

Dimensión de Hausdorff en el espacio de distribuciones de probabilidad con técnicas de teoría de la información



Álvaro de la Asunción Deza
Trabajo de fin de grado de Matemáticas
Universidad de Zaragoza

Directora del trabajo: Elvira Mayordomo
Directora del trabajo: Carmen Sanguesa
3 de diciembre de 2023

Abstract

Let's imagine a complex dataset. We can describe it in terms of the relevant data that provides information about the set and the secondary data that introduces redundancies and noise. This concept is particularly interesting in compression programs (such as zip, tar, ...), feature selection, or analyzing the complexity of a dataset. We will use the concept of effective dimension to quantify the relevant information.

Now, consider a figure that is not exactly 2D or 3D but occupies an intermediate space. These are fractals, objects that challenge intuition by having non-integer dimensions. For their study, we will introduce the concept of fractal dimensions.

Throughout the document, we will encounter concepts such as Kolmogorov complexity, along with different theorems that will lead us to prove that algorithmic dimensions (such as effective dimension) of points in a specific set can establish the fractal dimensions of that set.

The dataset we will work with is the set of probability measures. In other words, all the concepts developed are oriented towards probability measures. Therefore, it is necessary to begin the document with a chapter explaining the specific metric that will be used: the Prokhorov metric. We will demonstrate that the set of probability measures, together with this metric, forms a metric space, which also inherits some characteristics from the original set.

We will conclude by showing an example in which we compute the effective dimension of the measures from a specific set. Applying the theorems studied, we establish the fractal dimensions of that set. The purpose of this example is to provide an idea of the steps that need to be followed to define or bound fractal dimensions.

Índice general

Abstract	III
1. Convergencia débil de medidas de probabilidad y métrica de Prokhorov	1
1.1. Convergencia débil de medidas	2
1.2. Métrica de Prokhorov	3
2. Las dimensiones fractales y la teoría de la información	11
2.1. Dimensión de Hausdorff y de empaquetamiento	11
2.1.1. Dimensión con escala	13
2.2. Complejidad de Kolmogorov	13
2.3. Dimensión efectiva	16
2.4. Oráculos y relativización	16
2.5. El principio punto a conjunto	17
3. Ejemplos de cálculo de dimensión efectiva de medidas de probabilidad	19
3.1. Distribución uniforme	20
3.1.1. Mixtura de uniformes	22
3.2. Distribución exponencial	23
Bibliografía	25

Capítulo 1

Convergencia débil de medidas de probabilidad y métrica de Prokhorov

La métrica de Prokhorov tiene especial relevancia en varios campos de las matemáticas como la teoría de la probabilidad, la estadística o la teoría de la información. Proporciona una distancia entre dos distribuciones de probabilidad y nos permite cuantificar el grado de similitud entre ellas. A lo largo de este capítulo exploraremos sus propiedades y veremos que bajo ciertas condiciones, la convergencia bajo esta métrica es equivalente a la convergencia débil, que es un tipo de convergencia entre variables aleatorias muy utilizado.

Comenzamos viendo dos conceptos fundamentales que serán necesarios en los posteriores apartados: el espacio métrico y las medidas de Borel finitas. La idea de espacio métrico proporciona una estructura matemática que nos permite medir distancias entre puntos de un conjunto.

Definición 1. Sea X un conjunto. Una aplicación $d : X \times X \rightarrow \mathbb{R}$ es una *métrica* si $\forall x, y, z \in X$ cumple las siguientes condiciones:

- (1) $d(x, y) \geq 0$
- (2) $d(x, y) = 0 \iff x = y$
- (3) $d(x, y) = d(y, x)$
- (4) $d(x, y) \leq d(x, z) + d(z, y)$ (*desigualdad triangular*)

Denominamos *espacio métrico* al par (X, d) .

Las *medidas de Borel de probabilidad* asignan probabilidades a conjuntos borelianos, es decir, a conjuntos que pertenecen a la σ -álgebra generada por los conjuntos abiertos de un espacio métrico X . A dicha σ -álgebra la denotaremos $\mathcal{B}(X)$.

Definición 2. Sea $(X, \mathcal{B}(X))$ un espacio métrico. Una *medida de Borel finita* es una aplicación $\mu : \mathcal{B}(X) \rightarrow [0, \infty)$ nula en \emptyset y numéricamente aditiva, es decir, tal que:

1. $\mu(\emptyset) = 0$
2. Si $A_1, A_2, \dots \in \mathcal{B}(X)$ son conjuntos medibles disjuntos entre sí, entonces

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

Si además $\mu(X) = 1$ llamaremos a μ *medida de Borel de probabilidad*.

Nuestro interés será estudiar la distancia entre diferentes medidas de Borel de probabilidad. A partir de aquí, siempre que se nombren medidas serán medidas de Borel. Es importante conocer la siguiente propiedad de las medidas de probabilidad para alguno de los teoremas que mostraremos en este capítulo.

Definición 3. Sea $(X, \mathcal{B}(X))$ un espacio métrico. Decimos que una medida de probabilidad μ es *ajustada* si para todo $\epsilon > 0$ existe un compacto $K \subset X$ tal que $\mu(K) > 1 - \epsilon$. Un conjunto de medidas Π se dice que es ajustado si para todo $\epsilon > 0$ y para toda $\mu \in \Pi$ existe un compacto K tal que $\mu(K) > 1 - \epsilon$.

1.1. Convergencia débil de medidas

La convergencia débil de medidas describe cómo una sucesión de medidas se comporta al acercarse a una medida límite. La idea es que cuando la sucesión avanza, las medidas se vuelven más parecidas a la medida límite en términos de sus propiedades estadísticas o características de distribución.

Sea (X, d) un espacio métrico y denotemos

$$C_b(X) := \{f : X \rightarrow \mathbb{R} \mid f \text{ es continua y acotada}\}$$

Notar que toda $f \in C_b(X)$ es integrable respecto de cualquier medida de Borel finita en X .

Definición 4. Sean μ, μ_1, μ_2, \dots medidas finitas definidas sobre X . Decimos que $(\mu_n)_{n \in \mathbb{N}}$ *converge débilmente* a μ , y lo denotamos como $\mu_n \Rightarrow \mu$, si

$$\int_X f \cdot d\mu_n \longrightarrow \int_X f \cdot d\mu \quad \text{cuando } n \rightarrow \infty \quad \forall f \in C_b(X)$$

El siguiente resultado establece criterios útiles para verificar la convergencia débil de sucesiones de distribuciones de probabilidad. Por motivos de espacio lo enunciamos sin demostración, pero se puede encontrar la demostración en [1].

Teorema 1.1 (Teorema de Portmanteau). Sean (X, d) un espacio métrico y μ, μ_1, μ_2, \dots medidas de probabilidad definidas sobre X . Son equivalentes:

- i) $\mu_n \Rightarrow \mu$
- ii) $\int_X f \cdot d\mu_n \longrightarrow \int_X f \cdot d\mu$ para toda función f acotada y uniformemente continua.
- iii) $\limsup_{n \rightarrow \infty} \mu_n(C) \leq \mu(C)$ para todo cerrado $C \subseteq X$.
- iv) $\liminf_{n \rightarrow \infty} \mu_n(U) \geq \mu(U)$ para todo abierto $U \subseteq X$.
- v) $\mu_n(A) \longrightarrow \mu(A)$ para todo boreliano $A \subseteq X$ con $\mu(\partial A) = 0$, donde ∂A denota la frontera de A .

Veamos un ejemplo para ilustrar el significado de estas condiciones.

Ejemplo 1. Sea (X, ρ) un espacio métrico, y δ_x una medida de probabilidad sobre $\mathcal{B} = \mathcal{B}(X)$ definida como $\delta_x(A) = I_A(x)$, donde $I_A(\cdot)$ denota la función indicadora. Si $x_n \rightarrow x_0$ y f es continua, tenemos que

$$\int_X f \cdot d\delta_{x_n} = f(x_n) \longrightarrow f(x_0) = \int_X f \cdot d\delta_{x_0} \quad (1.1)$$

y por tanto $\delta_{x_n} \Rightarrow \delta_{x_0}$. De hecho vamos a comprobar que esta convergencia débil es equivalente a la convergencia puntual.

Denotamos la distancia entre x y un conjunto F como $\rho(x, F) = \inf\{\rho(x, a) : a \in F\}$. Tomamos $\epsilon > 0$ y para $C \in X$ cerrado definimos la función continua $f(x) = \max\{1 - \rho(x, C)/\epsilon, 0\}$. Suponemos que los x_n son todos diferentes de x_0 , por ejemplo, $x_0 = 0$ y $x_n = 1/n$. Entonces si tomamos $C = \{x_0\}$ se cumple la desigualdad (iii) de forma estricta:

$$0 = \limsup_{n \rightarrow \infty} \delta_{x_n}(\{x_0\}) < \delta_{x_0}(\{x_0\}) = 1$$

y si $U = \{x_0\}^c$ se cumple la desigualdad (iv) de forma estricta:

$$1 = \liminf_{n \rightarrow \infty} \delta_{x_n}(\{x_0\}^c) > \delta_{x_0}(\{x_0\}^c) = 0$$

Además notar que si tomamos $A = \{x_0\}$ no se cumple la convergencia $\delta_{x_n}(A) \rightarrow \delta_{x_0}(A)$, pero el teorema no se contradice porque $\delta_{x_0}(\partial\{x_0\}) = \delta_{x_0}(\{x_0\}) = 1 \neq 0$.

Por otro lado, si $x_n \not\rightarrow x_0$, $\exists N > 0$ tal que $\rho(x_0, x_n) > \epsilon$, $\forall n > N$. Entonces si $C = \{x_0\}$ se tiene que $f(x_0) = 1$ y $f(x_n) = 0$, $\forall n > N$. Luego no se cumple (1.1) y $\delta_{x_n} \not\Rightarrow \delta_{x_0}$.

En conclusión, $\delta_{x_n} \Rightarrow \delta_{x_0} \iff x_n \rightarrow x_0$.

Es interesante ver la relación entre la convergencia débil y la convergencia en distribución dado que en ocasiones son utilizadas indistintamente en el contexto de la teoría de la probabilidad y la estadística.

Ejemplo 2 (Equivalencia entre convergencia en distribución y convergencia débil cuando $S = \mathbb{R}$). Hay que recordar que dado un espacio de probabilidad (Ω, \mathcal{A}, P) y un espacio medible $(S, \mathcal{B}(S))$, una variable aleatoria es una función $Y : \Omega \rightarrow S$ tal que $Y^{-1}(B) \in \mathcal{A}$ para todo $B \in \mathcal{B}(S)$. Si $S = \mathbb{R}$ se llama función de distribución de la variable aleatoria a la función $F : \mathbb{R} \rightarrow [0, 1]$ tal que $F(x) = P(\omega \in \Omega : Y(\omega) \leq x)$.

Sea $(Y_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias tomando valores en $S = \mathbb{R}$ y $(F_n)_{n \in \mathbb{N}}$ su correspondiente sucesión de funciones de distribución. Diremos que Y_n converge a en distribución a una variable aleatoria Y con función de distribución F , y lo denotaremos como $Y_n \xrightarrow{D} Y$, si para todo punto de continuidad de F se tiene

$$\lim_{n \rightarrow \infty} F_n(x) = F(x). \quad (1.2)$$

En este caso μ_n representa la medida de probabilidad engendrada por Y_n sobre los borelianos de \mathbb{R} , esto es $\mu_n(A) = P(\omega \in \Omega : Y_n(\omega) \in A)$, para todo $A \in \mathcal{B}(\mathbb{R})$. Por (1.1.v) es claro que la convergencia débil implica la convergencia en distribución pues basta con tomar $A = (-\infty, x]$. Además, si Y_n converge en distribución a Y , entonces por el teorema de la convergencia dominada [9] (página 335) se tiene que la sucesión verifica el teorema (1.1.ii) y por tanto, sus medidas de probabilidad subyacentes convergen débilmente.

