Y) + b = L_{\mathbb{A}(\beta)}(AX + b, AY + b)$. Let us fix a $\omega \in \Omega$, If $X(\omega) - Y(\omega) \in\ < v_1 > \cup < v_2 >$, we can assume without lost of generality $X(\omega) - Y(\omega) = \lambda v_1$, so, we have that $(AX(\omega) + b) - (AY(\omega) + b) = AX(\omega) - AY(\omega) = A(X(\omega) - Y(\omega)) = A\lambda v_1 = \lambda A v_1$ so, $(AX(\omega) + b) - (AY(\omega) + b) \in\ < Av_1 > \cup < Av_2 >$. The inverse is also true.

In the other case (second part of (2.1)) , let $Q(\omega) = r(\omega) \cap s(\omega)$ where $r(\omega) := X(\omega) + < v_1 >$, $s(\omega) := Y(\omega) + < v_2 >$ and $\hat{Q}(\omega) = \hat{r}(\omega) \cap \hat{s}(\omega)$ where $\hat{r}(\omega) = (AX(\omega) + b) + < Av_1 >$ and $\hat{s}(\omega) = (AY(\omega) + b) + < Av_2 >$. Let us see that $\hat{Q}(\omega) = AQ(\omega) + b$.

In effect, $Q(\omega) = X(\omega) + t_1 v_1 = Y(\omega) + t_2 v_2$, then $A(X(\omega) + t_1 v_1) = A(Y(\omega) + t_2 v_2)$ and by linearity $AX(\omega) + At_1 v_1 = AY(\omega) + At_2 v_2$, so, $AX(\omega) + t_1 Av_1 = AY(\omega) + t_2 Av_2$, in conclusion, $(AX(\omega) + b) + t_1 Av_1 = (AY(\omega) + b) + t_2 Av_2 = \hat{Q}(\omega)$. In addition, as A is a nonsingular matrix then $Q(\omega)$ is the only point that satisfies this condition.

Finally, $\hat{Q}(\omega) + \alpha(AX(\omega) + b - \hat{Q}(\omega)) + \gamma(AY(\omega) + b - \hat{Q}(\omega))$ is equal to

$$AQ(\omega) + b + \alpha(AX(\omega) + b - (AQ(\omega) + b)) + \gamma(AY(\omega) + b - (AQ(\omega) + b))$$

and operating and by the linearity of A is identical to

$$A(Q(\omega) + \alpha(X(\omega) - Q(\omega)) + \gamma(Y(\omega) - Q(\omega))) + b.$$

Thus, $L_{\mathbb{A}(\beta)}(AX(\omega) + b, AY(\omega) + b) = AL_{\beta}(X(\omega), Y(\omega)) + b.$ □

We progressively build a family $\mathcal{P}$ where to apply Definition 2.1.1 . We denote by $\sigma(\mathcal{P})$ the set of distributions $P \in \mathcal{BP}_2$ such that there exists a reference $\mathcal{R}$ with $P = P_{\mathcal{R}}$. We prove in the following Proposition 2.3.3 that this family of distributions is closed under affine transformations. And therefore the Condition (i) of Definition 2.1.1 over these distributions makes sense.

Proposition 2.3.3 *Let $X : \Omega \longrightarrow \mathbb{R}^2$ a random vector, A a nonsingular real matrix and $b \in \mathbb{R}^2$. Let P be the distribution of X and P_{AX+b} the distribution of $AX + b$. If $P = P_{\mathcal{R}} \in \sigma(\mathcal{P})$ then $P_{AX+b} \in \sigma(\mathcal{P})$.*

Proof. Let $\mathcal{R} := \{\mu; v_1, v_2\}$. We have to prove that there exists a reference $\mathcal{R}' := \{\gamma; u_1, u_2\}$ such that $AX + b$, $\sigma^{\gamma}_{u_1,u_2}(AX + b)$, $\sigma^{\gamma}_{u_2,u_1}(AX + b)$ are identically distributed. We prove that $AX + b$ and $\sigma^{\gamma}_{u_1,u_2}(AX + b)$ are identically distributed, the other one proceeds equivalently. Firstly, we have that X and $\sigma^{\mu}_{v_1,v_2}(X)$ are identically distributed.

Thus, for any $\Gamma \in \mathcal{A}(\mathbb{R})^2$, we have $\mathbb{P}[AX + b \in \Gamma] = \mathbb{P}[X \in A^{-1}\Gamma - A^{-1}b] = \mathbb{P}[\sigma^{\mu}_{v_1,v_2}(X) \in A^{-1}\Gamma - A^{-1}b] = \mathbb{P}[A\sigma^{\mu}_{v_1,v_2}(X) + b \in \Gamma]$, that is, $AX + b$ and $A\sigma^{\mu}_{v_1,v_2}(X) + b$ are identically distributed. We prove the equality $A\sigma^{\mu}_{v_1,v_2}(X) + b = \sigma^{A\mu+b}_{Av_1,Av_2}(AX + b)$, i.e, $\gamma = A\mu + b$, $u_1 = Av_1$, $u_2 = Av_2$. Let $\mathbf{x} \in \mathbb{R}^2$. We know that $\sigma^{\mu}_{v_1,v_2}(\cdot)$ is a reflection, so, $\sigma^{\mu}_{v_1,v_2}(\mathbf{x}) - \mathbf{x} = \lambda v_2$

and $M - \mu = \alpha v_1$ for some $\lambda, \alpha \in \mathbb{R}$ and where $M := (\sigma^{\mu}_{v_1,v_2}(\mathbf{x}) + \mathbf{x})/2$. The following properties prove the result.

(i). $(A\sigma^{\mu}_{v_1,v_2}(\mathbf{x}) + b) - (A\mathbf{x} + b) = A\sigma^{\mu}_{v_1,v_2}(\mathbf{x}) - A\mathbf{x} = A(\sigma^{\mu}_{v_1,v_2}(\mathbf{x}) - \mathbf{x}) = A\lambda v_2 = \lambda A v_2 \in \ < Av_2 >$. So, $A\sigma^{\mu}_{v_1,v_2}(\mathbf{x}) + b \in r := (A\mathbf{x} + b)+ < Av_2 >$.

(ii). $\hat{M} := \frac{(A\sigma^{\mu}_{v_1,v_2}(\mathbf{x})+b)+(A\mathbf{x}+b)}{2} = \frac{A\sigma^{\mu}_{v_1,v_2}(\mathbf{x})+A\mathbf{x}}{2} + b = \frac{A(\sigma^{\mu}_{v_1,v_2}(\mathbf{x})+\mathbf{x})}{2} + b = A\left(\frac{\sigma^{\mu}_{v_1,v_2}(\mathbf{x})+\mathbf{x}}{2}\right) + b = AM + b$.

(iii). $\hat{M} - (A\mu + b) = (AM + b) - (A\mu + b) = A(M - \mu) = A(\alpha v_1) = \alpha A v_1$, so, $\hat{M} = (A\mu + b) + \alpha A v_1 \in (A\mu + b)+ \ < Av_1 >$. Thus, we have that $A\sigma^{\mu}_{v_1,v_2}(\mathbf{x}) + b \in s := (A\mu + b)+ < Av_1 >$.

Consequently, $\hat{M} = r \cap s$, and by definition of $\sigma^{A\mu+b}_{Av_1,Av_2}(\cdot)$ and $\hat{M}$, we conclude that $A\sigma^{\mu}_{v_1,v_2}(\mathbf{x}) + b = \sigma^{A\mu+b}_{Av_1,Av_2}(A\mathbf{x} + b)$. □

The following Collorary 2.3.1, is interesting because it justifies that in many of the results where we have the properties of the distributions $\sigma(\mathcal{P})$ it will suffice to concentrate on the canonical reference $\mathcal{R}_c$.

Corollary 2.3.1 *Let $P_{\mathcal{R}} \in \sigma(\mathcal{P})$ with $\mathcal{R} = \{\mu; \beta\}$ and $\beta = \{v_1, v_2\}$. By taking a matrix A such that $Av_1 = e_1$, $Av_2 = e_2$, which is the matrix whose columns are the coordinates of e_1 and e_2 in the base β respectively, and a $b := -A\mu$, we have that $RD(\mathbf{x}, P_{\mathcal{R}}, \beta) = RD(\mathbf{y}, P_{\mathcal{R}_c}, \beta_c)$, where $\mathbf{y} = A\mathbf{x} + b$.*

Proof. By the proof of Proposition 2.3.3 , we have that $P_{AX+b} = P_{\mathcal{R}_c}$. Then, the result follows from Proposition 2.3.2 . □

We have introduced the notion of statistical depth by reflections, Definition 2.3.1 from an intrinsic point of view, not mentioning coordinates. This point of view is more elegant and intuitive but not very useful for proof that correspond to the properties of Definition 2.1.1 that do not have to do with the Condition (i) of that definition. For this reason we introduce Proposicion 2.3.4 .

Proposition 2.3.4 *Let $P \in \mathcal{BP}_2$, and $\mathcal{R} := \{\mu; \beta\}$ a reference, with $\beta = \{v_1, v_2\}$. Then $RD(\mathbf{x}, P, \beta) = \mathbb{P}[\min(X_i, Y_i) \le x_i \le \max(X_i, Y_i), i = 1, 2\}]$, with* $\mathbf{x} = (x_1, x_2), X = (X_1, X_2), Y = (Y_1, Y_2)$ in the reference $\mathcal{R}$.

Proof. Let us fix $\omega \in \Omega$. If $X(\omega) - Y(\omega) \in < v_1 > \cup < v_2 >$ we have that $L_\beta\big(X(\omega), Y(\omega)\big) = (1-\alpha)X(\omega) + \alpha Y(\omega)$ for $\alpha \in [0,1]$, so, $L_\beta\big(X(\omega), Y(\omega)\big) = \big((1-\alpha)X_1(\omega) + \alpha Y_1(\omega), (1-\alpha)X_2(\omega) + \alpha Y_2(\omega)\big)$. As, $X_1(\omega), X_2(\omega), Y_1(\omega), Y_2(\omega)$ are real numbers and $\alpha \in [0,1]$ then,

$$L_\beta\big(X(\omega), Y(\omega)\big) = \{(x_1, x_2) : x_1 \in [\min(X_1(\omega), Y_1(\omega)), \max(X_1(\omega), Y_1(\omega))], \\ x_2 \in [\min(X_2(\omega), Y_2(\omega)), \max(X_2(\omega), Y_2(\omega))]\}.$$

On the other hand, $r(\omega) := X(\omega) + < v_1 >= \{(x_1, x_2) \in \mathbb{R}^2 : x_2 = X_2(\omega)\}$ and $s(\omega) := Y(\omega) + < v_2 >= \{(x_1, x_2) \in \mathbb{R}^2 : x_1 = Y_1(\omega)\}$. Then, $Q(\omega) = r(\omega) \cap s(\omega) = (Y_1(\omega), X_2(\omega))$. We compute: $X(\omega) - Q(\omega) = (X_1(\omega), X_2(\omega)) - (Y_1(\omega), X_2(\omega)) = (X_1(\omega) - Y_1(\omega), 0)$ and $Y(\omega) - Q(\omega) = (Y_1(\omega), Y_2(\omega)) - (Y_1(\omega), X_2(\omega)) = (0, Y_2(\omega) - X_2(\omega))$. Let $\alpha, \gamma \in [0,1]$ and for simplicity, let us denote $L_\beta := L_\beta\big(X(\omega), Y(\omega)\big)$, we have that,

$$\begin{aligned} L_\beta &= \big(Y_1(\omega), X_2(\omega)\big) + \alpha\big(X_1(\omega) - Y_1(\omega), 0\big) + \gamma\big(0, Y_2(\omega) - X_2(\omega)\big) \\ &= \big(\alpha X_1(\omega) + (1-\alpha)Y_1(\omega), \gamma Y_2(\omega) + (1-\gamma)X_2(\omega)\big) \\ &= \{(x_1, x_2) : x_1 \in [\min(X_1(\omega), Y_1(\omega)), \max(X_1(\omega), Y_1(\omega))], \\ &\qquad x_2 \in [\min(X_2(\omega), Y_2(\omega)), \max(X_2(\omega), Y_2(\omega))]\}. \end{aligned}$$

□

It is easy to see that if in previous Proposition 2.3.4 we take $\mathcal{R} = \mathcal{R}_c$ then we obtain the bivariate band depth [8]. We still do not have an easy way to calculate depth by reflections. The following Proposition 2.3.5 solves that problem for us for what we need in this work. It is a polynomial function evaluated on certain probabilities. Especially it will be useful for the main mathematical results of this work. It will also be important in Chapter **4** when we deal with computational complexity.

Proposition 2.3.5 *Let $P \in \mathcal{BP}_2$ be an absolutely continuous distribution. We have that*:

$$RD(\mathbf{x}, P, \beta) = 2P[(-\infty, x_1] \times (-\infty, x_2]]P[[x_1, \infty) \times [x_2, \infty)]$$
$$+2P[(-\infty, x_1] \times [x_2, \infty)]P[[x_1, \infty) \times (-\infty, x_2]]$$

where $\mathbf{x} = (x_1, x_2)$ in a reference $\mathcal{R}$ with center μ and base β.

