



UNIVERSIDAD DE MURCIA
Facultad de Matemáticas

TRABAJO FINAL DE GRADO

Métodos Probabilísticos en Análisis Funcional

*Realizado por
José Antonio Salmerón Garrido*

DIRIGIDO POR
MATÍAS RAJA BAÑO

JUNIO 2016

DECLARACIÓN DE ORIGINALIDAD

Yo, José Antonio Salmerón Garrido, declaro ser el autor original del trabajo “*Métodos Probabilísticos en Análisis Funcional*” bajo la tutela del profesor Matías Raja Baño, habiendo reflejado en la bibliografía y referenciado convenientemente todas las fuentes empleadas para su elaboración.

En Murcia, a 14 de junio de 2016

Nota: En la Secretaría de la Facultad de Matemáticas se ha presentado una copia firmada de esta declaración

“Es una gran verdad que cuando no está a nuestro alcance determinar lo que es verdadero, debemos aceptar aquello que sea más probable.”

René Descartes.

Índice general

Resumen	III
Abstract	VII
1. Introducción	1
1.1. Nociones básicas de Teoría de la Medida	1
1.2. Medida Gaussiana	2
1.3. Medida de Haar	3
1.3.1. Distancia Geodésica	5
1.4. Relación entre diferentes medidas	5
1.4.1. Medida Gaussiana y medida de Haar	6
1.4.2. Medida Gaussiana y Medida de Lebesgue	7
1.5. Otras definiciones básicas	8
1.6. Proyecciones Aleatorias	10
1.7. Objetivos	10
2. Desigualdades Isoperimétricas	11
2.0.1. El mejor de los problemas	11
2.0.2. El peor de los problemas	12
2.1. Problema Isoperimétrico	12
2.1.1. Espacio Euclídeo con medida de Lebesgue	14
2.1.2. Esfera n-dimensional con medida de Haar	14
2.1.3. Espacio Euclídeo con medida Gaussiana	19
2.2. Desigualdades Isoperimétricas	20
3. Concentración de la medida	23
3.1. Desigualdades Isoperimétricas y Concentración de la medida	23
3.2. Ejemplos de concentración: Variables aleatorias	25
3.3. Teorema de Dvoretzky	29
3.4. Ejemplos de concentración: Funciones Lipschitz	30
3.5. Un caso especial: El Lema de Johnson-Lindenstrauss	32
4. Lema de Johnson-Lindenstrauss	33
4.1. Introducción	33
4.2. Propiedades empleadas	33
4.2.1. Ortonormalidad	33
4.2.2. Simetría esférica	34
4.2.3. Subespacios Aleatorios	34
4.3. Prueba Original	34

4.4. Prueba íntegramente probabilística	37
4.5. Relajando hipótesis (I): Ortogonalidad	39
4.6. Relajando hipótesis (II): Simetría Esférica	41
4.7. Transformada Rápida de Johnson-Lindenstrauss	44
5. Aplicaciones: Compressed Sensing	47
5.1. Nociones Básicas	47
5.2. Matrices RIP	48
5.3. Algoritmos de recuperación	49
5.4. Relación con el Lema de Johnson-Lindenstrauss	49
6. Conclusiones y Propuestas de Futuro	51
6.1. Cuestiones de interés	51
6.2. Reflexiones finales	51
Bibliografía	53

Resumen

La demostración supone en matemáticas la acción diferencial con respecto a otras ciencias. Podemos conjeturar hipótesis, realizar afirmaciones e investigar sobre un cierto tema, pero todo quedará en vano si un resultado no se acompaña de una demostración rigurosa y válida. A lo largo de la historia de las matemáticas han surgido muchos tipos diferentes de demostraciones, maneras preestablecidas de proceder, tales como la inducción matemática o la reducción al absurdo. Métodos perfectamente válidos y aplicables a numerosos enunciados.

El trasfondo del presente trabajo supone proponer un nuevo método efectivo de demostración, basado en la Teoría de la Probabilidad: si bajo el paraguas de ciertas hipótesis nos encontramos en un espacio de medida probabilístico y queremos demostrar la veracidad de un enunciado el cual podemos describir mediante un suceso A, ¿Por qué no evaluar la probabilidad de que se cumpla dicho enunciado? Si conseguimos demostrar que es estrictamente positiva, repitiendo el experimento un número fijo de veces podemos hacer dicha probabilidad tan cercana a 1 como queramos y asegurar la validez del enunciado. El presente trabajo fija un resultado cuya demostración original se basa precisamente en este método y, por el camino, descubriremos nuevas aplicaciones de esta forma de proceder.

En el año 1984, William B. Johnson y Joram Lindenstrauss publican un artículo sobre el problema de extensión de funciones Lipschitz en espacios de Hilbert, en el que buscan y consiguen, generalizar la anterior cota propuesta por Marcus y Pisier para espacios L_p . Sin embargo, dicho artículo no ha pasado a la historia por conseguir demostrar el objetivo que se plantearon, sino por un lema previo que enuncian en el mismo trabajo y cuya utilidad suponía una herramienta para conseguir su acotación. Dicho lema geométrico es planteado en los siguientes términos:

Dados n puntos en un espacio Euclídeo d -dimensional, ¿cuál es la dimensión más pequeña $k=k(n)$ para la cual podemos mover dichos puntos a un espacio Euclídeo k -dimensional mediante una transformación que expanda o contraiga todos los pares de distancias con un factor de a lo más $1+\epsilon$?

Parémonos a analizar el significado del resultado. Se trata de encontrar una aplicación Lipschitz $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$, $k \ll d$ que verifique, para todos los elementos de un conjunto A de cardinal n ,

$$(1 - \epsilon)\|v_i - v_j\|_2^2 \leq \|f(v_i) - f(v_j)\|_2^2 \leq (1 + \epsilon)\|v_i - v_j\|_2^2, \forall v_i, v_j \in V$$

La prueba, exponente del método de demostración que mencionábamos anteriormente, se basa en elegir un subespacio aleatorio definido mediante una matriz de proyección y verificar que la probabilidad de que las distancias no se alteren sea estrictamente positiva. Nuestra conclusión nos lleva a asegurar la existencia de dicho subespacio y de dicha aplicación Lipschitz. Este es el conocido como Lema de Johnson Lindenstrauss, que será la columna vertebral del trabajo, y el eje alrededor del que orbitan los demás capítulos.

Es presentado al final del tercer capítulo y supone un antes y un después en la memoria. Antes, nos dedicamos a formalizar la base matemática que lo ampara y lo introduce. Después, nos centraremos en presentar sucesivas mejoras del enunciado desde varios puntos de vista y nombrar algunas de sus aplicaciones al campo de la ingeniería.

En el capítulo 1 comenzamos con una breve introducción a teoría de la medida, definiendo precisamente aquellas que protagonizarán nuestro estudio: la medida de Lebesgue, la medida Gaussiana y la medida de Haar. Entre las primeras, comprobaremos que quedan íntimamente relacionadas y sobre la última, la veremos como la única medida invariante por rotaciones que se puede definir en la esfera n -dimensional. Este capítulo finaliza con algunas definiciones básicas sobre teoría de la probabilidad, que serán empleadas en el trabajo.

El capítulo 2 aborda el problema isoperimétrico desde diferentes puntos de vista. Fijado un espacio de medida arbitrario, a saber $(\Omega, \mathcal{F}, \mu)$, dicho problema trata de, entre todos los conjuntos $A \subset \Omega$ de medida fija, $\mu(A) = a$, encontrar aquel que presenta menor perímetro. Se nos plantea entonces la pregunta de cómo definir de manera rigurosa la definición de perímetro, ya que existen varias formas (por ejemplo, la medida de Hausdorff). Definiremos la medida del perímetro cómo,

$$\mu^+(A) = \liminf_{\epsilon \rightarrow 0} \frac{\mu(A_\epsilon) - \mu(A)}{\epsilon} = \liminf_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \mu(A_\epsilon/A)$$

Donde el conjunto A_ϵ contiene a todos los elementos que quedan a distancia menor que ϵ de A . Con esta nueva formulación simplificamos notablemente el problema reduciéndolo a encontrar aquellos conjuntos que minimicen la medida de $\mu(A_\epsilon)$. A lo largo del capítulo, abordamos este problema en diferentes espacios de medida de interés: el espacio euclídeo con la medida de Lebesgue, la esfera unitaria con la medida de Haar y el espacio euclídeo con la medida Gaussiana. Probaremos, mediante los Teoremas 2.1.9 y 2.1.11 que en los dos primeros casos son las bolas, en su propia distancia, los conjuntos que cumplen la propiedad, mientras que en el tercero, como marca el Teorema 2.1.21 son un cierto tipo de semiespacios. Tras ello, nos centramos en la esfera con la medida de Haar como marco de trabajo y nos planteamos que, a parte de saber que conjuntos cumplen el mínimo, nos interesa conocer una acotación inferior para el valor de la medida de $\mu(A_\epsilon)$. Llevaremos a cabo una construcción que desemboca en un resultado en este sentido, Teorema 2.2.1, conocido como *Desigualdad Isoperimétrica*. Para terminar el capítulo, presentaremos un enunciado que nos afirma que, conforme la dimensión crece, la medida de la esfera unitaria se concentra en torno al ecuador. Introducimos y motivamos así los fenómenos de concentración de la medida, objeto del siguiente capítulo.

En el capítulo 3 analizaremos fenómenos de concentración de la medida. Ejemplos destacados son las funciones Lipschitz que condensan sus valores en torno a su mediana hasta tal punto que se pueden considerar “casi constantes”. Suelen ser fenómenos que aparecen cuando la dimensión del dominio de la función es elevado, como nos dicta la Proposición 3.4.2. En este capítulo se lleva a cabo una comparación de dichos fenómenos entre variables aleatorias y funciones Lipschitz. Empezamos relacionando el tema con el capítulo previo y demostrando, mediante las Proposiciones 3.1.2 y 3.1.1, que la aparición de fenómenos de concentración en funciones 1-Lipschitz definidas en la esfera unidad es condición necesaria y suficiente para la existencia de desigualdades isoperimétricas entre los conjuntos de medida mayor a $\frac{1}{2}$ del mismo espacio. Aportamos ejemplos de fenómenos de concentración de variables aleatorias, algunos sobradamente conocidos, como la desigualdad de Tchebychev y otros más sorprendentes, como el hecho de que la proyección a k coordenadas de un vector aleatorio gaussiano d -dimensional normalizado se concentra de manera muy estrecha en torno a su media, como probaremos en la Proposición 3.2.3.

Presentamos también en este capítulo los conocidos Teoremas de Dvoeretzky, parte fundamental del estudio de los fenómenos de concentración de la medida, que nos aseguran la existencia de una copia de un subespacio euclídeo dentro de cualquier espacio normado. Proponemos también una ‘versión Gaussiana’ del mismo, que nos reduce el problema de encontrar ese subespacio al de hallar una variable aleatoria gaussiana definida en dicho espacio de Banach. Mencionamos por último otro resultado de concentración de la medida: la existencia de una “matriz de proyección”, $\Phi \in \mathcal{M}^{k \times d}$, que consiga no perturbar las distancias entre los elementos de un conjunto finito y que reduzca su dimensión de manera considerable y con alta probabilidad.

$$P[(1 - \epsilon)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \epsilon)\|x\|_2^2] \geq 1 - 2e^{-c_o \epsilon^2 k}$$

Hemos recorrido un camino por la geometría y por el análisis matemático que, partiendo del problema isoperimétrico, nos ha conducido casi de forma natural a los fenómenos de concentración de la medida, y en particular al enunciado de Johnson Lindenstrauss. Llamativo resulta que una de las herramientas que nos hemos encontrado a lo largo de nuestro camino, sea precisamente la clave de la demostración que emplean los autores: la desigualdad isoperimétrica en la esfera. Cerrando así un círculo entre el problema isoperimétrico, los fenómenos de concentración de la medida y el lema de Johnson-Lindenstrauss.

El capítulo 4 comienza con un análisis a la demostración original del resultado, Lema 4.3.2, que como hemos dicho se basa en el hecho de buscar probabilidad positiva a un suceso. Estudiamos el proceso y nos damos cuenta de qué hipótesis son asumidas a la hora de efectuar la proyección aleatoria en un subespacio k -dimensional, tales como: la necesidad de que la proyección realizada de los datos sea ortonormal, que el subespacio sea elegido de manera aleatoria en un cierto sentido y que las variables que utilizamos para elegir dicha proyección cumplan la propiedad de poseer simetría esférica. El resto del capítulo tiene un doble objetivo: relajar dichas hipótesis y mejorar la cota de la dimensión de llegada.

A continuación aportamos la primera prueba realizara en su totalidad con métodos probabilísticos y debida a Gupta y Dasgupta, Teorema 4.4.2. Supone una mejora sustancial de la cota aunque siga partiendo de las mismas hipótesis que los autores originales. La gran novedad es identificar los subespacios aleatorios con los vectores aleatorios unitarios, y “modelizar” los primeros mediante variables gaussianas. Tras ella, analizamos el método propuesto por Indyk y Motwami en el Teorema 4.5.2, que nos muestran que no es necesario la condición de ortonormalidad en la matriz que proyectará los datos, sino que es suficiente con que sea “ortonormal en esperanza”. Proponen emplear una matriz cuyas entradas son $N(0, 1)$ independientes sin exigir la ortonormalidad. La siguiente hipótesis que dejamos caer, gracias a Achlioptas en el Teorema 4.6.4, es la de simetría esférica en la matriz de proyección. Propone emplear una matriz aleatoria cuyas entradas tengan una distribución discreta e independiente en el conjunto $\{-1, +1\}$ de manera equiprobable y demuestra que efectivamente sigue cumpliendo la propiedad deseada y además, su implementación es considerablemente más sencilla que las anteriores. Dado que no contemplamos la elección determinista del subespacio, ya hemos relajado todas las hipótesis que asumían los autores del lema. Una última mejora, la llamada Transformada rápida de Johnson Lindenstrauss, la presentamos en el teorema 4.7.4 y supone emplear matrices aleatorias que además tengan un número considerable de entradas nulas, con el fin de que la proyección de los datos sea lo más sencillo desde el punto de la complejidad algorítmica.

Finalmente, en el último capítulo presentamos algunas de las aplicaciones al mundo real que

tiene el lema. Nos centramos en el campo conocido como *Compressed Sensing*, cuyo objetivo es el de recuperar una señal que ha sufrido un cierto proceso de distorsión mediante un medio con ruido. Introduciremos el concepto de matrices RIP, muy utilizadas en este ámbito de la ingeniería y presentaremos un resultado, Teorema 5.4.2 que nos garantiza que las matrices que cumplen el lema son RIP.

Abstract

One of the main differences between mathematics and other sciences is the use of deductive reasoning and logic, along with axioms, to inference theories. Hypothesis can be guessed, statements can be made and certain topics can be researched, but every conjecture will remain unsolved without a valid and rigorous proof. Historically, mathematical statements have been proved by the use of different types of deductive arguments, such as reduction to the absurd, induction or even trial and error. Nowadays, these methods are still widely used when facing new hypothesis and conjectures.

In this research, another method of proof, based on Probability Theory, is proposed. Let us suppose we have an assumption in a probability space that we wish to prove: would it be possible to achieve this through the estimation of the probability of such event? If it were and we managed to deduce that its probability is positive, it would be possible to repeat the experiment a huge number of times and make its probability approach one, which would conclude the proof. This research derives from a result whose original proof is based on this method, and attempts to apply it to other statements.

In 1984, William B. Johnson and Joram Lindenstrauss posted an article about the extension of Lipschitz maps into a Hilbert space, generalizing previous research in L_p spaces made by Marcus and Pisier. Nevertheless, one of the most important parts in this article is a previous lemma they claim in it, using it to reach their objective. That geometric lemma is raised in the following terms:

“Given n points in the d -dimensional Euclidean space, what is the smallest $k = k(n)$ so that these points can be model into the k -dimensional Euclidean space via a transformation which expands or contracts all pairwise distances by a factor of at most $1 + \epsilon$?”

Right now, we ought to analyze the meaning of the result. It is found a Lipschitz map $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$, $k \ll d$, for every element of a finite set.

$$(1 - \epsilon)\|v_i - v_j\|_2^2 \leq \|f(v_i) - f(v_j)\|_2^2 \leq (1 + \epsilon)\|v_i - v_j\|_2^2, \forall v_i, v_j \in V$$

The proof of this lemma shows a great example of the use of the previous mentioned deductive method. It is based on the choice of a random subspace defined by a projection matrix with high probability of having the points at a “very” similar pairwise distance. Our research concludes the existence of such subspace and the Lipschitz map. This constitutes the Lemma of Johnson Lindenstrauss, the main point of the present research and the unifying thread that runs through the entire work. It is introduced at the end of the third chapter, after formalizing some mathematical concepts related to Measure and Probability Theory.

In the first chapter, we present a brief introduction to Measure Theory, defining some measures such as the Lebesgue measure, the Gaussian measure and the Haar measure. We explain the

relationship between the first two of them. Regarding Haar measure, we shall see that it is the unique rotational invariant measure defined on the n -dimensional sphere. We finish this chapter with some basic definitions that will prove themselves of importance along the following ones.

The second chapter consists of an approach to the Isoperimetric Problem from many points of view. Fixed a measure space, for instance $(\Omega, \mathcal{F}, \mu)$, the Isoperimetric Problem is to determine the set with the “smallest” perimeter among every set of the same measure. Thus, we encounter the problem of defining “the perimeter” of a set. There are several definitions, one of which involves the Hausdorff measure, but we will use the following one:

$$\mu^+(A) = \liminf_{\epsilon \rightarrow 0} \frac{\mu(A_\epsilon) - \mu(A)}{\epsilon} = \liminf_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \mu(A_\epsilon/A)$$

Where the set A_ϵ contains every $\omega \in \Omega$ whose distance to A is less than ϵ . By using this notation, the problem has been simplified: now, we wish to find the set which minimizes the measure of $\mu(A_\epsilon)$. We approach this problem from different angles: using the Euclidean space and the Lebesgue measure, the unit sphere and the Haar measure, or the Euclidean space and the Gaussian measure. Theorems 2.1.9 y 2.1.11 prove that the sets given by the first two cases are balls, whereas the set given by the third case is a certain kind of halfspace that will appear in Theorem 2.1.21. Afterwards, we focus on the unit sphere given by the Haar measure, and we wonder weather we could find a lower bound for the measure of the set A_ϵ . Theorem 2.2.1, known as Isoperimetric Inequality, is the answer to our question. At the end of the chapter, we present a statement which claims that the measure of the unit sphere is concentrated around its equator if the dimension is big enough. This proposition serves of introduction and motivation to the following chapter.

In the third chapter, we analyze concentration phenomena. One of the main examples are Lipschitz maps, whose values are tightly concentrated around the median, and therefore they can be considered as *almost constants*. Usually, this phenomena appears when the dimension of the space is big enough, as it is evidenced by 3.4.2. In this chapter, we also make a comparison between concentration measure in Lipschitz maps and in random variables. We first explain the relationship between this topic and the previous chapter. Afterwards, we give proof that the concentration in 1-Lipschitz maps defined in the sphere is a necessary and sufficient condition for the existence of isoperimetric inequalities for sets with measure greater than $\frac{1}{2}$. With respect to random variables, we discuss the most important examples regarding this kind of phenomena, such as the Tchebychev Inequality or Proposition 3.2.3, which establishes the fact that projection of k coordinates of a random Gaussian d -dimensional normalized vector is tightly concentrated around its mean.

Besides, in this chapter we present the Dvoeretzky’s Theorem, essential in the study of measure concentration. This theorem establishes that a copy of a Euclidean subspace can be found within any normed space. It is also enunciated a ‘Gaussian version’ of the same Theorem which verifies that finding the Euclidean subspace is equivalent to finding a random Gaussian variable defined in the Banach space. Finally, we introduce other phenomena of measure concentration: the existence of a projection matrix $k \times d$ which expands or contracts every pairwise distance by a factor of at most $1 + \epsilon$ with high probability in a finite set A reducing the dimension of the data:

$$P[(1 - \epsilon)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \epsilon)\|x\|_2^2] \geq 1 - 2e^{-c_o \epsilon^2 k}$$

In this way, throughout these chapters we have carried out a study in geometry and mathematical analysis. It started with the Isoperimetric Problem and it has led us to the study

of the measure concentration phenomena, in particular, the Johnson-Lindenstrauss lemma. It comes as a surprise that one of the statements found through our “construction” turns to be the key the authors use to prove the Isoperimetric problem in the sphere. Thus, we close the circle between the Isoperimetric Problem, the Concentration of Measure Phenomenon and the Johnson Lindenstrauss lemma.

Chapter four begins with an analysis of the original proof of Lemma 4.3.2. Its proof is based on a method previously mentioned, the one based on Probability Theory: it aims at proving that a certain event A has positive probability. There are three hypothesis that need to be assumed in the random projection: the orthogonality of the projection, the randomness of choice of the subspace and the spherical symmetry of the projection. The rest of the chapter focuses on the objectives of eliminating, if possible, these three hypothesis, as well as improving the dimension of the projection.

First, we proof the first improvement, which is completely based on a probabilistic method given by Gupta and Dasgupta in Theorem 4.4.2. They manage to improve the projection dimension, but make use of the same three hypothesis than Johnson and Lindenstrauss. After that, we focus on the version given by Indyk and Motwami in Theorem 4.5.2, which shows that the hypothesis of orthogonality in the projection is not necessary. Instead, they propose the use of a matrix whose inputs have a Gaussian independent distribution $N(0, 1)$. The next hypothesis that is going to be eliminated is the spherical symmetry of the variables. Its elimination is due to Achlioptas in Theorem 4.6.4. He proposes using a matrix whose inputs have a very simple discrete and independent distribution in the set $\{-1, 1\}$. He proves this matrix to fulfill the required conditions, being its implementation much easier than the previous one's. Since the random choice of projection remains necessary, we have already eliminated the original hypothesis assumed by the authors of the lemma. One last improvement, called the Johnson Lindenstrauss Fast Transform, is presented in Theorem 4.7.4. It assumes the use of random matrixes with a considerable number of null inputs, so that the projection of the data becomes as simple as possible.

Finally, we introduce some of the applications of our lemma into the real world. We focus on the field of Compressed Sensing, whose objective is the recovery of a signal which has suffered a certain process of distortion by means of a noisy channel. We introduce the concept of RIP matrixes, widely used in this engineering area, and we present one last statement, theorem 5.4.2, which guarantees that matrixes satisfying the lemma are also RIP.

Capítulo 1

Introducción

En este capítulo revisaremos conceptos previos a la vez que sentaremos las bases de nuestro trabajo. Fundamentalmente trataremos una breve introducción a la teoría de la medida, orientada a nuestros intereses posteriores, junto con ciertas nociones de teoría de probabilidad, otro de nuestros grandes protagonistas.

1.1. Nociones básicas de Teoría de la Medida

Vamos a comenzar aportando algunas definiciones básicas de Teoría de la medida.

Definición 1.1.1. Sea Ω un conjunto y \mathcal{F} un σ -álgebra sobre Ω . Diremos que una función $\mu : \mathcal{F} \rightarrow [0, \infty]$ es una **medida** si satisface las siguientes propiedades:

1. **Medida nula del conjunto vacío:** $\mu(\emptyset) = 0$
2. **Numerablemente Aditiva:** Para cualquier colección numerable de conjuntos disjuntos, $\{E_i\}_{i=1}^{\infty} \subset \mathcal{F}$, se cumple:

$$\mu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i)$$

Contar con una medida supone ir equipado de una regla que nos aporta una noción de longitud, de área, o de volumen. Para que dicha noción sea coherente con nuestra propia intuición de medir debemos exigir precisamente las condiciones de la definición anterior: que existan unos conjuntos susceptibles de ser medidos, que las medidas siempre sean positivas, que medir el conjunto vacío tenga resultado nulo y que la medida de conjuntos que no poseen intersección sea la suma de la medida de cada uno de ellos.

Definición 1.1.2. A la terna $(\Omega, \mathcal{F}, \mu)$ la llamaremos **espacio de medida** y al par (Ω, \mathcal{F}) , **espacio medible**. Diremos que un conjunto $E \subset \Omega$ es **medible** en (Ω, \mathcal{F}) si $E \in \mathcal{F}$. Si (X, Σ_X) y (Y, Σ_Y) son espacios de medida entonces diremos que la función $f : X \rightarrow Y$ es **medible** respecto dichos espacios si cumple que $f^{-1}(B) \in \Sigma_X, \forall B \in \Sigma_Y$. Diremos que el espacio de medida $(\Omega, \mathcal{F}, \mu)$ es **espacio de probabilidad** si $\mu(\Omega) = 1$.

Una función medible, que parte de un espacio dotado de una medida, induce de forma natural una medida en el espacio de llegada de la función. Surge así en concepto de medida imagen.

