ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

EXPLORANDO ASPECTOS DEL ANÁLISIS DESDE UN ENFOQUE PROBABILISTA: EL TEOREMA DE DENSIDAD DE LEBESGUE COMO UNA APLICACIÓN DE LA TEORÍA DE MARTINGALAS

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO MATEMÁTICO

PROYECTO DE INVESTIGACIÓN

CRISTIAN ANDRÉS GUACHAMIN ARGUELLO
cristian.guachamin01@gmail.com

Directora: ADRIANA UQUILLAS ANDRADE, PHD.
adriana.uquillas@epn.edu.ec

QUITO, MARZO 2020

DECLARACIÓN

Yo CRISTIAN ANDRÉS GUACHAMIN ARGUELLO, declaro bajo juramento que el trabajo aquí escrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo mis derechos de propiedad intelectual, correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su reglamento y por la normatividad institucional vigente.

Cristian Andrés Guachamin Arguello

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por CRISTIAN ANDRÉS GUACHAMIN ARGUELLO, bajo mi supervisión.

Adriana Uquillas Andrade, PhD.
Director del Proyecto

AGRADECIMIENTOS

A mis padres, por haberme enseñado mediante el ejemplo el valor del trabajo duro.

A mi hermano, quien es una de las razones que me empujaron a siempre buscar ser mejor.

A Michelle y a mis amigos, por siempre estar ahí.

A mis profesores, en particular a Andrés Merino, quien fue mi primer maestro y por quien aprendí a amar las matemáticas.

DEDICATORIA

A mis padres, Blanca y Víctor,
y a mi hermano, Kevin.

Índice general

Resumen

Dentro del Análisis Matemático, el Teorema de densidad de Lebesgue suele ser presentado como un corolario del Teorema de diferenciación de Lebesgue, el cual requiere del Teorema de recubrimientos de Vitali para su demostración. En este trabajo, se construye una demostración para el Teorema de densidad de Lebesgue sin hacer uso del Teorema de recubrimientos de Vitali. Para ello, se da un breve repaso de algunos resultados de Teoría de la medida y de Teoría de probabilidad, haciendo especial énfasis en el Teorema de convergencia de martingalas, mediante el cual se construyen las herramientas necesarias para proveer de una demostración probabilista tanto del Teorema de Radon-Nikodym, como del Teorema de densidad de Lebesgue.

Abstract

Within Mathematical Analysis, the Lebesgue Density Theorem is frequently presented as a corollary of the Lebesgue Differentiation Theorem, which requires the Vitali Covering Theorem for its proof. In this paper, a proof is constructed for the Lebesgue Density Theorem without using the Vitali Covering Theorem. For it, a brief review of some results of Measure Theory and Probability Theory is given, emphasizing specially the Martingale Convergence Theorem, through which the necessary tools are built to provide a probabilistic proof of both the Radon-Nikodym Theorem, as the Lebesgue Density Theorem.

Capítulo 1

Introducción

El Teorema de densidad de Lebesgue es un objeto de estudio importante en varias áreas de la Matemática como, por ejemplo, la Teoría descriptiva de conjuntos. En el trabajo *The descriptive set theory of the Lebesgue Density Theorem* de Andretta y Camerlo [1], se da una versión del Teorema de densidad de Lebesgue en el espacio de Cantor, y no sólo eso, sino que además estudian la complejidad de algunos conjuntos asociados al mismo. Por otra parte, en el Análisis Matemático, el Teorema de densidad de Lebesgue, así como su demostración a través del Teorema de recubrimientos de Vitali, es de utilidad al estudiar la derivación de integrales en espacios medidos abstractos, como se ve en [6].

Dentro de la Teoría de probabilidades, por una parte, como se puede ver en [7], el Teorema de densidad de Lebesgue es útil al momento de mejorar búsquedas sobre grafos aleatorios, y por otra, en [3], vemos que es de interés estudiar teoremas análogos al Teorema de densidad de Lebesgue sobre fractales y conjuntos asociados a un movimiento browniano. Esto último, nos lleva a pensar que el dar una demostración probabilista del Teorema de densidad de Lebesgue, puede ayudarnos a comprender mejor sus implicaciones dentro de la Teoría de probabilidades.

Este trabajo tiene como objetivo principal la construcción de una demostración probabilista del Teorema de densidad de Lebesgue a través de la Teoría de martingalas. Para ello, hemos empezado haciendo una recopilación y revisión profunda de resultados de Teoría de la medida y Teoría de martingalas presentados en [14] y los hemos distribuido en los tres primeros capítulos, mientras que las ideas utilizadas en el Capítulo 4 se basaron en el trabajo de Morayne y Solecki [11].

En el Capítulo 2, daremos un repaso a los principales resultados de la Teoría de la medida e integración y los relacionaremos con los espacios de probabilidad, finalizando con la revisión de algunas resultados sobre espacios $\mathcal{L}^p$, entre ellos, el Teorema de proyección.

En el Capítulo 3, probaremos la existencia de la esperanza condicional de una variable aleatoria, valiéndonos del Teorema de proyección visto en el Capítulo 2, y veremos algunas de sus propiedades, mismas que nos ayudarán al momento de trabajar con martingalas.

En el Capítulo 4, empezaremos a trabajar con martingalas, enfocándonos especialmente hacia el Teorema de convergencia de martingalas y pasando por el Lema de cruces. A partir del Teorema de convergencia de martingalas, se probarán tanto el Teorema de Lévy, como el Teorema de Radon-Nikodym.

Finalmente, en el Capítulo 5, haciendo uso de las herramientas provistas en el Capítulo 4, construiremos una filtración y una martingala adecuadas que, junto con el Teorema de Lévy, nos servirán para dar una demostración probabilista del Teorema de densidad de Lebesgue.

Capítulo 2

Medida e integración

Al plantearse la necesidad de *medir* subconjuntos de $\mathbb{R}^3$, una posible respuesta es asociar a cada conjunto lo que intuitivamente conocemos como su *volumen*. Inicialmente, la idea más básica que tenemos de volumen y que guarda relación con la geometría establece que el volumen debe cumplir las siguientes condiciones:

- el volumen de dos conjuntos disjuntos es igual a la suma de sus volúmenes;
- el volumen de un conjunto es invariante bajo traslaciones, rotaciones y reflexiones; y
- el volumen de un paralelepípedo se obtiene al tomar tres aristas ortogonales entre sí y multiplicar sus longitudes.