Proof. Let $\omega_0 \in \Omega$ and $X = (X_1, X_2), Y = (Y_1, Y_2)$ in the reference $\mathcal{R}$. We have that $X_1(\omega_0) \leq Y_1(\omega_0)$ or $X_1(\omega_0) \geq Y_1(\omega_0)$, and $X_2(\omega_0) \leq Y_2(\omega_0)$ or $X_2(\omega_0) \geq Y_2(\omega_0)$. Then, $\Omega = A_1 \cup A_2 \cup A_3 \cup A_4$ where $A_1 = \{\omega \in \Omega : X_1(\omega) \leq Y_1(\omega), X_2(\omega) \leq Y_2(\omega)\}$, $A_2 = \{\omega \in \Omega : X_1(\omega) \leq Y_1(\omega), X_2(\omega) \geq Y_2(\omega)\}$, $A_3 = \{\omega \in \Omega : X_1(\omega) \geq Y_1(\omega), X_2(\omega) \leq Y_2(\omega)\}$, $A_4 = \{\omega \in \Omega : X_1(\omega) \geq Y_1(\omega), X_2(\omega) \geq Y_2(\omega)\}$.

We define $B := \{\omega \in \Omega : \min(X_1(\omega), Y_1(\omega)) \leq x_1, \max(X_1(\omega), Y_1(\omega)) \geq x_1, \min(X_2(\omega), Y_2(\omega)) \leq x_2, \max(X_2(\omega), Y_2(\omega)) \geq x_2\}$.

$B = B \cap \Omega = B \cap (A_1 \cup A_2 \cup A_3 \cup A_4) = (B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3) \cup (B \cap A_4)$. On the other hand, we have that : (remember that P is absolutely continuous) $(B \cap A_i) \cap (B \cap A_j) = B \cap A_i \cap A_j \subseteq A_i \cap A_j$ where $\mathbb{P}[A_i \cap A_j] = 0$ with $i, j \in \{1, 2, 3, 4\}$. Then, by the Proposition 2.3.4 :

$RD(\mathbf{x}, P, \beta) = \mathbb{P}[B] = \mathbb{P}[B \cap A_1] + \mathbb{P}[B \cap A_2] + \mathbb{P}[B \cap A_3] + \mathbb{P}[B \cap A_4]$ and $\mathbb{P}[B \cap A_1] = \mathbb{P}[\{\omega \in \Omega : X_1(\omega) \leq x_1, X_2(\omega) \leq x_2, Y_1(\omega) \geq x_1, X_1(\omega) \geq x_2\}] = \mathbb{P}[\{\omega \in \Omega : X(\omega) \in (-\infty, x_1] \times (-\infty, x_2], Y(\omega) \in [x_1, \infty) \times [x_2, \infty)\}] = P[(-\infty, x_1] \times (-\infty, x_2]]P[[x_1, \infty) \times [x_2, \infty)]$ because X, Y are independent and identically distributed. Finally, reasoning with the rest of the probabilities $\mathbb{P}[B \cap A_i]$ in the same way and adding them up, the result is obtained. □

We finally build the family of distributions where we will apply Definition 2.1.1. We denote by $\sigma^*(\mathcal{P})$ the set of distributions $P = P_{\mathcal{R}} \in \sigma(\mathcal{P})$ that are absolutely continuous and whose density functions are locally positive at the center of the reference $\mathcal{R} = \{\mu; \beta\}$, i.e; μ.

We introduce some more notation and also a bit of geometric intuition to make it easier to understand Theorem 2.3.1. Let $(x, y) \in \mathbb{R}^2 - \{(0,0)\}$ with coordinates in the canonical reference $\mathcal{R}_c$. We define the partition of the plane given by the sets: $S_1 = [0, x] \times [0, y], S_2 = [-x, 0] \times [0, y], S_3 = [0, x] \times [-y, 0], S_4 = [-x, 0] \times [-y, 0], S_5 = [x, \infty) \times [y, \infty), S_6 = [x, \infty) \times (-\infty, -y]$, $S_7 = (-\infty, -x] \times (-\infty, -y]$, $S_8 = (-\infty, -x] \times [y, \infty)$, $S_9 = [x, \infty) \times [0, y]$, $S_{10} = [x, \infty) \times [-y, 0]$, $S_{11} = (-\infty, -x] \times [0, y]$, $S_{12} = (-\infty, -x] \times [-y, 0]$, $S_{13} = [0, x] \times [y, \infty)$, $S_{14} = [-x, 0] \times [y, \infty)$, $S_{15} = [0, x] \times (-\infty, -y]$, $S_{16} = [-x, 0] \times (-\infty, -y]$. If we take a distribution $P_{\mathcal{R}_c} \in \sigma^*(\mathcal{P})$ by reflective symmetry, Definition 2.2.1 , it is clear that: $P_{\mathcal{R}_c}[S_i] = z_1, i = 1,2,3,4$, $P_{\mathcal{R}_c}[S_i] = z_2, i = 5,6,7,8$, $P_{\mathcal{R}_c}[S_i] = z_3, i = 9,10,11,12$ and $P_{\mathcal{R}_c}[S_i] = z_4, i = 13,14,15,16$, for some $z_1, z_2, z_3, z_4 \in [0,1]$ such that $z_1 + z_2 + z_3 + z_4 = 1/4$. It is wise to have an image in mind for this situation. Figure 2.1 gives an idea of it. We can see that this is a good trick to calculate the statistical depth by reflections, Definition 2.3.1, for absolutely continuous distributions (see Proposition 2.3.5).

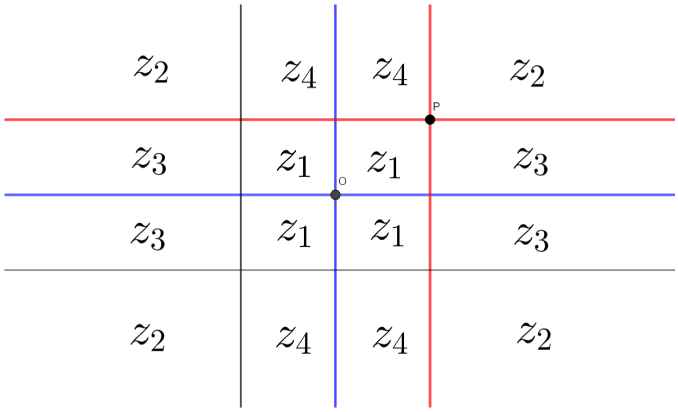


Figure 2.1: Probabilities $z_i's$ with reference $\mathcal{R}_c$. Blue lines to calculate $RD((0,0), P_{\mathcal{R}_c}, \beta_c)$ and red lines to calculate $RD((x,y), P_{\mathcal{R}_c}, \beta_c)$.

The following Theorem 2.3.1 proves Condition (ii) of Definition 2.1.1. On the other hand, this result is that it solves the statement we talked about in the introduction, Proposition 1.0.1. Indeed, it suffices to consider $\beta = \beta_c$, the canonical base of $\mathbb{R}^2$, and observe that $BD(\mathbf{x}, P) := RD(\mathbf{x}, P, \beta_c)$.

Theorem 2.3.1 *If $P = P_{\mathcal{R}} \in \sigma^*(\mathcal{P})$, with $\mathcal{R} = \{\mu; \beta\}$, then the function $Me(P, \beta) := \arg\max_{\mathbf{x}\in\mathbb{R}^2} RD(\mathbf{x}, P, \beta)$ is well defined, i.e; it exists and is unique. In addition, we have that $Me(P, \beta) = \mu$.*

Proof. Firstly, by Corollary 2.3.1 it is enough to prove the result for the case $\mathcal{R} = \mathcal{R}_c$. In the conditions of this corollary, we proof that $Me(P, \beta) := \mu$, the center of the reference $\mathcal{R}$, because we will see that $Me(P_{\mathcal{R}_c}, \beta_c) = \mathbf{0}$ and $\mu = A^{-1}(\mathbf{0}) - A^{-1}b$. Given $P_{\mathcal{R}_c} \in \sigma^*(P)$, let us see that $Me(P_{\mathcal{R}_c}, \beta_c) = (0, 0)$ or see that the point $(0, 0)$ is a global maximum. Let $(x, y) \in \mathbb{R}^2 - \{(0, 0)\}$. By symmetry we can take $x, y > 0$. We consider the sets S_i and the probabilities $z_i's$ of Figure 2.1.

By Proposition 2.3.5 we compute: $RD((0, 0), P_{\mathcal{R}_c}, \beta_c) = 4(z_1 + z_2 + z_3 + z_4)^2$ and $RD((x, y), P_{\mathcal{R}_c}, \beta_c) = 2z_2(4z_1 + 2z_3 + 2z_4 + z_2) + 2(2z_3 + z_2)(2z_4 + z_2)$. So,
$RD((0, 0), P_{\mathcal{R}_c}, \beta_c) - RD((x, y), P_{\mathcal{R}_c}, \beta_c) = 4(z_1^2 + z_3^2 + z_4^2 + 2z_1z_3 + 2z_1z_4) \geq 0$. Thus, $(0, 0)$ is a maximum.

Let us see that the function $RD(\cdot, P_{\mathcal{R}_c}, \beta_c)$ is uniquely maximized at $(0, 0)$. Let us suppouse for a contradiction that exists a point $(x_0, y_0) \neq (0, 0)$ such that $RD((x_0, y_0), P_{\mathcal{R}_c}, \beta_c) = RD((0, 0), P_{\mathcal{R}_c}, \beta_c)$. Thus, $z_1 = 0$, i.e, $P[S_1] = P[S_2] = P[S_3] = P[S_4] = 0$ and then:

$$\int_{-y_0}^{y_0} \int_{-x_0}^{x_0} f_{P_{\mathcal{R}_c}}(x, y) dx dy = \sum_{i=1}^{4} P[S_i] = 0. \tag{2.3}$$

We have that $f_{P_{\mathcal{R}_c}}$ is a density function of $P_{\mathcal{R}_c}$ in $(\mathbb{R}^2, \mathcal{A}(\mathbb{R}^2), P_{\mathcal{R}_c})$, then, $f_{P_{\mathcal{R}_c}} \geq 0$ and $f_{P_{\mathcal{R}_c}}$ is Borel measurable and so on, $f_{P_{\mathcal{R}_c}}$ is Lebesgue measurable. In addition , $[-x_0, x_0] \times [-y_0, y_0] \in \beta(\mathbb{R}^2)$, and by (2.3) we have that $f = 0$ in almost every point of $[-x_0, x_0] \times [-y_0, y_0]$ except in a set E of measure zero. However, as $P \in \sigma^*(\mathcal{P})$, there exists an $\epsilon > 0$ such that $f_{P_{\mathcal{R}_c}} > 0$ in $B((0, 0), \epsilon) \subset [-x_0, x_0] \times [-y_0, y_0]$. Denoting by $\mathfrak{m}$ the Lebesgue measure, we have that $\mathfrak{m}(B((0, 0), \epsilon)) = \pi\epsilon^2 > 0$. This leads to a contradiction. □

Proposition 2.3.6 *Under the conditions of Theorem 2.3.1 , we have that the maximum depth value is exactly $\frac{1}{4}$.*

Proof. Let P be a distribution such that $P = P_{\mathcal{R}_c} \in \sigma^*(\mathcal{P})$. By Proposition 2.2.1, $f_{P_{\mathcal{R}_c}}(x, y) = f_{P_{\mathcal{R}_c}}(-x, y) = f_{P_{\mathcal{R}_c}}(x, -y)$. Let $x, y \geq 0$ without loss of generality. Therefore,

$RD((0,0), P_{\mathcal{R}_c}, \beta_c) = 2\left(\int_0^\infty \int_0^\infty f(-x,-y)dxdy\right)\left(\int_0^\infty \int_0^\infty f(x,y)dxdy\right) +$
$2(\int_0^\infty \int_0^\infty f(-x,y)dxdy)(\int_0^\infty \int_0^\infty f(x,-y)dxdy) = 4(\int_0^\infty \int_0^\infty f(x,y)dxdy)^2$.
Since $f_{P_{\mathcal{R}_c}}$ is a density function then,

$$1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)dxdy = 4\int_0^\infty \int_0^\infty f(x,y)dxdy.$$

So, $\int_0^\infty \int_0^\infty f(x,y)dxdy = \frac{1}{4}$ and then, $RD((0,0), P_{\mathcal{R}_c}, \beta_c) = \frac{1}{4}$. □

As before we introduce a bit of geometric intuition to better understand the proof of the Theorem 2.3.2. Following the above ideas with the points $\mathbf{0}, t\mathbf{y}, \mathbf{y} \in \mathbb{R}^2$ where $t \in [0,1]$ we can build a natural partition of the plane like in Figure 2.1. So, it is useful to have an image in mind of this situation as shown in Figure 2.2. In following Theorem 2.3.2 we prove Condition (ii) of Definition 2.1.1 using a strategy similar to that of Theorem 2.3.1 with these ideas.

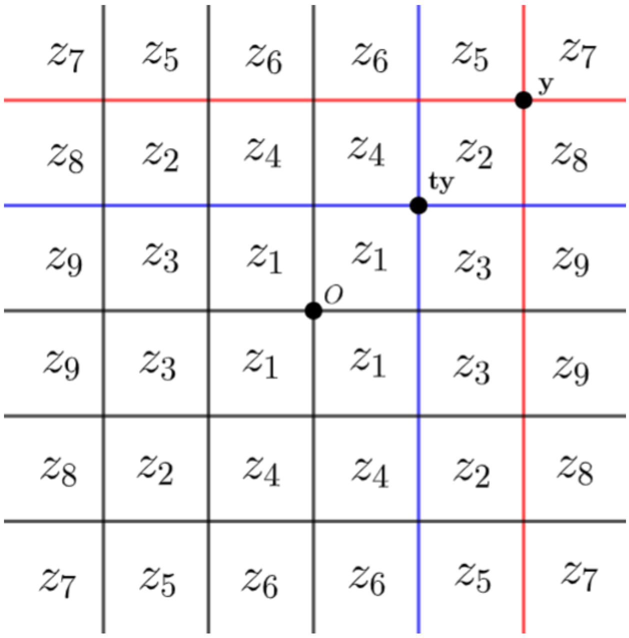


Figure 2.2: Probabilities $z_i's$ with reference $\mathcal{R}_c$ in the spirit of Figure 2.1.