1.2. Métrica de Prokhorov

Vamos a ver cómo una métrica d sobre un espacio X genera una distancia entre medidas de probabilidad definidas sobre X . Esta métrica se llama la métrica de Prokhorov. A continuación, veamos cómo se define.

Definición 5. Sea (X, d) un espacio métrico. Denotamos

$$\mathcal{P} = \mathcal{P}(X) := \text{medidas de Borel de probabilidad en } X$$

Tomamos $\mu, \nu \in \mathcal{P}$ y definimos la *métrica de Prokhorov* sobre \mathcal{P} (inducida por d) como la función

$$d_P(\mu, \nu) := \inf\{\alpha > 0 : \mu(A) \leq \nu(A_\alpha) + \alpha, \nu(A) \leq \mu(A_\alpha) + \alpha \quad \forall A \in \mathcal{B}(X)\}$$

donde

$$A_\alpha := \{x : d(x, A) < \alpha\} \text{ si } A \neq \emptyset \quad \text{y} \quad \emptyset_\alpha = \emptyset \quad \forall \alpha > 0$$

Notar que cualquier $\alpha \geq 1$ pertenece al conjunto que define d_P , luego el ínfimo está bien definido. Veamos que efectivamente es una métrica.

1. Es directo que $d_P(\mu, \nu) \geq 0$ y $d_P(\mu, \nu) = d_P(\nu, \mu) \quad \forall \mu, \nu \in \mathcal{P}$.
2. Por un lado, para todo $A \in \mathcal{B}(X)$ y $\alpha > 0$, $A \subseteq A_\alpha$, luego $\mu(A) \leq \mu(A_\alpha) + \alpha$, por tanto $d_P(\mu, \mu) \leq \alpha$ y se tiene que $d_P(\mu, \mu) = 0$.

Por otro lado, si $d_P(\mu, \nu) = 0$, entonces existe una sucesión $\alpha_n \searrow 0$ tal que $\mu(A) \leq \nu(A_{\alpha_n}) + \alpha_n$ y $\nu(A) \leq \mu(A_{\alpha_n}) + \alpha_n$ para todo n . Como $\bar{A} = \bigcap_n A_{\alpha_n}$, donde \bar{A} denota la clausura de A , tenemos que $\mu(A) \leq \nu(\bar{A})$ y $\nu(A) \leq \mu(\bar{A})$. De ahí se sigue que $\mu(A) = \nu(A)$ para todo cerrado A . Es trivial que los borelianos A tales que $\mu(A) = \nu(A)$ tienen estructura de σ -álgebra. Como esta σ -álgebra contiene a los abiertos, incluye a la σ -álgebra de los borelianos y tenemos que $\mu = \nu$.

3. Demostramos ahora la *desigualdad triangular*: sean $\mu, \nu, \eta \in \mathcal{P}$. Tomamos $\alpha > 0$ tal que

$$\mu(A) \leq \eta(A_\alpha) + \alpha, \quad \eta(A) \leq \mu(A_\alpha) + \alpha \quad \forall A \in \mathcal{B}(X)$$

y $\beta > 0$ tal que

$$\nu(A) \leq \eta(A_\beta) + \beta, \quad \eta(A) \leq \nu(A_\beta) + \beta \quad \forall A \in \mathcal{B}(X).$$

Luego para todo $A \in \mathcal{B}(X)$ tenemos que

$$\begin{aligned} \mu(A) &\leq \eta(A_\alpha) + \alpha \leq \nu((A_\alpha)_\beta) + \alpha + \beta \\ \nu(A) &\leq \eta(A_\beta) + \beta \leq \mu((A_\beta)_\alpha) + \beta + \alpha \end{aligned}$$

Notar que $(A_\alpha)_\beta \subseteq A_{\alpha+\beta}$. Dado que si $x \in (A_\alpha)_\beta \Rightarrow d(x, A_\alpha) < \beta \Rightarrow \exists y \in A_\alpha$ tal que $d(x, y) < \beta$ y como $y \in A_\alpha \Rightarrow \exists a \in A$ tal que $d(y, a) < \alpha$, luego $d(x, a) \leq d(x, y) + d(y, a) < \alpha + \beta$, por tanto $x \in A_{\alpha+\beta}$. De igual manera $(A_\beta)_\alpha \subseteq A_{\alpha+\beta}$.

Por tanto para todo $A \in \mathcal{B}$ tenemos que

$$\begin{aligned} \mu(A) &\leq \nu(A_{\alpha+\beta}) + \alpha + \beta \\ \nu(A) &\leq \mu(A_{\alpha+\beta}) + \alpha + \beta \end{aligned}$$

y por la definición, $d_P(\mu, \nu) \leq \alpha + \beta$. Tomando ínfimos sobre α y β obtenemos

$$d_P(\mu, \nu) \leq d_P(\mu, \eta) + d_P(\eta, \nu).$$

Observar que \mathcal{P} con esta métrica es acotado dado que la distancia entre dos medidas de probabilidad es siempre menor o igual que 1.

Ejemplo 3. Sea (X, d) un espacio métrico:

1. Dados $x_0, x_1 \in X$, la distancia de Prokhorov entre sus medidas indicadoras es

$$d_P(\delta_{x_0}, \delta_{x_1}) = \min \{d(x_0, x_1), 1\}$$

Si $d(x_0, x_1) \geq 1$, tomando un conjunto $A \in \mathcal{B}(X)$ tal que $x_0 \in A, x_1 \notin A$ y $d(x_1, A) \geq 1$, es claro que sólo con $\alpha \geq 1$ se cumple la desigualdad $\delta_{x_0}(A) \leq \delta_{x_1}(A_\alpha) + \alpha$.

Supongamos ahora que $d(x_0, x_1) = r < 1$. Si tomamos un conjunto $A \in \mathcal{B}(X)$ tal que $x_0 \in A$ y $x_1 \notin A$, la desigualdad $\delta_{x_1}(A) \leq \delta_{x_0}(A_\alpha) + \alpha$ se cumple para todo $\alpha \geq 0$. Si $d(x_1, A) = r$ (lo que se verifica en particular con $A = \{x_0\}$), tomando $\alpha = r$ se cumple la desigualdad $\delta_{x_0}(A) \leq \delta_{x_1}(A_\alpha) + \alpha$, dado que $x_1 \in A_\alpha$, y no se verifica para $\alpha < r$. En cambio si $d(x_1, A) < r$, basta tomar $\alpha = d(x_1, A)$ para que se cumpla dicha desigualdad. Análogo si tomamos un conjunto que contenga a x_1 y no a x_0 . En el caso de tomar conjuntos que contengan ambos puntos o ninguno, se cumple la definición de la métrica de Prokhorov para cualquier $\alpha \geq 0$. Ello demuestra que el ínfimo de la definición es r , y por tanto, el resultado.

2. Sean μ y ν las siguientes medidas de distribución uniforme con diferente soporte:

$$\mu : \mathcal{B}(X) \longrightarrow [0, 1] \text{ dada por } \mu(A) = \int_{A \cap [0,1]} dx$$

$$\nu : \mathcal{B}(X) \longrightarrow [0, 1] \text{ dada por } \nu(A) = \int_{A \cap [1,2]} dx$$

Si tomamos el conjunto $A = (0, \frac{1}{2})$, el mínimo α con el que se cumplen las desigualdades que definen la distancia de Prokhorov es $\frac{1}{2}$:

$$\mu(A) = \frac{1}{2} = 0 + \frac{1}{2} = \nu(A_{\frac{1}{2}}) + \frac{1}{2} \qquad \nu(A) = 0 \leq 1 + \frac{1}{2} = \mu(A_{\frac{1}{2}}) + \frac{1}{2}$$

Veamos que estas desigualdades se cumplen para cualquier conjunto. Sea $A \in \mathcal{B}(X)$. Si $\mu(A) = \nu(A)$, se cumplen las desigualdades para todo $\alpha \geq 0$. Suponemos ahora que $\mu(A) > \nu(A)$. Si $\mu(A) - \nu(A) \leq \frac{1}{2}$, es directo que las desigualdades se cumplen para $\alpha = \frac{1}{2}$. Si $\mu(A) - \nu(A) > \frac{1}{2}$, entonces necesariamente $\mu(A) = \frac{1}{2} + \beta$ para algún $\beta > 0$ y por tanto $\nu(A_{\frac{1}{2}}) \geq \beta$, ya que A debe concentrar al menos una probabilidad de β en el intervalo $[\frac{1}{2}, 1]$. Luego $\mu(A) \leq \nu(A_{\frac{1}{2}}) + \frac{1}{2}$, y la otra desigualdad es trivial. Análogo si $\mu(A) < \nu(A)$.

Notar que si tomamos los soportes separados a distancia mayor o igual que 1, la distancia entre las medidas sería 1.

A continuación veremos una interesante propiedad de esta métrica que nos facilitará los cálculos a la hora de calcular la distancia de Prokhorov entre medidas y nos servirá posteriormente para demostrar algunos teoremas.

Proposición 1. Sea (X, d) un espacio métrico y $\mu, \nu \in \mathcal{P}$. Si para todo $A \in \mathcal{B}(X)$ se cumple $\mu(A) \leq \nu(A_\alpha) + \alpha$, entonces $d_P(\mu, \nu) \leq \alpha$. Es decir, si para todo conjunto boreliano A se cumple una de las desigualdades de la definición de la métrica de Prokhorov, entonces se cumple la otra desigualdad también.

Demostración. Notar que los siguientes contenidos son equivalentes:

$$A \subseteq X \setminus B_\alpha \quad \text{y} \quad B \subseteq X \setminus A_\alpha \tag{1.3}$$

porque ambos son equivalentes a

$$d(x, y) > \alpha \quad \forall x \in A, \forall y \in B.$$

Dado A , tomamos $B = X \setminus A_\alpha$ que cumple la primera desigualdad de la definición de la métrica

$$\mu(B) \leq \nu(B_\alpha) + \alpha$$

luego

$$\mu(A_\alpha) = 1 - \mu(B) \geq 1 - \nu(B_\alpha) - \alpha = \nu(X \setminus B_\alpha) - \alpha \geq \nu(A) - \alpha.$$

donde la última desigualdad se deduce de (1.3). \square

Ejemplo 4. Sea μ una medida de distribución uniforme con soporte $[0, 1]$ y, dado $n \in \mathbb{N}$ lo suficientemente grande, ν una medida de distribución uniforme con soporte $[\frac{1}{n}, 1]$. Veamos que la distancia entre ambas medidas cumple que $d_P(\mu, \nu) \leq \frac{1}{n}$. Notar que fuera del intervalo $[0, 1]$ ambas medidas son nulas, luego podemos restringir el estudio a ese intervalo. Notar además que, por la proposición 1, basta comprobar la primera igualdad únicamente.

- Si $A = [0, 1]$, entonces $\mu(A) = 1$ y $\nu(A) = 1$.

- Si $A = [m, p]$ con $0 \leq m < p \leq \frac{1}{n}$, se cumple la primera desigualdad:

$$\mu(A) = p - m \leq \nu(A_{\frac{1}{n}}) + \frac{1}{n}$$

- Si $A = [m, p]$ con $\frac{1}{n} \leq m < p \leq 1$, tenemos que $\mu(A) = p - m$ y $\nu(A) = \frac{n}{n-1}(p - m)$. Aquí $\frac{n}{n-1} > 1$, $\nu(A) > \mu(A)$, y la primera desigualdad se satisface para cualquier α .
- Si $A = [m, p]$ con $0 \leq m < \frac{1}{n} < p \leq 1$, la desigualdad que define esta métrica se cumple para $\alpha = \frac{1}{n}$ tanto si $p < 1 - \frac{1}{n}$:

$$\mu(A) = p - m \leq \frac{n}{n-1}p + \frac{1}{n} = \nu(A_{\frac{1}{n}}) + \frac{1}{n},$$

como si $p \geq 1 - \frac{1}{n}$:

$$\mu(A) = p - m \leq p \leq 1 + \frac{1}{n} = \nu(A_{\frac{1}{n}}) + \frac{1}{n},$$

donde la segunda desigualdad se cumple para todo $n > 0$.