Basándonos en esto, el Cálculo integral nos brinda herramientas para extender la noción de volumen a un conjunto $A \subseteq \mathbb{R}^3$ de manera satisfactoria, siempre que A esté acotado y se encuentre determinado por superficies adecuadas. En este caso la idea para definir su volumen, consiste en aproximar al conjunto A por una cantidad finita de paralelepípedos disjuntos y, de esta manera, tomar el volumen de A como el límite de la suma de los volúmenes de los paralelepípedos mencionados, haciendo uso de la primera condición impuesta a la noción de volumen.

Para poder extender la definición de volumen a más conjuntos, por ejemplo, a conjuntos no acotados pero que tengan cierta regularidad, no bastará con aproximarlos con una cantidad finita de paralelepípedos, sino que será necesario una cantidad numerable de estos, por lo tanto, la primera condición que impusimos a la noción de volumen deberá ser extendida.

Por lo tanto, nos gustaría que existiese una función $\mu\colon \mathcal{P}(\mathbb{R}^3) \to [0,\infty]$, donde $\mathcal{P}(\mathbb{R}^3)$ representa al conjunto potencia de $\mathbb{R}^3$, tal que a cada subconjunto de $\mathbb{R}^3$ le asigne un valor no negativo que represente lo que entendemos por su volumen; es decir, que μ debería cumplir las siguientes propiedades:

a) Si $\{A_n\}_{n\in\mathbb{N}}$ es una familia de subconjuntos de $\mathbb{R}^3$, disjuntos dos a dos, entonces
$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mu(A_n).$$

b) Si $E, F \subseteq \mathbb{R}^3$ y E es congruente con F (es decir que E se puede transformar en F mediante rotaciones, traslaciones o reflexiones), entonces $\mu(E) = \mu(F)$.

c) Si $Q \subseteq \mathbb{R}^3$ es un paralelepípedo; es decir
$$Q = [a_1, b_1] \times [a_2, b_2] \times [a_3, b_3],$$
con a_i y $b_i \in \mathbb{R}$, para $i \in \{1, 2, 3\}$, entonces
$$\mu(Q) = (b_1 - a_1)(b_2 - a_2)(b_3 - a_3).$$

Existe una función que cumple con a), b) y c), sin embargo, tiene como limitante que no está definida sobre todo $\mathcal{P}(\mathbb{R}^3)$, es decir, no puede medir todos los subconjuntos de $\mathbb{R}^3$.

En 1924, Banach y Tarski [2], probaron que, si E y F son subconjuntos de $\mathbb{R}^n$ ($n \geq 3$), acotados y con interior no vacío, entonces existen $\{E_i\}_{i=1}^k$ y $\{F_i\}_{i=1}^k$, particiones de E y F, respectivamente, tales que E_i es congruente con F_i, para todo $i \in \{1, \ldots, k\}$. Esto nos dice que podemos partir una naranja en una cantidad finita de pedazos, reordenarlos y así formar una bola del tamaño de la Tierra. Así, si μ estuviese definida sobre todo $\mathbb{R}^3$, este resultado sería contradictorio, en otra palabras, existen conjuntos de los cuales no se puede medir su volumen.

Esto nos indica que debemos realizar una clasificación de los subconjuntos de $\mathbb{R}^n$ entre los que son susceptibles de asignarles una medida y los que no. Para realizar esta clasificación, nos valdremos de la idea de σ-álgebra, la cual nos ayudará a limitar los conjuntos que mediremos.

Las últimas dos propiedades que le exigimos a μ están directamente relacionadas con la geometría euclidiana; mientras que las funciones que cumplen a), llamadas medidas, aparecen en muchas otras situaciones, por ejemplo, en un problema de física que involucre distribuciones de masa, $\mu(E)$ puede representar la masa total en la región E. En cambio, en Teoría de probabilidades, que es el área en la que centraremos nuestro trabajo, uno considera un conjunto Ω que representa los posibles resultados de un experimento aleatorio, y para ciertos subconjuntos $E \subseteq \Omega$, la cantidad $\mu(E)$ representa la probabilidad de que el resultado del experimento pertenezca a E.

En este capítulo daremos un repaso por ciertos resultados básicos de Teoría de la medida e Integración y los iremos relacionando con la Teoría de probabilidades. Varios resultados serán presentados sin demostración, pero en caso de requerirla, se puede revisar [8, 9].

2.1. Espacios medidos

En esta sección consideraremos Ω como un conjunto no vacío, y estudiaremos un tipo particular de familias de subconjuntos de Ω que serán de gran importancia a lo largo de este trabajo.

DEFINICIÓN 2.1 (π-sistema, álgebra y σ-álgebra)**.** *Se dice que una colección no vacía $\mathcal{F} \subseteq \mathcal{P}(\Omega)$, donde $\mathcal{P}(\Omega)$ representa al conjunto potencia de Ω), es un π-sistema de subconjuntos de Ω si*

$$A, B \in \mathcal{F} \implies A \cap B \in \mathcal{F}. \tag{2.1}$$

Si, además, se tiene que

$$A \in \mathcal{F} \implies \Omega \setminus A \in \mathcal{F}, \tag{2.2}$$

diremos que $\mathcal{F}$ es una álgebra de subconjuntos de Ω.

Si $\mathcal{F}$ verifica (2.2) *y cumple que*

$$\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{F} \implies \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}, \tag{2.3}$$

entonces diremos que $\mathcal{F}$ es una σ-álgebra de subconjuntos de Ω.

DEFINICIÓN 2.2 (Espacio medible, conjunto medible)**.** *Si Ω es un conjunto no vacío y $\mathcal{F}$ es una σ-álgebra de subconjuntos de Ω, entonces al par $(\Omega, \mathcal{F})$ se lo llama espacio medible. A los elementos de $\mathcal{F}$ se los llama conjuntos $\mathcal{F}$-medibles o, si no hay riesgo de confusión, simplemente se los llama conjuntos medibles.*

En este trabajo, consideraremos Ω como el conjunto de todos los posibles resultados de un experimento aleatorio y lo llamaremos espacio muestral; a los elementos de Ω los llamaremos puntos muestrales. Si definimos una σ-álgebra $\mathcal{F}$ de subconjuntos de Ω, entonces podemos definir el espacio medido $(\Omega, \mathcal{F})$; a los elementos de

$\mathcal{F}$ los llamaremos eventos.