Theorem 2.3.2 *Let $P = P_{\mathcal{R}} \in \sigma(P)$, with $\mathcal{R} = \{\mu; \beta\}$ and $t \in [0,1]$, then $RD(\mathbf{x}, P, \beta) \leq RD(\mu + t(\mathbf{x} - \mu), P, \beta)$ for all $\mathbf{x} \in \mathbb{R}^2$.*

Proof. It is enough to prove it for the case $\mathcal{R} = \mathcal{R}_c$. In effect, in the conditions of Corollary 2.3.1 $RD(\mathbf{x}, P, \beta) = RD(\mathbf{y}, P_{\mathcal{R}_c}, \beta_c)$ where $\mathbf{y} = A\mathbf{x} + b = A(\mathbf{x} - \mu)$.
Note that, $RD(\mu + t(\mathbf{x} - \mu), P, \beta) = RD(A(\mu + t(\mathbf{x} - \mu)) + b, P_{\mathcal{R}_c}, \beta_c) = RD((A\mu + b) + At(\mathbf{x} - \mu), P_{\mathcal{R}_c}, \beta_c) = RD(tA(\mathbf{x} - \mu), P_{\mathcal{R}_c}, \beta_c) = RD(t\mathbf{y}, P_{\mathcal{R}_c}, \beta_c)$.
Then, $RD(\mathbf{x}, P, \beta) \leq RD(\mu + t(\mathbf{x} - \mu), P, \beta)$ if and only if $RD(\mathbf{y}, P_{\mathcal{R}_c}, \beta_c) \leq RD(t\mathbf{y}, P_{\mathcal{R}_c}, \beta_c)$.

Given $P_{\mathcal{R}_c} \in \sigma(\mathcal{P}), \mathbf{y} \in \mathbb{R}^2$, let us see that $RD(\mathbf{y}, P_{\mathcal{R}_c}, \beta_c) \leq RD(t\mathbf{y}, P_{\mathcal{R}_c}, \beta_c)$. Following the proof of Theorem 2.3.1 , by Proposition 2.3.5 we compute:

$RD(t\mathbf{y}, P_{\mathcal{R}_c}, \beta_c) = 2(4z_1 + 2z_3 + 2z_4 + 2z_6 + z_2 + 2z_9 + z_8 + z_7 + z_5)(z_2 + z_5 + z_7 + z_8) + 2(2z_3 + 2z_9 + z_2 + z_8 + z_5 + z_7)(2z_4 + 2z_6 + z_2 + z_5 + z_7 + z_8)$.

$RD(\mathbf{y}, P_{\mathcal{R}_c}, \beta_c) = 2(4z_1 + 4z_4 + 4z_3 + 4z_2 + 2z_6 + 2z_5 + 2z_8 + 2z_9 + z_7)z_7 + 2(2z_8 + 2z_9 + z_7)(2z_6 + 2z_5 + z_7)$.

$RD(t\mathbf{y}, P_{\mathcal{R}_c}, \beta_c) - RD(\mathbf{y}, P_{\mathcal{R}_c}, \beta_c) = 4z_2^2 + 4z_8^2 + 4z_5^2 + 8z_3z_4 + 8z_1z_2 + 8z_4z_2 + 8z_2z_9 + 8z_4z_9 + 8z_2z_8 + 8z_1z_8 + 8z_3z_8 + 8z_9z_8 + 8z_2z_5 + 8z_1z_5 + 8z_3z_5 + 8z_4z_5 + 8z_2z_6 + 8z_5z_6 + 8z_3z_6 + 8z_3z_2 + 8z_4z_8 \geq 0$.
So, $RD(t\mathbf{y}, P_{\mathcal{R}_c}, \beta_c) \geq RD(\mathbf{y}, P_{\mathcal{R}_c}, \beta_c)$. □

Finally, let us see that the depth by reflections 2.3.1 is a statistical depth function in the sense of Definition 2.1.1. We prove Condition (iv) of Definition 2.1.1. The proof of the following Proposition 2.3.7 is not original. It is an adaptation of the third item of Theorem 1 in [9] to the language that we have been introducing throughout this chapter.

Proposition 2.3.7 *Let $P \in \mathcal{BP}_2$, and a reference $\mathcal{R} = \{\mu; \beta\}$, we have the limit* $\lim_{||\mathbf{x}||_\infty \longrightarrow \infty} RD(\mathbf{x}, P, \beta) = 0$.

Proof. Let $\mathbf{x} = (x_1, x_2)$, $X = (X_1, X_2)$ and $Y = (Y_1, Y_2)$ in the reference $\mathcal{R}$. We take the infinity norm $||\cdot||_\infty$ in this reference. Firstly, let us see that $\{\mathbf{x} \in L_\beta(X, Y)\} \subset \{||X||_\infty \geq ||\mathbf{x}||_\infty\} \cup \{||Y||_\infty \geq ||\mathbf{x}||_\infty\}$. In effect, let $\omega \in \{\mathbf{x} \in L_\beta(X, Y)\}$, then by Proposition 2.3.4 we have that $\min(X_i(\omega), Y_i(\omega)) \leq x_i \leq \max(X_i(\omega), Y_i(\omega)), i = 1, 2$ so, $|x_i| \leq \max(|X_i(\omega)|, |Y_i(\omega)|), i = 1, 2$. Since $||\mathbf{x}||_\infty = \max(|x_1|, |x_2|)$, $||X||_\infty = \max(|X_1|, |X_2|)$, $||Y||_\infty = \max(|Y_1|, |Y_2|)$ then the inclusion follows.

Finally, $RD(\mathbf{x}, P, \beta) = \mathbb{P}[\{\mathbf{x} \in L_\beta(X, Y)\}] \leq \mathbb{P}[\{||X||_\infty \geq ||\mathbf{x}||_\infty\} \cup \{||Y||_\infty \geq ||\mathbf{x}||_\infty\}] \leq \mathbb{P}[\{||X||_\infty \geq ||\mathbf{x}||_\infty\}] + \mathbb{P}[\{||Y||_\infty \geq ||\mathbf{x}||_\infty\}]$.
So, if $||\mathbf{x}||_\infty \longrightarrow \infty$ then $RD(\mathbf{x}, P, \beta) \longrightarrow 0$. □

Corollary 2.3.2 *(Main result of the chapter). Let $P = P_\mathcal{R} \in \sigma^*(\mathcal{P})$, with $\mathcal{R} = \{\mu; \beta\}$, and $\beta \in \mathcal{B}_2$. Then the depth by reflections function,*

$$RD(\mathbf{x}, P, \beta) = \mathbb{P}[\{\mathbf{x} \in L_\beta(X, Y)\}],$$

is a $\mathcal{B}_2$-statistical depth function over $\sigma^*(\mathcal{P})$.

Chapter 3

Robustness

In the previous chapter we have proved that depth by reflections is a reasonable statistical depth function for some distributions, for example those in $\sigma^*(\mathcal{P})$. In this chapter we study some properties on robustness, on the concept of breakdown point in particular. For what, we give the formal definition of the empirical version and we prove some properties for its limiting breakdown point.

3.1 Sample depth by reflections

We work with **absolutely continuous** distributions unless otherwise stated. In this chapter we give the definitions with coordinates directly.
Let us denote a sample by $X^{(n)} := \{X_1, X_2, ..., X_n\}$ where $X_1, X_2, ..., X_n$ are independent and identically distributed random variables. Let $X^{(n)}$ being drawn from a distribution $P \in \mathcal{BP}_2$, a reference $\mathcal{R}$ and $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ with coordinates in $\mathcal{R}$. We denote:

$$q_j := q_j(x_1, x_2) := \sum_{i=1}^{n} \mathbf{I}_{C_j}(X_i),\ j = 1, ..., 4 \tag{3.1}$$

where

$$C_1 := C_1(x_1, x_2) := [x_1, \infty) \times [x_2, \infty), C_2 := (-\infty, x_1] \times [x_2, \infty),$$

$$C_3 := (-\infty, x_1] \times (-\infty, x_2], C_4 := [x_1, \infty) \times (-\infty, x_2].$$

Definition 3.1.1 Let $X^{(n)}$ being drawn from a distribution $P \in \mathcal{BP}_2$. We define the **sample (or empirical) depth by reflections** with respect to a reference $\mathcal{R} = \{\mu; \beta\}$ as:

$$RD_n(\mathbf{x}, X^{(n)}) := RD_n(\mathbf{x}, X^{(n)}, \beta) := \binom{n}{2}^{-1} (q_1 q_3 + q_2 q_4). \tag{3.2}$$

In general, we will use the short expressions, q_j,C_j and $RD_n(\mathbf{x}, X^{(n)})$. In the following propositions we prove some basic properties on the estimator that we have just defined. We first prove an immediate consequence of the *Glivenko-Cantelli Theorem*, which means that the empirical depth by reflections, Definition 3.1.1, almost surely approximates the theoretical depth by reflections. For simplicity we denote by a.s the almost surely convergence. The Proposition 3.1.1 is an expected and well-studied result for other depth functions. See for example Remark A.3 of [17].

Proposition 3.1.1 Let $X^{(n)}$ being drawn from an absolutely continuous distribution $P \in \mathcal{BP}_2$ and an arbitrary reference $\mathcal{R} = \{\mu, \beta\}$, then:

$$\sup_{\mathbf{x}\in\mathbb{R}^2} |RD_n(\mathbf{x}, X^{(n)}, \beta) - RD(\mathbf{x}, P, \beta)| \longrightarrow_{a.s} 0 \quad as \quad n \longrightarrow \infty$$

Proof. We note that $\binom{n}{2}^{-1} = \frac{2}{n(n-1)} \approx \frac{2}{n}$ if $n \gg 0$. Let $P_n^j := \frac{q_j}{n}$, $P_j := P[C_j]$. We have that: $\sup_{\mathbf{x}\in\mathbb{R}^2} |RD_n(\mathbf{x}, X^{(n)}) - RD(\mathbf{x}, P)| = 2 \sup_{\mathbf{x}\in\mathbb{R}^2} |P_n^1 P_n^3 + P_n^2 P_n^4 - P_1 P_3 - P_2 P_4| = 2 \sup_{\mathbf{x}\in\mathbb{R}^2} |P_n^1(P_n^3 - P_3) + P_3(P_n^1 - P_1) + P_n^2(P_n^4 - P_4) + P_4(P_n^2 - P_2)| \leq 2\sum_{j=1}^{2} \big(\sup_{\mathbf{x}\in\mathbb{R}^2} |P_n^j| \sup_{\mathbf{x}\in\mathbb{R}^2} |P_n^{j+2} - P_{j+2}| + \sup_{\mathbf{x}\in\mathbb{R}^2} |P_{j+2}| \sup_{\mathbf{x}\in\mathbb{R}^2} |P_n^j - P_j| \big)$.

As by the *Glivenko-Cantelli Theorem* $\sup_{\mathbf{x}\in\mathbb{R}^2} |P_n^j - P_j| \longrightarrow_{a.s} 0$ and $\sup_{\mathbf{x}\in\mathbb{R}^2} |P_n^j|$, $\sup_{\mathbf{x}\in\mathbb{R}^2} |P_j|$ are bounded so, the result follows. □

We need a theoretical result that justifies the calculation of a point of $\mathbb{R}^2$ that maximizes the sample depth by reflections, Definition 3.1.1, to approximate the theoretical deepest point and to be able to introduce the breakdown point later. We use the notation $\longrightarrow_p$ to indicate convergence in probability.