- Si A es cualquier boreliano con $\mu(A) = p$, tomamos $B = [0, p]$, que satisface $\nu(A_{\frac{1}{n}}) \geq \nu(B_{\frac{1}{n}})$, y por los casos anteriores tendríamos que $p = \mu(B) \leq \nu(B_{\frac{1}{n}}) + \frac{1}{n}$. por lo que $\nu(A_{\frac{1}{n}}) + \frac{1}{n} \geq \nu(B_{\frac{1}{n}}) + \frac{1}{n} \geq \mu(B) = \mu(A)$.

Por tanto, se cumple que $d_P(\mu, \nu) \leq \frac{1}{n}$, luego si $n \rightarrow \infty$ se tiene que $d_P(\mu, \nu) \rightarrow 0$ y ν converge en la métrica de Prokhorov a μ .

El siguiente lema sobre existencia de recubrimientos especiales nos ayudará a demostrar que si el espacio métrico es separable, la convergencia en la métrica de Prokhorov es equivalente a la convergencia débil. Esto concuerda con el último ejemplo que hemos visto dado que $[0, 1]$ es un conjunto separable y conocemos que $\mu_n \Rightarrow \mu$ (ya que se da la convergencia en distribución). Recordemos primero la definición de espacio métrico separable.

Definición 6. Sea (X, d) un espacio métrico, diremos que es *separable* si posee un subconjunto denso y contable.

A continuación, introducimos un lema técnico.

Lema 1. Sea (X, d) un espacio métrico separable y μ una medida de Borel finita definida sobre X . Para cada $\delta > 0$ existen bolas abiertas (o cerradas) contables B_1, B_2, \dots tales que

$$\bigcup_{i \in \mathbb{N}} B_i = X,$$

el radio de $B_i < \delta$ para todo $i \in \mathbb{N}$,

$$\mu(\partial B_i) = 0 \text{ para todo } i \in \mathbb{N}.$$

Demostración. Sea D un conjunto denso en X , $x \in D$ y $S(x, r) := \{y \in X : d(x, y) = r\}$. Notar que la frontera de las bolas abiertas o cerradas con centro en x y radio r está contenida en $S(x, r)$. Dado $\delta > 0$, tomamos $\mathcal{S} := \{S(x, r) : \delta/2 < r < \delta\}$. La colección \mathcal{S} es disjunta y como mucho un número contable de sus elementos tiene medida no nula. Puesto que \mathcal{S} es no contable existe un $r \in (\delta/2, \delta)$ tal que $\mu(S(x, r)) = 0$. De este modo para cada $x \in D$ podemos encontrar una bola abierta (o cerrada) $B(x, r)$ con radio $r \in (\delta/2, \delta)$ tal que $\mu(\partial B(x, r)) = 0$. Ahora, como D es denso, estas bolas cubren todo X , y como D es contable, tenemos un número contable de bolas B_1, B_2, \dots \square

Ahora ya estamos en condiciones de demostrar la equivalencia entre convergencia débil y convergencia en la métrica de Prokhorov.

Teorema 1.2. Sea (X, d) un espacio métrico separable y $\mu, \mu_1, \mu_2, \dots \in \mathcal{P}$, entonces

$$\mu_i \Rightarrow \mu \iff d_P(\mu_i, \mu) \rightarrow 0.$$

Demostración. \Leftarrow) Si $d_P(\mu_i, \mu) \rightarrow 0$, existe una sucesión $(\alpha_i)_{i \in \mathbb{N}} \searrow 0$ cumpliendo

$$\mu_i(A) \leq \mu(A_{\alpha_i}) + \alpha_i \quad \text{y} \quad \mu(A) \leq \mu_i(A_{\alpha_i}) + \alpha_i \quad \forall A \in \mathcal{B}(X).$$

Se sigue que $\forall A \in \mathcal{B}(X)$

$$\limsup_{i \rightarrow \infty} \mu_i(A) \leq \limsup_{i \rightarrow \infty} (\mu(A_{\alpha_i}) + \alpha_i) = \lim_{i \rightarrow \infty} \mu(A_{\alpha_i}) = \mu(\bar{A}),$$

donde la última desigualdad se deduce por el teorema de continuidad de las medidas de probabilidad (el límite de la medida de sucesos contenidos en sentido decreciente es la medida de la intersección de todos ellos). En particular, para cada conjunto cerrado $C \subseteq X$, $\limsup_{i \rightarrow \infty} \mu_i(C) \leq \mu(C)$, y por el Teorema de Portmanteau tenemos que $\mu_i \Rightarrow \mu$.

\Rightarrow) Suponemos ahora que $\mu_i \Rightarrow \mu$. Sea $\epsilon > 0$ y queremos ver que $\exists N > 0$ tal que para todo $i \geq N$ se tiene $d_P(\mu, \mu_i) \leq \epsilon$, es decir, $\mu_i(B) \leq \mu(B_\epsilon) + \epsilon$ y $\mu(B) \leq \mu_i(B_\epsilon) + \epsilon \quad \forall B \in \mathcal{B}(X)$. Tomamos $\delta > 0$ con $\delta < \epsilon/3$. Gracias al lema anterior, podemos tomar un conjunto de bolas abiertas disjuntas $\{B_i\}_{i \in \mathbb{N}}$ con radio $r < \delta/2$, tales que $\bigcup_{j \in \mathbb{N}} B_j = X$ y $\mu(\partial B_j) = 0$ para todo $j \in \mathbb{N}$. Elijo k de manera que se cumpla

$$\mu\left(\bigcup_{j=1}^k B_j\right) \geq 1 - \delta.$$

Definimos la colección finita de conjuntos abiertos que se pueden formar combinando $\{B_i\}_{i \leq k}$

$$\mathcal{A} := \left\{ \bigcup_{j \in J} B_j : J \subset \{1, \dots, k\} \right\}$$

Utilizaremos esa colección para aproximar los conjuntos borelianos. Notar que para cada $A \in \mathcal{A}$,

$$\partial A \subseteq (\partial B_1 \cup \dots \cup \partial B_k) \Rightarrow \mu(\partial A) \leq \mu(\partial B_1) + \dots + \mu(\partial B_k) = 0 \Rightarrow \mu(\partial A) = 0$$

Por el Teorema de Portmanteau (v), si $\mu_i \Rightarrow \mu$, entonces $\exists N > 0$ tal que para todo $i > N$ $\mu(A) < \mu_i(A) + \delta$ para todo $A \in \mathcal{A}$. Dado $B \in \mathcal{B}(X)$, sea A la unión de los conjuntos B_1, \dots, B_k que intersecan con B , y tenemos que:

$$B = [B \cap \bigcup_{j \leq k} B_j] \cup [B \cap \bigcup_{j > k} B_j] \subseteq A \cup \bigcup_{j > k} B_j \quad (1.4)$$

$$\mu\left(\bigcup_{j > k} B_j\right) < \delta \quad (1.5)$$

$$\mu(A) < \mu_i(A_\delta) + \delta \quad (1.6)$$

$$A \subset B_\delta = \{x : d(x, B) < \delta\} \text{ dado que el diámetro de cada } B_j \text{ es menor que } \delta \quad (1.7)$$

$$B_{2\delta} \subseteq B_\epsilon \quad (1.8)$$

Por lo tanto

$$\mu(B) \stackrel{(1.4)}{\leq} \mu(A) + \mu\left(\bigcup_{j > k} B_j\right) \stackrel{(1.5)}{\leq} \mu(A) + \delta \stackrel{(1.6)}{\leq} \mu_i(A_\delta) + 2\delta \stackrel{(1.7)}{\leq} \mu_i(B_{2\delta}) + 2\delta \stackrel{(1.8)}{\leq} \mu_i(B_\epsilon) + \epsilon$$

Dado que es cierto para todo $B \in \mathcal{B}(X)$, por la proposición anterior $d_P(\mu_i, \mu) \leq \epsilon \quad \forall i \geq N$. \square

Veamos ahora que el espacio métrico de las medidas de probabilidad sobre el espacio original junto con la métrica de Prokhorov hereda las buenas propiedades de separabilidad y completitud del espacio métrico original.

Teorema 1.3. *Sea (X, d) un espacio métrico separable, entonces $(\mathcal{P}(X), d_P)$ es separable.*

Demostración. Buscamos un conjunto en $\mathcal{P}(X)$ que sea contable y denso. Sea $\epsilon > 0$ y $\delta > 0$ con $\delta < \epsilon/3$, por el lema 1 tomamos un conjunto contable de bolas abiertas con centros disjuntos $\{B_i\}_{i \in \mathbb{N}}$ con radio $r < \delta/2$, tales que $\bigcup_{j \in \mathbb{N}} B_j = X$ y $\mu(\partial B_j) = 0$ para todo $j \in \mathbb{N}$. Para cada $B_i \neq \emptyset$ elegimos $x_i \in B_i$ disjuntos y definimos

$$\mathcal{M} := \left\{ r_1 \delta_{x_1} + \dots + r_n \delta_{x_n} : r_1, \dots, r_n \in \mathbb{Q} \cap [0, 1], \sum_{j=1}^n r_j = 1, n = 1, 2, \dots \right\}$$

Es claro que \mathcal{M} es contable por serlo el conjunto de bolas $\{B_i\}_{i \in \mathbb{N}}$. Dado $\mu \in \mathcal{P}(X)$, elegimos k de manera que $\mu(\bigcup_{i > k} B_i) < \delta$. Ahora podemos coger

$$\nu = r_1 \delta_{x_1} + \dots + r_k \delta_{x_k} \in \mathcal{M} \text{ cumpliendo } \sum_{i \leq k} |r_i - \mu(B_i)| < \delta$$

Tomamos $B \in \mathcal{B}(X)$ y sea I el conjunto de índices $i \leq k$ tales que $B_i \cap B \neq \emptyset$. Denotamos $A = \bigcup_{i \in I} B_i$. Entonces

$$\mu(B) \leq \mu(A) + \delta = \sum_{i \in I} \mu(B_i) + \delta \leq \sum_{i \in I} r_i + 2\delta = \nu(A) + 2\delta \leq \nu(B_\delta) + 2\delta \leq \nu(B_\epsilon) + \epsilon$$

Como es cierto para todo $B \in \mathcal{B}(X)$, por la proposición anterior $d_P(\mu, \nu) \leq \epsilon$, $\forall \epsilon > 0$. Luego \mathcal{M} es denso en $\mathcal{P}(X)$ y $(\mathcal{P}(X), d_P)$ es separable. \square

Antes del siguiente resultado sobre completitud del espacio de medidas de probabilidad, conviene recordar algunas definiciones así como enunciar un teorema del que haremos uso en la demostración.

Definición 7. Sea (X, d) un espacio métrico. Una sucesión $\{x_n\}$ es *d-fundamental o de Cauchy* si para todo $\epsilon > 0$, existe N tal que $d(x_n, x_m) < \epsilon$ para cualesquiera $n, m > N$. Si toda sucesión de Cauchy es convergente, diremos que el espacio métrico es *completo*.

Una familia \mathcal{V} de subconjuntos de X se dice *recubrimiento* de A si $A \subseteq \bigcup \mathcal{V}$. Si \mathcal{V} es recubrimiento de A y existe $\mathcal{W} \subseteq \mathcal{V}$ subfamilia de \mathcal{V} que es también recubrimiento de A , diremos que \mathcal{V} posee un subrecubrimiento \mathcal{W} de A .

Definición 8. Sea (X, d) un espacio métrico. Diremos que un subconjunto $K \subseteq X$ es compacto si todo recubrimiento abierto de K posee un subrecubrimiento finito.

Recordar que los conjuntos compactos tienen la propiedad de que cualquier sucesión del conjunto tiene una subsucesión convergente.

El siguiente teorema hace uso de la definición 3. La demostración se puede encontrar en [1].