Sean $\mathcal{F}$ una σ-álgebra de subconjuntos de Ω y $F \in \mathcal{F}$. Consideremos la siguiente colección de subconjuntos de F:

$$\mathcal{F}|_F := \{A \cap F : A \in \mathcal{F}\}.$$

No es complicado ver que $\mathcal{F}|_F$ es una σ-álgebra de subconjuntos de F. De hecho, a esta σ-álgebra se la conoce como la restricción de $\mathcal{F}$ al evento F y la retomaremos en el siguiente capítulo. Al par $(F, \mathcal{F}|_F)$ se lo conoce como subespacio medido de $(\Omega, \mathcal{F})$.

En ocasiones, nos interesa estudiar ciertos subconjuntos de Ω, los cuales pueden o no formar una σ-álgebra. En ese caso, buscaremos trabajar la σ-álgebra más pequeña que contenga a estos conjuntos.

DEFINICIÓN 2.3 (σ-álgebra generada)**.** *Sea $\mathcal{C} \subseteq \mathcal{P}(\Omega)$. Se define la σ-álgebra generada por la colección $\mathcal{C}$, denotada por $\sigma(\mathcal{C})$, como la σ-álgebra de subconjuntos de Ω más pequeña que contiene a $\mathcal{C}$. Es decir, que $\sigma(\mathcal{C})$ es la intersección de todas las σ-álgebras que contienen a $\mathcal{C}$.*

La σ-álgebra generada por una colección de subconjuntos de Ω es de especial utilidad cuando la colección en cuestión es un π-sistema, dado que hay propiedades que, al ser ciertas para un π-sistema, se extienden inmediatamente a la σ-álgebra más pequeña que lo contiene.

Como ejemplo, presentamos a continuación una σ-álgebra que nos acompañará a lo largo de todo este trabajo:

EJEMPLO 1 (σ-álgebra de Borel)**.** Sea (Ω, τ) un espacio topológico. Se define la *σ-álgebra de Borel* sobre Ω, denotada por $\mathcal{B}(\Omega)$, como la σ-álgebra generada por τ; es decir, $\mathcal{B}(\Omega) := \sigma(\tau)$. A los elementos de $\mathcal{B}(\Omega)$ se los llama *conjuntos de Borel* o *borelianos*.

NOTA. Denotaremos por $\mathcal{B} := \mathcal{B}(\mathbb{R})$ a la σ-álgebra de Borel sobre $\mathbb{R}$, dotado de su topología usual.

Los borelianos de un espacio topológico podrían llegar a tener estructuras bastante complejas, por lo tanto, nos gustaría tratar con alguna subclase de la σ-álgebra de Borel de tal forma que al estudiar propiedades sobre ella, estas se pudieran extender fácilmente a todos los borelianos; en este trabajo, esas subclases serán π-sistemas. En particular, buscaremos trabajar con π-sistemas que generen a $\mathcal{B}$.

En Teoría de probabilidades, es de particular interés calcular la probabilidad de conjuntos asociados a intervalos de la forma $(-\infty, x]$, con $x \in \mathbb{R}$. Es por esto, que trabajaremos con frecuencia con la colección

$$\pi(\mathbb{R}) := \{(-\infty, x] : x \in \mathbb{R}\},$$

misma que guarda relación con $\mathcal{B}$ en el sentido dado por la siguiente proposición:

PROPOSICIÓN 2.1. *Sea*

$$\pi(\mathbb{R}) := \{(-\infty, x] : x \in \mathbb{R}\}.$$

Se tiene que $\pi(\mathbb{R})$ es un π-sistema y

$$\mathcal{B} = \sigma(\pi(\mathbb{R})).$$

Al inicio de esta sección, hablamos de funciones que nos permitan medir conjuntos; naturalmente, buscaremos medir los conjuntos medibles de un espacio medible. Consideremos un espacio medible $(\Omega, \mathcal{F})$ y una función de conjuntos no negativa $\mu\colon \mathcal{F} \to [0, \infty]$. Se dice que μ es:

- **Aditiva**: si $\mu(\varnothing) = 0$ y, para todo $A, B \in \mathcal{F}$,

$$A \cap B = \varnothing \implies \mu(A \cup B) = \mu(A) + \mu(B);$$

- σ**-aditiva**: si $\mu(\varnothing) = 0$ y, para toda colección $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{F}$ de conjuntos disjuntos dos a dos tal que $\cup_{n \in \mathbb{N}} A_n \in \mathcal{F}$,

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mu(A_n);$$

- **finita**: si $\mu(\Omega) < \infty$;

- σ**-finita**: si existe una colección $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{F}$ tal que, para todo $n \in \mathbb{N}$, $\mu(A_n) < \infty$ y

$$\Omega = \bigcup_{n \in \mathbb{N}} A_n.$$

Podemos ver que si μ es σ-aditiva, también es aditiva. Y si μ es finita, entonces es σ-finita. La propiedad de σ-aditividad es la que nos permitirá definir a las funciones que serán llamadas medidas.

DEFINICIÓN 2.4 (Medida, espacio medido)**.** *Sea $(\Omega, \mathcal{F})$ un espacio medible. Una función $\mu\colon \mathcal{F} \to [0, \infty]$ es una* medida *sobre $(\Omega, \mathcal{F})$ si μ es σ-aditiva. A la tripleta*

$(\Omega, \mathcal{F}, \mu)$ se la llama espacio medido o espacio de medida. Si μ es finita, se dice que el espacio es de medida finita; y si μ es σ-finita, se dice que el espacio es σ-finito.

Si definimos una medida P sobre un espacio medible $(\Omega, \mathcal{F})$, que cumpla que $P(\Omega) = 1$, entonces diremos que el espacio medido $(\Omega, \mathcal{F}, P)$ es un *espacio de probabilidad*. Dado un evento $F \in \mathcal{F}$ y un punto muestral $\omega \in \Omega$, se define la cantidad $P(F)$ como la probabilidad de que $\omega \in F$.

A continuación veremos unas cuantas propiedades de las medidas.

PROPOSICIÓN 2.2. *Sea $(\Omega, \mathcal{F}, \mu)$ un espacio medido. Para todo $A, B \in \mathcal{F}$, se tienen las siguientes propiedades:*

a) $B \subseteq A \implies \mu(B) \leq \mu(A)$,

b) $B \subseteq A \wedge \mu(B) < \infty \implies \mu(A \setminus B) = \mu(A) - \mu(B)$,

c) $\mu(A \cup B) \leq \mu(A) + \mu(B)$,

d) $\mu(\Omega) < \infty \implies \mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$.