Proposition 3.1.2 *Let $X^{(n)}$ being drawn from a distribution $P = P_{\mathcal{R}} \in \sigma^*(\mathcal{P})$ with $\mathcal{R} = \{\mu; \beta\}$ and let $(M_n)_n$ be a sequence of random variables such that $RD_n(M_n, X^{(n)}) = \max_{\mathbf{x}\in r_v} RD_n(\mathbf{x}, X^{(n)})$ with $r_v = Me(P, \beta) + \alpha v$ where, $v \in S^1$ (unit circle) and $\alpha \geq 0$, then, $M_n \longrightarrow_p Me(P, \beta)$ as $n \longrightarrow \infty$.*

Proof. As $P \in \sigma^*(\mathcal{P})$ we know that $Me(P, \beta) = \mu$. Let us see that every subsequence $(M_{n_k})_k$ has a subsequence $M_{n_{k_j}} \longrightarrow_{a.s} \mu$. By Proposition 3.1.1 with probability 1, for all $\epsilon > 0$, exists $n_0 \in \mathbb{N}$ such that if $n \geq n_0$, for all $\mathbf{x} \in \mathbb{R}^2$, $|RD_n(\mathbf{x}, X^{(n)}) - RD(\mathbf{x}, P)| < \epsilon$. Let $k_0 \in \mathbb{N}$ such that $n_{k_0} \geq n_0$ and let $k \geq k_0$. If $M_{n_k} = \mu$ the result it is true. We suppose that $M_{n_k} \neq \mu$ and let us see that $|RD(M_{n_k}, P) - RD(\mu, P)| < 2\epsilon$. We have that

$|RD_n(M_{n_k}, X^{(n)}) - RD(M_{n_k}, P)| < \epsilon$, $|RD_n(\mu, X^{(n)}) - RD(\mu, P)| < \epsilon$ and by definition of M_{n_k} and μ, $RD_n(M_{n_k}, X^{(n)}) \geq RD_n(\mu, X^{(n)})$, $RD(M_{n_k}, P) < RD(\mu, P)$. So, as $RD_n(M_{n_k}, X^{(n)}) < RD(M_{n_k}, P) + \epsilon$ and $RD_n(\mu, X^{(n)}) >$

$RD(\mu, P) - \epsilon$, we have that:
$RD(\mu, P) - \epsilon < RD_n(\mu, X^{(n)}) \leq RD_n(M_{n_k}, X^{(n)}) < RD(M_{n_k}, P) + \epsilon$. Thus, $RD(\mu, P) - RD(M_{n_k}, P) = |RD(\mu, P) - RD(M_{n_k}, P)| < 2\epsilon$. In conclusion, $RD(M_{n_k}, P) \longrightarrow_{a.s} RD(\mu, P)$. We build the subsequence $M_{n_{kj}}$ as follows:
$\epsilon_1 = 1$, $k_{\epsilon_1} = \min\{k_0 : \forall k' \geq k_0, RD(\mu, P) - RD(M_{n_{k'}}, P) < \epsilon\}$, $M_{n_{k_1}} := M_{n_{k_{\epsilon_1}}}$.
$\epsilon_2 = RD(\mu, P) - RD(M_{n_{k_1}}, P)$, $k_{\epsilon_2} = \min\{k_0 > k_{\epsilon_1} : \forall k' \geq k_0, RD(\mu, P) - RD(M_{n_{k'}}, P) < \epsilon\}$, $M_{n_{k_2}} := M_{n_{k_{\epsilon_2}}}$.
Recursively for an arbitrary j:
$\epsilon_j = RD(\mu, P) - RD(M_{n_{k_{j-1}}}, P), k_{\epsilon_j} = \min\{k_0 > k_{\epsilon_{j-1}} : \forall k' \geq k_0, RD(\mu, P) - RD(M_{n_{k'}}, P) < \epsilon\}$, $M_{n_{k_j}} := M_{n_{k_{\epsilon_j}}}$.

We know that $M_{n_{kj}} = \mu + \alpha_{n_{kj}} v$ and by construction of $M_{n_{kj}}$ we have that : $RD(\mu, P) > RD(M_{n_{kj+1}}, P) > RD(M_{n_{kj}}, P)$.Then, by the Theorem 2.3.2 for all j, $0 < \alpha_{n_{kj+1}} < \alpha_{n_{kj}}$. Thereby, $(\alpha_{n_{kj}})$ is a decreasing sequence of real numbers bounded inferiorly by 0. Thus, $\alpha_{n_{kj}} \longrightarrow_{a.s} 0$ implies $M_{n_{kj}} \longrightarrow_{a.s} \mu$ then, $M_n \longrightarrow_p \mu$. □

Finally, we prove that the estimator we have defined is *unbiased*, which is usually a desirable property.

Proposition 3.1.3 *Let $X^{(n)}$ being drawn from an absolutely continuous distribution $P \in \mathcal{BP}_2$, then:* $\mathbb{E}[RD_n(\mathbf{x}, X^{(n)}, \beta)] = RD(\mathbf{x}, P, \beta)$.

Proof. The expectation of $RD_n(\mathbf{x}, X^{(n)})$, $\mathbb{E}[RD_n(\mathbf{x}, X^{(n)})]$, is equal to $\frac{2}{n(n-1)}\mathbb{E}[\sum_{i=1}^n \mathbf{I}_{C_1}(X_i) \sum_{i=1}^n \mathbf{I}_{C_3}(X_i) + \sum_{i=1}^n \mathbf{I}_{C_2}(X_i) \sum_{i=1}^n \mathbf{I}_{C_4}(X_i)]$, which is identical to $\frac{2}{n(n-1)}(n(n-1)P[C_1]P[C_3] + n(n-1)P[C_2]P[C_4]) = RD(\mathbf{x}, P)$. □

3.2 Breakdown point of depth by reflections

We have enough tools to define the breakdown point and prove the main results of the chapter. This concept will give us an idea of how robust is the estimator defined as sample depth by reflections. We follow the breakdown point definition in [3] which we reproduce below because we will use it in the main theorems of this chapter. Let $X^{(n)}$ be a sample of size n, and T a location estimator, then the **breakdown point** is:

$$\epsilon^*(T, X^{(n)}) := \min \left\{ \frac{m}{n+m} : \sup_{Y^{(m)}} ||T(X^{(n)} \cup Y^{(m)}) - T(X^{(n)})|| = +\infty \right\} \tag{3.3}$$

where the supremmum is taken over any sample $Y^{(m)}$ of size m.

Example. To familiarize ourselves with the previous concept, let us analyze the classic example of the sample mean. The breakdown point of the sample mean $\overline{X}_n$ is $(n+1)^{-1}$ because with only one point we can make tend to infinity $\overline{X}_n$. Thus, $\epsilon^* \longrightarrow_{a.s} 0$.

The following Lemma 3.2.1 is a technical result that we need for the proofs of the two theorems in this chapter. We assume that $0 \in \mathbb{N}$.

Lemma 3.2.1 *The following optimization problem:*

$$(Opt)\begin{cases} \max\ f = y_1x_2 + y_4x_1 + y_1y_3 + y_2y_4 \\ y_1 + y_2 + y_3 + y_4 = m \leq \frac{n}{2} \\ x_1 + x_2 = n \quad even \\ x_2 \geq \frac{n}{2} \\ x_i, y_i, n, m \in \mathbb{N} \end{cases}$$

Has a maximum value $\max f = nm$.

Proof. We prove it by induction in n . Let $n = 2$ and $m = 1$. The space of solutions is:

$$(x_1, x_2) \in \{(1,1),(0,2)\}$$

$$(y_1, y_2, y_3, y_4) \in \{(1,0,0,0),(0,1,0,0),(0,0,1,0),(0,0,0,1)\}.$$

It is easy to check that the maximum value is 2 for $x_1 = 0, x_2 = 2$ and $y_1 = 1, y_2 = y_3 = y_4 = 0$. The case, $m = 0$ is trivial. On then other hand, we suppose that it is true for n. Let us see that it is true for $n + 2$. If $n = 2k$ and $n + 2 = 2(k + 1)$ we have two cases:
1) The case $m \leq k$.
2) The case $m = k + 1$.

1) Let the space of solutions of $(Opt)_n$ to be $S_n = S_n(\mathbf{x}) \times S_n(\mathbf{y})$. We have that

$$S_{n+2} = (S_n(\mathbf{x}) + (1,1) \cup \{(0, n+2)\}) \times S_n(\mathbf{y})$$

Let $(\mathbf{x_{n+2}}, \mathbf{y_{n+2}}) \in S_{n+2}$.
1.1) $\mathbf{x_{n+2}} = \mathbf{x_n} + (1,1)$, $\mathbf{y_{n+2}} = \mathbf{y_n}$ with $\mathbf{x_n} = (x_n^1, x_n^2)$ and $\mathbf{y_n} = (y_n^1, y_n^2, y_n^3, y_n^4)$.

We compute $f_{n+2} = y_n^1(x_n^2 + 1) + y_n^4(x_n^1 + 1) + y_n^1y_n^3 + y_n^2y_n^4 = y_n^1x_n^2 + y_n^1 + y_n^4x_n^1 + y_n^4 + y_n^1y_n^3 + y_n^2y_n^4 = f_n + y_n^1 + y_n^4$. We know that $\max f_{n+2} \leq \max f_n + \max(y_n^1 + y_n^4) = nm + m < nm + 2m < (n+2)m$.

1.2) $\mathbf{x_{n+2}} = (0, n+2)$, $\mathbf{y_{n+2}} = \mathbf{y_n}$.

We compute $f_{n+2} = y_n^1(n+2) + y_n^1y_n^3 + y_n^2y_n^4 = ny_n^1 + 2y_n^1 + y_n^1y_n^3 + y_n^2y_n^4$. So, $\max f_{n+2} \leq \max(ny_n^1 + y_n^1y_n^3 + y_n^2y_n^4) + \max 2y_n^1 = nm + 2m = (n+2)m$.

On the other hand, if $y_n^1 = m, y_n^2 = y_n^3 = y_n^4 = 0$, then $f_{n+2} = (n+2)m$, so, $\max f_{n+2} \geq (n+2)m$. In conclusion, $\max f_{n+2} = (n+2)m$.

2) Let $n = 2k$, $m_n = k$, $n + 2 = 2(k+1)$ and $m_{n+2} = k + 1$. Firstly, the space of solutions is

$$S_{n+2} = (S_n(\mathbf{x}) + (1,1) \cup \{(0, n+2)\}) \times (S_n^k(\mathbf{y}) + \mathbf{e_i} \ ; i = 1,2,3,4)$$

Where if $\mathbf{y} \in S_n^k(\mathbf{y})$ then $y_n^1 + y_n^2 + y_n^3 + y_n^4 = k$.

2.1) $\mathbf{x_{n+2}} = \mathbf{x_n} + (1,1)$, $\mathbf{y_{n+2}} = \mathbf{y_n} + \mathbf{e_1} \quad (i = 1)$

We compute $f_{n+2} = (y_n^1 + 1)(x_n^2 + 1) + y_n^4(x_n^1 + 1) + (y_n^1 + 1)y_n^3 + y_n^2 y_n^4 = y_n^1 x_n^2 + y_n^1 + x_n^2 + 1 + x_n^1 y_n^4 + y_n^4 + y_n^1 y_n^3 + y_n^3 + y_n^2 y_n^4 = f_n + y_n^1 + y_n^3 + y_n^4 + x_n^2 + 1$. So, $\max f_{n+2} \leq \max f_n + \max(y_n^1 + y_n^3 + y_n^4) + \max x_n^2 + 1 \leq nm_n + m_n + n + 1 = 2k^2 + 2k + k + 1 = 2k^2 + 3k + 1 < 2k^2 + 4k + 2 = 2(k+1)^2$.

The cases $\mathbf{x_{n+2}} = \mathbf{x_n} + (1,1)$, $\mathbf{y_{n+2}} = \mathbf{y_n} + \mathbf{e_i}$ with $i = 2,3,4$ are similar.

2.2) $\mathbf{x_{n+2}} = (0, n+2)$, $\mathbf{y_{n+2}} = \mathbf{y_n} + \mathbf{e_1} \quad (i = 1)$

We compute $f_{n+2} = (y_n^1 + 1)(n+2) + (y_n^1 + 1)y_n^3 + y_n^2 y_n^4 = ny_n^1 + 2y_n^1 + n + 2 + y_n^1 y_n^3 + y_n^3 + y_n^2 y_n^4$. So, $\max f_{n+2} \leq \max(ny_n^1 + y_n^1 y_n^3 + y_n^2 y_n^4) + \max(2y_n^1 + y_n^3) + n + 2 = nm_n + 2m_n + n + 2 = 2k^2 + 2k + 2k + 2 = 2(k+1)^2$.

On the other hand, if $\mathbf{y_n} = (k, 0, 0, 0)$, i.e $\mathbf{y_{n+2}} = (k+1, 0, 0, 0)$, then $f_{n+2} = (k+1)(2(k+1)) = 2(k+1)^2$ then $\max f_{n+2} \geq 2(k+1)^2$. In conclusion, $\max f_{n+2} = 2(k+1)^2$.

The cases $\mathbf{x_{n+2}} = (0, n+2)$, $\mathbf{y_{n+2}} = \mathbf{y_n} + \mathbf{e_i}$ with $i = 2,3,4$ are similar. Thus, the result follows. □

In the following Theorem 3.2.1 we give a lower bound for the asymptotic breakdown point, (3.3), under the hypothesis of having a sample from a distribution $P \in \sigma^*(\mathcal{P})$ for an estimator T constructed from the empirical depth by reflections, Definition 3.1.1.

Theorem 3.2.1 *(Main result of the chapter). Let a sample $X^{(n)}$ from a distribution $P \in \sigma^*(\mathcal{P})$. For any estimator T such that*

$$T\left(X^{(n)}\right) \in \left\{M_n : RD_n\left(M_n, X^{(n)}\right) = \max_{\mathbf{x} \in \mathbb{R}^2} RD_n\left(\mathbf{x}, X^{(n)}\right)\right\}, \tag{3.4}$$

the (asymptotic) breakdowm point satisfies:

$$\epsilon^*\left(T, X^{(n)}\right) \longrightarrow_{a.s} \lambda, \quad \lambda \geq \frac{1}{7}. \tag{3.5}$$

Proof. Like in Theorems 2.3.1 and 2.3.2, it is enough to prove it for canonical reflective symmetry. Let us see that $\lim_n \epsilon^*\left(T, X^{(n)}\right) \geq \frac{1}{7}$ a.s. For that, let us suppose for a contradiction that m is smaller than $n/6$ for any $n \gg 0$ such that $\sup_{Y^{(m)}} ||T\left(X^{(n)} \cup Y^{(m)}\right) - T\left(X^{(n)}\right)||_\infty = +\infty$. Then, for any $M > 0$, there exists $Y_M^{(n)}$ such that $||T\left(X^{(n)} \cup Y_M^{(m)}\right) - T\left(X^{(n)}\right)||_\infty > M$. Let $M := \max_{X \in X^{(n)}} ||X||_\infty$. As $n \gg 0$ and $n \longrightarrow \infty$ by Proposition 3.1.2 we estimate $T\left(X^{(n)}\right) = \mathbf{0}$, so, exists $Y^{(m)}$ such that $||T\left(X^{(n)} \cup Y_M^{(n)}\right) - \mathbf{0})||_\infty = ||T\left(X^{(n)} \cup Y_M^{(m)}\right)||_\infty > M$.