Teorema 1.4 (Teorema de Prokhorov). *Sea (X, d) un espacio métrico completo separable y sea $E \subset \mathcal{P}$. Las siguientes afirmaciones son equivalentes:*

- (a) \overline{E} es compacto en \mathcal{P} .
- (b) E es un conjunto ajustado de medidas de probabilidad.

Ahora estamos ya en condiciones de enunciar nuestro resultado sobre separabilidad y completitud.

Teorema 1.5. *Sea (X, d) un espacio métrico separable y completo, entonces $(\mathcal{P}(X), d_P)$ es completo.*

Demostración. Sea $\{\mu_n\} \subset \mathcal{P}(X)$ una sucesión d_P -fundamental. Es suficiente ver que se trata de una sucesión ajustada de medidas, dado que por el teorema 1.4, eso implica que la clausura de dicha sucesión es compacta, y por tanto, contiene una subsucesión convergente con respecto a la distancia d_P . De lo que deducimos que μ_n es convergente, al ser d_P -fundamental.

Veamos que $\forall \epsilon > 0$ y $\forall \delta > 0$ existe una colección finita de δ -bolas B_i tales que

$$\mu_n(B_1 \cup \dots \cup B_m) > 1 - \epsilon \text{ para todo } n. \quad (1.9)$$

La propiedad de ser una sucesión ajustada de medidas se conseguiría tomando la clausura de $(B_1 \cup \dots \cup B_m)$. Elegimos η de manera que $0 < 2\eta < \min\{\epsilon, \delta\}$. Por otro lado, tomamos n_0 tal que para todo $n \geq n_0$ tenemos $d_P(\mu_{n_0}, \mu_n) < \eta$. Como X es separable contiene un subconjunto denso D , luego podemos cubrir X con bolas cerradas $A_i = B(x_i, \eta)$ con $x_i \in D$. Ahora elegimos m cumpliendo que $\mu_n(A_1 \cup \dots \cup A_m) > 1 - \eta$ para cada $n \leq n_0$. Denotamos $B_i = B(x_i, 2\eta)$. Entonces para $n \geq n_0$ se tiene

$$\mu_n(B_1 \cup \dots \cup B_m) \geq \mu_n((A_1 \cup \dots \cup A_m)_\eta) \geq \mu_{n_0}(A_1 \cup \dots \cup A_m) - \eta \geq 1 - 2\eta \geq 1 - \epsilon.$$

Y para $n \leq n_0$ se tiene

$$\mu_n(B_1 \cup \dots \cup B_m) \geq \mu_n(A_1 \cup \dots \cup A_m) \geq 1 - \eta \geq 1 - \epsilon.$$

Por tanto, B_1, \dots, B_m verifican (1.9). Si llamamos K a la clausura de la unión del conjunto de bolas B_i , tenemos que K es cerrado y totalmente acotado, luego es compacto y se cumple que

$$\mu_n(K) \geq 1 - \epsilon \text{ para todo } n.$$

□

En resumen, tenemos que si (X, d) es un espacio métrico separable y completo, entonces el espacio métrico $(\mathcal{P}(X), d_P)$ es separable, completo y la convergencia débil es equivalente a la convergencia en la métrica de Prokhorov.

Capítulo 2

Las dimensiones fractales y la teoría de la información

Las dimensiones fractales extienden la noción geométrica de dimensión entera de un objeto a otros valores reales. Un fractal es un objeto geométrico de dimensión fractal no entera, por ejemplo, los conjuntos autosimilares que exhiben autosimilitud a cualquier escala. Los fractales se encuentran en áreas de estudio muy variadas como las matemáticas, la física, la biología, la informática o el arte. En la naturaleza, muchos fenómenos y estructuras exhiben características fractales. Las dimensiones fractales son una medida de la complejidad de un objeto fractal que son utilizadas para describir cómo se distribuyen los detalles y la autosimilitud de estos objetos.

El objetivo de este capítulo es ver cómo se puede utilizar la dimensión algorítmica relativizada de los puntos de un conjunto para establecer las dimensiones fractales de dicho conjunto, aplicándolo al caso concreto del espacio de medidas de probabilidad. Las dimensiones fractales se pueden definir en cualquier espacio métrico, y las principales son la dimensión de Hausdorff y la dimensión de empaquetamiento. Tienen la ventaja de estar definidas para cualquier conjunto. Veremos cómo se definen para el conjunto de las medidas de probabilidad.

2.1. Dimensión de Hausdorff y de empaquetamiento

Para explicar la dimensión de Hausdorff primero es necesario definir la medida de Hausdorff. Esta medida generaliza las ideas clásicas de longitud, área y volumen. Se puede probar, que para subconjuntos de \mathbb{R}^n , la medida n -dimensional de Hausdorff es, salvo una constante multiplicativa, la medida n -dimensional de Lebesgue.

La dimensión de Hausdorff (y la de empaquetamiento) pueden ser definidas en cualquier espacio métrico, siendo no triviales en espacios métricos separables. Vamos a restringirnos aquí al espacio \mathcal{P} de las medidas de probabilidad que es un espacio métrico separable y acotado. El ser acotado simplifica ligeramente las definiciones de dimensión.

Sea $F \subseteq \mathcal{P}$ un conjunto de medidas de probabilidad. El diámetro de F se define como $|F| := \sup \{d_P(\mu, \nu) : \mu, \nu \in F\}$, es decir, la mayor distancia de Prokhorov entre dos medidas del conjunto. Si tomamos un conjunto contable de índices I , la colección $\{U_i\}_{i \in I}$ se denomina δ -recubrimiento de F si es un recubrimiento de F y $0 < |U_i| \leq \delta$ para todo $i \in I$. Sea entonces s un número no negativo. Para cualquier $\delta > 0$ definimos

$$\mathcal{H}_\delta^s(F) := \inf \left\{ \sum_{i \in I} |U_i|^s : \{U_i\}_{i \in I} \text{ es un } \delta\text{-recubrimiento de } F \right\} \quad (2.1)$$

Notar que si δ decrece, el número de δ -recubrimientos de F se reduce, luego el ínfimo crece.

Definición 9. Llamaremos *medida s -dimensional Hausdorff* del conjunto $F \subseteq \mathcal{P}$ al límite cuando $\delta \rightarrow 0$ en la ecuación (2.1), es decir,

$$\mathcal{H}^s(F) := \lim_{\delta \rightarrow 0} \mathcal{H}_\delta^s(F) \quad (2.2)$$

Ese límite existe para cualquier conjunto $F \subseteq \mathcal{P}$ y puede tomar los valores 0 o ∞ .

En la ecuación (2.1) podemos ver que dado un conjunto F y $\delta < 1$, $\mathcal{H}_\delta^s(F)$ es decreciente en s , luego por (2.2) se sigue que $\mathcal{H}^s(F)$ es igualmente no creciente. De hecho, si tomamos $\{U_i\}_{i \in I}$ un δ -recubrimiento de F y $t > s$ tenemos que

$$\sum_{i \in I} |U_i|^t \leq \delta^{t-s} \sum_{i \in I} |U_i|^s.$$

Tomando ínfimos, $\mathcal{H}_\delta^t(F) \leq \delta^{t-s} \mathcal{H}_\delta^s(F)$, y haciendo tender $\delta \rightarrow 0$ obtenemos que si $\mathcal{H}^s(F) < \infty$ entonces $\mathcal{H}^t(F) = 0$ para toda $t > s$. Por tanto, si existe un N para el que $\mathcal{H}^N(F) < \infty$, entonces existe un s_0 para el cual $\mathcal{H}^s(F) = \infty$ si $s < s_0$ y $\mathcal{H}^s(F) = 0$ si $s > s_0$.

Definición 10. Sea $F \subseteq \mathcal{P}$ un conjunto, se define su *dimensión de Hausdorff* como:

$$\dim_{\text{H}} F = \inf\{s : \mathcal{H}^s(F) = 0\} = \sup\{s : \mathcal{H}^s(F) = \infty\}$$

Por la definición anterior se tiene que

$$\mathcal{H}^s(F) = \begin{cases} \infty & \text{si } s < \dim_{\text{H}} F \\ 0 & \text{si } s > \dim_{\text{H}} F \end{cases}$$

Si $s = \dim_{\text{H}} F$, entonces $\mathcal{H}^s(F)$ puede valer 0, ∞ o cualquier valor intermedio. Notar también que si F es un conjunto contable su dimensión Hausdorff es 0. Esto se debe a que si F_i es un punto del espacio métrico, en nuestro caso una medida de probabilidad, se tiene que $\mathcal{H}^0(F_i) = 1$ y $\dim_{\text{H}} F_i = 0$. Luego si F es unión contable de F_i , entonces

$$\dim_{\text{H}} F = \dim_{\text{H}} \bigcup_{i=1}^{\infty} F_i = \sup_{1 \leq i \leq \infty} \{\dim_{\text{H}} F_i\} = 0.$$

La dimensión de empaquetamiento se define en términos de cómo un conjunto se puede cubrir con conjuntos más pequeños de manera óptima. Concretamente, se busca la tasa de crecimiento exponencial del número de conjuntos necesarios para cubrir un conjunto dado, a medida que los conjuntos que recubren decrecen.

Veamos cómo se define para el espacio de las medidas de probabilidad. Sea $F \subseteq \mathcal{P}$ un conjunto y $s > 0$. Tomamos un conjunto de bolas cerradas disjuntas dos a dos $\{B_i\}_{i \in I}$ tales que sus centros son los elementos de F y $|B_i| < \delta$.

Definición 11. La *pre-medida s -dimensional de empaquetamiento* del conjunto F se define como

$$P_0^s(F) = \lim_{\delta \rightarrow 0} \sup \left\{ \sum_{i \in I} |B_i|^s : \{B_i\}_{i \in I} \text{ disjuntas con centro en } F \text{ y } |B_i| < \delta \right\}. \quad (2.3)$$

Observar que no es una medida por no ser contablemente subaditiva. Ahora podemos definir una medida usando descomposiciones contables en función de esta pre-medida.

Definición 12. La *medida s -dimensional de empaquetamiento* del conjunto F se define como

$$P^s(F) = \inf \left\{ \sum_{j \in J} P_0^s(F_j) : F \subseteq \bigcup_{j \in J} F_j, \text{ con } J \text{ contable} \right\}. \quad (2.4)$$

Al igual que en la dimensión de Hausdorff existe un valor $s_0 \in [0, \infty]$, que llamaremos dimensión de empaquetamiento, tal que, si $s_0 < \infty$, para todo $s > s_0$, $P^s(F) = \infty$, y para todo $s < s_0$, $P^s(F) = 0$.

Definición 13. Sea $F \in \mathcal{P}$ un conjunto, se define su *dimensión de empaquetado* como

$$\dim_P(F) = \inf\{s : P^s(F) = 0\} = \sup\{s : P^s(F) = \infty\}.$$

2.1.1. Dimensión con escala

Podemos definir dimensión de Hausdorff y de empaquetamiento para cualquier espacio métrico, pero en muchos casos interesantes el espacio completo tiene dimensión infinita y la dimensión de un conjunto no suele ser útil para discriminar el tamaño. Podemos entonces ajustar la escala de la dimensión con el objetivo de que el espacio tenga dimensión finita. Para eso sustituiremos la función de escala habitual ($\varphi(s, \delta) = \delta^s$) por otra función que nos permita obtener un valor finito de la dimensión. De forma análoga si el espacio completo tiene dimensión 0 sustituiremos la función de escala habitual por otra función que nos permita obtener un valor positivo de la dimensión.