Definición 1.1.3. Dados dos espacios medibles, $(X, \Sigma_X), (Y, \Sigma_Y)$, una función $f : X \rightarrow Y$ medible y una medida μ sobre (X, Σ_X) . Llamaremos **medida imagen** de μ mediante f , como,

$$(f_*(\mu))(B) = \mu(f^{-1}(B)), \forall B \in \Sigma_Y$$

Observación 1.1.4. *Las buenas propiedades de las antiimágenes de funciones permiten probar de forma sencilla que efectivamente la medida imagen es una medida sobre (Y, Σ_Y) . Podemos interpretar que la función medible nos traslada la estructura de medida del espacio de partida al de llegada, y que esta nueva medida tiene una densidad determinada por f .*

De especial interés será el caso en el que la medida imagen no modifique la anterior medida, lo que nos motiva la siguiente definición:

Definición 1.1.5. *Sea (Ω, \mathcal{F}) un espacio medible y sea $f : \Omega \rightarrow \Omega$ una función medible. Diremos que una medida μ es **invariante** por f si $\forall A \in \mathcal{F}$,*

$$\mu(f^{-1}(A)) = \mu(A)$$

Diremos que es invariante por el conjunto G , si lo es $\forall f \in G$

Observación 1.1.6. *Siempre que tratemos con la medida de Lebesgue n -dimensional, emplearemos la notación μ_n y salvo que se diga lo contrario, consideraremos la σ -álgebra de Borel, la mínima generada por los abiertos de la topología en la que nos encontremos.*

No vamos a pararnos en detallar la construcción de la medida de Lebesgue, por no ser un tema central del trabajo, dicha construcción puede verse en [12, pág 28-51]. Sin embargo, y ya que trataremos con varias medidas, sí que conviene reflexionar sobre ella. La medida de Lebesgue es aquella que da respuesta y rigor a la noción de medida que posee el ser humano fruto de su intuición y pensamiento. En el espacio unidimensional \mathbb{R} , dicha medida corresponde al cálculo de longitudes, para el cual, un intervalo de la forma $I = (a, b)$ tiene medida $\mu_1(I) = b - a$. Fijando una escala adecuada, podemos verla como nuestra cinta métrica. En \mathbb{R}^2 y \mathbb{R}^3 supone el cálculo *uniforme* de áreas y volúmenes de cualquier objeto que consideremos medible. A un rectángulo de la forma $R = [a, b] \times [c, d]$ se le asigna una medida $\mu_2(R) = (b - a) \cdot (d - c)$ y a partir de estas nociones básicas, la medida de Lebesgue asigna un valor a cada conjunto susceptible de ser medido. Remarco la palabra uniforme precisamente por lo que veremos a continuación: si bien es cierto que Lebesgue plasma nuestra noción de medida de cuerpos de densidad uniforme, en la rama matemática de *Teoría de la Medida* este no es más que un caso particular de medida.

1.2. Medida Gaussiana

A continuación trataremos la medida Gaussiana, que cobrará un gran interés en nuestro trabajo. El objetivo es construir una medida con la *densidad* que nos proporciona la función gaussiana. En particular, el hecho de que $\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{x^2}{2}} dx = 1$ motiva las siguientes definiciones.

Definición 1.2.1. *Sea \mathcal{B} la σ -álgebra de Borel en \mathbb{R} . Entonces, se define la medida Gaussiana en \mathbb{R} como $\gamma : \mathcal{B} \rightarrow [0, 1]$:*

$$\gamma(A) = \frac{1}{\sqrt{2\pi}} \int_A e^{-\frac{x^2}{2}} dx, \forall A \in \mathcal{B}$$

Considerando $A = (a, b)$, notamos que, a diferencia de la anterior, esta medida no depende únicamente de la longitud del intervalo, sino de su posición en la recta real, pues a cada intervalo le corresponde una densidad que viene dada por la conocida función Gaussiana, $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Como podemos observar en la Figura 1.1, aunque ambos intervalos posean la misma longitud, su medida Gaussiana es sustancialmente diferente.

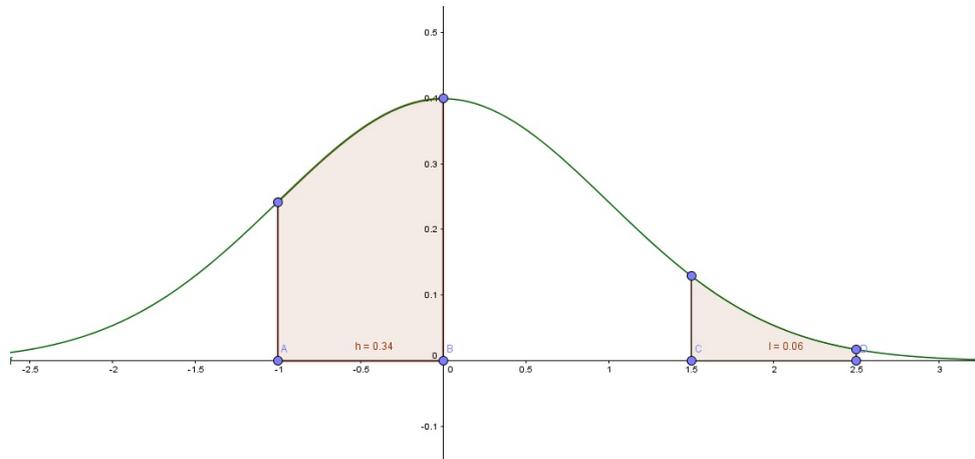


Figura 1.1: Medida Gaussiana de un intervalo. *Elaboración Propia*

Observación 1.2.2. Véase que del valor de la integral se deduce que $\gamma(\mathbb{R}) = 1$, lo cual hace que esta medida tenga mayor interés, pues es medida de probabilidad. Como veremos en este capítulo, la medida asignada a cada intervalo coincide con la probabilidad de dicho intervalo si consideramos una distribución normal standard.

Dicha medida puede ser generalizada a \mathbb{R}^n de forma análoga:

Definición 1.2.3. Sea $n \in \mathbb{N}$ y consideramos \mathcal{B} la σ -álgebra de Borel en \mathbb{R}^n . Entonces, se define la medida Gaussiana standard como $\gamma^n : \mathcal{B} \rightarrow [0, 1]$:

$$\gamma^n(A) = (2\pi)^{-\frac{n}{2}} \int_A e^{-\|x\|^2/2} d\mu_n(x), \forall A \in \mathcal{B}$$

Más generalmente, se define la medida gaussiana de media $\mu \in \mathbb{R}^n$ y varianza $\sigma^2 > 0$ mediante:

$$(\gamma^n)_{\mu, \sigma^2}(A) = (2\pi\sigma^2)^{-\frac{n}{2}} \int_A e^{-\frac{\|x-\mu\|^2}{2\sigma^2}} d\mu_n(x)$$

1.3. Medida de Haar

En esta sección vamos a introducir la Medida de Haar sobre la esfera unidad invariante por rotaciones. Para ello, utilizaremos [25, pag 44-48], donde además pueden consultarse ciertos detalles que omitiremos. A continuación daremos algunas definiciones básicas tanto en esta sección como en el resto del trabajo.

Definición 1.3.1. Llamaremos **esfera unitaria** n -dimensional, y la denotaremos por \mathbb{S}^{n-1} a la frontera de la bola Euclídea en el espacio \mathbb{R}^n , es decir,

$$\mathbb{S}^{n-1} = \{(x_1, \dots, x_n) : \|x\|_2^2 = x_1^2 + \dots + x_n^2 = 1\}$$

Llamaremos esfera centrada en el origen y de radio r^2 n -dimensional, y la denotaremos por $\Sigma_{r^2}^{n-1}$ a la frontera de la bola Euclídea del mismo radio en el espacio \mathbb{R}^n , es decir,

$$\Sigma_{r^2}^{n-1} = \{(x_1, \dots, x_n) : \|x\|_2^2 = x_1^2 + \dots + x_n^2 = r^2\}$$

El objetivo de esta sección es introducir la Medida de Haar sobre el grupo ortogonal $\mathcal{O}(n)$ en la esfera unitaria \mathbb{S}^{n-1} como la única medida sobre la dicha esfera invariante por rotaciones. Partimos del conocido hecho de que la medida de Lebesgue es invariante por traslaciones, es decir,

$$\int_{\mathbb{R}} f(x)d\mu = \int_{\mathbb{R}} f(x+a)d\mu, \quad \forall a \in \mathbb{R}$$

Nuestra intención es constatar la existencia de una medida θ , sobre grupos compactos para la que se cumpla,

$$\int_G f(x)d\theta = \int_G f(gx)d\theta$$

Con f una función integrable en G , y $g \in G$. El siguiente teorema nos responde de manera afirmativa a nuestro objetivo, nos aporta la existencia y unicidad de dicha medida.

Teorema 1.3.2. *Si G es un grupo topológico compacto, existe una única medida θ invariante sobre G de modo que $\theta(G) = 1$. Donde la invarianza de θ significa que $\forall A \subset G, g \in G$,*

$$\theta(A) = \theta(\{gh : h \in A\}) = \theta(\{hg : h \in A\})$$

Demostración. Puede consultarse en [26, pág. 44]. □

Hemos introducido dicha medida en general sobre grupos compactos. Ahora, enfocaremos dichos teorema al caso que nos atañe: la esfera unidad y el grupo ortonormal.

Definición 1.3.3. *Llamaremos **Grupo Ortogonal** de dimensión n sobre \mathbb{R} , denotado por $\mathcal{O}(n, \mathbb{R})$ al conjunto de aplicaciones lineales que preservan el producto escalar usual dotadas con la composición como operación, $(\mathcal{O}(n, \mathbb{R}), \circ)$. Es decir,*

$$\mathcal{O}(n, \mathbb{R}) = \{g : \mathbb{R}^n \longrightarrow \mathbb{R}^n : g(x) \cdot g(y) = x \cdot y, \forall x, y \in \mathbb{R}^n\}$$

Observación 1.3.4. *De modo equivalente puede definirse el grupo ortonormal como,*

$$\mathcal{O}(n, \mathbb{R}) = \{M \in \mathcal{M}^{n \times n} : |M| = \pm 1\}$$

Con el producto de matrices como operación. Esto es debido gracias a la identificación existente entre las aplicaciones lineales y las matrices, así como entre la composición de las primeras y el producto de las segundas.

Evidentemente, se trata de un grupo compacto. Por ejemplo, puede verse como un subconjunto cerrado y acotado de \mathbb{R}^{n^2} . Cerrado, por corresponder a las matrices que satisfacen la ecuación $MM^T = I_n$ y acotado dado que para cada vector se verifica $(MM^T)_{ii} = t_{1i}^2 + \dots + t_{ni}^2 = 1$, luego, $|t_{ij}| \leq 1$. Por tanto, el Teorema 1.3.2 nos determina la existencia y unicidad de una medida invariante por el grupo ortogonal sobre \mathbb{S}^{n-1} . Consideramos el grupo junto con la norma usual,

$$d(g, h) = \|g - h\|_{\infty} = \sup_{\|x\|=1} |g(x) - h(x)|$$

Además, el hecho de que preserve el producto escalar nos repercute en lo siguiente,

$$\|g(x)\|_2^2 = g(x) \cdot g(x) = x \cdot x = \|x\|_2^2 \implies g(\mathbb{S}^{n-1}) \subset \mathbb{S}^{n-1}, \forall g \in \mathcal{O}(n)$$

Con lo cual podemos restringir nuestro estudio del Grupo Ortogonal definido sobre la esfera unitaria. Denotamos por σ_{n-1} la restricción normalizada de la medida de Hausdorff $n-1$ dimensional en el conjunto \mathbb{S}^{n-1} . El siguiente teorema finaliza nuestro objetivo. Nos permite concluir que existe una única medida, además medida de probabilidad, que es invariante por las rotaciones del grupo $\mathcal{O}(n, \mathbb{R})$ y que nos será de gran utilidad en los siguientes capítulos cuando tratemos de medir áreas sobre la superficie de la esfera. Emplearemos θ_n para considerar la medida de Haar sobre \mathbb{S}^{n-1} .

Teorema 1.3.5. Para cualquier $x \in \mathbb{S}^{n-1}$, $A \subset \mathbb{S}^{n-1}$,

$$\theta_n(\{g \in \mathcal{O}(n) : g(x) \in A\}) = \sigma_{n-1}(A)$$

Demostración. Ver [25, pág. 47]. □

El teorema nos dice que, dado cualquier $x \in \mathbb{S}^{n-1}$, $A \subset \mathbb{S}^{n-1}$, la medida del conjunto A es igual a la medida de Haar del conjunto de elementos del grupo ortogonal que nos llevan x a algún elemento de A . Vuelvo a remarcar, esta medida de Haar es la única invariante por rotaciones y además, coincide con la restricción de la medida de Hausdorff.

1.3.1. Distancia Geodésica

Definición 1.3.6. Definimos la distancia geodésica entre $x, y \in \mathbb{S}^{n-1}$ como,

$$d(x, y) = \arccos \langle x, y \rangle$$

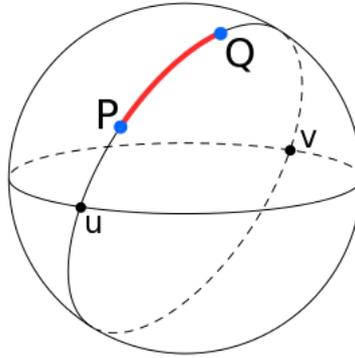


Figura 1.2: Representación de la distancia geodésica

Frecuentemente, cuando tratemos a la esfera como espacio de trabajo, la métrica que emplearemos será la distancia geodésica. Es precisamente aquella que nos dice qué distancia recorreremos por la superficie para ir de un punto a otro si tomamos el menor camino posible. Como sabemos, dicho camino son las curvas geodésicas, que en la esfera suponen circunferencias máximas, tales como los meridianos y el ecuador. El valor que nos aporta dicha distancia entre los puntos P y Q podemos verlo con claridad en la Figura 1.2. Se puede probar que efectivamente es una métrica y que coincide con la distancia recorrida mediante una curva geodésica, sin embargo, y al no ser un tema central del trabajo, no incluiremos los detalles.

1.4. Relación entre diferentes medidas

En este apartado trataremos de establecer diferentes relaciones entre las medidas que hemos introducido anteriormente: la Gaussiana y la de Lebesgue en \mathbb{R}^n y la medida de Haar sobre la esfera unidad. Emplearemos los resultados y las construcciones expuestos en [4].

1.4.1. Medida Gaussiana y medida de Haar

Ya hemos definido varias medidas, en esta sección nos centraremos en estudiar las posibles relaciones que existen entre dichas medidas. El siguiente teorema nos dice precisamente que la medida Gaussiana sobre \mathbb{R} se puede interpretar como la medida imagen mediante la proyección a una variable de la medida de Haar sobre Σ_n^{n-1} cuando n se hace lo suficientemente grande.

Teorema 1.4.1. *Sea $\Sigma_n \subset \mathbb{R}^n$ la esfera centrada en el origen y de radio \sqrt{n} y sea m_n la medida de Haar sobre dicho espacio. Consideramos la proyección*

$$\begin{aligned} \phi : \Sigma_n &\longrightarrow \mathbb{R} \\ (\xi_1, \dots, \xi_n) &\longrightarrow \xi_1 \end{aligned}$$

Y sea $\mathbf{v}_n(A) = m_n(\phi^{-1}(A))$, $\forall A \subset \mathbb{R}$ la medida imagen de m_n mediante ϕ . Entonces, $\forall A \subset \mathbb{R}$, borel medible, se verifica que

$$\lim_{n \rightarrow \infty} (\mathbf{v}_n(A)) = \gamma_1(A) = \frac{1}{\sqrt{2\pi}} \int_A e^{-\xi^2/2} d\xi$$

En particular, la densidad de \mathbf{v}_n converge a la densidad standard Gaussiana uniformemente sobre subconjuntos compactos de \mathbb{R}

Demostración. Tomemos $A = (a, b) \subset \mathbb{R}$ y veamos que efectivamente $\mathbf{v}_n(A) \rightarrow \gamma_1(A)$. Llamamos $B = \phi^{-1}(A) \subset \Sigma_n$. Por la forma en la que hemos definido la aplicación ϕ , el conjunto B está compuesto por los puntos que satisfacen $a < \xi_1 < b$, es decir,

$$B = \{(\xi_1, \dots, \xi_n) : a < \xi_1 < b\}$$

Para un valor fijo de ξ_1 , dicho conjunto es la esfera $n-2$ dimensional de radio $\sqrt{n - \xi_1^2}$. Además usaremos que el área de la esfera es $\hat{\sigma}_{n-1}(\mathbb{S}^{n-1}) = \frac{2\pi^{n/2}}{\Gamma(n/2)}$, donde $\hat{\sigma}_{n-1}$ representa la medida de Hausdorff $n-1$ dimensional sobre la esfera unitaria sin normalizar.

$$\begin{aligned} (1 + o(1))\mu_n(B) &= \frac{\hat{\sigma}_{n-1}(\mathbb{S}^{n-2})}{n^{(n-1)/2} \hat{\sigma}_{n-1}(\mathbb{S}^{n-1})} \int_a^b (n - \xi_1^2)^{(n-2)/2} d\xi_1 \\ &= \frac{\hat{\sigma}_{n-1}(\mathbb{S}^{n-2}) n^{(n-2)/2}}{n^{(n-1)/2} \hat{\sigma}_{n-1}(\mathbb{S}^{n-1})} \int_a^b \left(1 - \frac{\xi_1^2}{n}\right)^{(n-2)/2} d\xi_1 \\ &= \frac{1}{\sqrt{2\pi}} \frac{\Gamma(\frac{n}{2})}{\sqrt{n} \Gamma(\frac{n-1}{2})} \int_a^b \left(1 - \frac{\xi_1^2}{n}\right)^{(n-2)/2} d\xi_1 \end{aligned}$$

Mediante la fórmula de Stirling, deducimos,

$$\lim_{n \rightarrow \infty} \frac{\Gamma(\frac{n}{2})}{\sqrt{n} \Gamma(\frac{n-1}{2})} = \frac{1}{\sqrt{2}}$$

Transformamos la expresión,

$$\begin{aligned} \left(1 - \frac{\xi^2}{n}\right)^{(n-2)/2} &= e^{\frac{n-2}{2} \ln(1-\xi^2/n)} = e^{\frac{n-2}{2} [-\frac{\xi^2}{n} + O(n^{-2})]} \\ &= e^{\frac{n-2}{n} (-\frac{\xi^2}{2}) + O(n^{-2})} \rightarrow e^{-\frac{\xi^2}{2}}, n \rightarrow \infty \end{aligned}$$

Lo que concluye la prueba. □

Observación 1.4.2. *El mismo resultado es válido cuando la proyección se realiza con las k primeras coordenadas, es decir, la medida Gaussiana de \mathbb{R}^k puede ser vista como la medida imagen mediante la proyección a k variables de la medida de Haar sobre Σ_n^{n-1} cuando n se hace lo suficientemente grande. El enunciado preciso puede consultarse en [4, pág. 56]*

1.4.2. Medida Gaussiana y Medida de Lebesgue

El siguiente método nos proporciona una forma eficaz de estimar la medida de conjuntos de la forma $\{x : f(x) \geq a\}$ en un espacio de probabilidad arbitrario. A parte de ser de gran interés en sí mismo, nos dará una forma sencilla de probar una semejanza más que notable entre la medida Gaussiana γ_n y la de Lebesgue μ_n . Se conoce como el **método de la transformada de Laplace**.

Sea Ω un espacio con medida de probabilidad μ y consideramos $f : \Omega \rightarrow \mathbb{R}$ una función. Nuestro objetivo es estimar:

$$\mu\{x : f(x) \geq a\}, \quad a \in \mathbb{R}$$

Una forma de hacerlo es la siguiente, fijamos $\lambda > 0$, entonces:

$$f(x) \geq a \implies e^{\lambda f(x)} \geq e^{\lambda a}$$

$$\int_{\Omega} e^{\lambda f(x)} d\mu \geq \int_{\{x: f(x) \geq a\}} e^{\lambda f(x)} d\mu \geq \mu\{x : f(x) \geq a\} e^{\lambda a}$$

De donde deducimos que,

$$\mu\{x : f(x) \geq a\} \leq e^{-\lambda a} \int_{\Omega} e^{\lambda f(x)} d\mu$$

Válido $\forall \lambda > 0$, por tanto podremos elegir el valor que haga mínima dicha expresión. Con un razonamiento simétrico, obtenemos la desigualdad análoga,

$$\mu\{x : f(x) \leq a\} \leq e^{\lambda a} \int_{\Omega} e^{-\lambda f(x)} d\mu$$

La siguiente proposición nos muestra como la medida Gaussiana de \mathbb{R}^n se acumula cerca de la superficie de la esfera Euclídea Σ_n^{n-1} .

Proposición 1.4.3. *El espacio (\mathbb{R}^n, γ_n) condensa su medida en torno a la esfera Σ_n :*

$$(1) \forall \delta \geq 0, \quad \gamma_n\{x \in \mathbb{R}^n : \|x\|^2 \geq n + \delta\} \leq \left(\frac{n}{n+\delta}\right)^{-n/2} e^{-\delta/2}$$

$$(2) 0 < \delta \leq n, \quad \gamma_n\{x \in \mathbb{R}^n : \|x\|^2 \leq n - \delta\} \leq \left(\frac{n}{n-\delta}\right)^{-n/2} e^{\delta/2}$$

Demostración. Demostremos la primera parte, la segunda es análoga y puede verse en [4]. Fijo un valor $\lambda \in (0, 1)$. Entonces,

$$\|x\|^2 \geq n + \delta \implies e^{\lambda \|x\|^2/2} \geq e^{\lambda(n+\delta)/2}$$

Aplicamos el método expuesto anteriormente, de la transformada de Laplace, considerando $f(x) = \frac{\|x\|^2}{2}$, $a = \frac{n+\delta}{2}$:

$$\gamma_n\{x \in \mathbb{R}^n : \|x\|^2 \geq n+\delta\} \leq e^{-\lambda(n+\delta)/2} \int_{\mathbb{R}^n} e^{\lambda \|x\|^2/2} d\gamma_n = e^{-\lambda(n+\delta)/2} (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{(\lambda-1)\|x\|^2/2} d\mu_n$$

Donde en el último paso hemos empleado la derivación de Radon-Nikodym,

$$\frac{d\gamma_n}{d\mu_n}(x) = (2\pi)^{-n/2} e^{-\|x\|^2/2}$$

Por las propiedades de la norma, la última expresión puede efectuarse en cada variable, de modo que:

$$(2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{(\lambda-1)\|x\|^2/2} d\mu_n = \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{(\lambda-1)x^2/2} dx \right)^n = (1-\lambda)^{-n/2}$$

Dado que la última integral se puede calcular de manera sencilla, mediante el cambio de variable $x = \frac{y}{\sqrt{1-\lambda}}$, resulta una integral gaussiana:

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{(\lambda-1)x^2/2} dx = \frac{1}{\sqrt{2\pi}\sqrt{1-\lambda}} \int_{\mathbb{R}} e^{-y^2/2} dy = \frac{1}{\sqrt{1-\lambda}}$$

Obtenemos finalmente,

$$\gamma_n \{x \in \mathbb{R}^n : \|x\|^2 \geq n + \delta\} \leq (1-\lambda)^{-n/2} e^{-\lambda(n+\delta)/2}$$

Y tomando $\lambda = \frac{\delta}{n+\delta}$ concluye la demostración. \square

Corolario 1.4.4. Para cualquier $0 < \epsilon < 1$,

$$(1) \gamma_n \{x \in \mathbb{R}^n : \|x\|^2 \geq \frac{n}{1-\epsilon}\} \leq e^{-\epsilon^2 n/4}$$

$$(2) \gamma_n \{x \in \mathbb{R}^n : \|x\|^2 \leq (1-\epsilon)n\} \leq e^{-\epsilon^2 n/4}$$

Demostración. Los detalles pueden verse en [4], resulta de tomar $\delta = \frac{n\epsilon}{1-\epsilon}$ y emplear ciertas desigualdades para concluir la acotación. \square

El último corolario goza de una interpretación geométrica muy fuerte, y es que si $0 < \epsilon < 1$,

$$\gamma_n \{x \in \mathbb{R}^n : (1-\epsilon)n \leq \|x\|^2 \leq \frac{n}{1-\epsilon}\} \rightarrow 1, n \rightarrow \infty$$

Podemos concluir que el espacio de medida $(\mathbb{R}^n, \mathcal{B}, \gamma_n)$ se *aplata* en $(\Sigma_n^{n-1}, \mathcal{B}, \mu_n)$ cuando n sea lo suficientemente grande.

1.5. Otras definiciones básicas

Introduzcamos la noción de función Lipschitz que será de gran importancia a lo largo del trabajo.

Definición 1.5.1. Una función $f : X \rightarrow Y$ entre dos espacios métricos $(X, d_X), (Y, d_Y)$ diremos que es **L-Lipschitz** si existe una constante $L \in \mathbb{R}$ que verifique,

$$d_Y(f(x), f(y)) \leq L d_X(x, y), \forall x, y \in X$$

Observación 1.5.2. De forma análoga podemos referirnos a funciones L-Lipschitz entre espacios normados, tomando la distancia que define la norma, $d(x, y) = \|x - y\|$.

Demos ahora la prueba de un lema que nos asegura la unicidad en la mediana de las funciones Lipschitz, será de utilidad en este apartado, pues cuando recurramos a la mediana de f , sabremos que esta elección queda unívocamente determinada.

Lema 1.5.3. Sea $f : \Omega \rightarrow \mathbb{R}$ una función continua. Entonces tiene una única mediana. En particular, si f es L-Lipschitz, entonces la mediana es única.

Demostración. Supongamos que existen dos valores $a < b$ que cumplen la condición de mediana. Entonces trivialmente, todo el intervalo $[a, b]$ sería mediana. Para que este hecho no genere contradicción debe ocurrir que $\mu(f^{(-1)}(a, b)) = 0$, es decir, las antiimágenes de todo el intervalo de medianas debe tener medida nula. Sin embargo, las funciones Lipschitz tienen la propiedad de los valores intermedios, luego la función f toma todos los valores del intervalo (a, b) , luego $f^{(-1)}(a, b)$ es un conjunto no vacío y de probabilidad positiva, llegando así a una contradicción que proviene de suponer que f tiene dos medianas. \square

A continuación damos la definición de variable aleatoria, que supone asignar valores numéricos, usualmente de la recta real, al resultado de un experimento aleatorio. Dicho resultado vive en un espacio muestral Ω y la información de los posibles valores del resultado viene recogida en la σ -álgebra de sucesos \mathcal{F} .

Definición 1.5.4. Una **variable aleatoria** sobre un espacio de probabilidad (Ω, \mathcal{F}, P) es una aplicación $X : \Omega \rightarrow \mathbb{R}$ medible en la σ -álgebra \mathcal{F} , es decir, $X^{-1}([-\infty, s]) \in \mathcal{F}, \forall s \in \mathbb{R}$

El concepto de variable aleatoria supone traducir al lenguaje del análisis matemático experimentos probabilísticos como puede ser lanzar una moneda. Con la variable aleatoria y el espacio de probabilidad podemos operativizar dicho experimento y trabajar con él, desde un punto de vista matemático. Su importancia en Teoría de la Probabilidad y en Análisis Matemático es muy destacada, y en nuestro trabajo será de gran importancia. Conlleva una diferencia importante con el de aplicación. Mientras que la importancia de la función es el valor que toma en cada punto, para las variables nos fijamos en la distribución que siguen dichos valores.