Let $Q_M := T\big(X^{(n)} \cup Y_M^{(m)}\big)$. By the reflective symmetry of $X^{(n)}$ ($n \gg 0$) we can suppose that $Q_M \in [0,\infty) \times [0,\infty) = C_1((0,0))$ without loss of generality. (Note that $Q_M \neq \mathbf{0}$.)

On the other hand, we know that $\frac{1}{n}\sum_{i=1}^{n} \mathbf{I}_{C_j}(X_i) \longrightarrow_{a.s} \frac{1}{4}$ with $j = 1,2,3,4$ by the reflective symmetry and by classical asymptotic properties. See also Proposition 2.3.6. Then, if $n \gg 0$ we have that $\sum_{i=1}^{n} \mathbf{I}_{C_j}(X_i) = \frac{n}{4}$. So, by the Equation 3.1.1 It is easy to see $\binom{n}{2} RD_n\big(\mathbf{0}, X^{(n)}\big) = \frac{n^2}{8}$ and $\binom{n}{2} RD_n\big(Q_M, X^{(n)}\big) = 0$ without considering the m new points. It might be helpful to see Figure 3.1.

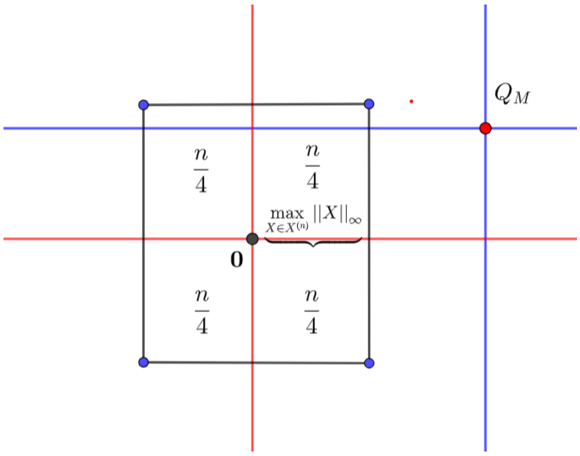


Figure 3.1: Geometric intuition of the situation of point Q_M in the plane in relation to the sample $X^{(n)}$ that is contained in the ball of radius M with the infinity norm.

If we consider the m points of $Y_M^{(m)}$ and we define:

$$w_i := \#\{Y \in Y_M^{(m)} : Y \in C_i(0,0)\}, \quad i = 1,2,3,4,$$

then, by Equation (3.2) we can compute: $\binom{n+m}{2} RD_{n+m}\big((0,0), X^{(n)} \cup Y_M^{(m)}\big) = \big(w_1 + \frac{n}{4}\big)\big(w_3 + \frac{n}{4}\big) + \big(w_2 + \frac{n}{4}\big)\big(w_4 + \frac{n}{4}\big) = w_1 w_3 + w_1 \frac{n}{4} + w_3 \frac{n}{4} + \frac{n^2}{16} + w_2 w_4 + w_2 \frac{n}{4} + w_4 \frac{n}{4} + \frac{n^2}{16} = \frac{n^2}{8} + (w_1 + w_2 + w_3 + w_4)\frac{n}{4} + (w_1 w_3 + w_2 w_4) = \frac{n^2}{8} + \frac{nm}{4} + (w_1 w_3 + w_2 w_4)$. On the other hand, if we define:

$$x_i := \#\{X \in X^{(n)} : X \in C_i(Q_M)\}, \quad i = 1,2,3,4$$

$$y_i := \#\{Y \in Y^{(m)} : Y \in C_i(Q_M)\}, \quad i = 1,2,3,4,$$

as before, it is possible to prove that: $\binom{n+m}{2} RD_{n+m}\big(Q_M, X^{(n)} \cup Y_M^{(m)}\big) = y_1 x_3 + y_4 x_2 + y_1 y_3 + y_2 y_4$. Moreover, by (3.4) we remember that :

$$\binom{n+m}{2} RD_{n+m}\big(Q_M, X^{(n)} \cup Y_M^{(m)}\big) \geq \binom{n+m}{2} RD_{n+m}\big((0,0), X^{(n)} \cup Y_M^{(m)}\big). \tag{3.6}$$

In addition, $y_1 + y_2 + y_3 + y_4 = m$, $x_2 + x_3 = n$ $x_3 \geq \frac{n}{2}$ ($Q_M \in C_1(0,0)$). As $n \gg 0$ without loss of generality we can assume that n is even. So, if we consider the optimization problem (O):

$$(O) \begin{cases} \max \;\; f_n = y_1x_3 + y_4x_2 + y_1y_3 + y_2y_4 \\ y_1 + y_2 + y_3 + y_4 = m \leq \frac{n}{2} \\ x_2 + x_3 = n \;\; even \\ x_3 \geq \frac{n}{2} \\ x_i, y_i, n, m \in \mathbb{N} \end{cases}$$

We have that $\binom{n+m}{2} RD_{n+m}\left(Q_M, X^{(n)} \cup Y_M^{(m)}\right) \leq \max f_n$ because by hypothesis $m < \frac{n}{6} < \frac{n}{2}$. So, by the Lemma 3.2.1 $\binom{n+m}{2} RD_{n+m}\left(Q_M, X^{(n)} \cup Y_M^{(m)}\right) \leq nm$.
Finally, $\max f_n - \binom{n+m}{2} RD_{n+m}\left((0,0), X^{(n)} \cup Y_M^{(m)}\right) = nm - \frac{n^2}{8} - \frac{nm}{4} - (w_1w_3 + w_2w_4) = \frac{3nm}{4} - \frac{n^2}{8} - (w_1w_3 + w_2w_4) = \frac{n}{4}\left(3m - \frac{n}{2}\right) - (w_1w_3 + w_2w_4) < 0$, because $m < \frac{n}{6}$. In conclusion , we have that:

$$\binom{n+m}{2} RD_{n+m}\left(Q_M, X^{(n)} \cup Y_M^{(m)}\right) \leq \max f_n < \binom{n+m}{2} RD_{n+m}\left((0,0), X^{(n)} \cup Y_M^{(m)}\right)$$

which contradicts (3.6). So, $m \geq \frac{n}{6}$ and $\lim_n \epsilon^*\left(T, X^{(n)}\right) \geq \frac{1}{7}$ a.s, because $(n/6)/(n + n/6) = 1/7$. □

In the general conditions of Theorem 3.2.1 it is not possible to proof that the limiting breakdown point is greater than $\frac{3-\sqrt{5}}{5} \approx \frac{1}{7} + \frac{1}{100}$. However, we have the following result.

Theorem 3.2.2 *If $X^{(n)}$ has uniform distribution $\mathcal{U}[-M, M]^2 \in \sigma^*(\mathcal{P})$ with $M > 0$, then exists an estimator T_0 such that:*
(i). $\epsilon^*\left(T_0, X^{(n)}\right) \longrightarrow_{a.s} \lambda, \quad \frac{1}{7} \leq \lambda \leq \frac{3-\sqrt{5}}{5}$.
(ii). $T_0\left(X^{(n)}\right) \in \{M_n : RD_n(M_n, X^{(n)}) = \max_{\mathbf{x} \in \mathbb{R}^2} RD_n(\mathbf{x}, X^{(n)})\}$.
We use symbolic calculator: https://www.dcode.fr/formal-calculator in the proof.

Proof. Let $n \gg 0$, and $Y^{(m)} \geq 0$. We take the function $f_m\left(Y^{(m)}\right) = Q \in \mathbb{R}^2$ such that: $Q := (Q_1, Q_2)$ and $Q_1 = \min_{Y \in Y^{(m)}} Y_1$, $Q_2 = \min_{Y \in Y^{(m)}} Y_2$. Let $Q \geq 0$. We define $Y_Q^{(m)} := f_m^{-1}(\{Q\})$ [1] with $m = (5\sqrt{5} - 11)n$.

We can estimate $\max\limits_{X \in X^{(n)}} ||X||_\infty \longrightarrow_{a.s} M$, and as we have done previously, it is possible to check that when $n \longrightarrow \infty$, then : $\binom{n+m}{2} RD_{n+m}\left(Q, X^{(n)} \cup Y_Q^{(m)}\right) = mn = (5\sqrt{5} - 11)n^2$ for all $Q \in \Gamma := \{(x,y) \in \mathbb{R}^2 : x > M, y > M\}$.

[1] This is an abuse of notation because in general f is not injective. So certainly it is a sample of m points such that $Y_Q^{(m)} \in f_m^{-1}(\{Q\})$.

We define $T_0(X^{(n)} \cup Y_Q^{(m)}) := Q$ if $Q \in \Gamma$. Let $Q' \in \mathbb{R}^2$. If $||Q'||_\infty > M$ by Lemma 3.2.1 and by previous optimization problem (O) we have that $\binom{n+m}{2} RD_{n+m}(Q', X^{(n)} \cup Y_Q^{(m)}) \leq nm$.

Let us see that if $||Q'||_\infty \leq M$ then $\binom{n+m}{2} RD_{n+m}(Q', X^{(n)} \cup Y_Q^{(m)}) \leq nm$.

Let $Q' \in \mathbb{R}^2 = \cup_{j=1}^4 C_j((0,0))$ such that $||Q'||_\infty \leq M$. We define:

$$w_i := \#\{X \in X^{(n)} : X \in S_i\}, i = 1, 2, ..., 16.$$

S_i was defined in the caption of Figure 2.1, $w_i \longrightarrow_{a.s} U[S_i]n$ with $U = U[-M, M]^2$ (as $n \gg 0$, we have only w_1, w_2, w_3, w_4) and we take:

$$D := \binom{n+m}{2} RD_{n+m}((0,0), X^{(n)} \cup Y_Q^{(m)}) - \binom{n+m}{2} RD_{n+m}(Q', X^{(n)} \cup Y_Q^{(m)})$$

We are interested in the minimum of D. As in a previous situations it is possible to check that: $(n \gg 0)$

Case $j = 1$ $(Q' \in C_1((0,0)))$

$$D_1 = 2w_1^2 + 2w_3^2 + 2w_4^2 + 4w_1w_3 + 4w_1w_4 - 3w_1m - w_3m - w_4m.$$

Case $j = 2$ $(Q' \in C_2((0,0)))$

$$D_2 = 2w_1^2 + 2w_3^2 + 2w_4^2 + 4w_1w_3 + 4w_1w_4 + w_1m + w_3m - w_4m.$$

Case $j = 3$ $(Q' \in C_3((0,0)))$

$$D_3 = 2w_1^2 + 2w_3^2 + 2w_4^2 + 4w_1w_3 + 4w_1w_4 + w_1m + w_3m + w_4m.$$

Case $j = 4$ $(Q' \in C_4((0,0)))$

$$D_4 = 2w_1^2 + 2w_3^2 + 2w_4^2 + 4w_1w_3 + 4w_1w_4 + w_1m - w_3m + w_4m.$$

So, it is clear that the important case under study is $Q' \in C_1((0,0))$ because $\min D = \min D_1$. On the other hand, let us see that it is enough to study the points of the segment $[\mathbf{0}, E]$ with $E = (M, M)$.

In effect, as $n \gg 0$, by standard asymptotic properties we have that almost surely:

$w_1 \longrightarrow \frac{xyn}{4}, w_2 \longrightarrow \frac{(1-x)(1-y)n}{4}, w_3 \longrightarrow \frac{(1-x)yn}{4}, w_4 \longrightarrow \frac{x(1-y)n}{4}$

where $x, y \in [0, 1]$. So, replacing in the previous polynomial D_1 we have that:

$$D_1(x,y) = \frac{n^2}{8}\big(x^2y^2 + (1-x)^2y^2 + (1-y)^2x^2 + 2x^2y(1-y) + 2y^2x(1-x)$$
$$-6(5\sqrt{5}-11)xy - 2(5\sqrt{5}-11)x(1-y) - 2(5\sqrt{5}-11)y(1-x)\big).$$

Thus, $D_1(x,y) = \frac{n^2}{8} D_0(x,y)$ and $\min D_1(x,y) = \frac{n^2}{8} \min D_0(x,y)$.

As D_0 is a polynomial, in particular a continuous function and $(x,y) \in [0,1]^2$ which is compact, by the *Weierstrass Theorem* a global minimum of D_1 exists. If the minimum is in the border, then it is clear that $\min D_1 = nm$ in $E = (M, M)$.