Además de las propiedades listadas anteriormente, tenemos los siguientes resultados sobre el comportamiento límite de medidas:

LEMA 2.3 (Continuidad de medidas)**.** *Sea $(\Omega, \mathcal{F}, \mu)$ un espacio medido. Supongamos que $\{F_n\}_{n\in\mathbb{N}} \subseteq \mathcal{F}$.*

a) Si $\{F_n\}_{n\in\mathbb{N}}$ es creciente, entonces

$$\lim_{n\to\infty} \mu(F_n) = \mu\left(\bigcup_{n\in\mathbb{N}} F_n\right).$$

b) Si $\{F_n\}_{n\in\mathbb{N}}$ es decreciente y existe $n_0 \in \mathbb{N}$ tal que $\mu(F_{n_0}) < \infty$, entonces

$$\lim_{n\to\infty} \mu(F_n) = \mu\left(\bigcap_{n\in\mathbb{N}} F_n\right).$$

Varias propiedades de interés, establecidas sobre elementos de un espacio medido, resultan no ser ciertas para todo el espacio. En Teoría de la medida y en Teoría de probabilidades, nos interesa conocer la medida del conjunto de puntos en el cual una propiedad no se cumple, pues si resultara que esta es cero, entonces podríamos decir que dicha propiedad se cumple prácticamente en todo el espacio y podríamos «despreciar» la parte en la que no. La siguiente definición establece la jerga que utilizaremos cuando tratemos con conjuntos de medida nula y las propiedades asociadas a ellos:

DEFINICIÓN 2.5 (Conjunto $\mu-$nulo, propiedad casi en todas partes)**.** *Sea* $(\Omega, \mathcal{F}, \mu)$ *un espacio medido. A un elemento* $F \in \mathcal{F}$ *se lo llama* μ*-nulo si* $\mu(F) = 0$*. Además, diremos que una propiedad* $Q(x)$ *se cumple en* μ*-casi todas partes en* Ω *si el conjunto* $\{x \in \Omega : Q(x)$ *es falsa*$\}$ *es* μ*-nulo.*

Cuando no haya ambigüedad sobre la medida definida en el espacio, hablaremos simplemente de conjuntos nulos o de propiedades que se cumplen en casi todas partes. Además, si el espacio medido es un espacio de probabilidad, en lugar de decir que una propiedad que se cumplen en casi todas partes, diremos que se cumple *casi seguramente*.

La frase *casi seguramente* antes mencionada cobra sentido al ver que, si $Q(x)$ se cumple casi seguramente en $(\Omega, \mathcal{F}, P)$, entonces

$$P(\{x \in \Omega : Q(x) \text{ es falsa}\}) = 0,$$

por lo tanto,

$$P(\{x \in \Omega : Q(x) \text{ es cierta}\}) = 1,$$

es decir que, casi con total certeza, el evento $\{x \in \Omega : Q(x) \text{ es cierta}\}$ ocurrirá.

A partir de este punto, trabajaremos en un espacio de probabilidad $(\Omega, \mathcal{F}, P)$.

Recordemos que para una sucesión de números reales $(x_n)_{n\in\mathbb{N}}$ se definen el límite superior y el límite inferior de la misma de la siguiente manera:

$$\limsup x_n = \inf_{n\in\mathbb{N}} \sup_{m\geq n} x_m \in [-\infty, \infty] \qquad \text{y} \qquad \liminf x_n = \sup_{n\in\mathbb{N}} \inf_{m\geq n} x_m \in [-\infty, \infty],$$

respectivamente. De manera análoga, dada una sucesión de eventos $(F_n)_{n\in\mathbb{N}}$, se define su límite superior y su límite inferior como

$$\limsup F_n := \bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} F_m \qquad \text{y} \qquad \liminf F_n := \bigcup_{n\in\mathbb{N}} \bigcap_{m\geq n} F_m,$$

respectivamente.

Antes de pasar a la siguiente sección, veamos unos cuantos lemas concernientes a sucesiones de eventos que serán muy útiles a lo largo de este trabajo.

LEMA 2.4 (Lema de Fatou para eventos)**.** *Sea* $(F_n)_{n\in\mathbb{N}}$ *una sucesión de eventos. Se* *tiene que*

$$P(\liminf F_n) \leq \liminf P(E_n).$$

LEMA 2.5 (Lema de Fatou inverso para eventos)**.** *Sea $(F_n)_{n\in\mathbb{N}}$ una sucesión de eventos. Se tiene que*

$$P(\limsup F_n) \geq \limsup P(E_n).$$

LEMA 2.6 (Lema de Borel-Cantelli)**.** *Sea $(F_n)_{n\in\mathbb{N}}$ una sucesión de eventos tal que*

$$\sum_{n\in\mathbb{N}} P(F_n) < \infty.$$

Se tiene que

$$P(\limsup F_n) = 0.$$

2.2. La medida de Lebesgue en $((0,1), \mathcal{B}(0,1))$

Hasta ahora hemos presentado definiciones y propiedades concernientes a los espacios medidos y a medidas en general. En esta sección veremos un ejemplo concreto: la medida de Lebesgue. Construiremos esta medida en el conjunto $(0,1)$ dotado de la σ-álgebra de Borel. Para ello, necesitaremos algunos resultados previos, cuyas demostraciones se pueden encontrar en [14].

LEMA 2.7 (Extensión de medidas finitas)**.** *Sean Ω un conjunto, $\mathcal{I}$ un π-sistema de subconjuntos de Ω y $\mathcal{F} = \sigma(\mathcal{I})$. Si μ_1 y μ_2 son medidas sobre $(\Omega, \mathcal{F})$ tales que $\mu_1(\Omega) = \mu_2(\Omega) < \infty$ y $\mu_1 = \mu_2$ en $\mathcal{I}$, entonces $\mu_1 = \mu_2$ en $\mathcal{F}$.*

TEOREMA 2.8 (Teorema de extensión de Carathéodory)**.** *Sean Ω un conjunto, $\mathcal{F}_0$ una álgebra de subconjuntos de Ω y $\mu_0 \colon \mathcal{F}_0 \to [0,\infty]$. Definamos $\mathcal{F} := \sigma(\mathcal{F}_0)$. Si μ_0 es σ-aditiva, entonces existe una medida μ sobre $(\Omega, \mathcal{F})$ tal que $\mu = \mu_0$ en $\mathcal{F}_0$. Si, además, $\mu(\Omega) < \infty$, gracias al Lema 2.7, esta extensión es única.*

Definamos $\Omega_1 := (0,1]$ y $\Omega_2 := (0,1)$. Vamos a construir una medida sobre el espacio medido $(\Omega_1, \mathcal{B}(\Omega_1))$ siguiendo las ideas propuestas en [14] para posteriormente usarla en el subespacio medido $(\Omega_2, \mathcal{B}(\Omega_2))$ que es el que nos interesa.