Definición 1.5.5. Una variable aleatoria diremos que es **Gaussiana**, de media $\mu \in \mathbb{R}$ y varianza $\sigma^2 \in \mathbb{R}^+$, y la expresaremos mediante $N(\mu, \sigma^2)$, si posee una distribución,

$$\Phi_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds$$

Observación 1.5.6. Como hemos dejado entrever anteriormente, las medidas Gaussianas quedan íntimamente relacionadas con las variables aleatorias Gaussianas standard. Apreciamos que un espacio de probabilidad definido mediante una variable aleatoria $N(0, 1)$ es esencialmente lo mismo que un espacio de medida gaussiano: la medida de un conjunto medible A será la probabilidad de que ocurra dicho suceso.

A continuación vamos a definir otras variables aleatorias que aparecerán en el texto y presentaremos algunas de sus relaciones. Las demostraciones de dichas proposiciones son sencillos ejercicios de cambios de variable.

Definición 1.5.7. Una variable aleatoria diremos que es **Gamma** de parámetros a y p y será denotada por $G(a, p)$, si posee una distribución,

$$G_{a,p}(x) = \frac{a^p}{\Gamma(p)} \int_0^x s^{p-1} e^{-as} ds, \quad \forall x \geq 0$$

Definición 1.5.8. Una variable aleatoria diremos que es **Exponencial** de parámetro λ y será denotada por $Exp(\lambda)$, si posee una distribución,

$$\Phi_\lambda(x) = 1 - e^{-\lambda x}, \quad x \geq 0$$

Vamos a enunciar unos sencillos resultados de probabilidad cuya demostración consiste en aplicar un cambio de variable sobre las funciones de densidad. Serán aplicados a lo largo del trabajo y nos relacionan las variables aleatorias que acabamos de definir.

Proposición 1.5.9. Si $X, Y \sim N(0, 1)$ son variables independientes, entonces $Z = X^2 + Y^2$ tiene una distribución exponencial con parámetro $\lambda = \frac{1}{2}$.

Demostración. Ver [35]. □

El siguiente resultado nos dice que variables exponenciales del mismo parámetro son reproductivas a una variable Gamma.

Proposición 1.5.10. Dadas k variables X_1, \dots, X_k independientes idénticamente distribuidas con distribución exponencial de parámetro λ , entonces $Y = X_1 + \dots + X_k$ sigue una distribución Gamma de parámetros λ y k .

Demostración. Ver [35]. □

1.6. Proyecciones Aleatorias

En álgebra lineal podemos definir una proyección de la siguiente manera.

Definición 1.6.1. Una **proyección** es una transformación lineal $Q : V \rightarrow V$ que verifica $V^2 = V$.

Por tanto, es una aplicación lineal que se puede identificar con la matriz que la define y además, esta es idempotente. Sin embargo, en el presente trabajo emplearemos proyecciones cuya imagen queda contenida en un subespacio de menor dimensión. En particular, con proyecciones de la forma,

$$Q : \mathbb{R}^d \rightarrow S \simeq \mathbb{R}^k \subset \mathbb{R}^d, \quad k \ll d$$

Por tanto, y empleando un cierto abuso de lenguaje, llamaremos proyecciones a este tipo de aplicaciones, que a menudo vendrán definidas por una matriz $M \in \mathcal{M}^{k \times d}$. Aunque formalmente no son “proyecciones”, pueden tratarse como tal si ampliamos la matriz M con filas nulas y aplicamos un cierto cambio de base. En particular, utilizaremos matrices aleatorias para generar nuestras proyecciones. Llamaremos **proyección aleatoria** a este tipo de aplicaciones, en el que el subespacio S donde viven los datos proyectados es elegido de manera aleatoria, en un cierto sentido que se aclarará más tarde. Podemos tratar estas matrices como una generalización de los vectores aleatorios $X = (X_1, \dots, X_d)$ y tienen por entradas variables aleatorias de una cierta distribución que usualmente tendrán un comportamiento Gaussiano standard independiente aunque veremos que podemos simplificar la distribución enormemente.

1.7. Objetivos

En esta sección vamos a plantear los siguientes objetivos a cumplir en el desarrollo del trabajo:

- Encontrar sendas aplicaciones de “demostraciones probabilísticas”, en las cuales evaluaremos una cierta probabilidad (o *medida*) para concluir la veracidad o existencia de un enunciado.
- Estudiar y vislumbrar la potencia de las Proyecciones Aleatorias como herramienta fundamental en la reducción de dimensión de un conjunto de datos junto con algunas de sus aplicaciones.
- Estudiar el Lema original de Johnson-Lindenstrauss. Analizar sus antecedentes desde el punto de vista de concentración de la medida y fijar la atención en sus predecesores: una batería de enunciados que mejoran la eficiencia del propio lema.

Capítulo 2

Desigualdades Isoperimétricas

2.0.1. El mejor de los problemas

*De Dido el esposo era Siqueo, el hombre más rico en oro
de los fenicios, y lo amó la infeliz sin medida,
desde que su padre la entregara sin mancha y la uniera a él en primeros
auspicios. Pero el poder de Tiro lo ostentaba su hermano
Pigmalión, terrible más que todos los otros por sus crímenes
Y vino a ponerse entre ambos la locura. Este a Siqueo,
impío ante las aras y ciego de pasión por el oro,
sorprende a escondidas con su espada, sin cuidarse
del amor de su hermana (...)
(...) Pero en sueños se le apareció
el propio fantasma de su insepulto esposo, (...)
(...) La anima luego a disponer la huida y salir de su patria,
y saca de la tierra antiguos tesoros escondidos (...)
Se van por el mar las riquezas del avaro Pigmalión;
una mujer dirige la empresa.
Llegaron a estos lugares, donde ahora ves enormes murallas
y nace el alcázar de una joven Carthago,
y compraron el suelo, que por esto llamaron Birsa,
cuanto pudieron rodear con una piel de toro...*

Virgilio. La Eneida, libro IV, [34].
Versos seleccionados por P.J. Herrero, [28]

Encontramos en estos preciosos versos de La Eneida, un origen histórico, legendario y, por qué no decir, casi mitológico que encierra un problema de corte geométrico. Sobre la Reina Dido, cuenta la leyenda que escapó del control, la avaricia y el terror de su hermano Pigmalión. Vagando por el norte de África, llegó a un punto en el que actualmente situamos Túnez y donde se encontraban asentados los gétulos. Jarbas, Rey de los gétulos, aceptó ceder generosamente una porción de tierra en la cual Dido pudiera establecerse, pero con una curiosa forma: Le regalaría tanta tierra como fuera capaz de encerrar con una piel de toro. Dido, buscando la mayor área de tierra posible, decidió cortar dicha piel en pequeñas tiras e ir uniéndolas por los extremos hasta obtener así algo parecido a una cuerda.

La belleza del planteamiento junto con su carácter geométrico supone una buena visión de la Antigua Grecia, pues no de forma desacertada, el frontispicio de la Academia de Atenas rezaba

“*Que no entre quien no sepa geometría*”. Hemos de recordar que la obra de Virgilio, momento inaugural de nuestro problema, fue publicada allá por el siglo I a.C. Los matemáticos, de la época y posteriores hemos tenido más de 2.000 años para buscarle una solución a la reina Dido. La pregunta que podemos plantearnos es... ¿Lo hemos conseguido?

La respuesta no es tan obvia como la pregunta. Si nos restringimos, de forma natural, a un área contenida en un plano, la solución nos la puede dictar nuestra intuición: el círculo es el objeto geométrico que ansiamos. Pero esto ya fue capaz la Reina Dido de deducirlo y Virgilio de contarlo. Si avanzamos situandonos en un espacio tridimensional (“Euclídeo”) quizá la lógica o la propia intuición sea de buena ayuda para llegar a la conclusión acertada: La esfera sería una solución óptima. ¿Podemos entonces afirmar que nuestro problema fue resuelto? Aunque quizá nadie hubiera podido vaticinarlo, salvo sanas excepciones, de dicho problema no sabemos mucho más que aquello que Virgilio y la intuición nos relatan.

Nuestro estudio comienza con una leyenda, un problema geométrico y miles de años para acercarnos a su solución. Podríamos ilustrarlo modificando ligeramente ciertos versos dickensianos y afirmar que: *Era el mejor de los problemas, era el peor de los problemas, la edad de la sabiduría y también de la locura.*

2.0.2. El peor de los problemas

Acaba aquí la leyenda, y nace el riguroso estudio matemático que se plantea la resolución del problema. El enunciado es evidente: Hemos de maximizar el área sujeta a un perímetro fijo. Dicho enunciado, que explica claramente el problema a resolver, podemos considerarlo como impreciso... ¿En qué espacio trabajamos?, ¿Qué medida empleamos para longitudes y áreas? En el presente trabajo, formalizaremos dicho enunciado en términos geométricos, pero rápidamente nos centraremos en nuestro objetivo y trataremos el problema desde un punto de vista probabilístico. Empezando a entrelazar cuestiones analíticas con resoluciones de corte probabilistas.

Veremos que en ambos casos la resolución no es total ni se puede generalizar. Abandonando el empeño de buscar el ‘mínimo absoluto’ del área, nos centraremos en averiguar cotas inferiores al problema. Es decir, relajamos nuestras exigencias y planteamos una versión más débil del problema. Dichas cotas inferiores vienen dadas mediante desigualdades que acotan inferiormente el valor del área, las conocemos como Desigualdades Isoperimétricas.

2.1. Problema Isoperimétrico

A continuación, pasamos a definir con mayor detalle en qué consiste nuestro problema. Comenzamos describiendo unos conjuntos que serán de gran utilidad, consisten en añadir a cualquier conjunto todos aquellos puntos que estén lo “suficientemente” cerca del original. Consideremos el espacio de medida $(\Omega, \mathcal{F}, \mu)$, junto con una métrica d sobre Ω .

Definición 2.1.1. Sea $A \subset \Omega$, y $\epsilon > 0$ se definen los ϵ -vecinos de A como,

$$A_\epsilon = \{\omega \in \Omega : d(\omega, A) < \epsilon\}$$

De considerable importancia es la siguiente definición que vamos a tratar. Estamos hablando sobre minimizar la medida del perímetro, cabe plantearse qué entendemos por el perímetro de un conjunto medible.

Definición 2.1.2. Si $A \subset \Omega$ es un conjunto medible Borel, definimos la **medida del perímetro** de A como,

$$\mu^+(A) = \liminf_{\epsilon \rightarrow 0} \frac{\mu(A_\epsilon) - \mu(A)}{\epsilon} = \liminf_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \mu(A_\epsilon/A)$$

El conjunto A_ϵ aglutina a todos los elementos de Ω suficientemente cercanos a A , si evaluamos el conjunto A_ϵ/A precisamente nos quedan todos aquellos elementos ϵ -ceranos a A , que constituyen, tomando límite, aquello que intuitivamente consideramos por perímetro. Esto nos replantea de nuevo el problema, y nos aporta una forma más sencilla de resolverlo: basta con estudiar aquellos conjuntos que, de medida fija, minimizan el valor de la medida de sus ϵ -vecinos y que a la postre, también harán mínimo el valor del perímetro.

Definición 2.1.3. Sea $(\Omega, \mathcal{F}, \mu, d)$ un espacio métrico y de medida, $0 < a < 1$, $\epsilon > 0$. Llamaremos **Problema Isoperimétrico**, en dicho espacio a encontrar el siguiente ínfimo:

$$\inf_{\epsilon > 0} \{ \mu(A_\epsilon) : A \in \mathcal{F}, \mu(A) = a \}$$

Como hemos mencionado, la solución exacta al problema la conocemos en casos muy concretos. Ante esa barrera, aparentemente insalvable, optaremos por estudiar una versión débil del problema, lo que hemos llamado Desigualdades Isoperimétricas: Buscamos cotas inferiores al valor $\mu(A_\epsilon)$ con $\mu(A) = a$ fijo. En las siguientes secciones trataremos de dar respuesta al problema planteado. Necesitaremos ciertas herramientas, tales como la Desigualdad de Brunn-Minkovski.

Definición 2.1.4. Sean $A, B \subset \Omega$. Definimos la suma de conjuntos de Minkovski como,

$$A + B := \{a + b : a \in A, b \in B\}$$

Teorema 2.1.5. (Desigualdad De Brunn-Minkovski) Para cualquier par de conjuntos A, B de medida no nula de \mathbb{R}^n y sea $0 < \lambda < 1$, entonces,

$$(1) \mu_n(\lambda A + (1 - \lambda)B) \geq \mu_n(A)^\lambda \mu_n(B)^{1-\lambda}$$

$$(2) \mu_n(A + B)^{1/n} \geq \mu_n(A)^{1/n} + \mu_n(B)^{1/n}$$

Demostración. Puede consultarse en [32] □

Definición 2.1.6. Llamamos **función isoperimétrica** a I_μ si es la mayor función que verifica:

$$\forall A \subset \Omega, \mu^+(A) \geq I_\mu(\mu(A))$$

Dado que nuestro objetivo es minimizar el valor de $\mu^+(A)$ de entre los conjuntos con $\mu(A)$ fijo, la función isoperimétrica nos pone límites a nuestros objetivos, diciendo que para cada valor de $\mu(A)$, la medida del perímetro de cualquier conjunto no podrá ser inferior a $I_\mu(\mu(A))$, por tanto, pedirle a la función I_μ que sea la mayor posible, es equivalente a pedir que la acotación sea la más fina posible: el conjunto que consiga la igualdad minimiza el perímetro.

Observación 2.1.7. No vamos a profundizar en el estudio de las funciones isoperimétricas. Simplemente destacar que es, en general, un objeto desconocido: únicamente sabemos de su valor exacto cuando tratamos en $\Omega = \mathbb{R}^n$ y la medida gaussiana o si el conjunto Ω es una bola Euclídea de \mathbb{R}^n y contamos con la medida de Lebesgue. Para ampliar el estudio puede consultarse [23].

2.1.1. Espacio Euclídeo con medida de Lebesgue

Empecemos por tratar el problema en el espacio que más podemos asemejar con una realidad física en bajas dimensiones. Consideremos el espacio de medida $(\mathbb{R}^n, \mathcal{B}, \mu_n)$, el siguiente teorema nos dice que los conjuntos que menos aumentan de medida al añadirle “los vecinos” son precisamente las bolas Euclídeas. Encontramos en este resultado una solución a nuestro problema: De entre los conjuntos de área fija, aquellos que minimizan el perímetro son las bolas con la norma Euclídea. Antes de probar el teorema necesitamos un lema previo que nos afirma la aditividad del operador $\mu_n^{1/n}$ respecto de las bolas. Emplearemos las ideas recogidas en [4, pág.81]

Lema 2.1.8. *El operador $\mu_n^{1/n}$ es aditivo en la suma de Minkovski para las bolas Euclídeas del mismo centro.*

Demostración. Consideramos B_r, B_s bolas de radios r y s respectivamente. Según la suma de Minkovski, $B_r + B_s = B_{r+s}$, entonces:

$$\begin{aligned} \mu_n^{1/n}(B_r + B_s) &= \mu_n^{1/n}(B_{r+s}) = \left(\frac{\pi^{n/2}(r+s)^n}{\Gamma(\frac{n}{2}+1)} \right)^{1/n} = \frac{\pi^{1/2}(r+s)}{\Gamma(\frac{n}{2}+1)^{1/n}} \\ &= \frac{\pi^{1/2}r}{\Gamma(\frac{n}{2}+1)^{1/n}} + \frac{\pi^{1/2}s}{\Gamma(\frac{n}{2}+1)^{1/n}} = \mu_n^{1/n}(B_r) + \mu_n^{1/n}(B_s) \end{aligned}$$

□

Teorema 2.1.9. *Sea $A \subset \mathbb{R}^n$ un conjunto de medida no nula y sea $B_r \subset \mathbb{R}^n$ una bola centrada en el origen de radio r que cumpla $\mu_n(A) = \mu_n(B_r)$. Entonces, $\forall \epsilon > 0$,*

$$\mu_n(A_\epsilon) \geq \mu_n(B_{r+\epsilon})$$

Demostración. Aplicamos la desigualdad de Brunn-Minkovski:

$$\mu_n^{\frac{1}{n}}(A_\epsilon) = \mu_n^{\frac{1}{n}}(A + B_\epsilon) \geq \mu_n^{\frac{1}{n}}(A) + \mu_n^{\frac{1}{n}}(B_\epsilon) = \mu_n^{\frac{1}{n}}(B_r) + \mu_n^{\frac{1}{n}}(B_\epsilon) = \mu_n^{\frac{1}{n}}(B_{r+\epsilon})$$

La primera igualdad es por definición de la suma de conjuntos de Minkovski, la segunda supone aplicar la desigualdad antes descrita y finalmente aditividad antes demostrada.

□

2.1.2. Esfera n-dimensional con medida de Haar

La versión probabilística del problema la plantearemos en el espacio de la esfera n-dimensional junto con la medida de Haar, antes definida. Es uno de los ejemplos más importantes, de largo el más analizado y por ello el mejor conocido. Su importancia radica en el hecho de que la esfera contenga todos los movimientos del grupo ortogonal, y que además la medida de Haar sea invariante mediante dicho grupo. En esta sección emplearemos las ideas reflejadas en [13].

Nos restringimos a partir de ahora al estudio de nuestro caso más destacado: $(\mathbb{S}^{n-1}, \mathcal{F}, \sigma_{n-1}, d)$. La esfera unitaria n-dimensional como conjunto, por la riqueza que nos aportan los giros, la σ -álgebra de Borel junto con la medida normalizada de Haar (única medida de probabilidad en la esfera n-dimensional que es invariante bajo el grupo ortogonal) y la distancia geodésica. A partir de ahora, y aunque no lo especifiquemos, este será nuestro espacio de trabajo. A continuación vamos a definir el concepto de casquete esférico, que juega el papel de las bolas euclídeas en el espacio \mathbb{R}^n .

Definición 2.1.10. Llamaremos **casquete esférico** de centro a y radio r a una bola, del mismo centro y radio, en la distancia geodésica sobre \mathbb{S}^{n-1} . Es decir, al conjunto:

$$C(r, a) = \{x \in \mathbb{S}^{n-1} : d(x, a) \leq r\}$$

El siguiente teorema nos arroja la misma situación que en $(\mathbb{R}^n, \mathcal{B}, \mu_n)$, y nos afirma que precisamente son los casquetes esféricos los conjuntos que menos aumentan de medida cuando les añadimos sus ϵ -vecinos. Recordemos que empleamos la notación σ_{n-1} para la medida de Hausdorff $n-1$ dimensional sobre \mathbb{S}^{n-1} que, como hemos visto, coincide con la medida de Haar.

Teorema 2.1.11. Sea $A \subset \mathbb{S}^{n-1}$ un conjunto cerrado y sea $\epsilon > 0$. Tomamos $C = C(r, a) \subset \mathbb{S}^{n-1}$ un casquete esférico que cumpla $\sigma_{n-1}(A) = \sigma_{n-1}(C)$. Entonces:

$$\sigma_{n-1}(A_\epsilon) \geq \sigma_{n-1}(C_\epsilon) = \sigma_{n-1}(C(r + \epsilon, a))$$

La demostración de este teorema es algo más complicada, vamos a separarla en tres lemas que nos conducen a probar nuestro resultado. Comencemos introduciendo algunos conceptos que intervienen en las pruebas. Aportamos la noción de radio de un conjunto cualquiera.

Definición 2.1.12. Llamaremos **radio** de $A \subset \mathbb{S}^{n-1}$ al valor,

$$r(A) = \min\{r : \exists x \in \mathbb{S}^{n-1}, A \subset B(x, r)\}$$

A continuación, vamos a definir una métrica entre conjuntos cerrados de \mathbb{S}^{n-1} , conocida como la *métrica de Hausdorff*.

Definición 2.1.13. Sean $A, B \subset \mathbb{S}^{n-1}$ conjuntos cerrados no vacíos. Definimos la **distancia** entre A y B como,

$$\delta(A, B) = \min\{r : A \subset B_r, B \subset A_r\}$$

Observación 2.1.14. Comprobemos que efectivamente se trata de una métrica:

- $\delta(A, B) \geq 0$, $\delta(A, B) = 0 \iff A \subset B, B \subset A \iff A = B$
- Obvio que $\delta(A, B) = \delta(B, A)$
- Sea $\delta(A, C) = r$ y $\delta(C, B) = s$, vamos a probar que $\delta(A, B) \leq r + s$. Por definición,

$$A \subset C_r, C \subset A_r \text{ y } C \subset B_s, B \subset C_s$$

Luego se verifican los siguientes contenidos,

$$A \subset C_r \subset B_{r+s}, B \subset C_s \subset A_{r+s}$$

De lo que deducimos trivialmente que $\delta(A, B) \leq \delta(A, C) + \delta(C, B)$

Definición 2.1.15. Sean $x, -x \in \mathbb{S}^{n-1}$ dos puntos antipodales y consideremos $\gamma \subset \mathbb{S}^{n-1}$ la semicircunferencia que los une. Para todo $y \in \gamma$ definimos H^y como el único hiperplano de \mathbb{R}^n ortogonal a γ (con el producto escalar usual) que contiene al punto y . La intersección $H^y \cap \mathbb{S}^{n-1}$ genera una esfera $n-2$ dimensional que denotaremos por $\mathbb{S}^{n-2, y}$ y la dotaremos con la medida de Haar correspondiente a su dimensión. Para $A \subset \mathbb{S}^{n-1}$ denotaremos por $A^y = A \cap H^y$. Sea $B^y \subset \mathbb{S}^{n-2, y}$ una bola de centro y y con medida $\sigma_{n-2, y}(B^y) = \sigma_{n-2, y}(A^y)$. Llamamos **simetrización** del conjunto A , y lo denotaremos por $\sigma_\gamma(A)$, al conjunto

$$B = \bigcup_{y \in \gamma} B^y$$

Observación 2.1.16. *Se sigue del Teorema de Fubini que $\sigma_{n-1}(A) = \sigma_{n-1}(B)$*

La demostración se llevará a cabo mediante inducción. El caso $n=1$ es trivial, y para n es consecuencia de los siguientes lemas.

Lema 2.1.17. *Sea $A \subset \mathbb{S}^{n-1}$ un conjunto cerrado y sea $M(A) = \{C \subset \mathbb{S}^{n-1} : C \text{ cerrado, } \sigma_{n-1}(C) = \sigma_{n-1}(A), \sigma_{n-1}(C_\epsilon) \leq \sigma_{n-1}(A_\epsilon) \forall \epsilon > 0\}$. Entonces existe $B \in M(A)$ de mínimo radio.*

Demostración. Usando la continuidad de la función que a cada conjunto le asigna su radio, basta con demostrar que $M(A)$ es un subconjunto cerrado de \mathbb{S}^{n-1} . Tomamos una sucesión de conjuntos que converge a B con la métrica de Hausdorff, $B^k \in M(A)$, $k \in \mathbb{N}$, se tiene que $\delta(B^k, B) \rightarrow 0$ y sea $\epsilon \geq 0$. Para ver que es cerrado, demostremos que el límite de la sucesión pertenece a $M(A)$. Por la definición de la métrica, para todo $\eta > 0$ encontramos un valor k de modo que $B \subset B_\eta^k$ y por tanto, $B_\epsilon \subset B_{\epsilon+\eta}^k$. En consecuencia,

$$\sigma_{n-1}(B_\epsilon) \leq \sigma_{n-1}(B_{\epsilon+\eta}^k) \leq \sigma_{n-1}(A_{\epsilon+\eta})$$

Donde la segunda desigualdad proviene de que $B^k \in M(A)$ y es válida $\forall \eta > 0$, por lo que podemos tomar límite en $\eta \rightarrow 0$.

$$\sigma_{n-1}(B_\epsilon) \leq \inf_{\eta} \sigma_{n-1}(A_{\epsilon+\eta}) = \sigma_{n-1}\left(\bigcap_{\eta>0} A_{\epsilon+\eta}\right) = \sigma_{n-1}(A_\epsilon)$$

En particular, para $\epsilon = 0$ tenemos que $\sigma_{n-1}(B) \leq \sigma_{n-1}(A)$. Con un razonamiento análogo concluimos que $\sigma_{n-1}(B) \geq \limsup_k \sigma_{n-1}(B^k) = \sigma_{n-1}(A)$. Por lo tanto, $B \in M(A)$. \square

El siguiente lema nos dice que el simetrizado de cualquier conjunto A pertenece a $M(A)$.

Lema 2.1.18. *Sea $A \subset \mathbb{S}^{n-1}$ un conjunto cerrado. Entonces, para todo semicírculo γ , $\sigma_\gamma(A) \in M(A)$.*

Demostración. Sea A un conjunto cerrado de \mathbb{S}^{n-1} y sea γ un semicírculo de la esfera que une x_0 con $-x_0$. Denotamos por u al punto medio de dicho semicírculo e identificamos la esfera $\mathbb{S}^{n-2,u}$ con \mathbb{S}^{n-2} . Para todo $y \in \gamma$, $y \neq \pm x_0$ defino una aplicación $\tau_y : \mathbb{S}^{n-2,y} \rightarrow \mathbb{S}^{n-2,u}$ que proyecta los valores de una esfera a la otra. Es decir, $\tau_y(x)$ es el punto de $\mathbb{S}^{n-2,u}$ que pertenece al semicírculo que une x_0 , x , $-x_0$. Dicha operación se puede ver claramente reflejada en la Figura 2.1. Es obvio que existe una función f que si $x_1 \in \mathbb{S}^{n-2,y_1}$, $x_2 \in \mathbb{S}^{n-2,y_2}$, entonces,

$$d(x_1, x_2) = f(y_1, y_2, d(\tau_{y_1}(x_1), \tau_{y_2}(x_2)))$$

Es decir, podemos expresar la distancia entre esos dos valores de la esfera en términos de la distancia de y_1 a y_2 junto con la distancia entre sus respectivas proyecciones. Esto se puede interpretar mediante un Teorema Esférico de Pitágoras, donde la hipotenusa une x_1 con x_2 y los catetos son las curvas que unen las proyecciones entre sí y los valores y_1, y_2 .