If the minimum is in the open subset, then as $D_1 \in C^\infty$ the gradient $\nabla D_0(x,y) = (0,0)$. So, it is easy to check,

$$\frac{\partial D_0}{\partial x} = 2x - 2xy^2 - 2(5\sqrt{5} - 11)y - 2(5\sqrt{5} - 11) = 0$$

$$\frac{\partial D_0}{\partial y} = 2y - 2x^2y - 2(5\sqrt{5} - 11)x - 2(5\sqrt{5} - 11) = 0.$$

Then, $0 = \frac{\partial D_0}{\partial x} - \frac{\partial D_0}{\partial y} = 2(x-y)(xy + 5\sqrt{5} - 10)$. As $x, y \in [0,1]$ and $5\sqrt{5} - 10 > 0$ we have that necessarily $x = y$. So, we study the points of the segment $[\mathbf{0}, E]$. If $x = y$, then we have to study: $D_0(x) = -x^4 + 2(12 - 5\sqrt{5})x^2 - 4(5\sqrt{5} - 11)x$.

We solve the problem $\min_{x \in [0,1]} D_0(x)$. If we study D_0' is easy to see that D_0 has a unique minimum in $(0,1)$, and it is possible to check with a symbolic software that:

$$x_0 := \arg\min_{x \in (0,1)} D_0(x) = \frac{1}{2}(1 - \sqrt{5(9 - 4\sqrt{5})})$$

$$D_0(x_0) = 67 - 30\sqrt{5} < 0.$$

We study the extreme $x = 1$, i.e the point $E = (M, M)$ and we have that $D_0(1) = 67 - 30\sqrt{5} = D_0(x_0)$. Thus,
$\binom{n+m}{2} RD_{n+m}(Q', X^{(n)} \cup Y_Q^{(m)}) \le \binom{n+m}{2} RD_{n+m}(Q, X^{(n)} \cup Y_Q^{(m)})$, for all $Q' \in \mathbb{R}^2$ and for all $Q \in \Gamma$, so, it is clear that:

$$RD_{n+m}(T_0(X^{(m)} \cup Y^{(m)}_{Q\in\Gamma}), X^{(m)} \cup Y^{(m)}_{Q\in\Gamma}) = \max_{Q' \in \mathbb{R}^2} RD_{n+m}(Q', X^{(m)} \cup Y^{(m)}_{Q\in\Gamma}).$$

Finally, we have that: $\sup_{Y^{(n)}} ||T_0(X^{(n)} \cup Y^{(m)}) - T_0(X^{(n)})||_\infty \ge$
$\ge \sup_{Q\in\Gamma} ||T_0(X^{(n)} \cup Y_Q^{(m)}) - \underbrace{T_0(X^{(n)})}_{\mathbf{0}}||_\infty = \sup_{Q\in\Gamma} ||Q||_\infty = +\infty.$

Lastly, $\lim \frac{n(5\sqrt{5}-11)}{n+n(5\sqrt{5}-11)} = \lim \frac{5\sqrt{5}-11}{5\sqrt{5}-10} = \frac{3-\sqrt{5}}{5}$, so, $\lim_n \epsilon^*(T_0, X^{(n)}) \le_{a.s} \frac{3-\sqrt{5}}{5}$. □

To end this chapter, we briefly mention that the breakdown point is known for some of the most famous depth functions in the bivariate case. It should be noted here that the treatment of the breakdown point that we have done in this chapter has been totally personal and therefore may differ considerably from the assumptions or hypotheses used for the results of the other depth functions, for example when talking about the breakdown point in general position. In the following table we summarize this information.

Depth function	Breakdown point
Halfspace depth [16]	$1/3$
Simplicial depth [6]	$< 1/3$
Oja depth [12]	0
Spatial depth [14]	$1/2$
Spherical depth [4]	$(\sqrt{2}-1)/\sqrt{2}$
Lens depth [7]	$(\sqrt{2}-1)/\sqrt{2}$

Table 3.1: Breakdown point of other depth functions.

In Table 3.1 next to the name of the depth function we put the reference where these depth functions were introduced. The reader interested in the references where the results on the breakdown point are proved can look at the references that appear in [7] and [13]. In particular, the reference [13] is a paper that is responsible for organizing the results about breakdown point and computational complexity of several famous depth functions. Specifically, the first four depth functions of Table 3.1 appear in that reference. The last two can be found in [7].

Chapter 4

Computational aspects

Motivation and other comments. In this chapter we focus on the computability part. Let us assume that we are working with data that comes from an absolutely continuous distribution $P \in \mathcal{BP}_2$ but not necessarily reflective symmetric. We focus on $\mathbb{R}^2$ because it is where there is more literature and for larger dimensions the analysis is much more complicated. Given a sample, the objective is to calculate the sample depth contours efficiently. On the other hand, in this section we do not follow any reference. We are not experts in computational geometry, so we do not follow an advanced strategy in this area. We only use strategies from a basic programming course and data structures. Furthermore, we will follow a language and notation typical of *object-oriented programming* (like Java, in particular) and we use the color blue in the attributes to differentiate them from local variables and methods. As motivation we are inspired by the result that appears in [10], which establishes that there is an algorithm such that it calculates the depth contours with temporal and spatial complexity $O(n^2)$ for the famous Tukey or Halfspace depth [16] and it is also known how to implement it. For simplicity we will work with the canonical reference $\mathcal{R}_c = \{\mathbf{0}, \beta_c\}$ (band depth), so we will write the short version $RD(\mathbf{x}, P) := RD(\mathbf{x}, P, \beta_c)$. Moreover, in the case that the sample comes from a distribution $P \in \sigma(\mathcal{P})$, as we have indicated before with other aspects, by Collorary 2.3.1 , there is no loss of generality.

4.1 Previous aspects

In [10] the **α-depth-contour** of distribution $P \in \mathcal{BP}_2$ is defined as the set:

$$RD^{-1}((\alpha, \infty), P) = \{\mathbf{x} \in \mathbb{R}^2 : RD(\mathbf{x}, P) > \alpha)\}. \qquad (4.1)$$

We need to discretize the plane. Let us see how to do it.

Let $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ a sample from distribution $P \in \mathcal{BP}_2$. We take the tuples $(X_{(1)}, X_{(2)}, ..., X_{(n)})$ and $(Y_{(1)}, Y_{(2)}, ..., Y_{(n)})$ where $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ and $Y_{(1)} \leq Y_{(2)} \leq ... \leq Y_{(n)}$. We denote the **state space** by $S := \{R(X_{(i)}, Y_{(j)}) \subseteq \mathbb{R}^2 : 1 \leq i \leq n-1, 1 \leq j \leq n-1\}$ where $R(X_{(i)}, Y_{(j)}) := \{(x, y) \in \mathbb{R}^2 : X_{(i)} \leq x \leq X_{(i+1)}, Y_{(j)} \leq y \leq Y_{(j+1)}\}$. Trivially, if $(x, y) \in \mathbb{R}^2 \setminus \cup_{R \in S} R$ then $RD_n((x, y), P) = 0$. In addition, it is clear that $\#S \leq (n-1)^2$, and for any $\mathbf{x}, \mathbf{y} \in \text{Int}(\ R(X'_i, Y'_j))$, we have that $q_h(\mathbf{x}) = q_h(\mathbf{y})$ for all $h = 1, 2, 3, 4$. $Int(\cdot)$ denotes the interior of a set.

If P is an absolutely continuous distribution, then the lines of the border of $R(X'_i, Y'_j)$ are irrelevant, (in the sense that they are sets of probability 0), an so, by the previous observations, we can concentrate in a one point of each $R(X'_i, Y'_j) \in S$ to study the α-depth-contour. Furthermore, we can consider that the coordinates of the points do not collapse, i.e; there are no collinear points. Thus, $\#S$ is $\Theta(n^2)$.

Due to the previous comments we have to traverse a space of size on the order of n^2. In each element of the state space S we have to calculate the empirical depth by reflections in the most efficient way possible. We also want to emphasize that if we had defined empirical depth by reflections Definition 3.1.1 in Chapter 3 as is normally done in literature, see for example [8] and [6], that is:

$$RD_n(\mathbf{x}, X^{(n)}) = \binom{n}{2}^{-1} \sum_{1 \leq i_1 \leq i_2 \leq n} \mathbf{I}\{\mathbf{x} \in L_\beta(X_{i_1}, X_{i_2})\}, \tag{4.2}$$

it is clear that the most obvious algorithm to calculate the α-depth-contours would have a time cost $\Theta(n^4)$ because calculating the empirical depth for each element of the state space S would have a time cost $\Theta(n^2)$. On the other hand, the properties tested in the previous chapter should be maintained for this other estimator because they are geometrically equivalent. A few properties of this estimator can be seen in [8].

On the other hand, to build and define our program we begin by describing the classes, in the sense of object-oriented programming, that compose it. For this, we give some explanations of the attributes of each class and we will detail the algorithms that are not obvious from the name of the method. For example, the "set( )" and "get( )" methods are what the reader expects, i.e; they are observer methods where in the first case we change the internal state of the variables of the objects and in the second we obtain the value that these variables have.

Class Point(X_1, X_2, x_1, x_2): It defines a point $X = (X_1, X_2)$ of the sample $X^{(n)}$ with coordinates X_1, X_2 whose quadrant is $i \in \{1, 2, 3, 4\}$ defined by a fixed point (x_1, x_2) following the notation $C_i(x_1, x_2)$ from the previous chapter. The methods it has are "getters" and "setters" for each attribute in addition to the constructor method that is in charge of calculating the quadrant i from the point (x_1, x_2).

Class Center(c_1,c_2, depth): It defines a representative point (the center) of each region of the state space S with coordinates c_1, c_2 and sample depth by reflections $RD_n((c_1, c_2), X^{(n)})$, the attribute depth. This class has only "getters", "setters" methods for each attribute in addition to the constructor.

Class Rectangle$(\mathbf{x}, \mathbf{y}, x_0, y_0, n)$: This is the class that contains the important methods. We will dedicate the next subsection of the chapter to describe them. Now we explain the attributes that this class has.

stackCol1: Stack of objects Point(X, Y, x_0, y_0) such that $X \in \mathbf{x}, Y \in \mathbf{y}$ whose quadrant is 1 or 4 ordered in increasing order by the first coordinate.

stackCol2: Stack of objects Point(X, Y, x_0, y_0) such that $X \in \mathbf{x}, Y \in \mathbf{y}$ whose quadrant is 2 or 3 ordered in decreasing order by the first coordinate.

stackRow1: Stack of objects Point(X, Y, x_0, y_0) such that $X \in \mathbf{x}, Y \in \mathbf{y}$ whose quadrant is 1 or 2 ordered in increasing order by the second coordinate.

stackRow2: Stack of objects Point(X, Y, x_0, y_0) such that $X \in \mathbf{x}, Y \in \mathbf{y}$ whose quadrant is 3 or 4 ordered in decreasing order by the second coordinate

ce: Center object. Initially, ce=c_0:=Center(x_0, y_0, sampleDepth( )).

qua1,qua2,qua3,qua4: Natural numbers following the notation $q_j(x_0, y_0)$ from the previous chapter, with sample $X^{(n)}$ such that $X = (X_1, X_2) \in X^{(n)}$ with $X_1 \in \mathbf{x}$ (x: List<Double>) and $X_2 \in \mathbf{y}$ (y: List<Double>).

sizeSample: Number of sample points. (sizeSample=parameter n.)

In the following Figure 4.1 we can see a class diagram that summarizes the structure of the program and the relationships between the classes that we have just introduced.

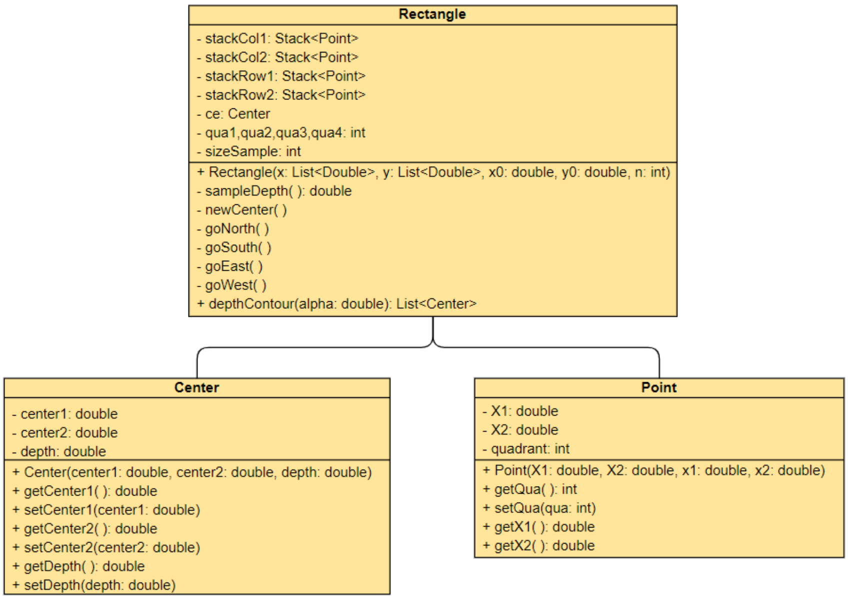


Figure 4.1: Unifield Modeling Language (UML) representation of classes. The attributes and methods of the classes are observed globally.