Diremos que $F \subseteq \Omega$ es un elemento de $\mathcal{I}$ si F se puede escribir como

$$F = (a_1, b_1] \cup (a_1, b_1] \cup \cdots \cup (a_r, b_r],$$

para algún $r \in \mathbb{N}$, donde

$$0 < a_1 \leq b_1 \leq \cdots \leq a_r \leq b_r \leq 1.$$

Tenemos que $\mathcal{I}$ es una álgebra, además,

$$\mathcal{F} := \sigma(\mathcal{I}) = \mathcal{B}((0,1]).$$

Definamos la función

$$\begin{aligned} \mu_0 \colon \mathcal{I} &\longrightarrow [0,\infty] \\ F &\longmapsto \mu_0(F) = \sum_{i=1}^{r}(b_i - a_i). \end{aligned}$$

En [14], podemos ver la demostración de que μ_0 es σ-aditiva, por lo tanto, gracias al Teorema de extensión de Carathéodory, existe una única medida μ en $(\Omega_1, \mathcal{B}(\Omega_1))$ que extiende a μ_0. A esta extensión se la conoce como la medida de Lebesgue en $(\Omega_1, \mathcal{B}(\Omega_1))$ y la denotaremos por λ. Ya que $(\Omega_2, \mathcal{B}(\Omega_2))$ es un subespacio medido de $(\Omega_1, \mathcal{B}(\Omega_1))$, entonces λ es también una medida en $(\Omega_2, \mathcal{B}(\Omega_2))$.

En [12], podemos encontrar una construcción de la medida de Lebesgue (distinta a la propuesta en [14]) que se puede aplicar a $((0,1), \mathcal{B}(0,1))$. De hecho, con el enfoque propuesto para esa construcción, podemos ver que la familia de los subconjuntos de $(0,1)$ que podemos medir con esta medida, denotada por $\mathbf{Leb}(0,1)$, es una es una σ-álgebra y, además es más grande que $\mathcal{B}(0,1)$; a $\mathbf{Leb}(0,1)$ la llamaremos la σ-álgebra de Lebesgue. Si denotamos por $\mathcal{N}$ a la familia de los subconjuntos de $(0,1)$ que tienen medida de Lebesgue nula, entonces tenemos que

$$\mathbf{Leb}(0,1) = \mathcal{B}(0,1) \cup \mathcal{N}.$$

Es decir, que podemos estudiar $\mathbf{Leb}(0,1)$ considerando primero los conjuntos λ-nulos y luego los borelianos de medida positiva. Esta idea será utilizada cuando estudiemos el Teorema de densidad de Lebesgue.

2.3. Integración

En esta sección estudiaremos el concepto de función medible y, a partir de ello, introduciremos el concepto de variable aleatoria, presentando además algunas de sus propiedades. Daremos la definición de integral para funciones medibles, y la de esperanza para variables aleatorias.

DEFINICIÓN 2.6. *Sean $(\Omega_1, \mathcal{F}_1)$ y $(\Omega_2, \mathcal{F}_2)$ dos espacio medibles. Diremos que una función $f\colon \Omega_1 \to \Omega_2$ es $(\mathcal{F}_1, \mathcal{F}_2)$-medible si, para todo $F \in \mathcal{F}_2$, se tiene que $f^{-1}(F) \in \mathcal{F}_1$. Es decir, que la pre imagen de todo conjunto $\mathcal{F}_2$-medible, a través de f, es $\mathcal{F}_1$-medible.*

Este concepto es familiar para nosotros, pues es análogo al de función continua en espacios topológicos. En este trabajo consideraremos $(\Omega_2, \mathcal{F}_2) := (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$, donde $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$, y en lugar de decir que una función es $(\mathcal{F}_1, \mathcal{B}(\overline{\mathbb{R}}))$-medible, simplemente diremos que es $\mathcal{F}_1$-medible.

Si $(\Omega, \mathcal{F})$ es un espacio medible, donde Ω representa un espacio muestral y $\mathcal{F}$ es la colección de eventos de Ω, entonces una función $\mathcal{F}$-medible $f\colon \Omega \to \overline{\mathbb{R}}$ será llamada una *variable aleatoria*.

Veamos algunas propiedades de las funciones medibles:

PROPOSICIÓN 2.9. *Sean $(\Omega, \mathcal{F})$ un espacio medible y $f\colon \Omega \to \overline{\mathbb{R}}$. Tenemos que, si $\mathcal{C} \subseteq \mathcal{B}(\overline{\mathbb{R}})$ es tal que $\sigma(\mathcal{C}) = \mathcal{B}(\overline{\mathbb{R}})$, y para todo $B \in \mathcal{C}$ se tiene que $f^{-1}(B) \in \mathcal{F}$, entonces f es $\mathcal{F}$-medible.*

A partir de esta proposición, podemos obtener inmediatamente el siguiente resultado:

PROPOSICIÓN 2.10. *Sea (Ω, τ) un espacio topológico. Si $f\colon \Omega \to \overline{\mathbb{R}}$ es continua, entonces es f es $\mathcal{B}(\Omega)$-medible.*

Además, tenemos un resultado muy útil al momento de probar que una función es una variable aleatoria:

PROPOSICIÓN 2.11. *Sean $(\Omega, \mathcal{F})$ un espacio medible. Una función $f\colon \Omega \to \mathbb{R}$ es $\mathcal{F}$ medible si, para todo $c \in \mathbb{R}$, se tiene que*

$$\{f \leq c\} := f^{-1}((-\infty, c]) = \{\omega \in \Omega : f(\omega) \leq c\} \in \mathcal{F}.$$

La Proposición 2.10 se deduce de la Proposición 2.9 al tener en cuenta que $\sigma(\tau) = \mathcal{B}(\Omega)$, y la Proposición 2.11, al considerar que $\sigma(\pi(\mathbb{R})) = \mathcal{B}$.