En nuestro caso de estudio, el espacio \mathcal{P} tiene dimensión de Hausdorff y de empaquetamiento infinitas. Modificando la forma de cuantificar el diámetro de un cubrimiento obtenemos medidas y dimensión con escala que resultan más útiles para \mathcal{P} . Utilizaremos la siguiente función de escala

$$\varphi(s, \delta) = 2^{-(1/\delta)^s}, \quad (2.5)$$

en lugar de la habitual. Por lo tanto ahora redefinimos la ecuación (2.1) como

$$\mathcal{H}_\delta^{s,\varphi}(F) := \inf \left\{ \sum_{i \in I} \varphi(s, |U_i|) : \{U_i\}_{i \in I} \text{ es un } \delta\text{-recubrimiento de } F \right\}. \quad (2.6)$$

Del mismo modo redefinimos la ecuación (2.3)

$$P_0^{s,\varphi}(F) = \limsup_{\delta \rightarrow 0} \left\{ \sum_{i \in I} \varphi(s, |B_i|) : \{B_i\}_{i \in I} \text{ disjuntas con centro en } F \text{ y } |B_i| < \delta \right\}. \quad (2.7)$$

Con estos cambios, denotaremos las medidas de Hausdorff y de empaquetamiento con escala como $\mathcal{H}^{s,\varphi}$ y $P^{s,\varphi}$. Además, los correspondientes conceptos \dim_H^φ y \dim_P^φ tienen un comportamiento análogo a los originales.

2.2. Complejidad de Kolmogorov

La complejidad de Kolmogorov de un objeto es la cantidad mínima de información necesaria para describirlo completamente, es decir, la longitud del programa más corto que puede generar ese objeto de manera exacta. Por tanto, en el contexto de medidas de probabilidad, la complejidad de Kolmogorov se puede entender como la longitud mínima de un programa que genera de forma computable la distribución de probabilidad en cuestión. Antes de pasar a formalizar dichas ideas, necesitamos definir el siguiente concepto.

Una máquina de Turing es un modelo teórico de computadora capaz de ejecutar cualquier algoritmo. Se trata de un sistema de entrada-salida determinado por una máquina finita de estados. Dada una secuencia finita p y un estado inicial, la máquina de Turing lee el primer símbolo de p , y mediante una serie de reglas internas produce unos símbolos de salida, cambia de estado y pasa a leer otro símbolo de p . Se repite este proceso hasta llegar a un estado de parada, y en ese momento la máquina devuelve el conjunto de símbolos de salida producidos.

Definición 14. Una máquina de Turing se puede definir como una cuádrupla $T = (Q, \Sigma, s, \delta)$ donde

1. Q es un conjunto finito de estados q .
2. Σ es un conjunto finito de símbolos. $\Sigma^{<\mathbb{N}}$ representa secuencias de 0 o más símbolos de Σ .
3. $s \in Q$ es el estado inicial.
4. $\{L, R\}$ denotan las direcciones izquierda y derecha respectivamente.
5. δ es una función transición que determina el siguiente movimiento:

$$\delta : (Q \times \Sigma) \rightarrow (\Sigma^{<\mathbb{N}} \times \{L, R\} \times Q)$$

La función de transición es una función entre estados de computación. Si $\delta(q_i, S_i) = (S_j, D, q_j)$, entonces cuando la máquina T está en el estado q_i leyendo la secuencia de símbolos S_i , reemplaza S_i por S_j , mueve en dirección $D \in \{L, R\}$ y va al estado q_j . Notar que se trata de una función parcial dado que no siempre está definida. Denotaremos como $T(p)$ a la salida producida por la máquina de Turing T con entrada p .

El problema con la máquina de Turing es que necesitamos una diferente para cada problema que tratemos de calcular. Por eso introducimos el siguiente concepto.

Definición 15. Diremos que una máquina de Turing U es universal si para cada máquina T existe una secuencia finita x tal que para cada secuencia finita y se tiene que $U(x, y) = T(y)$.

Es decir, una máquina de Turing universal es aquella que puede simular cualquier otra máquina de Turing. Veamos ahora la definición de la complejidad de Kolmogorov para secuencias finitas de bits. Sea $\{0, 1\}^{<\mathbb{N}}$ el conjunto de secuencias finitas de $\{0, 1\}$. Para cada $w \in \{0, 1\}^{<\mathbb{N}}$, la longitud de w , que denotaremos $|w|$, es el número de elementos de w . Sea U una máquina de Turing universal.

Definición 16. Dada $w \in \{0, 1\}^{<\mathbb{N}}$, la *complejidad de Kolmogorov de w* se define como

$$K_U(w) = \min_{p \in \{0, 1\}^{<\mathbb{N}}} \{|p| : U(p) = w\} \quad (2.8)$$

Observación. Dadas dos máquinas de Turing universales U, V , existe $c > 0$ tal que para toda secuencia w , $K_U(w) \leq K_V(w) + c$.

Basándonos en esa observación, fijamos una máquina de Turing universal U y la omitimos usando $K(w)$ en lugar de $K_U(w)$.

Ejemplo 5. La especificación de n requiere coste logarítmico. Por lo tanto si tenemos la cadena $x = \underbrace{1 \dots 1}_{n \text{ veces}}$, su complejidad será de orden logarítmico, es decir, $K(x) = O(\log n)$. También podemos encontrar cadenas con complejidad de menor orden. Por ejemplo si $n = 2^m$ para algún $m \in \mathbb{N}$, podemos describir n describiendo m y un programa que implemente la función $f(z) = 2^z$. La descripción de ese programa no depende de n , luego tiene complejidad constante y la descripción de m tiene complejidad logarítmica. Luego para esos valores de n tenemos

$$K(x) = O(\log(m)) = O(\log(\log(n)))$$

Otras cadenas de longitud n como las totalmente aleatorias tienen complejidad al menos n .

Puede también interesarnos considerar la complejidad de Kolmogorov de una secuencia condicionada a otra, que corresponde a la cantidad de información requerida para transformar una secuencia en otra.

Definición 17. Dadas $w, y \in \{0, 1\}^{<\mathbb{N}}$, la *complejidad de Kolmogorov de w condicionada a y* se define como

$$K(w|y) = \min_{p \in \{0, 1\}^{<\mathbb{N}}} \{|p| : U(p, y) = w\} \quad (2.9)$$

Incluimos a continuación algunas propiedades básicas de la complejidad de Kolmogorov. Podemos encontrar la demostración de estas propiedades en [8].

- Teorema 2.1.**
1. Existe una constante c tal que $K(w) \leq |w| + c$, para toda secuencia w .
 2. Para toda máquina de Turing U existe una constante c tal que $K(U(w)) \leq K(w) + c$, para toda cadena w tal que $U(w)$ esté definido.
 3. Existe una constante c tal que $K(wz) = K(z) + K(w|z) + O(\log n)$ para cualquier par de cadenas w, z con complejidad no mayor a n .
 4. Dado un entero n , existe una secuencia x con $|x| = n$ cumpliendo $K(x) \geq n$.
 5. La función $K : \{0, 1\}^{<\mathbb{N}} \rightarrow \mathbb{N}$ no es calculable, es decir, no existe una máquina de Turing T tal que $K(w) = T(w)$ para toda w .
 6. Si $g : \{0, 1\}^{<\mathbb{N}} \rightarrow \mathbb{N}$ es una función calculable creciente, entonces existen infinitos x para los que se cumple $K(x) < g(x)$.

Recientemente [7], la complejidad de Kolmogorov se ha definido en cualquier espacio métrico separable. Para ello se utiliza una enumeración de un subconjunto contable denso para el que se sigue la definición original de complejidad de Kolmogorov, y después se define la complejidad de Kolmogorov de un punto cualquiera aproximándolo con los elementos del denso contable.

Veamos aquí el caso del espacio de medidas de probabilidad. La complejidad de Kolmogorov de una medida de probabilidad se refiere a la cantidad de información necesaria para describir completamente esa medida. Si identificamos un programa con una secuencia finita de bits, podemos tomar una función que dado un programa nos devuelva una medida de probabilidad. Es decir, elegimos

$$f : \{0, 1\}^{<\mathbb{N}} \longrightarrow \mathcal{P}(X)$$

tal que la imagen de f es densa en $\mathcal{P}(X)$. Ahora podemos definir la complejidad de las medidas de probabilidad de la siguiente manera, definiendo la complejidad de una medida μ a partir de la complejidad de Kolmogorov de los elementos cercanos a μ en $\text{Im}(f)$.

Definición 18. Dada una medida de probabilidad $\mu \in \mathcal{P}$, la *complejidad de Kolmogorov de μ a precisión q* es

$$K_q(\mu) = \min \{K(w) : d_P(f(w), \mu) < 2^{-q}\} \quad (2.10)$$

Para el caso del espacio euclídeo \mathbb{R}^n elegimos como conjunto denso los puntos de coordenadas racionales no periódicos, y como función f la representación habitual que asigna a cada secuencia $w = \langle w_1, \dots, w_n \rangle$ el punto (q_1, \dots, q_n) donde q_i es el racional con representación w_i en binario.

Definición 19. Dado $x \in \mathbb{R}^n$, la *complejidad de Kolmogorov de x a precisión q* es

$$K_q(x) = \min \{K(w) : |f(w) - x| < 2^{-q}\} \quad (2.11)$$

2.3. Dimensión efectiva

La dimensión efectiva es un concepto utilizado para caracterizar la cantidad de información relevante o significativa contenida en un conjunto de datos. Es importante notar que dado un conjunto de datos, su dimensión efectiva puede ser mayor que su dimensión de Hausdorff. La dimensión efectiva se define primero para un punto (en nuestro caso, una medida) y después para conjuntos de puntos. Vamos a definir la dimensión efectiva dentro del espacio euclídeo \mathbb{R}^n y en \mathcal{P} , en este segundo caso usando la escala $\varphi(s, \delta) = 2^{-(1/\delta)^s}$.

Definición 20. Dado un punto $x \in \mathbb{R}^n$, su *dimensión efectiva* es

$$\dim(x) = \liminf_{r \rightarrow \infty} \frac{K_r(x)}{r}. \quad (2.12)$$

A partir de esta definición, podemos definir la dimensión efectiva de un conjunto de \mathbb{R}^n .

Definición 21. Dado $E \subseteq \mathbb{R}^n$, su *dimensión efectiva* es

$$\dim(E) = \sup_{x \in E} \dim(x). \quad (2.13)$$

Este concepto se considera una efectivización de la dimensión de Hausdorff dado que la dimensión efectiva puede caracterizarse a partir de la versión efectiva de la medida de Hausdorff, siendo la dimensión efectiva de un conjunto E el ínfimo de los s para los que la s -medida de E es 0 [6]. Esta caracterización es análoga a la definición original de dimensión de Hausdorff.

La dimensión de empaquetado también tiene una efectivización similar que llamaremos dimensión efectiva fuerte. Se puede definir tanto para puntos como para conjuntos de \mathbb{R}^n .

Definición 22. Dado un punto $x \in \mathbb{R}^n$, su *dimensión efectiva fuerte* es

$$\text{Dim}(x) = \limsup_{r \rightarrow \infty} \frac{K_r(x)}{r}. \quad (2.14)$$

Definición 23. Dado $E \subseteq \mathbb{R}^n$, su *dimensión efectiva fuerte* es

$$\text{Dim}(E) = \sup_{x \in E} \text{Dim}(x). \quad (2.15)$$

Observación. Del mismo modo, podemos definir la dimensión efectiva para medidas de probabilidad y conjuntos de estas. Sean $\mu \in \mathcal{P}$, y $E \subseteq \mathcal{P}$ sus dimensiones efectivas con escala φ se definen como

$$\dim^\varphi(\mu) = \liminf_{r \rightarrow \infty} \frac{\log(K_r(\mu))}{r} \quad \dim^\varphi(E) = \sup_{\mu \in E} \dim^\varphi(\mu)$$

Análogamente sus dimensiones efectivas fuertes son:

$$\text{Dim}^\varphi(\mu) = \limsup_{r \rightarrow \infty} \frac{\log(K_r(\mu))}{r} \quad \text{Dim}^\varphi(E) = \sup_{\mu \in E} \text{Dim}^\varphi(\mu)$$

2.4. Oráculos y relativización

Una máquina de Turing con oráculo es una variante teórica de la máquina de Turing, capaz de consultar un oráculo externo para resolver ciertos problemas sin coste alguno. Podemos visualizar el oráculo como un sistema de entrada-salida capaz de resolver un problema dado, que puede ser de cualquier complejidad, incluso indecidible.