4.2 The algorithms (of Rectangle)

At the head of the pseudocodes of the following algorithms is the functionality they have and their time complexity. We understand that they are easy to follow because they do not have anything exotic and do not deserve much extra explanation. We would simply like to point out as a summary that the Rectangle( ) constructor method is in charge of initializing the attributes and that its complexity comes from going through n sample data and ordering the stacks according to how they have been defined. We also have a method that is responsible for calculating the empirical depth by reflections, sampleDepth( ); and a method that is responsible for updating the center, or representative, for each movement we make in the state space S, newCenter( ). On the other hand, we have the methods that allow us to move through S, goNorth( ), goSouth( ), goEast( ), goWest( ) in the directions they indicate. The reader should think of these methods as the next( ), previous( ) methods of the classic iterator object in programming. Finally, we have the method that is responsible for calculating the α-depth-contour, depthContour(α). We denote by Tour(S,s_0) an ordered set of the form {goNorth( ), goSouth( ), goWest( ), ...} such that starting initially at the element $s_0 \in S$ and executing the methods "goDirection( )" of Tour(S,s_0) in the order in which they appear let us go through the state space S.

Algorithm 1 Constructor method. Initialize the attributes defined above, the complexity comes from going through the sample size n and ordering the stacks. Complexity $\Theta(n \log n)$

```
procedure RECTANGLE( )                                ▷ parameters in Figure 4.1
```

Algorithm 2 Method that calculates the sample depth. Complexity $\Theta(1)$

```
procedure SAMPLEDEPTH( )
    double aux = (sizeSample)(sizeSample - 1) / 2     ▷ (sizeSample choose 2)
    return (qua1qua3+qua2qua4)/aux                    ▷ See Definition 3.1.1
```

Algorithm 3 Method that calculates the center, or representative of the rectangle, which models a region of S. Complexity $\Theta(1)$

```
procedure NEWCENTER( )
    double depth = SAMPLEDEPTH( )
    ce.SETDEPTH(depth)
    double r1=stackCol1.PEEK( ).GETX1( )
    double r2=stackCol2.PEEK( ).GETX1( )
    double mean=MEAN(r1,r2)
    ce.SETCENTER1(mean)
    r1=stackRow1.PEEK( ).GETX2( )
    r2=stackRow2.PEEK( ).GETX2( )
    mean=MEAN(r1,r2)
    ce.SETCENTER2(mean)
```

Algorithm 4 Method that allows us to move vertically up through S. Complexity $\Theta(1)$

```
procedure GONORTH( )
    if stackRow1.PEEK( ).GETQUA( ) == 1 then
        qua1=qua1-1;
        qua4=qua4+1;
        stackRow1.PEEK( ).SETQUA(4)
    else
        qua2=qua2-1;
        qua3=qua3+1;
        stackRow1.PEEK( ).SETQUA(3)
    Point Q =stackRow1.POP( )
    stackRow2.PUSH(Q)
    if the stacks are not empty then
        NEWCENTER( )                                  ▷ Θ(1)
```

Algorithm 5 Method that allows us to move vertically down through S.
Complexity $\Theta(1)$

```
procedure GOSOUTH( )
    if stackRow2.PEEK( ).GETQUA( ) == 4 then
        qua4=qua4-1;
        qua1=qua1+1;
        stackRow2.PEEK( ).SETQUA(1)
    else
        qua3=qua3-1;
        qua2=qua2+1;
        stackRow2.PEEK( ).SETQUA(2)
    Point Q =stackRow2.POP( )
    stackRow1.PUSH(Q)
    if the stacks are not empty then
        NEWCENTER( )                                    ▷ Θ(1)
```

Algorithm 6 Method that allows us to move horizontally to the right through$S$.
Complexity $\Theta(1)$

```
procedure GOEAST( )
    if stackCol1.PEEK( ).GETQUA( ) == 1 then
        qua1=qua1-1;
        qua2=qua2+1;
        stackCol1.PEEK( ).SETQUA(2)
    else
        qua4=qua4-1;
        qua3=qua3+1;
        stackCol1.PEEK( ).SETQUA(3)
    Point Q =stackCol1.POP( )
    stackCol2.PUSH(Q)
    if the stacks are not empty then
        NEWCENTER( )                                    ▷ Θ(1)
```

Algorithm 7 Method that allows us to move horizontally to the left through S.
Complexity $\Theta(1)$

```
procedure GOWEST( )
    if stackCol2.PEEK( ).GETQUA( ) == 2 then
        qua2=qua2-1;
        qua1=qua1+1;
        stackCol2.PEEK( ).SETQUA(1)
    else
        qua3=qua3-1;
        qua4=qua4+1;
        stackCol2.PEEK( ).SETQUA(4)
    Point Q =stackCol2.POP( )
    stackCol1.PUSH(Q)
    if the stacks are not empty then
        NEWCENTER( )                                    ▷ Θ(1)
```

Algorithm 8 Method that receives **input** a positive real number α and returns as **output** the list of centers of the elements of S such that they have depth greater than α. Tour(S, s_0) is an ordered set. Complexity $\Theta(n^2)$.

```
procedure DEPTHCONTOUR(α: double)
    α-depth-contour= { }                                ▷ Empty List
    double c_1 = 0
    double c_2 = 0
    double depth= 0
    Center newC = null
    for each GODIRECTION( ) ∈ TOUR(S, s_0) do           ▷ Θ(n^2)
        GODIRECTION( )
        depth= ce.GETDEPTH( )
        if depth > α then
            c_1=ce.GETCENTER1( )
            c_2=ce.GETCENTER2( )
            newC= new CENTER(c_1, c_2, depth)
            α-depth-contour.ADD(newC)
    return α-depth-contour
```

For an example, see the Annex.

Theorem 4.2.1 *(Main result of the chapter). Let $X^{(n)}$ a sample from an absolutely continuous distribution $P \in \mathcal{BP}_2$. It is possible to compute the α-depth-contour with time-complexity $\Theta(n^2)$ and space-complexity $\Theta(n)$ where n is the sample size. In addition, the program is easily implementable.*

An important observation here is that with the algorithm above we calculate a single depth contour in time $O(n^2)$ and space $O(n)$. For the reference we mentioned at the beginning of Chapter **4** in the case of Halfspace depth, **all depth contours** in time and space $O(n^2)$ are calculated. We could also achieve this objective by using an $O(n^2)$ space. It would suffice to store the depth of each point in the state space S (the centers) and store them in a convenient data structure. Another observation is the fact that since calculating the convex hull of each depth region supposes a cost $O(n \log n)$, it follows that its calculation does not increase the complexity of the previous algorithm. On the other hand, it is trivially deduced that the temporal complexity of the bivariate median with the depth by reflections with the previous algorithm is $\Theta(n^2)$. As a final observation, it follows from the definition that the temporal complexity of the empirical depth calculation, $RD_n(\mathbf{x}, P)$, is $\Theta(n)$. We put in the Table 4.1 the complexity of other known depth functions in the bivariate case.

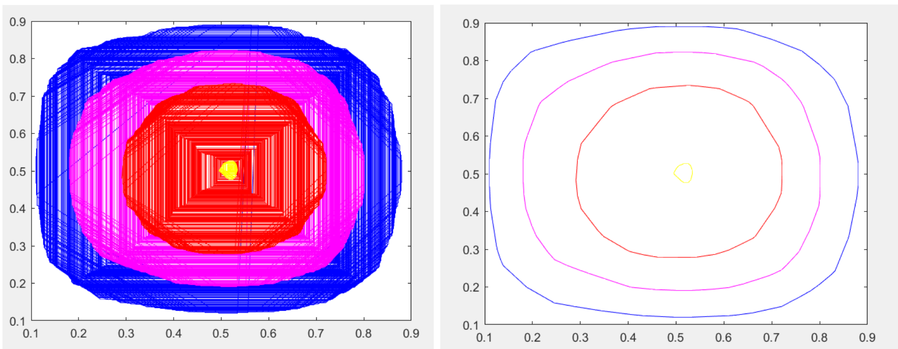
Depth function	Depth complexity	Median complexity
Halfspace depth [16]	$O(n \log n)$	$O(n \log^3 n)$
Simplicial depth [6]	$O(n^3)$	$O(n^4)$
Oja depth [12]	$O(n^2)$	$O(n^3 \log n)$
Spatial depth [14]	$O(n)$	$O(n)$
Spherical depth [4]	$O(n^2)$	$O(n^2)$
Lens depth [7]	$O(n^2)$	$O(n^2)$

Table 4.1: Time complexity of other depth functions.

As in Chapter **3**, in Table 4.1 next to the name of the depth function we put the reference where these depth functions were introduced. The reader interested in the references where the results about computational complexity are proved can look at the references that appear in [7] and [13]. In particular, the reference [13] is responsible for organizing the results on breakdown point and computational complexity of several famous depth functions. Specifically, the first four depth functions of Table 4.1 appear in that reference. The last two can be found in [7].

In Figure 4.2 we show a graphical representation of some α-depth-contours for the uniform distribution $U[0,1]^2$. The α-depth-contours are calculated with our Java program and the graph is made in MATLAB. Moreover, in Figure 4.2 to the right we represent the convex hull of the calculated α-depth-contours, that is, the intersection of all convex sets containing each α-depth-contour.

Example. *Let $X^{(1000)}$ a sample from uniform distribution $U[0,1]^2$. We graphic the α-depth-contour for different values of α.*



(a) α-depth-countors. (b) Convex hulls of α-depth-countors.

Figure 4.2: The α-depth-contours from the example above for $\alpha = 0.1, 0.15, 0.2, 0.25$, (blue, magenta, red, yellow).

Chapter 5

Independence test for two absolutely continuous variables

In this last part of the work we briefly sketch an application of the depth by reflections in the case of the canonical reference (band depth). We continue using the notation $RD(\mathbf{x}, P) := RD(\mathbf{x}, P, \beta_c)$. This application consists of a hypothesis test to check if two absolutely continuous variables, specifically with a density function, are independent. In this part of the work we will not be very rigorous and we will not give proof of the results, because after several attempts, lack of time and possibly experience, they have become quite complicated. Here we only explain the intuition, the operation of the method and some experimental checks.

Let X, Y be two absolutely continuous real random variables, the hypothesis test consists of:

$$H_0 : F_{(X,Y)}(x, y) = F_X(x)F_Y(y) \tag{5.1}$$

$$H_a : F_{(X,Y)}(x, y) \neq F_X(x)F_Y(y)$$

Where $F_{(X,Y)}$ is the joint distribution function and F_X, F_Y the marginal distribution functions. From Section **2.3.5** we know that,

$$RD((x, y), P) = 2P[(-\infty, x] \times (-\infty, y]]P[[x, \infty) \times [y, \infty)]+$$

$$2P[(-\infty, x] \times [y, \infty)]P[[x, \infty) \times (-\infty, y]].$$

Therefore, if we use the hypothesis that the coordinate variables (marginals) are independent we obtain that,

$$RD((x, y), P) = 4F_X(x)(1 - F_X(x))F_Y(y)(1 - F_Y(y)).$$

This motivates the following test statistic.

Let $X^{(n)} = \{(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)\}$ a simple random sample and we considerate the empirical depth of (3.2) and $RD_n^*((x,y), X^{(n)}) := \frac{4j_1j_2j_3j_4}{n^2(n-1)^2}$ where $j_1 := \sum_{i=1}^{n} \mathbf{I}_{(-\infty,x]}(X_i)$, $j_2 := \sum_{i=1}^{n} \mathbf{I}_{[x,\infty)}(X_i)$, $j_3 := \sum_{i=1}^{n} \mathbf{I}_{(-\infty,y]}(Y_i)$ and $j_4 := \sum_{i=1}^{n} \mathbf{I}_{[y,\infty)}(Y_i)$. We use the statistic,

$$T_n := T\left(X^{(n)}\right) := \sum_{i=1}^{n} \left(RD_n((X_i, Y_i), P) - RD_n^*((X_i, Y_i), P)\right)^2 \tag{5.2}$$

We **conjecture** the following statement: **For any absolutely continuous distributions** of X, Y (marginal distributions) if H_0 is true it follows that there exists a distribution parametrized by the sample size n, K_n, such that $T_n \sim K_n$. It can be computationally inspected that regardless of the distributions that have X, Y under the null hypothesis, for example, for a sample of size $n = 30$ there is a distribution of the form represented in Figure 5.1 :

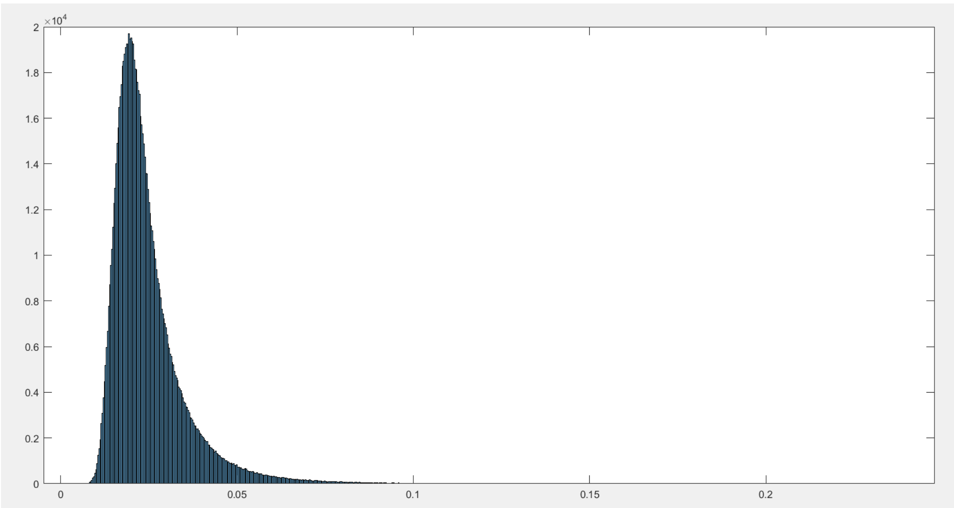


Figure 5.1: Histogram that converges to the K_{30} distribution.