TEOREMA 2.12. *Sea $(\Omega, \mathcal{F})$ un espacio medible. El conjunto de funciones $\mathcal{F}$-medibles es una álgebra sobre $\mathbb{R}$; es decir, que si f_1 y f_2 son funciones $\mathcal{F}$-medibles y $\lambda \in \mathbb{R}$, entonces las funciones λf_1, $f_1 + f_2$ y $f_1 f_2$ son también $\mathcal{F}$-medibles.*

Para complementar este teorema, tenemos la siguiente proposición:

PROPOSICIÓN 2.13. *Sea $(\Omega, \mathcal{F})$ un espacio medible. Si $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles, entonces las funciones $\inf_{n\in\mathbb{N}} f_n$, $\sup_{n\in\mathbb{N}} f_n$, $\liminf f_n$ y $\limsup f_n$ son $\mathcal{F}$-medibles.*

De esta proposición podemos agregar dos cosas: primero, si existe $\lim_{n\to\infty} f_n$, entonces este es $\mathcal{F}$-medible, y segundo, gracias a que toda función constante es medible, entonces, para toda función $\mathcal{F}$-medible f, tenemos que las funciones

$$f^+ := \max\{f, 0\} \qquad \text{y} \qquad f^- := \max\{-f, 0\},$$

llamadas *parte positiva* y *parte negativa* de f, respectivamente, son funciones $\mathcal{F}$-medibles.

A partir de este punto trabajaremos en un espacio medido $(\Omega, \mathcal{F}, \mu)$.

Las funciones candidatas a poseer integral son las funciones $\mathcal{F}$-medibles que van de Ω en $\overline{\mathbb{R}}$. Lastimosamente, no todas ellas se podrán integrar; de hecho, veremos enseguida las condiciones bajo las cuales podremos hablar de la integral de una función.

Sea $A \in \mathcal{F}$. Definimos la integral de la función $\mathbb{1}_A$, con respecto a la medida μ, como

$$\int_\Omega \mathbb{1}_A \, d\mu := \mu(A).$$

DEFINICIÓN 2.7. *Sea $f\colon \Omega \to \overline{\mathbb{R}}$. Se dice que f es una función simple si existe $\{a_i\}_{i=1}^n \subseteq [0, \infty]$, y $\{A_i\}_{i=1}^n \subseteq \mathcal{F}$ una colección de eventos disjuntos dos a dos tal que $\Omega = \bigcup_{i=1}^n A_i$, y*

$$f = \sum_{i=1}^n a_i \mathbb{1}_{A_i},$$

y, se define su integral con respecto a la medida μ, como

$$\int_\Omega f \, d\mu := \sum_{i=1}^n a_i \mu(A_i).$$

La integral de funciones simples cumple las siguientes propiedades:

PROPOSICIÓN 2.14. *Sean f, g funciones simples y $c \geq 0$. Se tiene que*

a) Las funciones $f + g$ y cf son simples. Además

$$\int_\Omega cf + g \, d\mu = c \int_\Omega f \, d\mu + \int_\Omega g \, d\mu.$$

b) Si $f \leq g$, entonces

$$\int_\Omega f\, d\mu \leq \int_\Omega g\, d\mu$$

c) Las funciones

$$f \wedge g := \text{mín}\{f, g\} \qquad y \qquad f \vee g := \text{máx}\{f, g\}$$

son simples

Veamos ahora la definición de integral para funciones medibles más generales:

DEFINICIÓN 2.8. *Sea f una función medible no negativa. Se define la integral de f con respecto a la medida μ, como*

$$\int_\Omega f\, d\mu := \sup\left\{\int_\Omega h\, d\mu : h \text{ es una función simple tal que } h \leq f\right\}.$$

El cálculo de integrales mediante esta definición no siempre es práctico, pero, afortunadamente, se puede simplificar gracias al uso de funciones simples, pues para cualquier función medible no negativa f, existe una sucesión creciente $(f_n)_{n\in\mathbb{N}}$ de funciones simples tal que $f_n \to f$ puntualmente. Para una demostración de este resultado, se puede consultar [12]. Esto, combinado con el siguiente resultado, es de gran utilidad al momento de calcular integrales.

TEOREMA 2.15 (Teorema de convergencia monótona)**.** *Sea $(f_n)_{n\in\mathbb{N}}$ una sucesión creciente de funciones medibles no negativas tal que $f_n \to f$, casi en todas partes. Se tiene que*

$$\int_\Omega f_n\, d\mu \to \int_\Omega f\, d\mu.$$

Este teorema es posiblemente el más importante en la Teoría de la integración, pues, a partir de él, se puede deducir los lemas de Fatou y el Teorema de convergencia dominada de Lebesgue, que veremos más adelante, así como la siguiente proposición:

PROPOSICIÓN 2.16. *Sean f y g dos funciones medibles no negativas. Si $f = g$ casi en todas partes, entonces*

$$\int_\Omega f\, d\mu = \int_\Omega g\, d\mu.$$

TEOREMA 2.17 (Lema de Fatou)**.** *Sea $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles no negativas. Se tiene que*

$$\int_\Omega \liminf f_n\, d\mu \leq \liminf \int_\Omega f_n\, d\mu.$$

TEOREMA 2.18 (Lema de Fatou inverso). *Sea $(f_n)_{n\in\mathbb{N}}$ como en el teorema anterior. Si existe g, una función medible no negativa tal que $f_n \leq g$, casi en todas partes, para todo $n \in \mathbb{N}$ y $\int_\Omega g\,d\mu < \infty$, entonces*

$$\int_\Omega \limsup f_n\,d\mu \geq \limsup \int_\Omega f_n\,d\mu.$$

Como una última propiedad de la integral para funciones medibles no negativas, tenemos la siguiente proposición:

PROPOSICIÓN 2.19. *Sean f y g dos funciones medibles no negativas. Sean $a, b \in [0,\infty]$. Se tiene que*

$$\int_\Omega af + bg\,d\mu = a\int_\Omega f\,d\mu + b\int_\Omega g\,d\mu$$

Nuevamente, esta propiedad se puede deducir a partir del Teorema 2.15.