En nuestro caso, un oráculo es un conjunto $A \in \{0, 1\}^{<\mathbb{N}}$. Denotaremos a las máquinas de Turing U con oráculo como U^A . La máquina con oráculo posee una cinta especial de consulta en la que se puede escribir una secuencia q y al pasar al estado de pregunta la máquina U^A cambia al estado afirmativo si $q \in A$ o al estado negativo si $q \notin A$. Por lo tanto, para cada oráculo A , la máquina de Turing U realiza un cálculo distinto. El caso de $A = \emptyset$ corresponde al cálculo sin oráculo.

Ahora, podemos definir la versión relativizada de la complejidad de Kolmogorov para cada oráculo $A \in \{0, 1\}^{<\mathbb{N}}$, que se refiere a la cantidad de información para describir una secuencia de bits cuando se proporciona sin coste la solución de pertenencia al oráculo A .

Definición 24. Dada $w \in \{0, 1\}^{<\mathbb{N}}$, la *complejidad de Kolmogorov de w relativa a A* es

$$K^A(w) = \min_{p \in \{0, 1\}^{<\mathbb{N}}} \{|p| : U^A(p) = w\}. \quad (2.16)$$

Análogamente definimos la versión relativizada para puntos en \mathbb{R}^n y para medidas de probabilidad.

Definición 25. Dado un punto $x \in \mathbb{R}^n$, la *complejidad de Kolmogorov de x a precisión q relativa a A* es

$$K_q^A(x) = \min \{K^A(w) : |f(w) - x| < 2^{-q}\} \quad (2.17)$$

Definición 26. Dada una medida de probabilidad $\mu \in \mathcal{P}$, la *complejidad de Kolmogorov de μ a precisión q relativa a A* es

$$K_q^A(\mu) = \min \{K^A(w) : d_P(f(w), \mu) < 2^{-q}\} \quad (2.18)$$

De la misma forma podemos definir las dimensiones efectivas relativizadas a un oráculo A .

Definición 27. Dado un punto $x \in \mathbb{R}^n$, su *dimensión efectiva* y su *dimensión efectiva fuerte* relativas a A son respectivamente

$$\dim^A(x) = \liminf_{r \rightarrow \infty} \frac{K_r^A(x)}{r} \quad \text{Dim}^A(x) = \limsup_{r \rightarrow \infty} \frac{K_r^A(x)}{r}. \quad (2.19)$$

Igualmente, dado $E \subseteq \mathbb{R}^n$, su *dimensión efectiva* y su *dimensión efectiva fuerte* relativas a A son respectivamente

$$\dim^A(E) = \sup_{x \in E} \dim^A(x) \quad \text{Dim}^A(E) = \sup_{x \in E} \text{Dim}^A(x). \quad (2.20)$$

Observación. Para el caso de las medidas de probabilidad volvemos a utilizar la función de escala en el cálculo de las dimensiones efectivas.

$$\dim^{\varphi, A}(E) = \sup_{\mu \in E} \dim^{\varphi, A}(\mu) \quad \text{Dim}^{\varphi, A}(E) = \sup_{\mu \in E} \text{Dim}^{\varphi, A}(\mu).$$

2.5. El principio punto a conjunto

En [1] Jack H. Lutz y Neil Lutz prueban que se puede utilizar la dimensión relativizada de cada punto de un conjunto E en un espacio Euclídeo para establecer un límite inferior de las dimensiones de Hausdorff y de empaquetamiento de E . Lo hacen a través del denominado principio punto a conjunto, que específicamente dice que para todo conjunto E de un espacio Euclídeo \mathbb{R}^n ,

$$\begin{aligned} \dim_H(E) &= \min_{A \subseteq \mathbb{N}} \dim^A(E) = \min_{A \subseteq \mathbb{N}} \sup_{x \in E} \dim^A(x) \\ \dim_P(E) &= \min_{A \subseteq \mathbb{N}} \text{Dim}^A(E) = \min_{A \subseteq \mathbb{N}} \sup_{x \in E} \text{Dim}^A(x) \end{aligned}$$

Posteriormente, en [7, 10] se extiende el principio de punto a conjunto para espacios métricos separables. En el anterior capítulo hemos visto que el espacio de medidas de probabilidad sobre un conjunto métrico separable es igualmente un espacio métrico separable. Por lo tanto podemos aplicar el principio de punto a conjunto para caracterizar la dimensión de Hausdorff y de empaquetamiento de conjuntos de este espacio, en base a las dimensiones relativizadas de sus medidas de probabilidad.

Teorema 2.2. *Sea $E \subseteq \mathcal{P}(X)$. Entonces*

$$\dim_H^\varphi(E) = \min_{B \subseteq \{0,1\}^{<\mathbb{N}}} \dim^{\varphi,B}(E). \quad (2.21)$$

$$\dim_P^\varphi(E) = \min_{B \subseteq \{0,1\}^{<\mathbb{N}}} \text{Dim}^{\varphi,B}(E). \quad (2.22)$$

Notar que, tanto en (2.21) como en (2.22), el lado izquierdo de las ecuaciones son las clásicas dimensiones de Hausdorff y de empaquetamiento, y el lado derecho de ambas ecuaciones es una propiedad puntual del conjunto que hace uso de teoría de la información algorítmica relativizada. Además, existe un oráculo para el cual $\dim^{\varphi,B}$ y $\text{Dim}^{\varphi,B}$ son mínimos, no ínfimos. Para dimensión sin escala se cumple un teorema análogo.

En resumen, el principio de punto a conjunto es una caracterización completa de las dimensiones de Hausdorff y de empaquetamiento en términos de relativización con oráculos de las dimensiones individuales de los puntos.

Para acabar el capítulo, veamos una aplicación del principio de punto a conjunto en espacios Euclídeos: el teorema de proyección de Marstrand. Este teorema analiza la dimensión de Hausdorff de las proyecciones ortogonales de un conjunto analítico. Fue demostrado en 1954 para \mathbb{R}^2 y extendido por Mattila a \mathbb{R}^n en 1975. Para $e \in S^{n-1}$ la proyección de un conjunto $E \subset \mathbb{R}^n$ sobre la línea L_e que pasa por el origen y por e es

$$\text{proj}_e E = \{e \cdot x \mid x \in E\}$$

Recordar que un subconjunto de un espacio polaco (espacio topológico separable completamente metrizable) es analítico si es la imagen continua de otro espacio polaco.

Teorema 2.3. *Sea $E \subseteq \mathbb{R}^n$ un conjunto analítico con $\dim_H(E) = s$. Entonces, para casi todo $e \in S^{n-1}$,*

$$\dim_H(\text{proj}_e E) = \min\{s, 1\}. \quad (2.23)$$

En 2022 este teorema fue generalizado en [5] para el caso de conjuntos regulares, es decir, conjuntos para los que coinciden la dimensión de Hausdorff y la de empaquetamiento. El principal ingrediente de la demostración fue la dimensión efectiva de Lutz y su principio de punto a conjunto, que permite cuantificar las dimensiones fractales de un conjunto a partir de sus dimensiones efectivas.

Capítulo 3

Ejemplos de cálculo de dimensión efectiva de medidas de probabilidad

A lo largo de este capítulo, trataremos de obtener la dimensión efectiva de medidas de probabilidad en $X = \mathbb{R}$ con diferentes distribuciones, así como sus dimensiones fractales. Para eso, primero deduciremos la complejidad de Kolmogorov de las distintas medidas, y finalmente aplicaremos el principio de punto a conjunto (2.2) para detectar una serie de casos de dimensión 0. Para el estudio de ejemplos de dimensión positiva es necesario utilizar herramientas de aleatoriedad de Martin-Löf que no hemos abordado en este trabajo inicial.

Siguiendo lo explicado en el capítulo anterior, para el cálculo de la complejidad de Kolmogorov necesitaremos un conjunto denso en $\mathcal{P}(X)$. Utilizaremos el conjunto \mathcal{M} construido en la demostración del teorema (1.3). Para ello, tomamos como conjunto denso en X los racionales, y fijamos una función $f : \{0, 1\}^{<\mathbb{N}} \rightarrow \mathcal{P}(X)$ cuya imagen es

$$\text{Im}(f) = \mathcal{M} := \{\alpha_1 \delta_{x_1} + \dots + \alpha_n \delta_{x_n} : \alpha_1, \dots, \alpha_n \in \mathbb{Q} \cap [0, 1], x_1, \dots, x_n \in \mathbb{Q}, \sum_{j=1}^n \alpha_j = 1, n = 1, 2, \dots\}.$$

Recordad que si identificamos un programa con una secuencia de bits, se trata de una función que dado un programa nos devuelve una medida de probabilidad. También es necesario elegir una codificación concreta de las medidas de dicho conjunto. Una codificación razonable es enumerar los $\alpha_1, \dots, \alpha_n$ y los x_1, \dots, x_n . Ahora, estamos en condiciones de calcular la dimensión efectiva de una medida de probabilidad. Dado que $\mathcal{P}(X)$ tiene dimensión de Hausdorff y de empaquetado infinitas, utilizaremos la función de escala (2.5), luego cuando hablemos de estas dimensiones fractales, será siempre con dicha escala φ .

Empezamos por uno de los casos más sencillos. Sea $X = [0, 1]$, tomamos una medida de probabilidad μ cuyo soporte es la sucesión de puntos $x_i = 2^{-i}$ para todo $i \in \mathbb{N}_{>0}$, y donde el valor de la medida en cada punto es el propio punto. Podemos aproximar esta medida con otra del conjunto \mathcal{M} con un número finito de puntos, que dependerá de la precisión r requerida. Fijamos $r = 2^{-k}$ con $k \in \mathbb{N}$. Por lo tanto, la medida $\nu \in \mathcal{M}$ tendrá 2^k puntos. Definimos la medida ν de la siguiente manera:

$$\nu = \sum_{i=1}^{2^k} \alpha_i \delta_{x_i}$$

con $x_i = 2^{-i} \forall i \in \{1, \dots, 2^k\}$, $\alpha_i = x_i \forall i \in \{1, \dots, 2^k - 1\}$ y $\alpha_{2^k} = 1 - \sum_{i=1}^{2^k-1} 2^{-i}$. Solo hemos necesitado el entero k para describir dicha medida, luego su complejidad de Kolmogorov es como mucho $\log k$. Para acotar la distancia entre μ y ν es importante notar que son coincidentes para todo $x_i > 2^{-2^k}$ y que $\alpha_{2^k} = \mu(\{2^{-2^k}\}) + \sum_{i=2^{k+1}}^{\infty} \mu(\{2^{-i}\})$, donde $\sum_{i=2^{k+1}}^{\infty} \mu(\{2^{-i}\}) \leq 2^{-r}$.

Por tanto, se tiene que para cualquier boreliano $A \in \mathcal{B}(X)$, se tiene que $|\nu(A) - \mu(A)| = |\nu(A \cap [0, 2^{-2^k}]) - \mu(A \cap [0, 2^{-2^k}])| \leq \max\{\nu(A \cap [0, 2^{-2^k}]), \mu(A \cap [0, 2^{-2^k}])\} \leq 2^{-r}$, luego se cumplen las ecuaciones de la métrica de Prokhorov y $d_P(\mu, \nu) \leq 2^{-r}$. Entonces podemos acotar la complejidad de Kolmogorov de μ :

$$K_r(\mu) \leq \log k.$$

Con este resultado tenemos que

$$\dim^\varphi(\mu) = \liminf_{r \rightarrow \infty} \frac{\log(K_r(\mu))}{r} \leq \liminf_{k \rightarrow \infty} \frac{\log(\log k)}{2^k} = 0$$

es decir, la dimensión efectiva de esta medida es cero.

3.1. Distribución uniforme

En la distribución uniforme continua, la probabilidad se reparte por igual a lo largo de un intervalo. Se utiliza para extender la regla de Laplace para calcular la probabilidad en un espacio muestral con sucesos equiprobables, al caso en el que el espacio muestral sea un intervalo de \mathbb{R} . Como bien sabemos, una medida μ con distribución uniforme en un intervalo $[a, b]$ es aquella que verifica $\mu((a_1, a_2]) = (a_2 - a_1)/(b - a)$, si $a \leq a_1 < a_2 \leq b$.