Por lo tanto, $\forall y_1, y_2 \in \gamma$, $\epsilon > 0$ con $d(y_1, y_2) \leq \epsilon$ existe $\eta > 0$ de modo que para cualquier $C \subset \mathbb{S}^{n-2,y_1}$, $C_\epsilon^{y_2} = C_\epsilon \cap \mathbb{S}^{n-2,y_2} = \tau_{y_2}^{-1}((\tau_{y_1} C)_\eta)$. Es obvio que si los valores están más alejados de ϵ , esta intersección es vacía y nos impide el razonamiento.

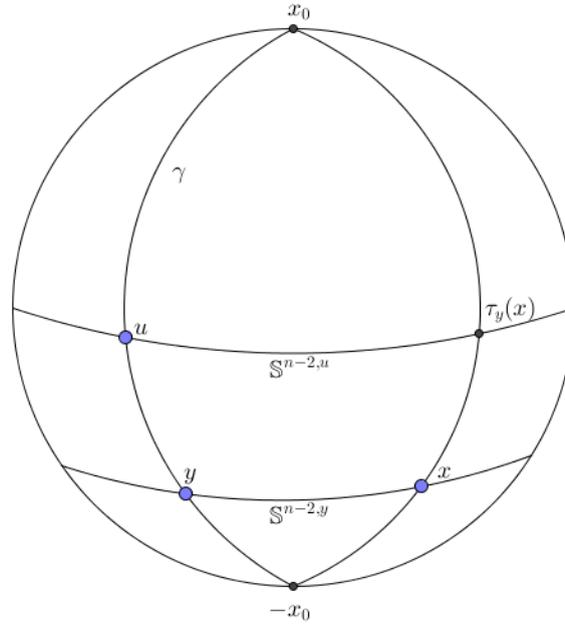


Figura 2.1: Representación gráfica de las distintas esferas. *Elaboración Propia*

En general, aplicando τ_{y_2} a ambos miembros y considerando $A \subset \mathbb{S}^{n-1}$, obtenemos,

$$\tau_y((A_\epsilon)^y) = \bigcup_{z \in \gamma, d(z,y) \leq \epsilon} (\tau_z A^z)_\eta \quad (2.1)$$

Tomo $B = \sigma_\gamma(A)$ la simetrización del conjunto A, y aplico la igualdad anterior,

$$\tau_y((B_\epsilon)^y) = \bigcup_{z \in \gamma, d(z,y) \leq \epsilon} (\tau_z B^z)_\eta$$

Para todo $z \in \gamma$ el conjunto $\tau_z B^z$ es, por definición, un casquete esférico de $\mathbb{S}^{n-2,z}$ para el cual $\sigma_{n-2}(\tau_z B^z) = \sigma_{n-2}(\tau_z A^z)$. Por la hipótesis de inducción,

$$\sigma_{n-2}((\tau_z B^z)_\eta) \leq \sigma_{n-2}((\tau_z A^z)_\eta)$$

Como todos los conjuntos de 2.1 son casquetes del mismo centro, obtenemos,

$$\sigma_{n-2}((\tau_y(B_\epsilon)^y)) = \sup_{z \in \gamma, d(z,y) \leq \epsilon} \sigma_{n-2}((\tau_z B^z)_\eta) \leq \sigma_{n-2}((\tau_z A^z)_\eta) \leq \sigma_{n-2}((\tau_y(A_\epsilon)^y))$$

Hemos probado que, $\forall y \neq \pm x_0 \in \gamma$

$$\sigma_{n-2,y}((B_\epsilon)^y) \leq \sigma_{n-2,y}((A_\epsilon)^y)$$

Aplicando el teorema de Fubini concluimos que $\sigma_{n-1}(B_\epsilon) \leq \sigma_{n-1}(A_\epsilon)$, y por tanto, $\sigma_y(A) \in M(A)$ \square

Lema 2.1.19. *Sea $B \subset \mathbb{S}^{n-1}$ un conjunto cerrado que no es un casquete esférico. Entonces existen un número finito de semicírculos $\{\gamma_i\}_{i=1}^n$ de modo que $r(\sigma_{\gamma_n}(\sigma_{\gamma_{n-1}} \dots \sigma_{\gamma_1}(B))) < r(B)$.*

Demostración. En esta prueba intervienen conceptos que escapan al objetivo del presente trabajo, para ver los detalles se puede consultar [13, pág.92]. □

Tal y como se plantea, el objetivo es probar que existe un casquete esférico $C(a, r) \in M(A)$, para todo $A \subset \mathbb{S}^{n-1}$. El Lema 2.1.17 nos habla de la existencia de un conjunto en $M(A)$ de mínimo radio, que a la postre, será nuestro casquete esférico. El Lema 2.1.19 nos dice precisamente que dicho conjunto debe ser un casquete, pues al contrario, existiría un número finito de simetrizaciones de menor radio e igual medida, que por el Lema 2.1.18 pertenece a $M(A)$, llegando a una contradicción pues partíamos de un conjunto de mínimo radio.

Como es evidente, $C(a, r)_\epsilon = C(a, r + \epsilon)$ y son precisamente dichos conjuntos, los casquetes esféricos los que determinan una cota para los ϵ -vecinos de cualquier $A \subset \mathbb{S}^{n-1}$. Por lo tanto, es de considerable utilidad conocer la medida de Haar de las bolas en la distancia geodésica. Omitiremos los detalles de las operaciones, que nos conducen a la expresión:

$$\sigma_{n-1}(C(a, r)) = \rho_n \int_{-\pi/2}^{r-\pi/2} \cos^{n-2} t dt, \quad \rho_n^{-1} = \int_{-\pi/2}^{\pi/2} \cos^{n-2} t dt = \sqrt{\pi} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})}$$

Si consideramos un hemisferio de la esfera, dado por $C(x_0, \frac{\pi}{2})$, es claro que su medida es $\frac{1}{2}$. Por otra parte, como hemos mencionado, el conjunto de ϵ -vecinos viene dado por $C(x_0, \frac{\pi}{2} + \epsilon)$, cuya medida es:

$$\sigma_{n-1}(C(x_0, \frac{\pi}{2} + \epsilon)) = \rho_n \int_{-\pi/2}^{\epsilon} \cos^{n-2} t dt = \rho_n \left(\int_{-\pi/2}^0 \cos^{n-2} t dt + \int_0^{\epsilon} \cos^{n-2} t dt \right)$$

Por la simetría de la función $\cos^{n-2}(x)$, es claro que $\rho_n \int_{-\pi/2}^0 \cos^{n-2} t dt = \frac{1}{2}$

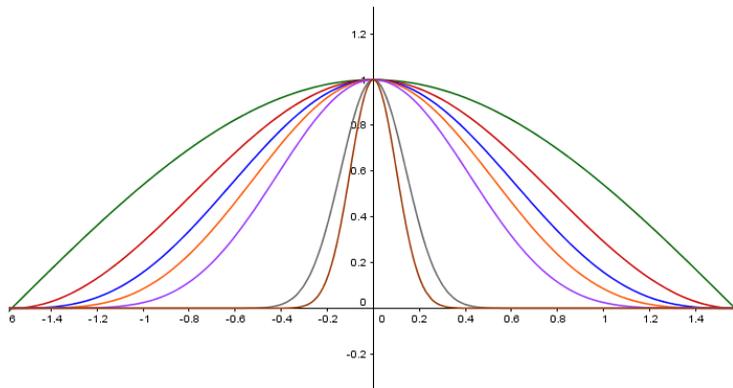


Figura 2.2: $f_n(x) = \cos^n(x)$, $n = 1, 2, 3, 4, 5, 50, 100$. *Elaboración Propia.*

Atendiendo a la Figura 2.2, es evidente que el área de las gráficas de $f_n(x) = \cos^n(x)$ se acumula, conforme n crece, a una cierta banda $(-\epsilon, \epsilon)$. Por tanto, para un ϵ fijo, el valor de $\int_0^\epsilon \cos^{n-2} t dt \rightarrow \int_0^{\pi/2} \cos^{n-2} t dt$ si $n \rightarrow \infty$. Con ello, $\rho_n \int_0^\epsilon \cos^{n-2} t dt \rightarrow \frac{1}{2}$. Y podemos afirmar, de manera intuitiva y sin concretarlo analíticamente, que la medida de $\mu_{n-1}(C(x_0, \frac{\pi}{2} + \epsilon))$ se acerca a 1. El teorema anterior nos dice precisamente que para cualquier $A \subset \mathbb{S}^{n-1}$, el valor $\sigma_{n-1}(A_\epsilon)$ está acotado inferiormente por $\sigma_{n-1}(C(x_0, \frac{\pi}{2} + \epsilon))$, que se acerca a 1. Con este comentario, hemos querido dejar entrever un resultado que será formalizado a final de este capítulo, en el Corolario 2.2.6

2.1.3. Espacio Euclídeo con medida Gaussiana

En esta sección trataremos de llegar a un resultado similar, pero con la medida Gaussiana definida sobre \mathbb{R}^n . Seguiremos [4] como base para los resultados.

Definición 2.1.20. Llamaremos semiespacio de punto α al conjunto $H(\alpha) = \{(\xi_1, \dots, \xi_n) : \xi_1 \leq \alpha\}$, $\alpha \in \mathbb{R}$.

Para el siguiente resultado tenemos una pequeña variación de lo que se presentaba hasta ahora, ya no son las bolas, con una cierta métrica, aquellos conjuntos que tienen la propiedad de ser los que menos aumentan de medida al añadirles los ϵ -vecinos, sino que ahora son los conocidos como semiespacios.

Teorema 2.1.21. Sea $A \subset \mathbb{R}^n$ un conjunto cerrado y $\epsilon > 0$. Sea $H = H(\alpha) \subset \mathbb{R}^n$ un semiespacio que verifica $\gamma_n(A) = \gamma_n(H)$, entonces,

$$\gamma_n(A_\epsilon) \geq \gamma_n(H_\epsilon(\alpha)) = \gamma_n(H(\alpha + \epsilon))$$

Demostración. La demostración se basa en las siguientes ideas:

- Mediante el Teorema 1.4.1 y su posterior comentario, la medida de Gauss se identifica como límite de la proyección de la medida de Haar, reduciendo así el problema al caso ya probado.
- Los conjuntos cuyas antiimágenes producen casquetes esféricos son precisamente los del tipo $H = H(\alpha)$.

Los detalles pueden consultarse en [4, pág. 61]

□

El siguiente resultado nos muestra como aumenta exponencialmente la medida, en términos Gaussianos, del semiespacio $H(t)$ cuando t crece. No es de extrañar, pues sabemos cómo es la distribución Gaussiana, que condensa la medida en una esfera en torno al origen, pero sí que es digno de destacar la rapidez con la que aumenta.

Lema 2.1.22. Sea el semiespacio $H = H(t)$, $t \geq 0$. Entonces,

$$\gamma_n(H) \geq 1 - e^{-t^2/2}$$

Demostración. Empleamos el método de la transformada de Laplace, $\forall \lambda > 0$

$$\begin{aligned} \gamma_n\{x : \xi_1 \geq t\} &\leq e^{-\lambda t} E[e^{\lambda x}] = e^{-\lambda t} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\lambda \xi} e^{-\xi^2/2} d\xi = e^{-\lambda t} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(\xi/\sqrt{2} - \sqrt{2}\lambda/2)^2} e^{\lambda^2/2} d\xi \\ &= e^{-\lambda t + \lambda^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(\xi/\sqrt{2} - \sqrt{2}\lambda/2)^2} d\xi = e^{-\lambda t + \lambda^2/2} \end{aligned}$$

Tomamos $\lambda = t$ y obtenemos así la acotación deseada.

□

2.2. Desigualdades Isoperimétricas

Nos quedamos con \mathbb{S}^{n-1} dotada con la medida de Haar como área de trabajo. En esta sección nos encargaremos de, dado un conjunto A cualquiera, encontrar una cota inferior para el valor de la medida de sus ϵ -vecinos. En esta dirección, el objetivo del apartado es demostrar, de modo constructivo, el teorema conocido por Desigualdad Isoperimétrica. Emplearemos [31] para guiarnos en la prueba. Muchos autores expresan este tipo de desigualdades isoperimétricas sólo para conjuntos de medida mayor a $\frac{1}{2}$ y a menudo varían las constantes que lo acotan, sin embargo, la idea que se persigue es siempre la misma: dado un conjunto medible, aportar una cota inferior sobre la medida de sus ϵ -vecinos.

Teorema 2.2.1. (Desigualdad Isoperimétrica) *Sea $A \subset \mathbb{S}^{n-1}$, $\epsilon > 0$. Entonces,*

$$\sigma_{n-1}(A_\epsilon) > 1 - 2\sigma_{n-1}(A)^{-1}e^{-n\epsilon^2/16}$$

Para probarlo, necesitaremos la siguiente definición:

Definición 2.2.2. *Sea $(X, \|\cdot\|)$ un espacio normado. Llamamos **módulo de convexidad** al valor:*

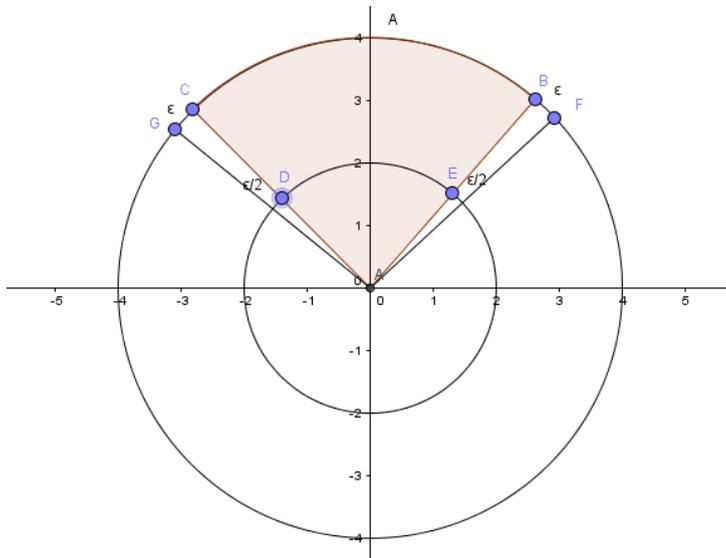
$$\delta(\epsilon) = \inf\{1 - \frac{1}{2}\|x + y\|; \|x\|, \|y\| \leq 1, \|x - y\| \geq \epsilon\}$$

Este valor mide cuánto de la curvatura de la esfera es debido a la convexidad del espacio en el que vive. Es un valor entre 0 y 1, que cuanto más cercano a 1 significa que los segmentos de gran longitud permanecen poco curvados en el cuerpo convexo.

Consideramos $\|\cdot\|$ una norma cualquiera en el espacio $(\mathbb{R}^n, \mathcal{B}_n, \mu_n)$ y la esfera unitaria S con la métrica $d(x, y) = \|x - y\|$. Además, considero la medida de Borel de probabilidad,

$$\lambda(A) = \frac{\mu_n\{tA, 0 \leq t \leq 1\}}{\mu_n\{x, \|x\| \leq 1\}}$$

Que supone asignarle al conjunto A la medida, normalizada en la bola unidad de la norma, del conjunto proyectado desde la frontera hasta el centro del espacio. Sea $K = \{x : \|x\| \leq 1\}$ y consideremos el conjunto $A \subset S$ sobre la superficie de la esfera el delimitado por los puntos C y B en la figura 2.3. De acuerdo con la expresión antes descrita, su medida $\lambda(A)$ será igual a la parte sombreada de la figura de manera normalizada. Además, llamaremos $A^* = \{tA : \frac{1}{2} \leq t \leq 1\}$, que en nuestra figura es la corona circular delimitada por los puntos $CBED$ y sobre dicho conjunto volvemos a considerar la medida de Lebesgue normalizada por la bola unidad, que denotaremos por $\mathbf{v}(A^*)$. Nótese que orlar el conjunto A considerando sus ϵ -vecinos supone extender dicho conjunto hasta los puntos GF , manteniéndose en la superficie de la esfera. Asimismo, podemos considerar el conjunto $(A_\epsilon)^*$, que corresponde a la corona circular orlada por ϵ . Véase que por las longitudes de la circunferencia, el conjunto A^* debe ser “extendido” $\epsilon/2$ para que $\mathbf{v}(A_{\epsilon/2}^*)$ pueda estar acotado superiormente por $\lambda(A_\epsilon)$. Debemos también considerar que $A_{\epsilon/2}^*$ no queda enteramente contenido en K , sin embargo, con un ϵ suficientemente pequeño, la medida del conjunto que escapa es más pequeña que la medida del conjunto delimitado por ADE , manteniendo así la acotación $\lambda(A_\epsilon) \geq \mathbf{v}(A_{\epsilon/2}^*)$.


 Figura 2.3: Representación de los conjuntos en S

Teorema 2.2.3. Sea $(\mathbb{R}^n, \|\cdot\|)$ espacio normado y sea δ su módulo de convexidad. Entonces para cualquier conjunto medible Borel $A \subset S$, y $\epsilon > 0$ se tiene,

$$\lambda(A_\epsilon) > 1 - 2\lambda(A)^{-1}e^{-2n\delta(\epsilon/2)}$$

Demostración. Con la notación antes descrita, vamos a probar que para cualquier $A \subset S$ se tiene que $\mathbf{v}(A_\epsilon^*) > 1 - \mathbf{v}(A^*)^{-1}e^{-2n\delta(\epsilon)}$. Empleemos ahora el conjunto $C = \{x \in K : d(x, B) \geq \epsilon\}$. Es obvio que si $x \in B, y \in C$ entonces $\|\frac{x+y}{2}\| \leq 1 - \delta(\epsilon)$, pues $\delta(\epsilon)$ recordemos que es el ínfimo de los valores $\{1 - \|\frac{x+y}{2}\|\}$ con x e y verificando $\|x\|, \|y\| \leq 1, \|x - y\| \geq \epsilon$ condiciones que se satisfacen por la definición de los conjuntos. Con dicha afirmación concluimos que se da la siguiente inclusión:

$$\frac{B + C}{2} \subset (1 - \delta(\epsilon))K$$

Empleemos la desigualdad de Brunn-Minkovski vista en Teorema 2.1.5, tomando $\lambda = \frac{1}{2}$ y los conjuntos B, C . Entonces,

$$\mathbf{v}(C)^{\frac{1}{2}}\mathbf{v}(B)^{\frac{1}{2}} \leq \mathbf{v}\left(\frac{1}{2}C + \frac{1}{2}B\right) = \mathbf{v}\left(\frac{C + B}{2}\right) \leq (1 - \delta(\epsilon))^n$$

Donde la última desigualdad se obtiene usando el contenido previo. Finalmente, empleando la desigualdad $(1 - x)^n \leq e^{-nx}, x \in [0, 1]$ deducimos que,

$$\mathbf{v}(C)\mathbf{v}(B) \leq (1 - \delta(\epsilon))^{2n} \leq e^{-2n\delta(\epsilon)}$$

Nótese que por definición del conjunto C , $\mathbf{v}(C) = 1 - \mathbf{v}(B_\epsilon)$. Por tanto, despejando la expresión, hemos probado que $\mathbf{v}(B_\epsilon) > 1 - \mathbf{v}(B)^{-1}e^{-2n\delta(\epsilon)}$.

Finalmente nos conduce a la expresión del enunciado para cualquier $A \subset S$,

$$\lambda(A_\epsilon) \geq \mathbf{v}(A_{\epsilon/2}^*) > 1 - \frac{e^{-2n\delta(\epsilon/2)}}{\mathbf{v}(A^*)} \geq 1 - 2\frac{e^{-2n\delta(\epsilon/2)}}{\lambda(A^*)}$$

Donde hemos empleado $\lambda(A) \leq 2\mathbf{v}(A^*)$

□

El siguiente lema nos aporta una cota para el módulo de convexidad en Espacios de Hilbert.

Lema 2.2.4. *Para cualquier Espacio de Hilbert, se cumple que $\delta(\epsilon) \geq \epsilon^2/8$. En particular, se verifica para $(\mathbb{R}^n, \|\cdot\|)$ con la norma Euclídea usual.*

Demostración. Empleamos la identidad del paralelogramo. Si consideramos, para todo $\epsilon > 0$ el conjunto $\{x, y : \|x\|, \|y\| \leq 1, \|x - y\| \geq \epsilon\}$, entonces:

$$\left\| \frac{x+y}{2} \right\|^2 = 2 \left\| \frac{x}{2} \right\|^2 + 2 \left\| \frac{y}{2} \right\|^2 - \left\| \frac{x-y}{2} \right\|^2 \leq 1 - \frac{\epsilon^2}{4}$$

Tomando ínfimos en dicho conjunto, deducimos,

$$\delta(\epsilon) = 1 - \sqrt{1 - \frac{\epsilon^2}{4}} \geq \frac{\epsilon^2}{8}$$

Para la última desigualdad hemos empleado el siguiente desarrollo de Taylor

$$\sqrt{1+x} = 1 + \frac{x}{2} - \frac{x^2}{8} + \frac{x^3}{16} + \dots \leq 1 + \frac{x}{2}, \quad x \in [-1, 0]$$

□

El anterior teorema junto con este lema nos aporta una demostración de la desigualdad isoperimétrica, Teorema 2.2.1, objetivo de esta sección.

Presentamos el siguiente resultado, que además servirá de motivación para el próximo capítulo, sobre fenómenos de concentración de la medida. Seguimos tomando la esfera n -dimensional como espacio de trabajo. El resultado nos muestra que si tomamos un conjunto de la esfera cuya medida sea $\frac{1}{2}$, por ejemplo un hemisferio, la medida de los ϵ -vecinos será cercana a 1 si n es lo suficientemente grande. Dicho de otra forma, la esfera n dimensional concentra toda su medida en torno al ecuador cuando n es grande. Aunque el resultado se puede refinar y hacer que la acotación sea más precisa, asimismo nos vale para llegar a nuestra conclusión

Corolario 2.2.5. *Sea $A \subset \mathbb{S}^{n-1}$ que cumpla $\mu(A) = \frac{1}{2}$. Entonces $\mu(A_\epsilon) > 1 - 4e^{-n\epsilon^2/16}$*

Demostración. Directo de lo anterior. □

Además este resultado tiene una interpretación geométrica muy clara. Podemos tomar el conjunto A como un hemisferio de la esfera, entonces la medida del conjunto de ϵ -vecinos de A se acerca a 1 cuando n es grande. En otras palabras, cuando la dimensión de la esfera crece, casi toda su medida se acumula en torno al ecuador, en una franja de altura ϵ .

Observación 2.2.6. *El corolario anterior puede tener una expresión más fina todavía. Se puede demostrar que si $A \subset \mathbb{S}^{n-1}$ de modo que $\mu(A) \geq \frac{1}{2}$ entonces $\mu(A_\epsilon) \geq 1 - 2e^{-n\epsilon^2/2}$. Sin embargo, requiere una prueba con mayor detalle que puede ser consultada en [23].*

Capítulo 3

Concentración de la medida

La concentración de la medida es un fenómeno, también llamado observación, por el cual funciones Lipschitz o variables aleatorias aglutinan sus valores cercanos a sus medianas o medias respectivamente. Tanto se acercan los valores a dichos parámetros, que incluso podemos hablar de que son casi constantes. Suele ser un fenómeno que ocurre conforme subimos la dimensión del espacio.

Uno de los ejemplos más importantes es el mencionado en el final del capítulo anterior. La esfera unitaria de un espacio euclídeo condensa su medida cerca del ecuador si la dimensión es lo suficientemente grande. En este capítulo estudiaremos la relación que existe entre dichos fenómenos con la existencia de Desigualdades isoperimétricas y además probaremos importantes resultados de concentración sobre funciones Lipschitz o variables aleatorias, que se convertirán en los actores principales del presente capítulo.

Remarquemos las palabras que emplea Michel Talagrand para referirse a los fenómenos de concentración de la medida: *“Un aspecto probabilístico de concentración de la medida es que una variable aleatoria que depende (en un modo suave, Lipschitz) sobre la influencia de ciertas variables independientes (pero no demasiado de alguna de ellas) es esencialmente constante”*

3.1. Desigualdades Isoperimétricas y Concentración de la medida

Vamos a comenzar nuestro estudio discutiendo la relación entre los fenómenos de concentración de la medida y la existencia de desigualdades isoperimétricas. En efecto, veremos que de una cierta manera, se caracterizan mutuamente. Los fenómenos de concentración serán tratados mediante funciones 1-Lipschitz cuyos valores se aglutinan cerca del valor de su mediana. Nuestro espacio de trabajo será un espacio de medida cualquiera, a saber $(\Omega, \mathcal{F}, \mu)$. En este apartado seguiremos el hilo conductor marcado en [31], donde aporta ligeras indicaciones para las demostraciones que han sido desarrolladas en el presente trabajo.