Consideremos ahora una función medible cualquiera f. Podemos descomponer f de modo que se la pueda ver como diferencia de dos funciones medibles no negativas de la siguiente manera:

$$f = f^+ - f^-,$$

donde f^+ y f^- son la parte positiva y la parte negativa de f, respectivamente. Así mismo, se tiene que

$$|f| = f^+ + f^-.$$

Con esta propiedad en mente, podemos pasar a la siguiente definición:

DEFINICIÓN 2.9. *Sea f una función medible tal que*

$$\min\left\{\int_\Omega f^+\,d\mu, \int_\Omega f^-\,d\mu\right\} < \infty.$$

Se define la integral de f con respecto a la medida μ, como

$$\int_\Omega f\,d\mu := \int_\Omega f^+\,d\mu - \int_\Omega f^-\,d\mu.$$

Además, para todo $F \in \mathcal{F}$ se define la integral de f sobre el conjunto F, con respecto a la medida μ, como

$$\int_F f\,d\mu = \int_\Omega f \cdot \mathbb{1}_F\,d\mu.$$

Otras notaciones equivalentes para la integral de una función f, que se usarán a lo largo de este trabajo son

$$\int_\Omega f\,d\mu = \int_\Omega f(x)\,d\mu = \int_\Omega f(x)\,d\mu(x).$$

DEFINICIÓN 2.10 (Función integrable, espacio de funciones integrables)**.** *Sea f una función medible. Si*

$$\int_\Omega |f|\,d\mu = \int_\Omega f^+\,d\mu + \int_\Omega f^-\,d\mu < \infty,$$

diremos que f es una función μ-integrable y, a la clase de todas las funciones integrables, la denotaremos por $\mathcal{L}^1(\Omega, \mathcal{F}, \mu)$. Si no hay riesgo de confusión en cuanto a la medida definida en el espacio, hablaremos simplemente de funciones integrables.

Notemos que si $f \in \mathcal{L}^1(\Omega, \mathcal{F}, \mu)$, entonces $\int_\Omega f\,d\mu < \infty$, pues

$$\left|\int_\Omega f\,d\mu\right| \leq \int_\Omega |f|\,d\mu.$$

De manera general, definimos, para $p \in [1, \infty)$, la clase $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ de funciones p-integrables, como la clase de funciones medibles f que verifican

$$\int_\Omega |f|^p\,d\mu < \infty.$$

Denotaremos por $(\mathcal{L}^p(\Omega, \mathcal{F}, \mu))^+$ a la clase de funciones p-integrables con valores en $[0, \infty]$.

Como último resultado de esta sección, presentamos el Teorema de convergencia dominada de Lebesgue:

TEOREMA 2.20. *Sea $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles convergente, casi en todas partes, a una función f. Si existe $g \in (\mathcal{L}^1(\Omega, \mathcal{F}, \mu))^+$ tal que, para todo $n \in \mathbb{N}$,*

$$|f_n| \leq g,$$

casi en todas partes, entonces

$$\int_\Omega |f_n - f|\,d\mu \to 0;$$

y, así,

$$\int_\Omega f_n\,d\mu \to \int_\Omega f\,d\mu.$$

2.4. Esperanza

En esta sección presentaremos y exploraremos algunas propiedades de la esperanza de una variable aleatoria, por lo tanto pasaremos a trabajar sobre un espacio de probabilidad $(\Omega, \mathcal{F}, P)$ y, a diferencia de la sección anterior, ya no hablaremos de funciones medibles, sino de variables aleatorias.

DEFINICIÓN 2.11. *Sea X una variable aleatoria. Se define la esperanza o valor esperado de X como la integral de X con respecto a la medida de probabilidad P, siempre que esta exista, y se la denota por*

$$\boldsymbol{E}(X) := \int_\Omega X\, dP.$$

Ya que una variable aleatoria es una función medible, todos los resultados presentados en la sección anterior siguen siendo válidos. Presentaremos algunos de ellos, pero ahora con la notación que acabamos de introducir en la definición anterior, además aprovecharemos la finitud de la medida P para presentar un nuevo resultado.

TEOREMA 2.21. *Sea $(X_n)_{n\in\mathbb{N}}$ una sucesión de variables aleatorias. Se tienen los siguientes resultados:*

- ***Convergencia monótona:*** *Si $(X_n)_{n\in\mathbb{N}}$ es creciente y converge casi seguramente a una variable aleatoria X, entonces $\boldsymbol{E}(X_n) \to \boldsymbol{E}(X)$.*
- ***Lema de Fatou:*** *Si $X_n \geq 0$, para todo $n \in \mathbb{N}$, entonces*

$$\boldsymbol{E}(\liminf X_n) \leq \liminf \boldsymbol{E}(X_n).$$

- ***Convergencia dominada:*** *Si $(X_n)_{n\in\mathbb{N}}$ converge casi seguramente a una variable aleatoria X y existe $Y \in (\mathcal{L}^1(\Omega, \mathcal{F}, P))^+$ tal que, para todo $n \in \mathbb{N}$,*

$$|X_n| \leq Y$$

casi seguramente, entonces

$$\boldsymbol{E}(|X_n - X|) \to 0;$$

y así

$$\boldsymbol{E}(X_n) \to \boldsymbol{E}(X).$$

- ***Convergencia acotada:*** *Si* $(X_n)_{n\in\mathbb{N}}$ *converge casi seguramente a una variable aleatoria* X *y existe una constante* $K \in \mathbb{R}$ *tal que, para todo* $n \in \mathbb{N}$,

$$|X_n| \leq K$$

casi seguramente, entonces

$$\boldsymbol{E}(|X_n - X|) \to 0;$$

y, así,

$$\boldsymbol{E}(X_n) \to \boldsymbol{E}(X).$$

Es claro que el Teorema de convergencia acotada se obtiene del Teorema de convergencia dominada al tomar $Y = K$, puesto que $\mathbf{E}(K) = K < \infty$.

Antes de avanzar, veamos una desigualdad que será útil en capítulos posteriores:

TEOREMA 2.22 (Desigualdad de Jensen)**.** *Sean* $I \subseteq \mathbb{R}$ *un intervalo abierto y* $c\colon I \to \mathbb{R}$ *una función convexa. Sea* $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ *tal que*

$$P(X^{-1}(I)) = 1 \qquad y \qquad \boldsymbol{E}(|c(X)|) < \infty.$$

Se tiene que

$$\boldsymbol{E}(c(X)) \geq c(\boldsymbol{E}(X)).$$

Habiendo estudiado la integral en la sección anterior, no queda mucho que decir de la esperanza de una variable aleatoria; sin embargo, aprovecharemos esta sección para estudiar un poco más las clases $\mathcal{L}^p(\Omega, \mathcal{F}, P)$. Los resultados que veremos hasta el final de esta sección se pueden revisar a profundidad en [4].