Sea $[a, b] = [0, 1]$. Tomamos μ una medida de distribución uniforme con soporte X . Primero calculamos su complejidad de Kolmogorov a cierta precisión r . Fijamos $r = 2^k$ con $k \in \mathbb{N}$. Buscamos la medida en \mathcal{M} con menor complejidad y cuya distancia a μ sea menor que 2^{-r} . Parece lógico pensar que para aproximar una medida con distribución uniforme, los puntos x_i deben estar repartidos uniformemente en X , y el valor de los α_i debe ser igual para toda i . Elegimos

$$\nu_k = \sum_{i=1}^{2^k} \alpha_i \delta_{x_i}, \text{ donde } x_i = i \cdot 2^{-2^k} \text{ y } \alpha_i = 2^{-2^k},$$

esto es, con todos los puntos separados a distancia 2^{-2^k} . Calculamos la complejidad de ν_k utilizando el siguiente programa:

“Dado un número k fijo, enumera todos los puntos de la forma

$$x_i = i \cdot 2^{-2^k} \text{ y } \alpha_i = 2^{-2^k}, \text{ hasta } x_i = 1.”$$

Notar que solo hemos necesitado k para describir todos los puntos. Luego la complejidad de esta medida es de orden logarítmico: $K_r(\nu_k) \leq \log k + O(1)$.

Probemos ahora que $d_P(\mu, \nu_k) \leq 2^{-2^k}$. Tomamos $\alpha = 2^{-2^k}$ y veamos que $\nu_k(A) \leq \mu(A_\alpha) + \alpha$ para todo conjunto de Borel A , que sin pérdida de generalidad, puede considerarse incluido en $[0, 1]$. Por la proposición (1), tendremos el resultado.

Dado un conjunto $A \in \mathcal{B}([0, 1])$, denotamos por n_A el número de puntos x_i contenidos en A . Por un lado se tiene que $\nu_k(A) = n_A \cdot \alpha$. Por otro lado, al ser μ una medida uniforme en el intervalo $[0, 1]$, $n_A \cdot \alpha = \mu(\bigcup_{k|x_k \in A} [x_k, x_k + \alpha])$. Además $\bigcup_{k|x_k \in A} [x_k, x_k + \alpha] \subseteq A_\alpha$. De lo que se deduce inmediatamente la acotación, ya que

$$\nu_k(A) = n_A \cdot \alpha = \mu\left(\bigcup_{k|x_k \in A} [x_k, x_k + \alpha]\right) \leq \mu(A_\alpha) + \alpha.$$

Por lo tanto, $d_P(\mu, \nu_k) \leq 2^{-2^k}$, es decir, cuando $n \rightarrow \infty$ se tiene que $d_P(\mu, \nu_k) \rightarrow 0$. Entonces podemos afirmar que $K_r(\mu) \leq \log k + O(1)$. Luego la dimensión efectiva de esta medida es

$$\dim^\varphi(\mu) = \liminf_{r \rightarrow \infty} \frac{\log(K_r(\mu))}{r} \leq \liminf_{k \rightarrow \infty} \frac{\log(\log k + O(1))}{2^k} = 0 \quad (3.1)$$

Tratamos ahora de generalizar este ejemplo para cualquier medida con distribución uniforme cuyo soporte sea un único intervalo. Sea ahora $\mu_{a,b}$ una medida de distribución uniforme con soporte el intervalo $[a, b]$ con $a, b \in \mathbb{R}$ y $a < b$. Veamos como construir una medida $\nu_{a,b} \in \mathcal{M}$ cumpliendo $d_P(\mu_{a,b}, \nu_{a,b}) \leq 2^{-r}$.

Primero tomamos una partición diádica donde cada intervalo tiene una longitud de $l = 2^{-k_1}$ con k_1 a elegir para conseguir la precisión deseada. Elegimos x_1 y x_n entre esos diádicos de manera que $x_1 \leq a < x_1 + l$ y $x_n - l < b \leq x_n$, y de esta manera, $[a, b] \subseteq [x_1, x_n]$. Notemos que el número de intervalos contenidos en $[x_1, x_n]$ será, como mucho, $[b-a]/l + 1$ dependiendo de si a y b coinciden con x_1 y x_n respectivamente, donde $[x]$ es la parte entera de x .

Asignamos a cada intervalo diádico una probabilidad de $l/(x_n - x_1)$ en su extremo inferior. Es decir, definimos

$$\nu_{a,b} = \sum_{i=1}^{n-1} \alpha_i \delta_{x_i},$$

con $\alpha_i = l/(x_n - x_1)$ y $x_i = u_1 + m_i \cdot 2^{-k_1}$ con $u_1 = [a]$ y $m_i \in \mathbb{Z}$ de manera que x_i está contenido en $[x_1, x_n]$ para todo $i \in \{1 \dots n-1\}$. Veamos ahora cómo elegir k_1 de manera que mantengamos la precisión en 2^{-r} como en el ejemplo anterior.

Sea $A \in \mathcal{B}(X)$, denotamos como I al conjunto de subíndices i tales que $x_i \in A$. Es decir, $|I| = n_A$ representa el número de puntos x_i contenidos en A .

- Si $1, n \notin I$, tenemos que $\nu_{a,b}(A) = n_A \cdot \frac{l}{x_n - x_1}$. Por otro lado $\mu_{a,b}([x_n, x_n + l)) = \frac{l}{b-a}$. Además $\bigcup_{i \in I_1} [x_i, x_i + l) \subseteq A_l$. De lo que se deduce inmediatamente la acotación, ya que

$$\nu_{a,b}(A) = n_A \cdot \frac{l}{x_n - x_1} \leq n_A \cdot \frac{l}{b-a} = \mu\left(\bigcup_{i \in I} [x_i, x_i + l)\right) \leq \mu(A_l) + l.$$

- Si x_1 o x_n pertenecen al conjunto A , dado que la discrepancia entre ambas medidas en los intervalos extremos $([x_1, x_2], [x_{n-1}, x_n])$ se puede acotar por $\alpha = \frac{2l}{b-a}$, si denotamos como I_2 al conjunto de índices i tales que $x_i \in A \cap [a, b]$ tenemos que

$$\nu_{a,b}(A) = n_A \cdot \frac{l}{x_n - x_1} \leq \mu_{a,b}\left(\bigcup_{i \in I_2} [x_i, x_i + l)\right) + \frac{2l}{b-a} \leq \mu_{a,b}\left(\bigcup_{i \in I_2} [x_i, x_i + \beta)\right) + \beta \leq \mu_{a,b}(A_\beta) + \beta,$$

donde $\beta = \max\{l, \frac{2l}{b-a}\}$.

Por tanto, $d_P(\mu_{a,b}, \nu_{a,b}) \leq \beta$. Si queremos que la precisión sea 2^{-r} , tenemos que elegir l de manera que $\beta \leq 2^{-r}$. Sea $s \in \mathbb{N}$ el menor posible tal que $1/(b-a) \leq 2^{-s}$. Tomamos $k_1 = r+1+s$. Notamos que s es una constante que no depende de r . Hay que recordar que $l = 2^{-k_1}$, luego llegados a este punto, para describir la medida $\nu_{a,b}$ necesitamos los puntos x_1, x_n y el entero k_1 . Podemos definir x_1 y x_n como

$$x_1 = u_1 + m_1 \cdot 2^{-k_1} \text{ con } u_1 = [a] \text{ y } m_1 \in \mathbb{Z} \text{ tal que } u_1 + m_1 \cdot 2^{-k_1} \leq a \text{ y } u_1 + (m_1 + 1) \cdot 2^{-k_1} > a$$

$$x_n = u_n + m_n \cdot 2^{-k_1} \text{ con } u_n = [b] \text{ y } m_n \in \mathbb{Z} \text{ tal que } u_n + m_n \cdot 2^{-k_1} \geq b \text{ y } u_n + (m_n - 1) \cdot 2^{-k_1} < b$$

Por lo tanto la complejidad de Kolmogorov de $\nu_{a,b}$ depende de la complejidad de m_1, u_1, m_n, u_n y k_1 . Como son números enteros, su complejidad no depende de r . En el peor de los casos, la complejidad de x_1 y x_n es

$$K(x_1) \leq K(k_1) + K(u_1) + K(m_1) + \mathcal{O}(1) \leq \log k_1 + \log u_1 + \log m_1 + \mathcal{O}(1)$$

$$K(x_n) \leq K(k_1) + K(u_n) + K(m_n) + \mathcal{O}(1) \leq \log k_1 + \log u_n + \log m_n + \mathcal{O}(1)$$

Dado que $d_P(\mu_{a,b}, \nu_{a,b}) \leq 2^{-r}$, la dimensión efectiva de $\mu_{a,b}$ con escala φ es cero:

$$\dim^\varphi(\mu_{a,b}) = \liminf_{r \rightarrow \infty} \frac{\log(K_r(\mu_{a,b}))}{r} \leq \liminf_{r \rightarrow \infty} \frac{\log(K(k_1) + K(x_1) + K(x_n))}{r} = 0 \quad (3.2)$$

Por tanto, podemos afirmar que todas las medidas de probabilidad con distribución uniforme cuyo soporte es un intervalo con extremos reales, tienen dimensión efectiva cero. Gracias al principio punto a conjunto que hemos visto en el capítulo anterior, tenemos que la dimensión de Hausdorff de dicho conjunto, que denotaremos como \mathcal{U} , cumple

$$\dim_H^\varphi(\mathcal{U}) = \min_{A \subseteq \mathbb{N}} \dim^{\varphi, A}(\mathcal{U}) = \min_{A \subseteq \mathbb{N}} \sup_{\mu \in \mathcal{U}} \dim^{\varphi, A}(\mu) = 0. \quad (3.3)$$

Observar que la dimensión efectiva que hemos calculado no incluye oráculo, lo que se corresponde con $A = \emptyset$ en (3.3).

3.1.1. Mixtura de uniformes

Supongamos ahora que tenemos una medida escalonada con n intervalos de igual longitud donde el valor de la medida se distribuye uniformemente en cada intervalo. La medida total de cada intervalo es diferente y proporcional a su posición en el conjunto de intervalos, es decir, la medida del primer intervalo será la menor, y del último será la mayor. Formalizamos la definición.

Sea μ_i una medida con distribución uniforme en el intervalo $(i-1, i)$, y sea $p_i = i/(\sum_{j=1}^n j)$ la medida total de cada intervalo para $i \in \{1, \dots, n\}$. Entonces nuestra medida se define como

$$\mu_n(A) = \sum_{i=1}^n p_i \cdot \mu_i(A) \quad (3.4)$$

Para buscar la medida en \mathcal{M} con menor complejidad y que aproxime μ con precisión $r \in \mathbb{N}$, elegimos

$$\nu_r = \sum_{i=1}^n \left(\alpha_i \cdot \sum_{j=1}^{2^r} \delta_{x_{i,j}} \right)$$

con $x_{i,j}$ el punto número j del intervalo i y $\alpha_i = p_i/2^r$ para cada $i \in \{1, \dots, n\}$. Notar que cada intervalo tiene 2^r puntos, luego la medida cuenta con $n \cdot 2^r$ puntos totales. Utilizamos el siguiente programa para describir la medida:

“Dados r y n números fijos, enumera todos los puntos de la forma $x_{i,j} = j-1 + i \cdot 2^{-r}$

desde $i = 1, j = 1$ hasta $i = 2^r, j = n$, y enumera todos los $\alpha_i = \frac{i}{\sum_{j=1}^n j} \cdot 2^{-r}$ hasta $i = n$.”