Veamos primeramente que la existencia de desigualdades isoperimétricas implica que las funciones 1-Lipschitz se concentran en torno a su mediana. Mantendremos la siguiente notación sobre funciones Lipschitz y su mediana: consideramos m_f la única mediana de la función f y definimos los conjuntos,

$$A^+ = \{x : f(x) \geq m_f\}, A^- = \{x : f(x) \leq m_f\}$$

Por la propia definición de mediana se verifica, $\mu(A^+) \geq \frac{1}{2}$, $\mu(A^-) \geq \frac{1}{2}$. Para simplificar notación usaremos $(A^+)_\epsilon = A_\epsilon^+$

Proposición 3.1.1. *Sea $f : \Omega \rightarrow \mathbb{R}$ una función 1-Lipschitz. Si existe $\eta > 0$ de modo que para todo $B \subset \Omega$ con $\mu(B) \geq \frac{1}{2}$ se tiene que $\mu(B_\epsilon) \geq 1 - \frac{\eta}{2}$. Entonces,*

$$\mu(\{x : |f(x) - m_f| \geq \epsilon\}) \leq \eta$$

Demostración. Podemos aplicar la hipótesis a los conjuntos A^+ , A^- y resulta,

$$\mu(A_\epsilon^+) \geq 1 - \frac{\eta}{2}, \quad \mu(A_\epsilon^-) \geq 1 - \frac{\eta}{2}$$

Es obvio que $|f(x) - m_f| \geq \epsilon, \forall x \in (A_\epsilon^+ \cap A_{-\epsilon}^-)^c$ y finalmente calculemos la medida de ese conjunto,

$$\mu((A_\epsilon^+ \cap A_\epsilon^-)^c) = \mu(A_\epsilon^{+c} \cup A_\epsilon^{-c}) = \mu(A_\epsilon^{+c}) + \mu(A_\epsilon^{-c}) - \mu(A_\epsilon^{+c} \cap A_\epsilon^{-c}) < \frac{\eta}{2} + \frac{\eta}{2} = \eta$$

$$\text{Por tanto, } \mu(\{x : |f(x) - m_f| \geq \epsilon\}) \leq \eta$$

□

El siguiente resultado supone un recíproco del anterior. Nos dice que si trabajamos en un espacio de medida donde todas las funciones 1-Lipschitz se concentran cerca de la mediana, entonces podremos deducir desigualdades isoperimétricas sobre conjuntos de medida mayor o igual a $\frac{1}{2}$.

Proposición 3.1.2. *$\forall \epsilon > 0, \exists \eta > 0$ de modo que si se cumple $\mu(\{x : |f(x) - m_f| \geq \epsilon\}) \leq \eta$, para cualquier función 1-Lipschitz, entonces $\forall A \subset \Omega$ que cumpla $\mu(A) \geq \frac{1}{2}$, se tiene,*

$$\mu(A_\epsilon) \geq 1 - \eta$$

Demostración. Tomamos $A \subset \Omega$ de modo que $\mu(A) \geq \frac{1}{2}$ y sea f la función $f(x) = d(x, A)$ distancia al conjunto A . Veamos en primer lugar que efectivamente es 1-Lipschitz, debemos comprobar que $|d(x, A) - d(y, A)| \leq d(x, y), \forall x, y \in \Omega$. Separando el valor absoluto, el problema se traduce en,

$$\begin{cases} d(x, A) - d(y, A) \leq d(x, y), \forall x, y \in \Omega \\ d(y, A) - d(x, A) \leq d(x, y) = d(y, x), \forall x, y \in \Omega \end{cases}$$

Podemos reducirlo, por simetría, a demostrar la primera desigualdad. Aplicamos la definición de ínfimo en $d(x, A)$ y para todo $\epsilon > 0$ encontramos un $z \in A$ que se aproxima a $d(x, A)$ tanto como queramos. Es decir,

$$\forall \epsilon > 0 \exists z \in A, d(y, z) \leq d(y, A) + \epsilon$$

Empleando esta propiedad y utilizando la desigualdad triangular,

$$d(x, A) \leq d(x, z) \leq d(x, y) + d(y, z) \leq d(x, y) + d(y, A) + \epsilon$$

Tomando límite cuando $\epsilon \rightarrow 0$, obtenemos la desigualdad buscada. Concluimos que $f(x) = d(x, A)$ es 1-Lipschitz, y podemos aplicarle la hipótesis del enunciado.

Lo siguiente será probar que 0 es mediana de la función $f(x) = d(x, A)$ Efectivamente consideramos los conjuntos,

$$\hat{A}^+ = \{x : f(x) \geq 0\} = \{x : d(x, A) \geq 0\} = \Omega$$

$$\hat{A}^- = \{x : f(x) \leq 0\} = \{x : d(x, A) = 0\} \supseteq A$$

Cuya medida evidentemente es mayor o igual a $\frac{1}{2}$,

$$\mu(\hat{A}^-) \geq \mu(A) \geq \frac{1}{2}, \quad \mu(\hat{A}^+) = \mu(\Omega) = 1 \geq \frac{1}{2}$$

Finalmente aplicamos la hipótesis del enunciado a la función f con mediana 0, el conjunto,

$$\mu(A_\epsilon) = \mu(\{x \in \Omega : d(x, A) \leq \epsilon\}) = 1 - \mu(\{x \in \Omega : f(x) > \epsilon\}) \geq 1 - \mu(\{x \in \Omega : f(x) \geq \epsilon\}) \geq 1 - \eta$$

□

El siguiente resultado es un caso de concentración de medida conocido por *Lema de Levy*. Lo presentamos en esta sección por ser su “lugar natural”, sin embargo, cobrará gran importancia en el siguiente capítulo. Aportamos la versión dada en [13].

Lema 3.1.3. (Levy) Sea $f : \mathbb{S}^{n-1} \rightarrow \mathbb{R}$. Sea M_f mediana de f y definimos $B = \{x : f(x) = M_f\}$. Entonces $\forall \epsilon > 0$,

$$\mu_{n-1}(B_\epsilon) \geq 2\gamma_n \int_0^\epsilon \cos^{n-2} t dt$$

Demostración. Para todo $\epsilon > 0$ se cumple $A_{+\epsilon} \cap A_{-\epsilon} = B_\epsilon$ Aplicando la desigualdad isoperimétrica del Teorema 2.1.11 y su comentario posterior,

$$\begin{aligned} \mu_{n-1}(B_\epsilon^c) &= \mu_{n-1}(A_{+\epsilon}^c) + \mu_{n-1}(A_{-\epsilon}^c) = 2 - \mu_{n-1}(A_{+\epsilon}) - \mu_{n-1}(A_{-\epsilon}) \\ &\leq 2 \left(1 - \rho_n \int_{-\pi/2}^\epsilon \cos^{n-2} t dt \right) \end{aligned}$$

De donde deducimos que,

$$\mu_{n-1}(B_\epsilon) \geq 2\rho_n \int_{-\pi/2}^\epsilon \cos^{n-2} t dt - 1 = 2\rho_n \left(\int_{-\pi/2}^0 \cos^{n-2} t dt + \int_0^\epsilon \cos^{n-2} t dt \right) - 1$$

La simetría de $\cos(x)$ nos permite concluir que $\rho_n \int_{-\pi/2}^0 \cos^{n-2} t dt = \frac{1}{2}$, de lo que se sigue el resultado. □

3.2. Ejemplos de concentración: Variables aleatorias

Presentamos ahora un resultado clásico de concentración de la medida, la Desigualdad de Tchebychev. Nos afirma que una variable aleatoria se concentra en torno a su media de forma tan ajustada como dicte la varianza. Otras formulaciones del teorema pueden verse en [35].

Teorema 3.2.1. (Desigualdad de Tchebychev) Sea X una variable aleatoria con esperanza matemática finita y sea $\epsilon > 0$. Entonces,

1. $P(|X - E[X]| \geq \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}$
2. $P(|X| \geq a) \leq \frac{E[|X|^k]}{a^k}$, $a, k > 0$

Demostración. Sea Z una variable aleatoria no negativa cualquiera, con $E[Z] = \mu$ finita y $a \in \mathbb{R}$. Empleamos la notación I_A para la función indicador en el conjunto A . Entonces,

$$E[Z] = E[ZI_{Z \geq a}] + E[ZI_{Z < a}] \geq E[ZI_{Z \geq a}] \geq E[aI_{Z \geq a}] = aP(Z \geq a)$$

Tomando $Z = (X - E[X])^2$, $a = \epsilon^2$, obtenemos (1),

$$P[|X - E[X]| \geq \epsilon] = P[(X - E[X])^2 \geq \epsilon^2] \leq \frac{E[(X - E[X])^2]}{\epsilon^2} = \frac{Var(X)}{\epsilon^2}$$

Para obtener (2),

$$P[|X| \geq a] = P[|X|^k \geq a^k] = P[|X^k| \geq a^k] \leq \frac{E[|X^k|]}{a^k}$$

□

A continuación vamos a tratar un fenómeno de concentración de la medida, que aunque ya de por sí tiene gran interés, lo emplearemos en el siguiente capítulo. Podemos encontrar su formulación original en [8]. Retomamos nuestro espacio habitual, la esfera \mathbb{S}^d ahora mediante vectores aleatorios normalizados y consideremos X_1, \dots, X_d variables independientes e idénticamente distribuidas, $X_i \sim N(0, 1)$. Con ellas formamos el vector aleatorio:

$$Y = \frac{1}{\|X\|_2} (X_1, \dots, X_d)$$

donde $X = (X_1, \dots, X_d)$. Es obvio que el vector Y pertenece a la esfera d -dimensional, pues tiene norma 1. Es decir, el vector aleatorio Y representa un punto de \mathbb{S}^{d-1} escogido de manera aleatoria y uniforme. Consideramos ahora el vector aleatorio $Z \in \mathbb{R}^k$ que surge de proyectar Y sobre sus k primeras coordenadas, es decir,

$$Z = \frac{1}{\|X\|_2} (X_1, \dots, X_k)$$

Definimos $L := \|Z\|_2^2$. Con esta notación, resulta que

$$L = \frac{X_1^2 + \dots + X_k^2}{X_1^2 + \dots + X_d^2}$$

Mediante un argumento de simetría, podemos comprobar que $E[L] = \frac{k}{d}$. Efectivamente,

$$E[L] = E \left[\frac{X_1^2 + \dots + X_k^2}{X_1^2 + \dots + X_d^2} \right] = E \left[\frac{X_1^2}{X_1^2 + \dots + X_d^2} \right] + \dots + E \left[\frac{X_k^2}{X_1^2 + \dots + X_d^2} \right] = kE \left[\frac{X_1^2}{X_1^2 + \dots + X_d^2} \right]$$

Donde la esperanza de dicho término podemos calcularla de manera sencilla,

$$1 = E \left[\frac{X_1^2 + \dots + X_d^2}{X_1^2 + \dots + X_d^2} \right] = E \left[\frac{X_1^2}{X_1^2 + \dots + X_d^2} \right] + \dots + E \left[\frac{X_d^2}{X_1^2 + \dots + X_d^2} \right] = dE \left[\frac{X_1^2}{X_1^2 + \dots + X_d^2} \right]$$

El siguiente resultado nos muestra que la variable L se concentra sobre el valor de su media. Para ello necesitaremos un lema técnico sobre variables normales standard

Lema 3.2.2. Si $X \sim N(0, 1)$, entonces $E[e^{sX^2}] = \frac{1}{\sqrt{1-2s}}$, $s \in (-\infty, \frac{1}{2})$

Demostración. Efectivamente,

$$E[e^{sX^2}] = \int_{-\infty}^{\infty} e^{sx^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{x^2(s-\frac{1}{2})} dx$$

Realizamos el cambio de variable, $x^2(s - \frac{1}{2}) = -t$. Derivando, obtenemos: $2xdx = \frac{dt}{\frac{1}{2}-s}$ y finalmente, $dx = \frac{1}{2} \frac{dt}{\sqrt{t(\frac{1}{2}-s)}}$, con $s \in (-\infty, \frac{1}{2})$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{x^2(s-\frac{1}{2})} dx = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\frac{1}{2}-s}} \int_0^{\infty} e^{-t} t^{-\frac{1}{2}} = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\frac{1}{2}-s}} \Gamma\left(\frac{1}{2}\right) = \frac{1}{\sqrt{\frac{1}{2}-s}}$$

□

Pasemos ya a enunciar el lema que nos afirma que la variable L se concentra en torno a su media.

Proposición 3.2.3. *Con la notación anterior, sea $k < d$:*

1. Si $\beta < 1$, entonces: $Pr[L \leq \beta \frac{k}{d}] \leq \beta^{k/2} (1 + \frac{(1-\beta)k}{d-k})^{\frac{d-k}{2}}$
2. Si $\beta > 1$, entonces: $Pr[L \geq \beta \frac{k}{d}] \leq \beta^{k/2} (1 + \frac{(1-\beta)k}{d-k})^{\frac{d-k}{2}}$

Demostración. Reescribimos la variable L en términos de X_1, \dots, X_d y estimamos la probabilidad del enunciado:

$$Pr[d(X_1^2 + \dots + X_k^2) \leq k\beta(X_1^2 + \dots + X_d^2)] = Pr[k\beta(X_1^2 + \dots + X_d^2) - d(X_1^2 + \dots + X_k^2) \geq 0]$$

A continuación aplicamos la función exponencial a ambos miembros,

$$= Pr[e^{k\beta(X_1^2 + \dots + X_d^2) - d(X_1^2 + \dots + X_k^2)} \geq 1] \quad \forall t \geq 0$$

Acotamos dicha probabilidad mediante la versión de la Desigualdad de Tchebychev (2) con los valores $a=k=1$

$$\leq E[e^{k\beta(X_1^2 + \dots + X_d^2) - d(X_1^2 + \dots + X_k^2)}]$$

Separamos la exponencial de una suma como producto de exponenciales y agrupamos por variables aleatorias

$$= E[e^{X_1^2(tk\beta-d)} \dots e^{X_k^2(tk\beta-d)} e^{X_{k+1}^2(tk\beta)} \dots e^{X_d^2(tk\beta)}]$$

Empleando la hipótesis de independencia de las variables podemos expresar la esperanza del producto como producto de esperanzas:

$$= E[e^{X_1^2(tk\beta-d)}] \dots E[e^{X_k^2(tk\beta-d)}] E[e^{X_{k+1}^2(tk\beta)}] \dots E[e^{X_d^2(tk\beta)}]$$

Usamos que todas las variables están idénticamente distribuidas, y por tanto, los mismos momentos. Sea $X \sim N(0, 1)$:

$$= E[e^{tk\beta X^2}]^{d-k} E[e^{t(k\beta-d)X^2}]^k$$

Finalmente empleamos el lema anteriormente demostrado,

$$= (1 - 2tk\beta)^{-\frac{d-k}{2}} (1 - 2t(k\beta - d))^{-\frac{k}{2}}$$

Definimos la función $g(t) = (1 - 2tk\beta)^{-\frac{d-k}{2}} (1 - 2t(k\beta - d))^{-\frac{k}{2}}$, con lo cual hemos probado que $\forall t \geq 0$,

$$\Pr[L \leq \beta \frac{k}{d}] \leq g(t)$$

Con la intención de que la cota sea la menor posible, nos planteamos el objetivo de buscar el valor $t_0 \geq 0$ que minimice la función g . Nótese que en dicha expresión aparecen los exponentes $-\frac{k}{2}, -\frac{d-k}{2}$ donde $k, d - k$ son valores enteros positivos. Para que tenga sentido “la raíz”, los valores $1 - 2tk\beta$ y $1 - 2t(k\beta - d)$ deben ser estrictamente positivos. Es decir,

$$\begin{cases} 1 - 2tk\beta > 0 \Rightarrow tk\beta < \frac{1}{2} \Rightarrow 0 < t < \frac{1}{2k\beta} \\ 1 - 2t(k\beta - d) > 0 \Rightarrow t(k\beta - d) < \frac{1}{2} \Rightarrow 0 < t < \frac{1}{2(k\beta - d)}, \text{ si } k\beta \neq d \end{cases}$$

Apreciamos que la primera restricción domina a la segunda, pues $k\beta > k\beta - d$. Por comodidad, en lugar de minimizar $g(t)$, maximizaremos $f(t) = g(t)^{-2}$. Nuestro problema se ha simplificado en:

$$\begin{cases} \max (1 - 2tk\beta)^{d-k} (1 - 2t(k\beta - d))^k \\ \text{s.a. } 0 < t < \frac{1}{2k\beta} \end{cases}$$

Que resulta ser un sencillo problema de análisis real de una variable. Derivamos la función f e igualamos a cero. Obtenemos que f alcanza su máximo en $t_0 = \frac{1-\beta}{2\beta(d-k\beta)}$, que se encuentra en el intervalo.

Por tanto hemos encontrado la cota $g(t_0) = \frac{1}{\sqrt{f(t_0)}} = \beta^{k/2} (1 + \frac{(1-\beta)k}{d-k})^{\frac{d-k}{2}}$

□

Corolario 3.2.4. Con la notación anterior, sea $k < d$:

$$1. \text{ Si } \beta < 1, \text{ entonces: } \Pr[L \leq \beta \frac{k}{d}] \leq e^{\frac{k}{2}(1-\beta+\ln\beta)}$$

$$2. \text{ Si } \beta > 1, \text{ entonces: } \Pr[L \geq \beta \frac{k}{d}] \leq e^{\frac{k}{2}(1-\beta+\ln\beta)}$$

Demostración. Para demostrar la primera parte, hemos de comprobar que se verifica,

$$\beta^{k/2} \left(1 + \frac{(1-\beta)k}{d-k} \right)^{\frac{d-k}{2}} \leq e^{\frac{k}{2}(1-\beta+\ln\beta)}, \forall \beta < 1$$

Empleando que $e^{\frac{k}{2} \ln \beta} = \beta^{k/2}$ y elevando al cuadrado, al tratarse de bases positivas, hemos de comprobar:

$$\left(1 + \frac{(1-\beta)k}{d-k} \right)^{d-k} \leq e^{k(1-\beta)}, \forall \beta < 1$$

Donde el primer término podemos reescribirlo como,

$$\left(1 + \frac{1}{\frac{d-k}{(1-\beta)k}} \right)^{d-k} = \left[\left(1 + \frac{1}{\frac{d-k}{(1-\beta)k}} \right)^{\frac{d-k}{(1-\beta)k}} \right]^{(1-\beta)k} \leq e^{k(1-\beta)}$$

Donde hemos empleado que $(1 + \frac{1}{x})^x \leq e, x \geq 0$

□

3.3. Teorema de Dvoretzky

Dedicaremos esta sección del capítulo a estudiar los teoremas de Dvoretzky. Dada su complejidad, no podremos detenernos en los detalles de las demostraciones, sin embargo, no sería concebible tratar el tema de los fenómenos de concentración de la medida y no mencionar dichos teoremas, pieza fundamental dentro de esta teoría. Los detalles omitidos podrán consultarse en [29].

Como sabemos, los espacios normados pueden ser vistos desde el punto de vista de los espacios vectoriales, espacios topológicos y espacios métricos. La siguiente definición, de isomorfismo topológico, nos determina cuando dos espacios normados son indistinguibles desde el punto de vista topológico y vectorial.

Definición 3.3.1. Sean $(E, \|\cdot\|_E)$, $(F, \|\cdot\|_F)$ dos espacios de Banach. Diremos que una aplicación $T : E \rightarrow F$ es un **isomorfismo topológico** si es un isomorfismo algebraico (lineal y biyectivo) y además las aplicaciones T, T^{-1} son continuas.

Observación 3.3.2. Habitualmente diremos que los espacios son isomorfos, refiriéndonos a la existencia de un isomorfismo topológico entre ellos.

Sin embargo esto no afecta a la estructura de espacio métrico. Para ello necesitaríamos que la aplicación T conservara la norma. Las siguientes definiciones siguen la línea de dar la identificación total entre espacios.

Definición 3.3.3. Sean $(E, \|\cdot\|_E)$, $(F, \|\cdot\|_F)$ dos espacios de Banach y sea $\lambda \geq 1$. Diremos que los espacios E, F son λ -**isomorfos** y lo expresaremos mediante $E \overset{\lambda}{\sim} F$, si existe un isomorfismo $T : E \rightarrow F$ de modo que se verifique $\|T\| \|T^{-1}\| \leq \lambda$

Definición 3.3.4. Se define la “distancia” de Banach-Mazur entre dos espacios normados,

$$d(E, F) = \inf\{\lambda : E \overset{\lambda}{\sim} F\}$$

En base a lo que acabamos de definir, los espacios quedan totalmente identificados si la distancia de Banach-Mazur entre ellos es igual a 1.

Pasamos ya a enunciar el teorema de Dvoretzky en su versión habitual. Este resultado nos dice que de cualquier espacio de Banach, podemos extraer un subespacio que sea copia del espacio Euclídeo ℓ_2^n .

Teorema 3.3.5. (Dvoretzky) Para todo $\epsilon > 0$ existe $\eta(\epsilon) > 0$ de modo que para cualquier espacio de Banach $(W, \|\cdot\|)$ N -dimensional podemos encontrar un subespacio $V \subset W$ de dimensión $n = \lfloor \eta(\epsilon) \log N \rfloor$ que verifica $V \overset{1+\epsilon}{\sim} \ell_2^n$.

Demostración. Puede verse en [29, pág. 174] □

Observación 3.3.6. Nótese que este enunciado es igualmente válido si la dimensión del espacio original $(W, \|\cdot\|)$ fuera infinita. Basta con tomar un subespacio finito E y aplicarle el resultado.

A continuación vamos a dar la “traducción” del teorema de Dvoretzky a una formulación probabilística. En esta sección vamos a considerar variables aleatorias que toman valores en un espacio de Banach B separable. Por tanto, la definición de variable aleatoria sería la de aplicación $X : \Omega \rightarrow B$ que satisface la misma condición de medibilidad.

Necesitamos ahora traducir el concepto de variable Gaussiana para un espacio de Banach cualquiera, para esta definición nos llevaremos la variable a \mathbb{R} mediante todos los elementos del dual de B .

Definición 3.3.7. Diremos que la variable aleatoria $X : \Omega \rightarrow B$ definida en un espacio de Banach B , es Gaussiana si $\forall x^* \in B^*$ aplicación continua del dual de B , $x^*(X)$ es una variable aleatoria Gaussiana.

Definición 3.3.8. Sea la variable aleatoria $X : \Omega \rightarrow B$ definida en un espacio de Banach B . Definimos la **dimensión** de la variable mediante,

$$d(X) = \frac{E\|X\|^2}{\sigma(X)^2}$$

Donde $\sigma(X) = \sup\{E[x^*(X)^2]^{1/2} : x^* \in B^*, \|x^*\| \leq 1\}$ es una versión “débil” de los momentos de la variable aleatoria.

Con estos nuevos conceptos, estamos ya en posición de introducir la versión Gaussiana del teorema de Dvoretzky. Dicho resultado nos dice que la existencia de variables aleatorias Gaussianas en un espacio de Banach B nos determina si en dicho espacio somos capaces de encontrar una copia de ℓ_2^n

Teorema 3.3.9. (Dvoretzky, Versión Gaussiana) Para todo $\epsilon > 0$, $\exists \eta(\epsilon) > 0$ de modo que se cumple la siguiente propiedad: Sea X una variable aleatoria Gaussiana que toma valores en un espacio de Banach B . Entonces B contiene un subespacio F de dimensión $n = [\eta(\epsilon)d(X)]$ que es $(1 + \epsilon)$ -isomorfo a ℓ_2^n . Recíprocamente, si B contiene un subespacio F que es $(1 + \epsilon)$ -isomorfo a ℓ_2^n entonces existe una variable aleatoria Gaussiana que toma valores en B y que cumple $d(X) \geq (1 + \epsilon)^{-2}n$.

Demostración. Puede consultarse en [29, pág. 171] □

Observación 3.3.10. Por este resultado, el problema de buscar un subespacio copia de ℓ_2^n se reduce a buscar una variable aleatoria Gaussiana de dimensión n .

3.4. Ejemplos de concentración: Funciones Lipschitz

En esta sección trataremos otro de los fenómenos de concentración más conocidos: el de las funciones Lipschitz. El objetivo central es probar que si tenemos $f : \mathbb{S}^{n-1} \rightarrow \mathbb{R}$ una función Lipschitz en la esfera con dimensión suficientemente grande, entonces encontramos un subespacio en el cual la función se condensa en torno al valor de su mediana. Seguiremos los resultados que aparecen en [5, pág. 281-285].

Lema 3.4.1. Consideremos E un Espacio de Banach k -dimensional, $0 < \epsilon < 1$ y $\{x_j\}_{j=1}^N$ en la bola unidad del espacio ϵ -separados, es decir, $\|x_i - x_j\| \geq \epsilon$, $i \neq j$. Entonces se verifica,

$$N \leq \left(1 + \frac{2}{\epsilon}\right)^k$$

Demostración. Ver [5, pág. 283] □

Proposición 3.4.2. Sea $f : \mathbb{S}^{n-1} \rightarrow \mathbb{R}$ una función L -Lipschitz respecto de la norma Euclídea y sea $0 < \epsilon < \frac{L}{2}$. Entonces existe un subespacio k -dimensional $F \subset \mathbb{R}^n$ con $k \geq \frac{Cn\epsilon^2}{L^2|\log(\frac{\epsilon}{L})|}$ y existe un valor λ_0 de modo que se verifica:

$$m_f - \epsilon \leq f(x) \leq m_f + \epsilon \quad \forall x \in F \cap \mathbb{S}^{n-1}$$

Demostración. Llamamos m_f a la mediana de la función f y defino el conjunto $A = \{x : f(x) = m_f\}$. Nótese que $A \neq \emptyset$ ya que las funciones Lipschitz tienen la propiedad de los valores intermedios. Como f es L -Lipschitz, resulta que $|m_f - f(x)| \leq \frac{\epsilon}{2} \forall x \in A_{\epsilon/2L}$. Efectivamente,

$$|\lambda_0 - f(x)| = |f(y) - f(x)| \leq L|x - y| \leq L \frac{\epsilon}{2L} = \frac{\epsilon}{2}$$

Dado que $x \in A_{\epsilon/2L}$, podemos encontrar un elemento $y \in A$ de modo que se verifique la última desigualdad. Aplicamos la desigualdad isoperimétrica a los conjuntos A^+, A^- vista en la Observación 2.2.6. En efecto,

$$\sigma_{n-1}(A^+), \sigma_{n-1}(A^-) \geq \frac{1}{2} \implies \sigma_{n-1}(A_{\epsilon/2L}^+), \sigma_{n-1}(A_{\epsilon/2L}^-) \geq 1 - 2e^{-n\epsilon^2/8L^2}$$

Expresamos el conjunto $A_{\epsilon/2L} = A_{\epsilon/2L}^+ \cap A_{\epsilon/2L}^-$, pudiendo estimar de forma sencilla su medida,

$$\sigma_{n-1}(A_{\epsilon/2L}^c) \leq \sigma_{n-1}((A_{\epsilon/2L}^+)^c) + \sigma_{n-1}((A_{\epsilon/2L}^-)^c) \leq 4e^{-n\epsilon^2/8L^2}$$

Con lo que finalmente resulta,

$$\sigma_{n-1}(A_{\epsilon/2L}) \geq 1 - 4e^{-n\epsilon^2/8L^2}$$

Identificando la medida de Haar invariante por $\mathcal{O}(n)$ con la normalizada sobre la esfera resulta, $\forall x \in \mathbb{S}^{n-1}$, $\theta_n\{U \in \mathcal{O}(n) : Ux \in A_{\frac{\epsilon}{2L}}\} = \sigma_{n-1}(A_{\frac{\epsilon}{2L}}) \geq 1 - 4e^{-n\epsilon^2/8L^2} > 0 \iff 4 < e^{n\epsilon^2/8L^2}$.