The most important part of this hypothesis test is that the K_n distribution does not depend on the distribution of the variables X, Y under the null hypothesis. Consequently, although it would be much better to have a formula or an analytical expression to define K_n to calculate the *p-values*, $\mathbb{P}(K_n \geq T(X^{(n)})|H_0)$, it is enough to empirically construct K_n for each n by means of Monte Carlo simulations and to tabulate it. This last idea seems important to us, and it is not at all new or original, since John Napier built the famous table of logarithms long before the correct analytical expression (Euler) of this function was provided. However, the intuition and practical use were equally valid. We believe that in statistics today, with the computational power that exists, we should return to that idea of constructing tables more, since waiting to develop a precise mathematical theory that explains each event involves a time and human cost that is not always practicable. Let us not forget that statistics is a practical discipline. We have tabulated K_n for n between 30 and 330 to make small examples. It is included as an additional file located in the folder, **independentTest**, with the name **spvalues.mat**. This table works in conjunction with the file whose name is **sdatos.mat**. But with enough time it can be done for larger values of n. With this, in addition, the complexity of the test is linear in the size of the data once the table has been constructed for K_n, which is the cost of the statistic T_n. As it is a non-parametric hypothesis test, it did not seem reasonable to consider too small samples, say less than 30 because it could lead to faulty results.

Finally, we compare our test with another that performs the same task. The most famous is the Hoeffding independence test [5], which curiously has a philosophy similar to ours, which somehow studies the parameter $\Delta = \int (F_{(X,Y)} - F_X F_Y)^2 dF_{(X,Y)}$. Although being honest, the idea, or the inspiration to define our statistic came from the Pearson's chi square test studied in the basic statistic course. However, we consider that our hypothesis contrast may be more powerful than Hoeffding since it uses more complete and global information. We will try to reflect these observations with the following computational experiments. The strategy to measure the performance of both tests consists of setting a model for the variables X, Y first and then carrying out $10,000$ simulations where in each simulation a sample X, Y is calculated from the fixed model and the p-value of each test computed.

Then, we compared the mean and variance of the p-values of both tests. In addition, we look at the proportion of rejections obtained in the $10,000$ simulations, that is, the proportion of times that a p-value $< .05$ is obtained. The cases in which X, Y are independent, X, Y follow a linear dependence and finally X, Y follow a quadratic dependence have been investigated. We use sample size $n = 100$. The case under the null hypothesis is exemplified in the first part of Table 5.1 where X follows a normal distribution and Y

a uniform distribution. There, our test obtains a proportion of rejections over the nominal level. We consider this is due to not having an analytical expression of K_n. In the second part of Table 5.1 we represent the case for X a normal distribution and Y, dependent of X, is X plus an uniform. Although the results are similar, we obtain a higher rejection rate and the p-values obtained are smaller in average and with a smaller variance. The last case studied is represented in the third part of Table 5.1, where X is again a normal distribution and Y is X to the square plus a uniform. There we can observe that our test clearly outperforms the Hoeffding independences test. The explanation can be found in the structure of the statistics, where ours being wider, informally speaking. The Hoeffding independence test is based on a parameter whose fundamental ingredient is the cumulative distribution function $F_{(X,Y)}(x,y) = \mathbb{P}[(X,Y) \in C_3(x,y)]$ (following the notation C_i of Chapter 3), while in our case, the statistic T in its definition includes the calculations of $\mathbb{P}[(X,Y) \in C_i(x,y)]$ with $i = 1,2,3,4$. Therefore, as $C_3(x,y) \subset \bigcup_{i=1}^{4} C_i(x,y)$, we say that T is wider since it studies the distribution from "more angles".

$X \sim \mathcal{N}(0,1)$, $Y \sim \mathcal{U}(0,1)$ ind.	Δ	T
Mean of p-values	.421752	.384278
Variance of p-values	.050028	.072373
Proportion of rejections	.052400	.099200
$X \sim \mathcal{N}(0,.1)$, $Y \sim X + \mathcal{U}(0,1)$	Δ	T
Mean of p-values	.035871	.015232
Variance of p-values	.006257	.002445
Proportion of rejections	.819300	.926100
$X \sim \mathcal{N}(0,.3)$, $Y \sim X^2 + \mathcal{U}(0,1)$	Δ	T
Mean of p-values	.184396	.016142
Variance of p-values	.028073	.001581
Proportion of rejections	.252500	.919200

Table 5.1: Parameters of the p-values in 10000 simulations of samples of size $n = 100$ according to the model that appears in the upper left part of each subtable.

References

[1] Anirvan Chakraborty and Probal Chaudhuri, (2014). On data depth in infinite dimensional spaces. *Annals of the Institute of Statistical Mathematics, Springer; The Institute of Statistical Mathematics, vol. 66(2), pages 303-324, April.*
[2] Donoho, D.L. (1982). Breakdown properties of multivariate location estimators. *Ph. D. qualifying paper, Dept. Statistics, Harvard University.*
[3] Donoho,D.L.,and Gasko,M. (1992). Breakdown properties of localitation estimates based on halfspace depth and projected outlyingness. *The Annals of Statistics, Vol.20, No.4, 1803-1827.*
[4] Elmore, R.T., Hettmansperger, T.P., and Xuan, F. (2006). Spherical Data Depth and a Multivariate Median. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science (Vol. 72, Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications), ed. R. Liu, R. Serfling and D. Souvaine, Providence, Rhode Island: AMS, pp. 87–101.*
[5] Wassily Hoeffding. (1948). A Non-Parametric Test of Independence. *Ann. Math. Statist. 19 (4) 546 - 557.*
[6] Liu R. (1990). On a notion of data depth based on random simplices. Annals of Statistics 18,405–414.
[7] Zhenyu Liu and Reza Modarres (2011). Lens data depth and median. *Journal of Nonparametric Statistics, 23:4, 1063-1074.*
[8] Lopez-Pintado S. and Romo J. (2009). On the Concept of Depth for Functional Data. *American Statistical Association, Vol.104,No.486.*
[9] Lopez-Pintado S. and Romo J. (2006). On the concept of depth for functional data, 1–32. https://e-archivo.uc3m.es/bitstream/handle/10016/243/ws063012.pdf?sequence=1&isAllowed=y
[10] K. Miller, S. Ramaswwami, P. Rousseeuw, J. Sellarés, D. Souvaine, I. Streinu, and A. Struyf. (2003). Efficient computation of location depth contours by methods of computational geometry. *Statistics and Computing, vol. 13, pp. 153–162.*

[11] Nieto-Reyes A. and Battey H. (2016). A Topologically Valid Definition of Depth for Functional Data. Statistical Science, Vol.32 No.1, 61-79.
[12] H. Oja. (1983). Descriptive statistics for multivariate distributions. *Statist. Probab. Lett.,1:327–332.*
[13] Rousseeuw, P. J., and Struyf, A. (2017). Computation of robust statistics: depth, median, and related measures. *In The Handbook of Discrete and Computational Geometry, J. E. Goodman and J. O'Rourke, Eds. Chapman & Hall/CRC, Boca Raton, pp. 1279–1292.*
[14] R. Serfling. (2002). A depth function and a scale curve based on spatial quantiles. *In Y. Dodge, editor, Statistical Data Analysis Based on the L1-Norm and Related Methods, pages 25–38, Birkhäuser, Basel.*
[15] Serfling, R. (2004). Multivariate symmetry and asymmetry. *Encyclopedia of statistical sciences 8.*
[16] Tukey, J.W., (1975). Mathematics and picturing of data. *Proceedings of ICM, Vancouver, 2 523-531.*
[17] Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *Ann. Statist. 28 461–482.*
[18] Zuo, Y. and Serfling, R. (2000). On the performance of some robust nonparametric location measures relative to a general notion of multivariate symmetry. *J.Static.Plann.Inference 84 55-79.*

Annex

The objective of this annex is to visually and didactically show the operation of the program described in Chapter 4 for a few iterations of a 4-point data set.

Let $X^{(4)} = \{(2,5),(4,3),(6,1),(1,-1)\}$, we execute the previous methods for some iterations.

$S = \{R(1,-1), R(1,1), R(1,3), R(2,-1), R(2,1), R(2,3), R(4,-1), R(4,1), R(4,3)\}$

$$s_0 = R(2,1). \quad Tour(S, s_0) = \{goEast(), goSouth(), goWest(), goWest(), goNorth(),$$
$$goNorth(), goEast(), goEast()\}$$

Rectangle $(\{2,4,6,1\},\{5,3,1,-1\},3,2,4)$ with $(3,2)$ the middle point of s_0.

Q_1 =Point$(2,5,3,2)$, Q_2 =Point$(4,3,3,2)$, Q_3 =Point$(6,1,3,2)$, Q_4 =Point$(1,-1,3,2)$.

The graphical representation of the state space S in Figure 5.2.

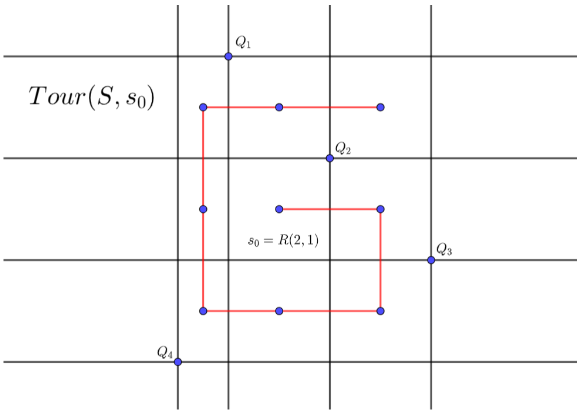


Figure 5.2: $Tour(S, s_0)$, canonical reference $\mathcal{R}_c$

The representation of the points types is in Figure 5.3.

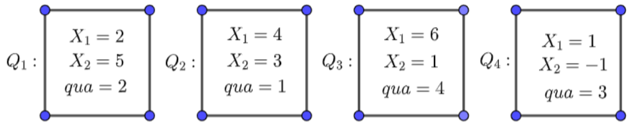


Figure 5.3: Points, Q_1, Q_2, Q_3, Q_4. Attribute qua:=quadrant

In the figures 5.4, 5.5 and 5.6, the situation of the rectangular class attributes is graphically represented for each iteration indicated. In particular the situation of the stacks is represented in each figure and above them we indicate the center of the rectangle for the corresponding iteration.

$$ce(0) = Center(3, 2, \frac{1}{3}), \quad sizeSample = 4$$

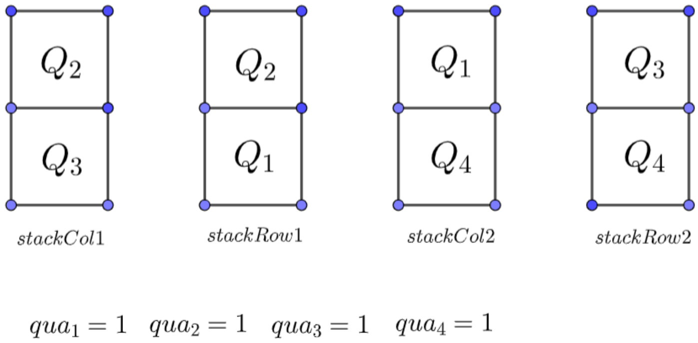


Figure 5.4: Attributes of Rectangle class. Iteration=0.

$$ce(1) = Center(5, 2, \frac{1}{3}), \quad sizeSample = 4$$

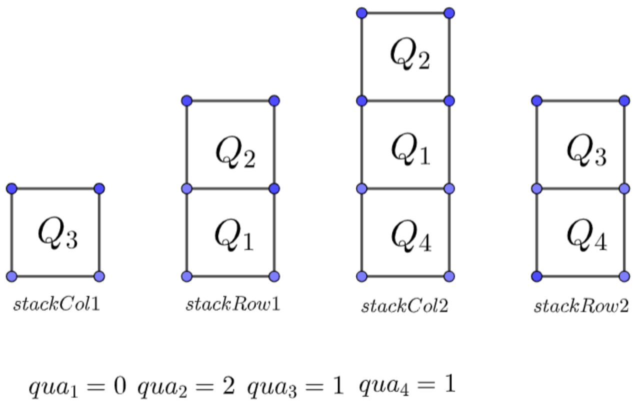


Figure 5.5: Attributes of Rectangle class. Iteration=1.

$$ce(2) = Center(5, 0, \frac{1}{6}), \quad sizeSample = 4$$

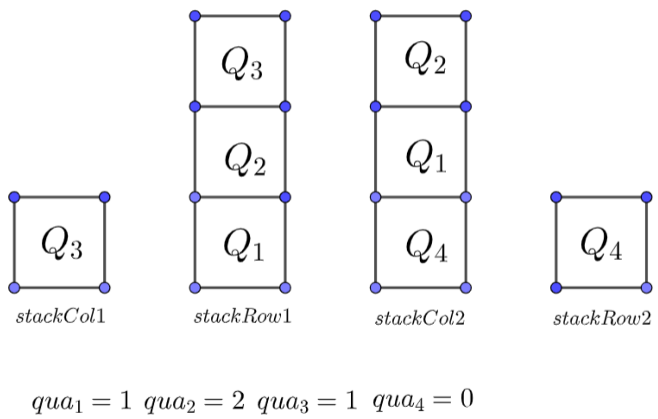


Figure 5.6: Attributes of Rectangle class. Iteration=2.