Notar que solo hemos necesitado r y n , ambos números enteros, luego el orden de la complejidad de ν_r es logarítmico y la complejidad de Kolmogorov de μ_n se puede acotar de la siguiente forma:

$$K_r(\mu_n) \leq \log r + \log n + O(1).$$

Con este resultado, podemos afirmar que la dimensión efectiva de esta medida también es nula:

$$\dim^\varphi(\mu_n) = \liminf_{r \rightarrow \infty} \frac{\log(K_r(\mu_n))}{r} \leq \liminf_{r \rightarrow \infty} \frac{\log(\log r + \log n + O(1))}{r} = 0.$$

Siguiendo un razonamiento similar al anterior, podríamos probar que el conjunto de medidas con distribución uniforme en varios soportes tiene dimensión de Hausdorff nula si los extremos de los intervalos son números racionales.

3.2. Distribución exponencial

La distribución exponencial sirve para modelar el tiempo de espera Y para la ocurrencia de un evento aleatorio. Tiene un parámetro característico $\lambda > 0$ que indica el número de veces que se espera que ocurra el evento estudiado durante un periodo de tiempo concreto. Se utiliza habitualmente para el cálculo de la fiabilidad de un producto. Su función de densidad está dada por $f(x; \lambda) = \lambda e^{-\lambda x}$ si $x \geq 0$. La probabilidad asociada a cualquier boreliano se calcula mediante la integral de la función de densidad sobre ese conjunto. Además, su función de distribución acumulada se define como $F(x; \lambda) = 1 - e^{-\lambda x}$ si $x \geq 0$.

Sea μ una medida de probabilidad que sigue una distribución exponencial de parámetro $\lambda = 1$. Buscamos una medida $\nu \in \mathcal{M}$ que se aproxime a μ a precisión r . Suponemos $r \in \mathbb{N}$ fijo. La intuición nos dice que en este caso cada α_i de ν tendrá un valor diferente. Además, a pesar de que μ está definida para todo $x \geq 0$, nuestra función tendrá un número finito de puntos.

Tomamos una partición diádica d_i con $i = 0, \dots, n$ de longitud $l = 2^{-k_1}$ con k_1 a elegir para conseguir la precisión deseada. Elegiremos también n en base a la precisión buscada, de modo que d_n cumpla $P(Y > d_n) = e^{-d_n} \leq 2^{-r}$. La probabilidad acumulada en cada intervalo diádico $[d_i, d_{i+1}]$ es

$$p_i = P(Y \leq d_{i+1}) - P(Y \leq d_i) = (e^{-l})^i (1 - e^{-l}) \quad \text{para todo } i \in \{0, \dots, n-1\}.$$

Nuestra intención es asignar al extremo inferior de cada intervalo la probabilidad correspondiente al mismo y definir una medida ν como en el anterior ejemplo. Pero hay que recordar que los α_i de la medida que definamos han de ser racionales. Por tanto necesitamos aproximar la probabilidad de cada intervalo a un número racional. Para ello, tomamos las siguientes aproximaciones:

$$e^{-l} \approx 1 - l \quad \text{y} \quad 1 - e^{-l} \approx l.$$

En consecuencia, la probabilidad que asignamos a cada intervalo diádico es $p_i^* = l(1 - l)^i$ para todo $i \in \{0, \dots, n-1\}$. Estamos aproximando las probabilidades de una exponencial por las de una geométrica que modela el número de fracasos antes del primer éxito con probabilidad de éxito l , luego se cumple $\sum_{i=1}^{n-1} p_i^* \leq 1$. Recordemos que la función de masa de probabilidad de la distribución geométrica con probabilidad p es $P(Z = i) = p(1 - p)^i$, con $i = 0, 1, \dots$.

Notar que si desarrollamos e^{-l} en serie de Taylor en 0 con resto integral tenemos que

$$e^{-l} = 1 - l + \frac{e^{-x}}{2} l^2, \quad \text{con } x \text{ punto intermedio entre } 0 \text{ y } l.$$

Por tanto $|e^{-l} - (1 - l)| \leq l^2/2$ y $1 - l \leq e^{-l}$. Además $1 - e^{-l} \leq l$, luego no tenemos garantía de que $p_i^* \leq p_i$. Pasamos a definir la siguiente medida:

$$\nu = \sum_{i=0}^n \alpha_i \delta_{x_i},$$

con $x_i = d_i$ para todo $i \in \{0, \dots, n\}$, $\alpha_i = p_i^*$ para todo $i \in \{0, \dots, n-1\}$ y $\alpha_n = 1 - \sum_{i=0}^{n-1} p_i^*$. Notar que α_n acumula la medida de μ en el intervalo $[x_n, +\infty)$ sumada a la discrepancia total entre las probabilidades aproximadas de los anteriores intervalos $[x_i, x_{i+1}]$, es decir, $\alpha_n = e^{-d_n} + \sum_{i=0}^{n-1} (p_i - p_i^*)$. Buscamos una cota de la distancia de Prokhorov entre ambas medidas. Sin pérdida de generalidad, podemos suponer $A \subseteq [0, \infty)$.

- Sea $A \in \mathcal{B}([0, \infty))$ tal que $d_n \notin A$, y sea $I_1 = \{i | x_i \in A\}$. Dado que $\bigcup_{i \in I_1} [x_i, x_i + l) \subseteq A_l$, se tiene que $\mu(\bigcup_{i \in I_1} [x_i, x_i + l)) = \sum_{i \in I_1} p_i \leq \mu(A_l)$. Por lo tanto, $\nu(A) = \sum_{i \in I_1} p_i^* \leq \sum_{i \in I_1} p_i + \sum_{i \in I_1} |p_i - p_i^*| \leq \mu(A_l) + \sum_{i \in I_1} |p_i - p_i^*| \leq \mu(A_l) + \sum_{i=0}^{n-1} |p_i - p_i^*|$.
- Sea ahora $A \in \mathcal{B}([0, \infty))$ tal que $d_n \in A$. Si tomamos $B = A \setminus \{d_n\}$, entonces $A = B \cup \{d_n\}$. Sea $I_2 = \{i | x_i \in B\}$, tenemos que $\bigcup_{i \in I_2} [x_i, x_i + l) \subseteq A_l$, y por tanto $\mu(\bigcup_{i \in I_2} [x_i, x_i + l)) = \sum_{i \in I_2} p_i \leq \mu(A_l)$. Como $d_n \in A$, $\nu(A) = \sum_{i \in I_2} p_i^* + \alpha_n \leq \sum_{i \in I_2} p_i + \sum_{i \in I_2} |p_i - p_i^*| + \alpha_n$. De lo que deducimos que $\nu(A) \leq \mu(A_l) + \sum_{i \in I_2} |p_i - p_i^*| + \alpha_n \leq \mu(A_l) + \sum_{i=0}^{n-1} |p_i - p_i^*| + \alpha_n$.

Por lo tanto, $\forall A \in \mathcal{B}(X)$ se cumple $\nu(A) \leq \mu(A_\beta) + \beta$, con $\beta = \max\{l, \sum_{i=0}^{n-1} |p_i - p_i^*| + \alpha_n\}$. Recordad que la precisión fijada era r , luego necesitamos que $\beta \leq 2^{-r}$. Es decir, se debe cumplir: $l \leq 2^{-r}$ y $\sum_{i=0}^{n-1} |p_i - p_i^*| + \alpha_n \leq 2^{-r}$. Para ello podemos imponer que $\sum_{i=0}^{n-1} |p_i - p_i^*| \leq 2^{-(r+1)}$ y elegimos n de manera que $\alpha_n \leq 2^{-(r+1)}$. Veamos cómo elegir l para cumplir todas las cotas. Para todo $i \in \{0, \dots, n-1\}$ se tiene que

$$|p_i - p_i^*| = |e^{-li}(1 - e^{-l}) - (1 - l)^i l| \leq l|e^{-li} - (1 - l)^i| + e^{-li}|1 - e^{-l} - l|$$

Desarrollamos la función x^i en serie de Taylor en torno a $(1 - l)$ y tenemos que

$$|e^{-li} - (1 - l)^i| \leq i\xi^{i-1}|e^{-l} - 1 + l| \leq i e^{-l(i-1)}|e^{-l} - 1 + l|$$

con ξ un punto intermedio entre $1 - l$ y e^{-l} . En consecuencia,

$$|p_i - p_i^*| \leq (e^{-l(i-1)}li + e^{-li})|1 - e^{-l} - l|$$

Acotando por las sumas infinitas obtenemos lo siguiente:

$$\sum_{i=0}^{n-1} |p_i - p_i^*| \leq \left(l \sum_{i=0}^{n-1} i e^{-l(i-1)} + \sum_{i=0}^{n-1} e^{-li} \right) |e^{-l} - 1 + l| \leq \left(l \frac{1}{(1 - e^{-l})^2} + \frac{e^l}{e^l - 1} \right) |e^{-l} - 1 + l|$$

Por lo tanto, si escogemos $l \leq 2^{-r}$ de manera que cumpla (3.5), tenemos que $d_P(\mu, \nu) \leq 2^{-r}$.

$$\left(l \frac{1}{(1 - e^{-l})^2} + \frac{e^l}{e^l - 1} \right) |e^{-l} - 1 + l| \leq 2^{-(r+1)} \quad (3.5)$$

Finalmente podemos describir la medida ν con el siguiente programa:

“Dado r número fijo, calcula el mayor $l = 2^{-m}$ con $m \in \mathbb{N}, m \geq r$ y l cumpliendo (3.5).

Enumera todos los puntos de la forma $x_i = il, \alpha_i = l(1 - l)^i$ desde $i = 0$

hasta cumplir $1 - \sum_{j=0}^i \alpha_j \leq 2^{-(r+1)}$. Añade $x_{i+1} = (i+1)l$ y $\alpha_{i+1} = 1 - \sum_{j=0}^i \alpha_j$.”

Notar que solo hemos necesitado r , es decir, no es necesario conocer l o n . Sabemos que r es un número natural, luego la complejidad de su descripción es de orden logarítmico, lo que implica que $K_r(\mu) \leq \log r + O(1)$. Finalmente la dimensión efectiva de μ será cero dado que

$$\dim^\varphi(\mu) = \liminf_{r \rightarrow \infty} \frac{\log(K_r(\mu))}{r} \leq \liminf_{r \rightarrow \infty} \frac{\log(\log r + O(1))}{r} = 0.$$

Bibliografía

- [1] ONNO VAN GAANS, *Probability measures on metric spaces*, notes of the seminar "Stochastic Evolution Equations", Delft University of Technology, Winter 2002/2003.
- [2] ELVIRA MAYORDOMO, *Una generalización del teorema de Proyección de Marstrand*, La Gaceta de la RSME, Vol. 25, Núm 2, Págs 343-352, 2022.
- [3] PATRICK BILLINGSLEY, *Convergence of Probability Measures*. The University of Chicago. 1999.
- [4] KENNETH FALCONER, *Fractal Geometry: Mathematical Foundations and Applications*, University of Bristol, 1952.
- [5] DONALD STULL, NEIL LUTZ, *Projection theorems using effective dimension*, Computability 11, pp. 85-112, 2022.
- [6] JAN REIMANN, *Computability and Fractal Dimension*, Doctoral Dissertation, Universität Heidelberg, 2004.
- [7] J.H. LUTZ, N. LUTZ, AND E. MAYORDOMO, *Extending the Reach of the Point-to-Set Principle*, 39th International Symposium on Theoretical Aspects of Computer Science (STACS 2022), 2022.
- [8] MING LI, PAUL VITÁNYI, *An Introduction to Kolmogorov Complexity and Its Applications*, Third Edition, Springer Verlag 2008.
- [9] PATRICK BILLINGSLEY, *Probability and Measure*, The University of Chicago, 1995.
- [10] BEN KOCH, *Exploring developments in effective fractal dimension and applications of the point-to-set principle*. Master dissertation, University of Chicago, 2023
- [11] FACULTAD DE CIENCIAS DE LA UNIVERSIDAD DE ZARAGOZA, *Directrices propias para la elaboración del trabajo fin de grado en Matemáticas*, disponible en <https://ciencias.unizar.es/trabajo-fin-de-grado-en-matematicas>.