Por tanto, si elegimos $N \in \mathbb{N}$ de modo que verifique $4N < e^{n\epsilon^2/8L^2}$ garantizamos la existencia de $U_0 \in \mathcal{O}(n)$ que cumple $U_0x_j \in A_{\epsilon/2L}$ para todo elemento de $\{x_j\}_{j=1}^N \subset \mathbb{S}^{n-1}$

Consideramos ahora el subespacio $F_0 \subset \mathbb{R}^n$ generado por los primeros k -vectores unitarios de \mathbb{R}^n y sea $\{x_j\}_{j=1}^N$ un conjunto maximal $\frac{\epsilon}{2L}$ -separado en $F_0 \cap \mathbb{S}^{n-1}$. Del lema mencionado anteriormente, se tiene que $N \leq \left(1 + \frac{4L}{\epsilon}\right)^k$ y por el hecho de ser maximal, se cumple:

$$4 \left(1 + \frac{4L}{\epsilon}\right)^k \leq e^{n\epsilon^2/8L^2}$$

Por lo dicho previamente, encontramos $U_0 \in \mathcal{O}(n)$ de modo que $U_0x_j \in A_{\epsilon/2L}$ y por tanto se verifica $|m_f - f(U_0x_j)| \leq \frac{\epsilon}{2} \forall j \in \{1, \dots, N\}$. Dado $x \in F_0 \cap \mathbb{S}^{n-1}$, como la sucesión $\{x_j\}_{j=1}^N$ es $\frac{\epsilon}{2L}$ -separada existirá un índice j_0 de modo que se cumple $\|x_{j_0} - x\| \leq \frac{\epsilon}{2L}$ y finalmente:

$$|f(U_0x) - m_f| \leq |f(U_0x) - f(U_0x_{j_0})| + |f(U_0x_{j_0}) - m_f| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Concluyendo que la función f es ϵ -cercana a la constante λ_0 sobre la esfera unidad de $F = U_0F$. Para la estimación de la dimensión de F resta ver que $k = \lceil \frac{Cn\epsilon^2}{L^2 \log(\frac{\epsilon}{L})} \rceil + 1$ satisface el lema previamente enunciado. \square

Observación 3.4.3. *Dicha prueba supone la primera aparición del método probabilístico de demostración. Hemos calculado la medida (probabilidad) de que exista una cierta rotación que cumpla una característica determinada y al resultar estrictamente positiva, concluimos su existencia.*

El siguiente resultado supone una generalización del anterior, en un espacio de Banach arbitrario B , donde la función Lipschitz parte de su bola unidad a un cierto cuerpo \mathbb{K} . De nuevo encontramos un subespacio para el cual la función f se concentra en torno a su mediana de modo que podemos decir que es *casi constante* en dicho subespacio.

Proposición 3.4.4. *Sea $f : B_E \rightarrow \mathbb{K}$ una función L -Lipschitz definida sobre la esfera unidad de un espacio de Banach $(E, \|\cdot\|)$ y sea $0 < \epsilon < \frac{L}{2}$. Entonces existe un valor λ_0 y un subespacio $F \subset E$ de dimensión $k \geq C\epsilon^2 \frac{\log n}{|\log(\epsilon/L)|L^2}$ de modo que se cumple:*

$$|f(x) - \lambda_0| \leq \epsilon \quad \forall x \in F, \|x\| = 1$$

Demostración. Sigue del teorema de Dvoretzky y del resultado previo. \square

3.5. Un caso especial: El Lema de Johnson-Lindenstrauss

A continuación vamos a introducir un fenómeno de concentración de la medida, conocido en su origen como Lema de Johnson-Lindenstrauss, que además nos servirá de enlace entre este capítulo y el siguiente.

Lema 3.5.1. *Sea $0 < \epsilon < 1$ y $V = \{v_1, \dots, v_n\} \subset \mathbb{R}^d$. Entonces encontramos un entero $k = k(\epsilon, n)$ y una función Lipschitz $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$ de modo que se verifica,*

$$(1 - \epsilon)\|v_i - v_j\|_2^2 \leq \|f(v_i) - f(v_j)\|_2^2 \leq (1 + \epsilon)\|v_i - v_j\|_2^2, \quad \forall i, j \in \{1, 2, \dots, n\}$$

Donde estamos considerando la norma Euclídea en $\mathbb{R}^d, \mathbb{R}^k$

Observación 3.5.2. *Como veremos más adelante, este enunciado puede reformularse con una aplicación lineal mediante el producto de los vectores de V por una cierta matriz aleatoria Φ . Tampoco hemos querido detallar el valor de $k(\epsilon, n)$ a sabiendas de que será el tema central del trabajo a partir de ahora, sin embargo, cabe destacar que dicho valor no depende de la dimensión del espacio de partida, sino que únicamente de la precisión ϵ y del número de valores a proyectar. Nótese que podemos tomar el conjunto $E = \{v_i - v_j : v_i, v_j \in V\}$. Entonces podemos reformular el lema, asegurando que existe una constante universal C_0 de modo que,*

$$P[(1 - \epsilon)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \epsilon)\|x\|_2^2] \geq 1 - 2e^{-c_0\epsilon^2 k}$$

Que expresado de esta manera podemos ver claramente que se trata de un fenómeno de concentración de la medida.

No vamos a dar ahora la prueba de este enunciado, pues de ello nos encargaremos en el próximo capítulo. Este lema nos afirma que dado cualquier conjunto discreto de puntos del espacio \mathbb{R}^d , con d arbitrario, encontramos una aplicación Lipschitz de modo que nos proyecta los puntos a un espacio de menor dimensión garantizando que las distancias se distorsionan únicamente mediante un valor $(1 \pm \epsilon)$. Como hemos mencionado, es un claro ejemplo de concentración de la medida que será estudiado como fenómeno geométrico y desde el punto de vista de las proyecciones aleatorias. Este enunciado motiva el siguiente capítulo y fija definitivamente el rumbo que toma el presente trabajo.

Capítulo 4

Lema de Johnson-Lindenstrauss

4.1. Introducción

El resultado conocido como Lema de Johnson-Lindenstrauss fue publicado en 1984 en un artículo realizado por William B. Johnson y Joram Lindenstrauss. En dicho artículo atacan el problema de controlar las normas de las extensiones de funciones Lipschitz desde un punto de vista geométrico. En el camino, encuentran una más que poderosa forma de proyectar un conjunto de datos reduciendo su dimensión de forma considerable sin perturbar demasiado las distancias. Precisamente, y como veremos a continuación, la llave que emplean para conseguirlo es la desigualdad isoperimétrica en la esfera n dimensional, resultado ya visto en este trabajo.

A la postre veremos que se asumen ciertas condiciones a la hora de demostrar el resultado: La necesidad de que la proyección realizada de los datos sea ortonormal, que el subespacio sea elegido de manera aleatoria en un cierto sentido y que las variables que utilizamos para elegir dicha proyección cumplan la propiedad de poseer simetría esférica.

A lo largo del capítulo nos detendremos en concretar qué significa cada una de estas propiedades y nos centraremos en ir analizando las sucesivas mejoras del lema original, que supone ir relajando las hipótesis antes descritas a la vez que conseguimos disminuir la dimensión a la cual proyectamos los datos.

4.2. Propiedades empleadas

4.2.1. Ortonormalidad

Una de las propiedades que asumimos en la proyección de los datos es la de ortonormalidad. Si pensamos que el subespacio viene determinado por la matriz que tiene por filas los vectores que lo generan, proyectar un vector equivale a multiplicarlo por dicha matriz. La noción de ortonormalidad que emplean los autores en sus demostraciones viene prefijada por las siguientes definiciones.

Definición 4.2.1. Una proyección mediante una matriz Q será **ortogonal** si verifica:

$$\langle Qx, y \rangle = \langle x, Qy \rangle$$

Definición 4.2.2. Una matriz $M \in \mathcal{M}^n$ diremos que es **ortonormal** si verifica que $MM^T = M^T M = I_n$

Como veremos, los autores emplearán proyecciones que verifiquen ambas propiedades. Cabe destacar, tal y como mencionabamos en la Introducción, sección 1.6 cuando avancemos en el tema, emplearemos matrices que no son cuadradas y que por tanto carece de sentido la Definición 4.2.2. Sin embargo, no supone un problema dado que a esas alturas ya habremos dejado caer esta hipótesis asumida.

4.2.2. Simetría esférica

La siguiente hipótesis que empleamos es la de simetría esférica. Esta propiedad nos quiere decir, en un cierto sentido, que la proyección empleada es invariante bajo rotaciones, que podemos expresarlo mediante producto por matriz ortogonal. Dicha idea motiva las siguientes definiciones.

Definición 4.2.3. Diremos que un vector aleatorio $X \in \mathbb{R}^d$ es **ortogonalmente invariante** si $\forall A \in \mathcal{O}(n)$ se verifica que $AX \sim X$ tienen la misma distribución.

Definición 4.2.4. Diremos que un vector aleatorio $X \in \mathbb{R}^d$ tiene **simetría esférica** respecto su vector de medias μ si $X - \mu$ es ortogonalmente invariante.

Generalmente emplearemos variables de esperanza nula, con lo cual su simetría será en torno al origen. Atendiendo a la definición, es obvio que una variable Gaussiana posee simetría esférica y de hecho, se caracterizan en un cierto sentido reflejado en [6]

4.2.3. Subespacios Aleatorios

Otra hipótesis fundamental en nuestro trabajo es la sección aleatoria del subespacio de proyección. Si, como hemos dicho, proyectar equivale a multiplicar los vectores por una matriz cuyos vectores generan el nuevo subespacio, podemos dar la siguiente definición.

Definición 4.2.5. El subespacio diremos que es **aleatorio** si la matriz de proyección tiene un comportamiento de variable aleatoria.

Dicha aleatoriedad será vista mediante dos formas diferentes. Bien como un cambio de base aleatorio sobre una proyección de k coordenadas o bien como la proyección por una matriz cuyas entradas siguen una variable aleatoria, más específicamente, trataremos con matrices cuyas componentes siguen una distribución $N(0, 1)$ e independientes o distribuciones equiprobables del conjunto $\{-1, +1\}$. La siguiente definición cobrará especial interés en la prueba de Gupta y Dasgupta. Nos define el concepto de vector unitario aleatorio.

Definición 4.2.6. Sea $Z = (Z_1, \dots, Z_d)$ un vector aleatorio cuyas componentes son independientes e idénticamente distribuidas tal que $Z_i \sim N(0, 1)$. Llamaremos **vector unitario aleatorio** d -dimensional a un vector aleatorio de la forma $\vec{x} = \frac{1}{\|Z\|} (Z_1, \dots, Z_d) \in \mathbb{S}^{d-1}$.

4.3. Prueba Original

En esta sección vamos a tratar la prueba original del Lema de Johnson-Lindenstrauss. Dichos autores publicaron a mediados de los años 80 un artículo llamado “*Extensions of Lipschitz maps into a Hilbert space*” donde atacaban el siguiente problema geométrico:

Consideremos (X, d) un espacio métrico cualquiera, y $n \in \mathbb{N}$, $n \geq 2$. Sea $f : A \subset X \rightarrow \ell_2$ una aplicación Lipschitz definida en algún subconjunto que cumpla $|A| = n$. El problema trata de buscar la mínima constante $L = L(X, n)$, válida para cualquier A , de modo que,

$$\|\hat{f}\|_{lip} \leq L \|f\|_{lip}$$

Donde $\hat{f} : A \subset X \rightarrow \ell_2$ es una aplicación Lipschitz que extiende a f en el conjunto A , es decir, que $\hat{f}(x) = f(x)$, $\forall x \in A$. Denotamos por $\|f\|_{lip}$ la constante Lipschitz de la aplicación f . Es obvio que, como \hat{f} extiende a f , tenemos que $L \geq 1$. En el artículo, mencionan un resultado clásico de Kirschbraun en el que se prueba que $L(\ell_2, n) = 1$, $\forall n \geq 2$. El objetivo del artículo es demostrar que existe una constante universal C de modo que la constante L verifica,

$$L(X, n) \leq C(\log n)^{1/2}$$

Para cualquier espacio métrico (X, d) . Su investigación, aunque interesante, parece a priori algo alejado de nuestro problema de interés. De mayor sorpresa resulta que lo que ha pasado a la historia como el “*Lema de Johnson-Lindenstrauss*” es un pequeño resultado previo que demuestran como herramienta para probar su objetivo, la anterior acotación. En palabras de los propios autores:

“*La herramienta principal para probar el Teorema es un simple y elemental lema geométrico, el cual describimos ahora: Dados n puntos en un espacio Euclídeo, ¿cuál es la dimensión más pequeña $k=k(n)$ para la cual podemos mover dichos puntos a un espacio Euclídeo k -dimensional mediante una transformación que expanda o contraiga todos los pares de distancias con un factor de a lo más $1+\epsilon$? La respuesta, que $k \geq C(\epsilon) \log n$, es una simple consecuencia de la Desigualdad Isoperimétrica en \mathbb{S}^{n-1} ”*

Realmente sorprendente resulta la narrativa tan sencilla con la que son capaces de presentar un fenómeno de tal magnitud, aunque no es eso lo primero que nos llama la atención al leer esas líneas, sino el hecho de que la llave de la prueba de ese enunciado la tiene la desigualdad isoperimétrica que veíamos con detalle en el Capítulo Segundo. A continuación damos el enunciado exacto del lema geométrico al que se refieren los autores, como podemos ver, dicho enunciado viene en términos de la perturbación de las normas de la aplicación Lipschitz extendida. Antes, enunciemos una proposición técnica que será empleada en la prueba.

Proposición 4.3.1. *Sea X un espacio vectorial de dimensión n . Sea $\|\cdot\|$ una norma proveniente de un producto escalar sobre X y $\|\cdot\|_X$ una seminorma de modo que existe b tal que para todo $x \in X$, $\|x\|_X \leq b\|x\|$. Sea M_r la mediana de $r(x) = \|x\|_X$ sobre $\{x : \|x\| = 1\}$ y $0 < \epsilon \leq 1$. Consideremos $\{y_i\}_{i=1}^m$ unitarios sobre ℓ_2^n donde $4m \leq e^{n\epsilon^2/2}$. Entonces existe una isometría $U : \ell_2^n \rightarrow X$ que verifica,*

$$M_r - b\epsilon \leq \|Uy_i\| \leq M_r + b\epsilon, \quad i = 1, \dots, m$$

Demostración. La prueba se basa en el Lema de Levy referido a una seminorma, ver [13, pág 58]. □

Lema 4.3.2. (Johnson-Lindenstrauss, 1984) *Para todo $0 < \epsilon < 1$ existe una constante $K = K(\epsilon)$ de modo que si $A \subset \ell_2^d$, $|A| = d$ con $d \in \mathbb{N}$ entonces existe una aplicación Lipschitz $f : A \rightarrow \ell_2^k$, $k = k(d, \epsilon) = \lceil K \log d \rceil$ que satisface,*

$$\|\hat{f}\|_{lip} \|\hat{f}^{-1}\|_{lip} \leq \frac{1 + \tau}{1 - \tau}$$

Donde $\hat{f} : \ell_2^d \rightarrow \ell_2^k$ extiende a f .

Observación 4.3.3. *La demostración del Lema posee cierta complejidad y algunos de sus detalles escapan al objetivo del presente trabajo. Sin embargo, y por ser el resultado central, daremos*

el esquema de la prueba y por supuesto, la aplicación del “método probabilístico” de encontrar probabilidad positiva. Para algún paso intermedio aportaremos la referencia donde se pueda contrastar.

Demostración. Consideremos Q la proyección a las primeras k coordenadas de ℓ_2^d y denotemos por θ_d la medida normalizada de Haar sobre $\mathcal{O}(d)$. Definimos la variable aleatoria $f : (\mathcal{O}(d), \theta_d) \rightarrow C(\ell_2^d, \|\cdot\|_\infty)$ definida por,

$$f(U) = U^*QU$$

Que a cada matriz ortogonal proporciona un “cambio de base” con respecto a una proyección sobre k coordenadas, aportando así la noción de proyección aleatoria k dimensional. Para comprobar que esta proyección tiene la propiedad que nosotros deseamos basta con aplicar la desigualdad de Levy vista en Lema 3.1.3 convertido en el resultado previo. Dado un vector $x \in \mathbb{R}^d$, definimos,

$$r(x) = \sqrt{d} \left(\sum_{i=1}^k x_i^2 \right)^{1/2} = \sqrt{d} \|Qx\|_2$$

Vista como aplicación del espacio ℓ_2^d , es una seminorma que cumple $r(x) \leq \sqrt{d} \|x\|$. Consideremos el conjunto $B = \left\{ \frac{x-y}{\|x-y\|_2} : x, y \in A, x \neq y \right\} \subset \mathbb{S}^{d-1}$ Nuestro objetivo ahora es el de encontrar $U \in \mathcal{O}(d)$ y una constante M para la que se cumpla:

$$M(1 - \epsilon) \leq r(Ux) \leq M(1 + \epsilon)$$

Consideramos M_r la mediana de la aplicación $r(\cdot)$ en \mathbb{S}^{d-1} . La proposición 4.3.1, junto con el denominado Lema de Levy, nos conduce a la desigualdad,

$$\theta_d(U \in \mathcal{O}(d) : M_r - \sqrt{d}\epsilon \leq r(Uy) \leq M_r + \sqrt{d}\epsilon) \geq 1 - 4e^{-\frac{d\epsilon^2}{2}}$$

Seguidamente, y estos detalles los omitimos (pueden verse en [19]), debemos estimar la mediana de $r(x)$ en términos de \sqrt{n} , para ello, y de modo muy técnico, hay que emplear la *Desigualdad de Khintchine* junto con el Lema 1.7 de [13] y hacer elecciones adecuadas del término ϵ y k . Finalmente,

$$\theta_d(U \in \mathcal{O}(d) : M_r(1 - \tau) \leq r(Uy) \leq M_r(1 + \tau), \forall y \in B) \geq 1 - 2d(d+1)e^{-\frac{\tau^2 K \log d}{18}}$$

Estrictamente positivo si tomamos K suficientemente grande. □

Observación 4.3.4. *Sobre la anterior demostración cabe realizar ciertas reflexiones:*

- Véase la relación entre la construcción de la prueba con la aplicación f definida al comienzo, pues

$$r(Uy) = \sqrt{d} \|QUy\|_2^2 = \sqrt{d} \|U^*QUy\|_2^2 = \sqrt{d} \|f(U)y\|_2^2$$

- Hemos de mencionar, como aportación fundamental al trabajo, la presencia de nuestra llamada “demostración probabilística”. Para demostrar que efectivamente existe una aplicación Lipschitz que no distorsiona en gran medida las distancias, comprobamos que la probabilidad de que ocurra ese suceso, tomando una proyección aleatoria, es estrictamente positiva: concluimos su prueba.

- Debemos destacar también que los autores toman un conjunto de cardinalidad igual a la dimensión del espacio de partida. Esto será generalizado en la siguiente versión ya para un conjunto V de n elementos. Con dicha generalización veremos que la dependencia del valor k no es con la dimensión del espacio de partida sino con el número de elementos proyectados.

La arbitrariedad en la elección del subespacio k -dimensional proviene de tomar Q la matriz que proyecta un vector sobre sus k primeras coordenadas y efectuarle un cierto cambio de base ortonormal, de modo que nuestra matriz de proyección resulta, U^*QU cumple nuestra serie de “buenas propiedades” que vamos a analizar a continuación.

- **Proyección Ortogonal:**

$$\begin{aligned} \langle U^*QUx, y \rangle &= \langle UU^*QUx, Uy \rangle = \langle QUx, Uy \rangle = \langle Ux, QUy \rangle \\ &= \langle U^*Ux, U^*QUy \rangle = \langle x, U^*QUy \rangle \end{aligned}$$

Donde hemos empleado la propiedad de las matrices ortogonales respecto del producto escalar, junto con el hecho de que, para el producto escalar, es indiferente proyectar a k coordenadas el primer vector o el segundo. Además, la matriz de proyección es claramente ortonormal.

- **Subespacio Aleatorio:** Por la propia construcción de la matriz de proyección, el subespacio k dimensional al que son proyectados los vectores es elegido de manera aleatoria.
- **Simetría Esférica:** Por el empleo de las matrices ortogonales en la proyección, deducimos que ésta posee simetría esférica.

A partir de aquí, en el trabajo realizaremos una comparativa personal entre los diferentes artículos que muestran sucesivas “mejoras” de la presente versión del enunciado. Trataremos de mostrar el esquema común que representan y el cómo se consiguen los diferentes progresos y el significado de cada uno.

4.4. Prueba íntegramente probabilística

La siguiente versión del lema que vamos a tratar es la demostrada por Gupta y Dasgupta en 1999, cuyo artículo puede consultarse en [8]. Siguen empleando todas las hipótesis antes descritas, pero supone una enorme simplificación con respecto a la demostración original además de una sustancial mejora de la dimensión del espacio de llegada. Abre la puerta a las demostraciones de corte probabilístico en la cuál se emplean variables aleatorias, casi siempre Gaussianas, para determinar el subespacio sobre el cual se proyectarán los datos.

Continúan con la misma idea de proyección que emplean Johnson y Lindenstrauss, tomando la matriz U^*QU donde $U \in \mathcal{O}(d)$ arbitraria, con el objetivo de evaluar la proyección $\|U^*QUx\|$ donde podemos suponer, sin pérdida de generalidad, que el vector x es unitario. Entonces,

$$\|U^*QUx\|_2 = \|QUx\|_2$$

En vista a la expresión de la derecha, y remarcando que $U \in \mathcal{O}(d)$ se ha elegido de manera aleatoria, nos encontramos con que el vector Ux es unitario y aleatorio, luego, en base a su propia definición, puede venir modelizado mediante,

$$Ux = \frac{(X_1, \dots, X_d)}{\|X\|_2}$$

Con las variables $X_i \sim N(0,1)$ e independientes. Y con ello, su proyección a k variables mediante Q vendría dada por,

$$QUx = \frac{(X_1, \dots, X_k)}{\|X\|_2}$$

Finalmente, podríamos evaluar dicha norma de la forma,

$$\|QUx\|_2^2 = \frac{X_1^2 + \dots + X_k^2}{X_1^2 + \dots + X_d^2}$$

Atendiendo a estas expresiones nos viene a la memoria el lema de concentración de la medida, visto en Proposición 3.2.3, cuya construcción era idéntica. Concluimos que es equivalente estimar la proyección arbitraria de un vector fijo unitario a estimar la norma de un vector aleatorio proyectado sobre un subespacio fijo de k coordenadas. Bajo esta idea se apoya el siguiente teorema, pero antes, daremos un lema técnico sencillo que será empleado en la demostración de dicho teorema.

Lema 4.4.1. *Se verifica,*

1. $\ln(1-x) \leq -x - \frac{x^2}{2}, \forall x \in [0, 1)$
2. $\ln(1+x) \leq x - \frac{x^2}{2} + \frac{x^3}{3}, \forall x \geq 0$

Demostración. Para probar 1, defino $f(x) = \ln(1-x) + x + \frac{x^2}{2}, x \in [0, 1)$. Entonces $f(0) = 0$, luego si probamos que es decreciente en todo el dominio, en particular será no positiva y tendremos el resultado,

$$f'(x) = -\frac{1}{1-x} + 1 + x = \frac{x^2 - 2}{1-x} < 0$$

De manera análoga se puede probar 2. □

Teorema 4.4.2. *(Gupta y Dasgupta, 1999) Sea $0 < \epsilon < 1, n \in \mathbb{N}$ y $k \in \mathbb{N}$ que verifica*

$$k \geq 4 \left(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3} \right)^{-1} \ln n$$

Entonces para cualquier conjunto $V = \{v_1, \dots, v_n\} \subset \mathbb{R}^d$ existe una aplicación Lipschitz $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$ de modo que $\forall v_i, v_j \in V,$

$$(1 - \epsilon)\|v_i - v_j\|_2^2 \leq \|f(v_i) - f(v_j)\|_2^2 \leq (1 + \epsilon)\|v_i - v_j\|_2^2$$

Además, esta aplicación la podemos determinar el tiempo polinómico.

Demostración. Supongamos que $k < d$, pues de lo contrario basta con tomar los mismos puntos o, su inmersión en un espacio de dimensión mayor. Tomamos S un subespacio aleatorio ortonormal k -dimensional de la forma antes detallada, tomo dos puntos fijos del conjunto, $v_i, v_j \in V$ y llamaremos v'_i a la proyección de cada $v_i \in V$ en S . Entonces consideramos la variable $L = \|v'_i - v'_j\|_2^2$ y $\mu = \frac{k}{d}\|v_i - v_j\|_2^2$. Por lo detallado antes del enunciado, podemos aplicar la Proposición 3.2.3 de concentración de la medida,

$$P[L \leq (1 - \epsilon)\mu] \leq e^{\frac{k}{2}(1 - (1 - \epsilon) + \ln(1 - \epsilon))} \leq e^{\frac{k}{2}(\epsilon - (\epsilon + \frac{\epsilon^2}{2}))} = e^{-k\epsilon^2/4} \leq e^{-2 \ln n} = \frac{1}{n^2}$$

En el segundo paso hemos empleado el lema previo y el tercero es consecuencia de la elección que hacemos para k . Efectivamente,

$$k \geq 4 \left(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3} \right)^{-1} \ln n \geq 4 \left(\frac{\epsilon^2}{2} \right)^{-1} \ln n \implies k \frac{\epsilon^2}{4} \geq 2 \ln n$$

De modo análogo, aplicando de nuevo el resultado de concentración, podemos estimar la siguiente probabilidad,

$$P[L \geq (1 + \epsilon)\mu] \leq e^{\frac{k}{2}(1-(1+\epsilon)+\ln(1+\epsilon))} \leq e^{\frac{k}{2}(-\epsilon+(\epsilon-\frac{\epsilon^2}{2}+\frac{\epsilon^3}{3}))} = e^{-\frac{k(\epsilon^2/2-\epsilon^3/3)}{2}} \leq e^{-2 \ln n} = \frac{1}{n^2}$$

Donde en el segundo paso hemos empleado el lema previo y el tercer paso es consecuencia directa de la elección que hacemos del valor de k . Démonos cuenta de que estamos estimando la probabilidad del siguiente suceso,

$$A = \{(1 - \epsilon)\mu \leq L \leq (1 + \epsilon)\mu\} = \{(1 - \epsilon) \frac{k}{d} \|v_i - v_j\|_2^2 \leq \|v'_i - v'_j\|_2^2 \leq (1 + \epsilon) \frac{k}{d} \|v_i - v_j\|_2^2\}$$

$$A = \{(1 - \epsilon) \|v_i - v_j\|_2^2 \leq \|\sqrt{d/k} (v'_i - v'_j)\|_2^2 \leq (1 + \epsilon) \|v_i - v_j\|_2^2\}, \text{ con } i, j \text{ fijos}$$

Resta tomar como aplicación $f(v_i) = \sqrt{\frac{d}{k}} v'_i$, que además es Lipschitz,

$$\|f(u) - f(v)\|_2^2 = \frac{d}{k} \|u' - v'\|_2^2 = \frac{d}{k} \|U^*QU(u - v)\|_2^2 = \frac{d}{k} \|QU(u - v)\|_2^2 \leq \frac{d}{k} \|QU\|^2 \|u - v\|_2^2$$

Según la estimación de las probabilidades, la aplicación f así definida perturba las distancias mediante un factor mayor a $(1 \pm \epsilon)$ sólo con, a lo sumo, una probabilidad de $\frac{2}{n^2}$

Este proceso se ha evaluado fijando dos puntos $v_i, v_j \in V$. De entre un conjunto de n puntos, podemos establecer $\binom{n}{2}$ parejas diferentes. Así pues la probabilidad de que la perturbación de la distancia entre cualquier par de puntos sea mayor de lo deseado es de a lo sumo $\binom{n}{2} \cdot \frac{2}{n^2} = 1 - \frac{1}{n}$. Leyendo el resultado al revés, la probabilidad de que ocurra que la aplicación f posee la propiedad del enunciado es de al menos $\frac{1}{n} > 0$, luego concluimos la veracidad del enunciado. Necesitaríamos repetir la proyección un número $O(n)$ de veces, con lo que obtenemos la segunda parte del teorema: el tiempo polinómico en conseguir encontrar f . \square

Observación 4.4.3. *Al final de la prueba hemos vuelto a aplicar un razonamiento de corte probabilístico. Encontramos que un suceso posee probabilidad estrictamente positiva y deducimos su existencia, quizá repitiendo el experimento un cierto número de veces.*

4.5. Relajando hipótesis (I): Ortogonalidad

La siguiente versión del teorema supone de nuevo un avance considerable. Conseguimos eliminar una de nuestras hipótesis, en concreto, la de ortonormalidad en la matriz de proyección. Puede consultarse el artículo de los autores originales Indyk y Motwami en [18] aunque también tendremos en consideración el análisis de dicho artículo que se hace en [17].

La idea que persigue el nuevo método es la de crear una matriz R cuyas entradas son aleatorias con distribución standard $N(0, 1)$. Es evidente que de esta forma sólo conseguimos “la

esperanza” de que las filas sean ortonormales entre sí, es decir, si $R_{ij} \sim N(0, 1)$ independientes, entonces,

$$E[R_i R_j] = \sum_{l=1}^k E[R_{il} R_{jl}] = \sum_{l=1}^k E[R_{il}] E[R_{jl}] = 0, \quad i \neq j$$

$$E[R_i R_i] = \sum_{l=1}^k E[R_{il} R_{il}] = \sum_{l=1}^k E[R_{il}^2] = k$$

Así pues, proyectar mediante la matriz $\frac{1}{\sqrt{k}}R \in \mathcal{M}^{k \times d}$ consigue mantenerla la “esperanza” de ortonormalidad, pero sin imponer la condición propiamente dicha. Pasamos a enunciar, y demostrar parcialmente, un lema técnico que emplearemos en la prueba del teorema.

Lema 4.5.1. *La distribución normal d -dimensional $N^d(0, 1)$ verifica las siguientes propiedades:*

1. *Tiene simetría esférica con respecto al origen.*
2. *Si $X \sim N^d(0, 1)$ y u es un vector unitario, entonces $X \cdot u \sim N(0, 1)$.*

Demostración. (1) Es directo usando que si $Z \sim N_d(\mu, \Sigma)$ entonces $AZ \sim N_d(A\mu, A^T \Sigma A)$. Efectivamente, si tomo $M \in \mathcal{O}(d)$, $MX \sim N(0, M^T I_d M) \equiv N(0, I_d)$.

(2) Considero el vector canónico $e_1 = (1, 0, \dots, 0) \in \mathbb{R}^n$. Claramente, $X \cdot e_1 = X_1$ posee distribución $N(0, 1)$. Consideremos ahora una matriz R de rotación que lleve el vector u a e_1 . Por el punto anterior, sabemos que RX queda distribuido conforme a $N^d(0, 1)$. Deducimos que $X \cdot u$ posee la misma distribución que $RX \cdot Ru$ que a su vez comparte distribución con $X \cdot e_1$, el cual sabemos que sigue una $N(0, 1)$. \square

Como hemos comentado anteriormente, la idea es emplear como matriz de proyección una cuyas entradas sigan un comportamiento aleatorio $N(0, 1)$ e independientes, y luego reescalar por un factor $\frac{1}{\sqrt{k}}$. Olvidémonos momentáneamente de dicho factor, y veamos la matriz R como una matriz que en la fila i -ésima tiene un vector aleatorio R_i que sigue una distribución $N^d(0, I_d)$. Supongamos, sin pérdida de generalidad que tenemos un vector unitario $u \in \mathbb{R}^d$ que queremos proyectar a través de nuestra matriz, dicha proyección resulta,

$$\hat{f}(u) = R \cdot u = \begin{pmatrix} \text{---} & R_1 & \text{---} \\ \text{---} & R_2 & \text{---} \\ \vdots & & \vdots \\ \text{---} & R_k & \text{---} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_d \end{pmatrix} = \begin{pmatrix} R_1 \cdot u \\ R_2 \cdot u \\ \vdots \\ R_k \cdot u \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{pmatrix}$$

Definimos las variables $X_j = R_j \cdot u$, que por el lema previo sabemos que tienen un comportamiento $N(0, 1)$. Llamamos W al vector proyección, es decir, $\hat{f}(u) = L = W(u) = (X_1, \dots, X_k)^T$. Así mismo, $\|\hat{f}(u)\|_2^2 = X_1^2 + \dots + X_k^2$ es la norma del vector proyectado. El objetivo del lema de Johnson-Lindenstrauss es que este valor quede cerca de la norma del vector u . En esa misma línea, presentamos el siguiente teorema,

Teorema 4.5.2. (Indyk y Motwami, 1998) *Con la notación antes definida, tomamos $k = O(\log n)$ y para cualquier $\beta > 1$, tenemos:*

- $E[L] = k$

- $P[L \geq \beta k] \leq O(k) \cdot e^{-\frac{k}{2}[\beta - (1 + \ln \beta)]}$
- $P[L \leq \frac{k}{\beta}] \leq O(k) \cdot e^{-\frac{k}{2}[\beta^{-1} - (1 - \ln \beta)]}$

Demostración. Por el lema previo sabemos que cada X_i sigue una $N(0, 1)$ y además, X_1, \dots, X_k son independientes. Definimos la variable $Y_i = X_{2i-1}^2 + X_{2i}^2$, $i = 1, \dots, \tau$, con $\tau = k$. Como hemos visto en el capítulo introductorio, $Y_i \sim \text{Exp}(\frac{1}{2})$. Sea $L = \sum_{i=1}^{\tau} Y_i$. De nuevo el mismo enunciado nos afirma que $L \sim G(k/2, 1/2)$, cuya esperanza es $E[L] = \sum_{i=1}^{\tau} E[Y_i] = 2\tau = k$, lo que prueba el primer epígrafe.

Para ver la segunda parte, definimos $\eta = \lambda\beta k$ y tenemos:

$$P[L \geq \beta k] = 1 - P[L < \beta k] = 1 - G_L(\beta k) = \sum_{i=0}^{\tau} e^{-\eta} \frac{\eta^i}{i!} \leq (\tau + 1)e^{-\eta} \frac{\eta^{\tau}}{\tau!}$$

Para la última desigualdad, véase que en la expresión de la izquierda hay $\tau + 1$ sumandos, luego basta con probar que cada sumando está acotado por $\frac{\eta^{\tau}}{\tau!}$. Es decir, si $i \in \{0, \dots, \tau\}$

$$\frac{\eta^i}{i!} \leq \frac{\eta^{\tau}}{\tau!} \iff \frac{\eta^{\tau-i}}{\tau(\tau-1)\dots(i+1)} \geq 1 \iff \frac{\eta}{\tau} \frac{\eta}{\tau-1} \dots \frac{\eta}{i+1} \geq 1$$

Pero la última expresión es obvia, pues todos los productos son mayores o iguales a 1 ya que $\eta = \beta\tau > \tau$, $\beta > 1$. Ahora, usando que $\tau! \geq (\tau e)^{\tau}$, se cumple,

$$\begin{aligned} P[L \geq \beta k] &\leq (\tau + 1)e^{-\eta} \frac{\eta^{\tau}}{\tau^{\tau}/e^{\tau}} = (\tau + 1)e^{-\beta\tau} \left(\frac{e\beta\tau}{\tau}\right)^{\tau} = (\tau + 1)e^{-\beta\tau} e^{\tau + \tau \ln(\beta)} \\ &= (\tau + 1)e^{-\tau[\beta - (1 + \ln \beta)]} \leq \frac{k + 2}{2} e^{-\frac{k}{2}[\beta - (1 + \ln \beta)]} \end{aligned}$$

La tercera parte se realiza de modo similar al anterior, los detalles se pueden consultar en [17]. \square

Observación 4.5.3. *Concluimos que si reescalamos la matriz con un factor $\frac{1}{\sqrt{k}}$, como ya habíamos mencionado, conseguimos que las normas se conserven en esperanza,*

$$E[\|f(u)\|_2^2] = E[\|\frac{1}{\sqrt{k}}Ru\|_2^2] = \frac{1}{k}E[\|Ru\|_2^2] = \frac{1}{k}k = 1 = \|u\|^2$$

y además el segundo y tercer apartado del teorema nos da la probabilidad de que se mantenga la norma con distorsión de β , $\frac{1}{\beta}$ respectivamente, $\beta > 1$. El lema en versión general sobre un conjunto V de n puntos de \mathbb{R}^d proviene de aplicar este resultado a los valores $u = \frac{v_i - v_j}{\|v_i - v_j\|}$, con $v_i, v_j \in V$

4.6. Relajando hipótesis (II): Simetría Esférica

De nuevo damos otro paso al frente mejorando nuestra versión, o equivalentemente, dejando caer otra de nuestras hipótesis. Es el turno de la simetría esférica en las variables de proyección. Este resultado, sorprendente por su sencillez nos asegura que no es necesario recurrir a la construcción de variables Gaussianas independientes para conseguir nuestro resultado, basta con una sencilla distribución discreta que asigna valores equiprobables al conjunto $\{-1, +1\}$, cuya complejidad a la hora de llevarlo a cabo es sustancialmente menor. Dicho avance se debe

a Achlioptas, cuyo artículo se puede consultar en [2].

Por lo tanto, y a falta de indicar las constantes que reescalan la matriz, la nueva proyección sería de la siguiente manera,

$$f(u) = R \cdot u = \begin{pmatrix} \text{---} & c_1 & \text{---} \\ \text{---} & c_2 & \text{---} \\ \vdots & & \vdots \\ \text{---} & c_k & \text{---} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ \vdots \\ u_d \end{pmatrix} = \begin{pmatrix} c_1 \cdot u \\ c_2 \cdot u \\ \vdots \\ c_k \cdot u \end{pmatrix}$$

Donde los vectores c_j están formados por d componentes independientes e con la distribución discreta que asigna de manera equiprobable los valores $\{-1, +1\}$.

Lema 4.6.1. *Sea $w = (w_1, \dots, w_d) \in \mathbb{R}^d$ un vector aleatorio cuyas componentes son independientes e idénticamente distribuidas con $E[w_j] = 0, \forall j \in \{1, \dots, d\}$, y $Var[w_j] = 1, \forall j \in \{1, \dots, d\}$, entonces:*

- $E[w_j \cdot u] = 0$
- $E[(w_j \cdot u)^2] = \|u\|_2^2$

Demostración. Basta con emplear de forma sencilla las propiedades de linealidad del operador esperanza.

$$\begin{aligned} E[c_i \cdot u] &= E\left[\sum_{j=1}^d c_{ij} u_j\right] = \sum_{j=1}^d u_j E[c_{ij}] = 0 \\ E[(c_i \cdot u)^2] &= E\left[\left(\sum_{j=1}^d c_{ij} u_j\right)^2\right] \\ &= E\left[\sum_{j=1}^d (c_{ij} u_j)^2 + \sum_{l=1}^d \sum_{m=1}^d c_{il} u_l c_{im} u_m\right] \\ &= \sum_{j=1}^d u_j^2 E[c_{ij}^2] + \sum_{l=1}^d \sum_{m=1}^d u_l u_m E[c_{il}] E[c_{im}] \\ &= \sum_{j=1}^d u_j^2 = \|u\|_2^2 \end{aligned}$$

□

Observación 4.6.2. *Nótese la generalidad en la que se ha enunciado el anterior lema. Si bien es cierto que la distribución que tratamos, que distribuye de manera independiente y equiprobable el conjunto $\{-1, +1\}$, cumple esas propiedades, el lema deja abierta la opción a numerosas distribuciones discretas sencillas.*

Lema 4.6.3. *Para todo $h \in [0, d/2]$, $d \geq 1$ y u vector unitario,*

$$E[e^{h(c \cdot u)^2}] \leq \frac{1}{\sqrt{1 - 2h/d}}$$

Donde c es un vector aleatorio en las condiciones del lema previo.

Demostración. Puede consultarse en [2]. □

Teorema 4.6.4. (Achlioptas, 2003) Sea $V = \{v_1, \dots, v_n\} \subset \mathbb{R}^d$. Dados $\epsilon, \beta > 0$ sea $k \in \mathbb{N}$ que verifique, $k \geq k_0 = \frac{4+2\beta}{\epsilon^2/2-\epsilon^3/3} \log n$. Definimos la matriz aleatoria $R \in \mathcal{M}^{k \times d}$ con $R(i, j) = r_{ij}$ donde r_{ij} son variables independientes con la siguiente distribución,

$$r_{ij} = \begin{cases} +1, & \text{con probabilidad } \frac{1}{2} \\ -1, & \text{con probabilidad } \frac{1}{2} \end{cases}$$

Consideremos la matriz $E = \frac{1}{\sqrt{k}}R$ y sea $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$ la aplicación que proyecta el espacio \mathbb{R}^d mediante dicha matriz. Con probabilidad de al menos $1 - 2n^{-\beta}$,

$$(1 - \epsilon)\|v_i - v_j\|^2 \leq \|f(v_i) - f(v_j)\|^2 \leq (1 + \epsilon)\|v_i - v_j\|^2, \forall v_i, v_j \in V$$

Demostración. Realicemos una construcción de esquema similar a la anterior. Con la notación previa, nuestra función f candidata a cumplir la propiedad, queda definida de la siguiente forma,

$$f(u) = \frac{1}{\sqrt{k}}(c_1 \cdot u, \dots, c_k \cdot u)$$

En primer lugar, notemos que, al menos en esperanza, se mantienen las normas, es decir,

$$E[\|f(u)\|_2^2] = E\left[\left\|\frac{1}{\sqrt{k}}(c_1 \cdot u, \dots, c_k \cdot u)\right\|_2^2\right] = \frac{1}{k}E\left[\sum_{i=1}^k (c_i \cdot u)^2\right] = \frac{1}{k} \sum_{i=1}^k E[(c_i \cdot u)^2] = \|u\|_2^2$$

Sin pérdida de generalidad, supongamos que tratamos con un vector u unitario. Vamos a probar que se cumple,

- $P[\|f(u)\|_2^2 > (1 + \epsilon)\frac{k}{d}] < e^{-\frac{k}{2}(\epsilon^2/2-\epsilon^3/3)}$
- $P[\|f(u)\|_2^2 < (1 - \epsilon)\frac{k}{d}] < e^{-\frac{k}{2}(\epsilon^2/2-\epsilon^3/3)}$

Veamos al primera expresión, la segunda es análoga. Usando la desigualdad de Tchebychev vista en Teorema 3.2.1,

$$P[\|f(u)\|_2^2 > (1 + \epsilon)\frac{k}{d}] = P[e^{h\|f(u)\|_2^2} > e^{h(1+\epsilon)\frac{k}{d}}] < E[e^{h\|f(u)\|_2^2}]e^{-h(1+\epsilon)\frac{k}{d}}$$

Donde el primer producto puede reescribirse, empleando la independencia de las variables,

$$E[e^{h\|f(u)\|_2^2}] = E[e^{\frac{h}{k} \sum_{j=1}^k (c_j \cdot u)^2}] = \prod_{j=1}^k E[e^{\frac{h}{k} (c_j \cdot u)^2}] = (E[e^{\frac{h}{k} (c_1 \cdot u)^2}])^k$$

Aplicamos el lema anterior y tomamos $h = \frac{d}{2} \frac{\epsilon}{1+\epsilon}$ resulta,

$$P[\|f(u)\|_2^2 > (1 + \epsilon)\frac{k}{d}] < \left(\frac{1}{\sqrt{1-2h/d}}\right)^k e^{-h(1+\epsilon)\frac{k}{d}} = ((1 + \epsilon)e^{-\epsilon})^{k/2} < e^{-\frac{k}{2}(\epsilon^2/2-\epsilon^3/3)}$$

Para la última desigualdad hemos empleado la segunda parte del Lema 4.4.1. Finalmente, resta estimar la probabilidad total del suceso, pero esto es directo de la elección de k . Efectivamente,

$$2e^{-\frac{k}{2}(\epsilon^2/2-\epsilon^3/3)} \leq 2e^{-\frac{1}{2}(4+2\beta)(\epsilon^2/2-\epsilon^3/3)^{-1}(\epsilon^2/2-\epsilon^3/3) \log n} = 2e^{-(2+\beta) \log n} = 2n^{-2-\beta} \leq 2n^{-\beta}$$

□

Observación 4.6.5. Aunque hemos enunciado el teorema para una cierta distribución, atendiendo a la demostración y a la generalidad de los lemas previos podemos concluir que dicha versión es válida para cualquier matriz cuyas componentes sigan un comportamiento independiente, de esperanza nula y de varianza 1. Podemos por tanto emplear la siguiente distribución,

$$r_{ij} = \sqrt{3} \cdot \begin{cases} +1, & \text{con probabilidad } \frac{1}{6} \\ 0, & \text{con probabilidad } \frac{2}{3} \\ -1, & \text{con probabilidad } \frac{1}{6} \end{cases}$$

Que mantiene la cota de la dimensión y la estimación de la probabilidad de que se cumpla. Además, como veremos a continuación, abre la puerta a la “siguiente mejora”, pues la matriz de proyección resulta tener un gran número de entradas nulas con lo que la complejidad algorítmica de efectuar los productos desciende notablemente.

La importancia de la distribución detallada en la observación anterior radica en el hecho de que la matriz de proyección tiene, es esperanza, $\frac{2}{3}$ de sus entradas nulas. Ello nos provoca que sus cálculos sean mucho más sencillos de llevar a cabo.

4.7. Transformada Rápida de Johnson-Lindenstrauss

Llegados a este punto la pregunta que podemos hacernos es si aún queda algo que mejorar. Hemos “perdido” en el camino la hipótesis de ortonormalidad y la de simetría esférica dejando únicamente la aleatoriedad de proyección en el subespacio, aparentemente necesaria. Nuestro interés, al hilo de la observación en la última sección, será de hacer el método lo más rápido computacionalmente posible. Ello supone simplificar el número de productos que intervienen al multiplicar la matriz por los vectores.

Definición 4.7.1. Decimos que una matriz es **escasa** (*sparse*, en inglés) si la mayoría de sus entradas son nulas. Por el contrario, si la mayoría de elementos son no nulos diremos que es **densa**.

Con esta nomenclatura, hasta ahora hemos estado tratando con matrices densas, cuya complejidad algorítmica es elevada. Una de las distribuciones propuestas por Achlioptas nos permite trabajar con una matriz relativamente escasa, con esperanza de que $\frac{2}{3}$ de sus valores sean nulos. El método que vamos a presentar a continuación, conocido como la transformada rápida de Johnson-Lindenstrauss intenta generar una matriz escasa que nos sirva de proyección y no perturbe las distancias más de un factor $(1 \pm \epsilon)$. El método es el siguiente, consideramos la matriz, $\Phi = PHD \in \mathcal{M}^{k \times d}$, donde:

- **Sparse Johnson-Lindenstrauss:** $P \in \mathcal{M}^{k \times d}$ es una matriz aleatoria cuyos elementos son independientes y con la siguiente distribución.

$$P_{ij} = \begin{cases} 1, & \text{con probabilidad } 1 - q \\ N(0, \frac{1}{q}), & \text{con probabilidad } q \end{cases}$$

Donde $q = \{\Theta(\frac{\epsilon^{p-2} \log^p n}{d}), 1\}$

- **Walsh-Hadamard:** $H \in \mathcal{M}^{d \times d}$ es una matriz determinista, llamada de Walsh-Hadamard con la siguiente expresión,

$$H_{ij} = d^{-1/2} (-1)^{\langle i-1, j-1 \rangle}$$

Donde emplamos un producto escalar binario: Transformamos los valores $i-1, j-1$ a sus vectores binarios y efectuamos el producto escalar usual.

- $D \in \mathcal{M}^{d \times d}$ es una matriz diagonal y aleatoria, cuya distribución viene dada por,

$$D_{ii} = \begin{cases} +1, & \text{con probabilidad } \frac{1}{2} \\ -1, & \text{con probabilidad } \frac{1}{2} \end{cases}$$

Observación 4.7.2. Sin pérdida de generalidad estamos suponiendo que $n \geq d = 2^m > k$

Sin embargo, trabajar a priori con matrices escasas puede ser un arma de doble filo en “casos extremos”. La matriz P es escasa de modo que sólo una fracción q de sus elementos son no nulos. Pueden surgirnos de forma natural las siguientes preguntas: ¿Por qué no usar dicha matriz como proyección?, ¿Es necesario el uso de H y D ?

La razón de que efectivamente no podemos emplear P es que el operador $\|Px\|_2$ es muy susceptible de variación en casos que el vector $x \in \mathbb{R}^d$ sea escaso. Vayamos al caso extremo en el que dicho vector consta únicamente de una coordenada. Entonces al operador $\|Px\|_2$ sólo contribuyen las entradas no nulas que queden alineadas con la única coordenada distinta de cero del vector x , dejando de controlar a priori el valor de $\|Px\|_2$.

Para dar solución a este problema entran en juego las matrices H y D , y con ella la respuesta a la segunda pregunta. No podemos evitar ni controlar que ciertos valores de \mathbb{R}^d , que pueden llegar a molestarnos, queden fuera de nuestro conjunto de proyección. Pero sí es legítimo efectuar un acondicionamiento sobre ellos para evitar el anterior problema. Antes mencionábamos que ocurriría si el vector concentra su medida en unas pocas coordenadas, el caso opuesto es cuando la norma del vector viene repartida de forma casi uniforme entre las distintas componentes del vector, y en ese caso diremos que es un vector suave (smooth, en inglés). Precisamente ese es el objetivo del producto $y = HDx$: para cualquier vector x , podemos afirmar que y es “suave”.

Observación 4.7.3. La matriz HD es ortogonal y por tanto no afecta a la norma del vector resultante, efectivamente,

$$(HD)^T HD = D^T H^T HD = D^T D = I_d$$

$$HD(HD)^T = HDD^T H^T = HH^T = I_d$$

Donde $D^T D$ es claramente la identidad, pues supone multiplicar los elementos de la diagonal, que siempre tienen el mismo signo y por tanto su producto siempre es 1. Y $H^T H$ también lo es, pues H es una matriz simétrica que al multiplicar fila i -ésima por columna i -ésima obtenemos la suma de d cuadrados de la forma $\left(\pm \frac{1}{\sqrt{d}}\right)^2$ que suman 1. La fila i -ésima por la columna j -ésima resulta 0, pues es simétrica y tiene el mismo número de elementos positivos que negativos.

Teorema 4.7.4. (Ailon-Chazelle, 2009) Sea $V = \{v_1, \dots, v_n\} \subset \mathbb{R}^d$ y sea $0 < \epsilon < 1$ y consideremos $\Phi \sim FJLT$. Con probabilidad de al menos $\frac{2}{3}$ ocurre,

$$(1 - \epsilon)k\|u\|_2 \leq \|\Phi u\|_2 \leq (1 + \epsilon)k\|u\|_2, \quad \forall u \in V$$

Donde $k = c\epsilon^{-2}$, para alguna constante c .

Demostración. Para este teorema, vamos a aportar las ideas que hay en la demostración, obviando ciertos detalles. Para una demostración completa, ver \square .

- Paso 1: Sin pérdida de generalidad, suponemos que $\|x\|_2 = 1$. Debemos comprobar que el vector HDx es “suave” en los términos anteriores para verificar que efectivamente multiplicar por HD supone un preprocesamiento de los vectores. Llamamos al vector aleatorio $u = HDx = (u_1, \dots, u_d)^T$, donde las coordenadas presentan una distribución,

$u_1 = \sum_{i=1}^d a_i x_i$ donde cada $a_i = \pm d^{-1/2}$ independientes. Además, ambos vectores tienen la misma norma, como se ha mencionado en la observación previa.

Aplicando un argumento de acotación de Chernoff (similar a la desigualdad de Tchebychev), podemos demostrar que,

$$\max_{x \in V} \|HDx\|_\infty = O(d^{-1/2} \sqrt{\log n})$$

y que $\forall s > 0$, $P[|u_j| \geq s] \leq \frac{1}{20nd}$.

- Paso 2: Considero el vector $y = (y_1, \dots, y_k)^T = Pu = \Phi x$. Por la forma en la que queda definido, $y_1 = \sum_{j=1}^d r_j b_j u_j$, donde b_1, \dots, b_d son un experimento Bernouilli de probabilidad q y las variables $r_1, \dots, r_d \sim N(0, q^{-1})$. Nuestro objetivo es estimar el valor de la variable $y_1^2 + \dots + y_k^2$. Ahora consideramos la variable $Z_1 = \sum_{j=1}^d b_j u_j^2$ que nos ayudará a estimar el valor de y_1^2 . De forma sencilla, $E[Z_1] = \sum_{j=1}^d u_j^2 E[b_j] = q$
- Paso 3: Consideramos el conjunto $\mathcal{P} = \{(a_1, \dots, a_d) : 0 \leq a_j \leq \frac{1}{m}, \forall j, \sum_{j=1}^d a_j = 1\}$, por el paso 1, $u^2 = (u_1^2, \dots, u_d^2) \in \mathcal{P}$. Denotemos por u^* un vector tal que u^{*2} es un vértice de \mathcal{P} , sin pérdida de generalidad suponemos que $u^* = (m^{-1/2}, \dots, m^{-1/2}, 0, \dots, 0)$ con las primeras m coordenadas no nulas.
- Paso 4: Definimos la variable Z^* de igual modo que Z referida al vector u^* . Se puede probar que $Z^* \sim B(m, q)/m$. Acotamos los momentos de $S = \sum_{i=1}^k Z_i$ por los de $S^* \sim B(mk, q)/m$ (donde $Z_i^* \sim Z^*$) y finalmente podemos probar que,

$$kq(1 - \epsilon) \leq \sum_{i=1}^k Z_i \leq kq(1 + \epsilon)$$

□

Capítulo 5

Aplicaciones: Compressed Sensing

5.1. Nociones Básicas

La aplicación del lema de Johnson-Lindenstrauss que vamos a estudiar con mayor detenimiento es el área de la ingeniería o matemática aplicada que se conoce por *Compressed Sensing*, cuya traducción podría ser “Detección comprimida de señales”. Es un área moderna, nacida en 2006 con el medallista Fields Terence Tao como uno de sus máximos precursores, cuyo objetivo trata la reconstrucción de señales consideradas escasas (en el sentido visto en el anterior capítulo) que han sido alteradas o pueden estar incompletas. Además, transmitir la señal suele ser un proceso muy costoso en términos de energía y que además suele verse afectado por la aplicación de un ruido externo que modifica nuestra señal original. El objetivo, será reconstruir dicha señal a partir de la información que nos llega empleando técnicas de optimización convexa y bajo la hipótesis de tratar con vectores y matrices de tipo sparse. En esta sección emplearemos las ideas fijadas en [9].

Reflejado el objetivo y el por qué del tema a tratar, vamos a darle un mayor sentido matemático. El concepto de señal, largamente estudiado desde un punto de vista físico o de ingeniería, supone tomar una serie de valores discretizados de un fenómeno continuo físico tal como pueden ser los hercios del análisis de una onda sonora. Para nosotros, una señal no es más que un vector con valores reales que definimos de la siguiente forma.

Definición 5.1.1. Llamaremos *señal* a un vector $x = (x_i)_{i=1}^d \in \mathbb{R}^d$.

Al hilo de la última parte del capítulo previo, nos centraremos en el estudio de aquellas señales cuyas entradas son mayoritariamente nulas. Definimos la noción de vector *sparse* de forma análoga al de matriz *sparse*.

Definición 5.1.2. Diremos que un vector $x \in \mathbb{R}^d$ es *t-escaso* (en inglés, *t-sparse*), si tiene un número a lo sumo t de coordenadas no nulas, es decir,

$$\|x\|_0 := \#\{i : x_i \neq 0\} \leq t$$

El conjunto de los vectores *t-escasos* se denotará por Ψ_t .

Definición 5.1.3. Denotaremos como el **Problema de Detección Comprimida** (en inglés *Compressed Sensing problem*) a recuperar una señal $x \in \Psi_t$ conociendo el siguiente vector y ,

$$y = Ax$$

Donde la matriz A es conocida como **matriz de detección** (*Sensing matrix*).

Observación 5.1.4. *Los procesos de deformación de la onda junto con el ruido, suele tener un comportamiento gaussiano por lo tanto es habitual que la modelización de la matriz de detección venga dada por una matriz aleatoria con entradas $N(0, 1)$.*

Por la forma de los vectores, con un gran número de entradas nulas, supone resolver un sistema de ecuaciones indeterminado. A continuación, podemos preguntarnos sobre cómo podríamos recuperar la señal mediante un algoritmo eficiente, y sobre el acierto en esos algoritmos. Tratemos de atacar el problema de detección comprimida mediante la resolución de,

$$(P_1) \min_x \|x\|_1 \text{ sujeto a, } y = Ax$$

Definición 5.1.5. *Si existe solución a este problema diremos que el vector es recuperable en norma $\|\cdot\|_1$*

Dado que nuestro espacio de trabajo son los vectores de Ψ_t , es razonable, dado un vector cualquiera $x \in \mathbb{R}^d$ tratar de aproximarlos por uno de nuestro conjunto de t -sparse. Con esa idea aportamos la siguiente definición.

Definición 5.1.6. *Sea $r > 0$. Llamaremos **señal compresible** en norma $\|\cdot\|_1$ $x \in \mathbb{R}^d$, con constante C y tipo r si,*

$$\sigma_t(x) := \min_{y \in \Psi_t} \|x - y\|_1 \leq C \cdot t^{-r}, \text{ con } t \in \{1, \dots, d\}$$

5.2. Matrices RIP

Vamos a introducir un caso muy importante de matrices de detección, las llamadas matrices RIP.

Definición 5.2.1. *Diremos que $A \in \mathcal{M}^{k \times d}$ tiene la **propiedad de la isometría restringida**, en inglés *Restricted Isometry Property (RIP)*, de orden t si existe $\delta_t \in (0, 1)$ de modo que,*

$$(1 - \delta_t) \|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_t) \|x\|_2^2, \forall x \in \Psi_t$$

En base a la expresión que define este tipo de matrices, podemos vislumbrar un parecido más que evidente con las matrices aleatorias que tratábamos en el Lema de Johnson-Lindenstrauss, aunque ahora únicamente lo dediquemos a vectores t -sparse y no se restringe a un conjunto finito. El siguiente teorema nos estima el error cometido al aproximar la señal x mediante la solución del problema de minimizar en norma $\|\cdot\|_1$ cuando tratamos con matrices de detección RIP.

Teorema 5.2.2. *Sea $A \in \mathcal{M}^{k \times d}$ que satisface la condición RIP de orden $2t$ con $\delta_{2t} < \sqrt{2} - 1$. Sea $x \in \mathbb{R}^d$ y tomamos \hat{x} una solución del problema (P_1) . Entonces*

$$\|x - \hat{x}\|_2 \leq C \cdot \left(\frac{\sigma_t(x)}{\sqrt{t}} \right)$$

Demostración. Ver [7]. □

Observación 5.2.3. *En particular, si el vector es t -sparse, entonces $\sigma_t(x) = 0$. Con la notación empleada, concluimos que para todo vector t -sparse recuperable en norma $\|\cdot\|_1$, la solución del problema de minimización coincide con la señal que buscamos.*

5.3. Algoritmos de recuperación

Hemos de mencionar que la teoría antes desarrollada nos aporta la existencia de solución al problema de detección comprimida pero no un método efectivo para encontrarla. El problema en norma $\|\cdot\|_1$ puede tratarse como,

$$\min_x \|x\|_1, \text{ sujeto a } \|Ax - y\|_2^2 \leq \epsilon$$

Considerando la existencia de un ruido que altera la condición $Ax = y$. Para un parámetro $\lambda > 0$ adecuado podemos resolver el siguiente problema,

$$\min_x \frac{1}{2} \|Ax - y\|_2^2 + \lambda \|x\|_1$$

Que se trata de un problema de optimización convexa para el cual podemos recurrir al método del gradiente proyectado, que puede verse en [14, pág. 586-597]

5.4. Relación con el Lema de Johnson-Lindenstrauss

Hasta ahora hemos dado una ligera introducción a un tema en auge y realmente interesante, como es la recuperación de señales a través de transmisiones. El hecho de que ocupe un lugar destacado en el capítulo de aplicaciones es debido a su conexión con nuestro objeto de estudio, el lema de Johnson-Lindenstrauss. En esta sección nos serviremos de [3] como guía en las demostraciones.

Al final de la sección anterior hemos introducido las matrices *RIP* que además nos ayudaban a reconstruir en norma $\|\cdot\|_1$ los vectores *t*-sparse. A continuación estudiaremos la relación entre las matrices *RIP* y aquellas $A \in \mathcal{M}^{k \times d}$ que satisfagan el siguiente fenómeno de concentración para todo $x \in \mathbb{R}^d$ fijo pero arbitrario,

$$P[(1 - \epsilon)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \epsilon)\|x\|_2^2] \geq 1 - 2e^{-c_0 \epsilon^2 k}$$

En otras palabras, aquellas matrices que verifiquen el lema de Johnson-Lindenstrauss. La relación entre ambas propiedades queda plasmada en el siguiente teorema, cuyo significado nos arroja que

$$JL \implies RIP$$

Fijaremos nuestro interés en aquellos vectores que tienen un número fijo de coordenadas no nulas dentro del espacio \mathbb{R}^d . A continuación introducimos la notación respectiva a este subespacio, es un concepto similar al de *k*-sparse, pero ahora fijamos de antemano las coordenadas nulas.

Definición 5.4.1. Sea $T = \{i_1, \dots, i_s\}$ un conjunto de índices de cardinal $s \leq N$. Denotamos por X_T el conjunto de vectores de \mathbb{R}^n que son nulos en los índices que no pertenecen al conjunto T , es decir,

$$X_T = \{x \in \mathbb{R}^d : x_j = 0 \text{ si } j \notin T\}$$

Teorema 5.4.2. Sea Φ una matriz de tamaño $k \times d$ de modo que satisface el fenómeno de concentración anterior. Entonces, para cualquier $0 < \delta < 1$ tenemos,

$$(1 - \delta)\|x\|_2 \leq \|\Phi x\|_2 \leq (1 + \delta)\|x\|_2, \forall x \in X_T$$

con probabilidad de al menos $1 - 2\left(\frac{12}{\delta}\right)^s e^{-c_0(\delta/2)^2 k}$, con $s = \#T$ y c_0 una constante universal.

Demostración. Sin pérdida de generalidad, por la linealidad de multiplicar por una matriz, podemos suponer que $\|x\|_2 = 1$. Sea s el cardinal del conjunto T , y escogemos un conjunto finito $Q_T \subset X_T$ de modo que $\|q\|_2 = 1$, $\forall q \in Q_T$ y que además verifique que si $x \in X_T$, entonces:

$$\min_{q \in Q_T} \|x - q\|_2 \leq \frac{\delta}{4}$$

Como se demuestra en [24], podemos elegir un conjunto Q_T de cardinal acotado por $(\frac{12}{\delta})^s$. Como el conjunto Q_T es finito, aplicamos la hipótesis de concentración para $\epsilon = \delta/2 \in (0, 1)$, y resulta que, con una probabilidad superior a $1 - 2(\frac{12}{\delta})^s e^{-c_0(\delta/2)^{2k}}$, se verifica,

$$\left(1 - \frac{\delta}{2}\right) \|q\|_2^2 \leq \|\Phi q\|_2^2 \leq \left(1 + \frac{\delta}{2}\right) \|q\|_2^2, \forall q \in Q_T$$

Ahora definimos A como el mínimo valor para el cuál se verifica,

$$\|\Phi x\|_2 \leq (1 + A)\|x\|_2, \forall x \in X_T$$

Nuestro objetivo será demostrar que $A \leq \delta$ pudiendo así concluir la acotación superior de la norma $\|\Phi x\|_2$. Para ello tomamos $q \in Q_T$,

$$\|\Phi x\|_2 \leq \|\Phi q\|_2 + \|\Phi(x - q)\|_2 \leq 1 + \frac{\delta}{2} + (1 + A)\frac{\delta}{4}$$

Para acotarlo empleamos la propiedad de concentración junto con la definición de A y la distancia entre los elementos de Q_T y X_T . Además, A es el valor mínimo que verifica la acotación, por tanto,

$$A \leq \frac{\delta}{2} + (1 + A)\frac{\delta}{4} = \frac{3}{4}\delta + A\frac{\delta}{4} \iff A\left(1 - \frac{\delta}{4}\right) \leq \frac{3}{4}\delta \iff A \leq \frac{(3/4)\delta}{(1 - \delta/4)} = \frac{3}{4 - \delta}\delta \leq \delta$$

Con lo que hemos probado la cota superior de la desigualdad. Para la inferior,

$$\|\Phi x\|_2 \geq \|\Phi q\|_2 - \|\Phi(x - q)\|_2 \geq 1 - \frac{\delta}{2} - (1 + \delta)\frac{\delta}{4} \geq 1 - \delta$$

□

Observación 5.4.3. Véase que hemos conseguido extender la propiedad de JL a un conjunto infinito con un número fijo de coordenadas no nulas, en otras palabras, hemos probado la propiedad RIP.

De hecho, la propiedad de ser matriz RIP y de cumplir el lema de concentración de Johnson-Lindenstrauss, se caracterizan en el siguiente sentido.

Teorema 5.4.4. Sea $\eta > 0$, $0 < \epsilon < 1$. Consideremos un conjunto finito $V \subset \mathbb{R}^d$ de cardinalidad n . Sea $p > 40 \log \frac{4n}{\eta}$ y supongamos que $\Phi \in \mathcal{M}^{k \times d}$ satisface la condición RIP de orden p y $\delta \leq \frac{\epsilon}{4}$. Sea $\xi \in \mathbb{R}^d$ una secuencia de variables independiente de distribución equiprobable sobre $\{-1, +1\}$. Entonces con probabilidad de al menos $1 - \eta$,

$$(1 - \epsilon)\|x\|_2^2 \leq \|\Phi D_\xi x\|_2^2 \leq (1 + \epsilon)\|x\|_2^2, \forall x \in V$$

Demostración. Ver [21].

□

Capítulo 6

Conclusiones y Propuestas de Futuro

6.1. Cuestiones de interés

Vamos a comenzar este capítulo nombrando algunos de los resultados más importantes vistos en el trabajo pero que por falta de extensión y tiempo no han tenido cabida en la redacción de la memoria, y que quizá puedan ser abordadas en un futuro trabajo:

1) Respecto al capítulo referido al problema isoperimétrico, cabe decir que M. Gromov establece en su artículo *“Paul Levy’s Isoperimetric Inequality”* una generalización de las desigualdades isoperimétricas estudiadas en la Esfera, pero ahora sobre Variedades Riemannianas compactas y comprueba además que son consistentes sobre \mathbb{S}^{n-1} .

2) A lo largo del trabajo hemos estudiado el Lema de Johnson-Lindenstrauss sobre el espacio Euclídeo \mathbb{R}^d , cabe preguntarse si en un espacio normado cualquiera $(X, \|\cdot\|)$ se verifica la existencia de una aplicación Lipschitz que proyecte los elementos de un cierto conjunto sobre un subespacio aleatorio conservando las distancias de los pares de elementos. Assaf Naor y el propio William B. Johnson, en su artículo *“The JL lemma almost characterizes Hilbert Space, but not quite”* demuestran que si un espacio cumple dicha propiedad, entonces es *“casi Euclídeo”* en el sentido de que para todo subespacio de dimensión n podemos establecer un homeomorfismo sobre su imagen a un espacio Hilbert con distorsión $2^{2^{O(\log n)}}$.

3) Sobre aquellas hipótesis que asumíamos y tratábamos de relajar, hemos mencionado que la de proyección aleatoria era fundamental en nuestro trabajo y por tanto no podíamos eliminarla. Sin embargo, lo cierto es que en la actualidad se buscan formas de determinar el subespacio de proyección sin recurrir a la aleatoriedad. En esta dirección nos encontramos con el artículo *“Derandomized dimensionality reduction with applications”* de Lars Engebretsen junto con el propio Indyk. En él, construyen una proyección determinista que cumple la propiedad de no distorsionar los pares de distancias de un conjunto finito.

6.2. Reflexiones finales

En el trabajo hemos intentado entrelazar herramientas de la Teoría de la Probabilidad junto con Análisis Matemático, primero mencionando el método probabilístico de demostración y después sentando bases analíticas de resultados a la postre enteramente desarrollados mediante cálculo de probabilidades. Hemos podido comprobar su aparición en resultados tan fundamentales como los Teoremas de Dvoretzky y en esta dirección queremos reflejar una cierta reflexión.

Aunque no ha aparecido en el trabajo, el teorema central del límite se enuncia de la siguiente manera,

Teorema 6.2.1. Teorema Central del Límite Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes e idénticamente distribuidas con $E[X_j] = \mu$, $Var[X_j] = \sigma^2$. Llamamos $S_n = \sum_{i=1}^n X_i$. Entonces,

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \rightarrow N(0, 1)$$

Donde la convergencia se expresa en distribución.

En otras palabras, nos dice que dada una variable aleatoria X de media y varianza finita, mediante un número de copias independientes de ella misma podemos obtener una variable aleatoria Gaussiana. Similar resulta al Teorema de Dvoretzky visto en el Teorema 3.3.5, cuyo enunciado nos dice que dado un Espacio normado cualquiera, de él podemos extraer una copia del espacio Euclídeo ℓ_2^n . Supone un primer paralelismo entre las variables Gaussianas y los espacios Euclídeos que toma mayor expresión en la versión Gaussiana del Teorema de Dvoretzky, visto en Teorema 3.3.9, que nos equipara la búsqueda de la variable aleatoria Gaussiana en un espacio de Banach con la de la copia del subespacio Euclídeo. De manera intuitiva y quizá de modo poco riguroso, podemos decir que: “Las variables Gaussianas son a la Teoría de la Probabilidad lo que los Espacios Euclídeos al Análisis Funcional”.

Por otra parte, cabe retomar el “círculo natural” del que hablábamos sobre el Lema de Johnson-Lindenstrauss, el Problema Isoperimétrico en \mathbb{S}^{n-1} y los Fenómenos de concentración. Atendiendo a la demostración original del Lema podemos establecer las siguientes relaciones:

- El Lema de Johnson-Lindenstrauss es un ejemplo de fenómeno de concentración de la medida mediante funciones Lipschitz.
- La concentración en funciones 1-Lipschitz en la esfera implica la existencia de desigualdades isoperimétricas.
- La existencia de desigualdades isoperimétricas en la esfera implica la concentración de funciones 1-Lipschitz en dicho espacio (*Recíproco*).
- El Lema de Levy relaciona la concentración de funciones continuas en la esfera con la desigualdad isoperimétrica.
- Finalmente, la desigualdad isoperimétrica es la llave fundamental en la prueba del lema de Johnson-Lindenstrauss.

Ahora, atendiendo a la demostración de Gupta y Dasgupta,

- Se relaciona la concentración en \mathbb{S}^{n-1} con la concentración en subespacios aleatorios k -dimensionales mediante el estudio de la norma de la variable,

$$\frac{(X_1, \dots, X_k)}{\|X\|_2^2}$$

Entra en juego con la relación $\|U^*QUx\|_2^2 = \|QUx\|_2^2$, con Q proyección a k coordenadas y $U \in \mathcal{O}(n)$. Que nos dice que desde el punto de vista de la norma, es lo mismo proyectar a un subespacio un vector fijo, que proyectar a un subespacio fijo un vector aleatorio de \mathbb{S}^{n-1} .

Bibliografía

- [1] N. AILON y B. CHAZELLE, *The fast Johnson-Lindenstrauss Transform and approximate nearest neighbors*, Siam J. comput., Vol 39, (2009), pp. 302-322.
- [2] D. ACHLIOPTAS, *Database-friendly random projections: Johnson-Lindenstrauss with binary coins*, Journal Computer System Science, Vol. 66 (2003), pp. 671-687.
- [3] R. BARANIUK et al, *A Simple Proof of the Restricted Isometry Property for Random Matrices*, Springer Science+Business Media, (2006).
- [4] A. BARVINOK, *Measure Concentration*, (2005).
- [5] Y. BENYAMINI y J. LINDENSTRAUSS, *Geometric nonlinear functional analysis, Volume 1*, American Mathematical Society-Colloquium Publications, Vol 48, (1991).
- [6] A. BOISBUNON, *The class of multivariate spherically symmetric distributions*, Technical Report, (2005).
- [7] A. COHEN, W. DAHMEN y R. DE VORE, *Compressed sensing and best k-term approximation*, Journal American Mathematical Society, Vol 22, (2009), pp. 211-231.
- [8] S. DASGUPTA y A. GUPTA, *An elementary proof of the Johnson-Lindenstrauss Lemma*, Technical report, Vol 99-006, (1999).
- [9] M. DUARTE y Y. C. ELДАР, *Compressed sensing*, IEEE, (2011).
- [10] D. L. DONOHO, *Structured Compressed Sensing: From Theory to Applications*, IEEE Trans. Info. Theory, vol. 52, (2006), pp. 1289–1306.
- [11] A. DVORETZKY, *Some results on convex bodies and banach spaces*, Proc. Internat. Sympos. Linear Spaces, (1961), pp. 123-160.
- [12] J.A. FACENDA y F.J. FRENICHE, *Integración en funciones de varias variables*, Pirámide, (2002), pp. 21-52.
- [13] T. FIGIEL, J. LINDENSTRAUSS y V.D. MILMAN, *The dimension of almost spherical sections of convex bodies*, Bulletin American Mathematical Society, Vol 82, (1988), pp. 575-578.
- [14] M. FIGUEIREDO, R. NOWAK y S. WRIGHT, *Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems*, IEEE J. Sel. Top. Signa., (2007), pp. 586-5797.
- [15] W. FELLER, *An introduction to probability theory and its applications*, Wiley, 3ed edition, Vol 2, (1971), pp. 525-527.

- [16] P. FRANKL y H. MAEHARA, *The Johnson-Lindenstrauss lemma and the sphericity of some graphs*, Journal Combinatorial Theory, Vol 44, (1988), pp. 355-362.
- [17] S. HAR-PELED, *Geometric Approximation Algorithms*, (2006).
- [18] P. INDYK y R. MOTWANI, *Approximate nearest neighbors: Towards removing the curse of dimensionality*, Proceedings of the 30th Annual ACM Symposium on Theory of Computing, (1998), pp. 604-613.
- [19] W. B. JOHNSON y J. LINDENSTRAUSS, *Extensions of Lipschitz mappings into a Hilbert space*, Contemporary Mathematics, Vol 26, (1984).
- [20] S. JANSON, *Gaussian Hilbert Space*, Cambridge University Press, (1997).
- [21] F. KRAHMER y R. WARD *New and improved Johnson-Lindenstrauss embeddings via the Restricted Isometry Property*, Cornell University Library, (2011).
- [22] H. KUO, *Gaussian measures in banach spaces*, Springer, (1975), pp. 1-36.
- [23] M. LEDOUX, *The concentration of measure phenomenon*, American Mathematical Society-Board, (2000), pp. 1-44.
- [24] G. LORENTZ, M. V GOLITSCHKEK y Y. MAKOVOZ *Constructive Approximation: Advanced Problems*, Springer, (1996), pp. 399-445.
- [25] P. MATTILA, *Geometry of sets and measures in euclidean spaces, fractals and rectifiability*, Cambridge University Press, (1995), pp. 44-48.
- [26] M.E. MUNROE, *Introduction to measure and integration*, Addison-Wesley, (1953).
- [27] A. NAOR, *Concentration of Measure*, (2008).
- [28] P.J. HERRERO, *Una historia del problema isoperimétrico clásico, con geometría elemental*, (2011).
- [29] G. PISIER, *Probabilistic Methods in the Geometry of Banach Spaces*, (1986), pp. 168-184.
- [30] A. SAHA, *A survey of Johnson Lindenstrauss transform - methods, extensions and applications*, (2009).
- [31] G. SCHECHTMAN, *Concentration, results and applications*, Elsevier preprint, (1999), pp. 1-15.
- [32] R. SCHNEIDER, *Convex Bodies: the Brunn-Minkowski Theory*, Cambridge University Press, (1993), pp. 369-440.
- [33] S. VEMPALA, *The random projection method*, American Mathematical Society, Vol. 65, (2004).
- [34] P. VIRGILIO, *La Eneida*, EDAF, (1985).
- [35] P. ZOROA y N. ZOROA, *Elementos de probabilidades*, Diego Marin, (2008), pp. 248-289.