Sea $p \in [1, \infty)$. El conjunto $\mathcal{L}^p(\Omega, \mathcal{F}, P)$, dotado de la suma y producto de funciones usuales es un espacio vectorial sobre $\mathbb{R}$. Si definimos, en $\mathcal{L}^p(\Omega, \mathcal{F}, P)$, la función

$$\begin{aligned} \|\cdot\|_p\colon\ \mathcal{L}^p(\Omega, \mathcal{F}, P) &\longrightarrow [0, \infty] \\ X &\longmapsto \left(\mathbf{E}\left(|X|^p\right)\right)^{\frac{1}{p}}, \end{aligned} \tag{2.4}$$

tenemos que $\|0\|_p = 0$; sin embargo, si $\mathbf{E}(|X|^p) = 0$, no tenemos la certeza de que X sea la variable aleatoria idénticamente nula, pues si $X = 0$, salvo en un conjunto de probabilidad nula, entonces $\mathbf{E}(|X|^p) = 0$. Es decir, que la aplicación definida en (2.4) no es una norma en $\mathcal{L}^p(\Omega, \mathcal{F}, P)$. La manera con la que los analistas lidian con este problema es cocientando el espacio $\mathcal{L}^p(\Omega, \mathcal{F}, P)$ bajo la siguiente relación de equivalencia: para todo $X, Y \in \mathcal{L}^p(\Omega, \mathcal{F}, P)$,

$$X \sim Y \iff X = Y \text{ casi seguramente.}$$

El espacio resultante se suele denotar por $L^p(\Omega, \mathcal{F}, P)$ y, ahora, ya podemos decir que (2.4) define una norma en el mismo. De hecho, el espacio $(L^p(\Omega, \mathcal{F}, P), \|\cdot\|_p)$ es completo, y para $p = 2$, es un espacio de Hilbert, dotado del producto interno definido por

$$(X, Y) \mapsto \langle X, Y \rangle := \mathbf{E}(XY).$$

Vimos que (2.4) no define una norma en $\mathcal{L}^p(\Omega, \mathcal{F}, P)$, sin embargo, gracias a que, en $L^p(\Omega, \mathcal{F}, P)$,

$$\mathbf{E}(|X|^p) = 0 \iff X = 0 \text{ casi seguramente,}$$

podemos "trasladar" varias propiedades del espacio $L^p(\Omega, \mathcal{F}, P)$ a $\mathcal{L}^p(\Omega, \mathcal{F}, P)$; entre ellas, la de completitud. Abusando del lenguaje, diremos que el espacio $(\mathcal{L}^p(\Omega, \mathcal{F}, P), \|\cdot\|_p)$ es un espacio normado completo y que $(\mathcal{L}^2(\Omega, \mathcal{F}, P), \langle\cdot,\cdot\rangle)$ es un espacio de Hilbert.

Con esto en mente, podemos empezar a revisar ciertas propiedades de los espacios $\mathcal{L}^p(\Omega, \mathcal{F}, P)$, empezando por la relación de contenencia que tienen con respecto al índice p.

Proposición 2.23. *Sean $1 \leq p \leq q < \infty$. Se tiene que $\mathcal{L}^q(\Omega, \mathcal{F}, P) \subseteq \mathcal{L}^p(\Omega, \mathcal{F}, P)$. Así, para todo $X \in \mathcal{L}^q(\Omega, \mathcal{F}, P)$, tenemos que*

$$\|X\|_p \leq \|X\|_q.$$

Como ya dijimos, $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ será considerado como un espacio de Hilbert, por lo tanto, tenemos la siguiente desigualdad:

Proposición 2.24 (Desigualdad de Cauchy-Schwarz)**.** *Sean $X, Y \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$. Se tiene que $XY \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ y*

$$|\boldsymbol{E}(XY)| \leq \boldsymbol{E}(|XY|) \leq \|X\|_2 \|Y\|_2$$

Por último, revisemos el que quizás sea el teorema más importante de esta sección gracias a la relevancia que cobrará en el siguiente capítulo.

Teorema 2.25. *Sea $\mathcal{K}$ un subespacio vectorial completo de $\mathcal{L}^2(\Omega, \mathcal{F}, P)$. Si $X \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$, entonces existe $Y \in \mathcal{K}$ tal que* *a)* $\|X - Y\| = \text{ínf}\{\|X - W\| : W \in \mathcal{K}\}$*, y*

b) $\langle X - Y, Z \rangle = 0$*, para todo* $Z \in \mathcal{K}$*.*

A la variable Y *se la conoce como la proyección ortogonal de* X *sobre* $\mathcal{K}$*. Si* $\tilde{Y} \in \mathcal{K}$ *cumple con las mismas propiedades que* Y*, entonces* $Y = \tilde{Y}$*, casi seguramente.*

Capítulo 3

Esperanza condicional

En este capítulo revisaremos y generalizaremos el concepto de probabilidad condicional elemental al relacionarlo con la esperanza condicional. Además daremos una demostración de la existencia de la esperanza condicional sin hacer uso del Teorema de Radon-Nikodym, aprovechando la completitud de los espacios $\mathcal{L}^p$ en el sentido dado en el capítulo anterior.

3.1. Un caso simple

Consideremos un espacio de probabilidad $(\Omega, \mathcal{F}, P)$ y fijemos un evento $B \in \mathcal{F}$ tal que $P(B) > 0$. La definición de probabilidad condicional elemental nos dice que la probabilidad de que ocurra cualquier evento $A \in \mathcal{F}$, dado que antes ha ocurrido el evento B está dada por

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{3.1}$$

Intuitivamente, si pensamos en P como una medida de área, entonces la ecuación (3.1) representaría la cantidad de área de B que es ocupada por A.

El hecho de que haya ocurrido el evento B, nos da información adicional sobre el experimento aleatorio modelado por $(\Omega, \mathcal{F}, P)$. Es decir, que ahora las probabilidades de ocurrencia de los demás eventos se han visto modificadas debido a la ocurrencia de B. De hecho, si nos fijamos en (3.1), ya que $P(B)$ es una cantidad conocida por nosotros, lo que nos interesa calcular ahora es $P(A \cap B)$; es decir, necesitamos medir únicamente eventos de la forma $A \cap B$, con $A \in \mathcal{F}$. Esto nos lleva a hacer una modificación en el modelo para nuestro experimento aleatorio, pasando de $(\Omega, \mathcal{F}, P)$ a $(B, \mathcal{F}|_B, \tilde{P})$, donde

$$\mathcal{F}|_B = \{A \cap B : A \in \mathcal{F}\}$$

y

$$\begin{aligned} \tilde{P}\colon \ \mathcal{F}|_B &\longrightarrow [0,1] \\ A \cap B &\longmapsto \tilde{P}(A \cap B) = \frac{P(A \cap B)}{P(B)}. \end{aligned} \tag{3.2}$$

Esta idea se ajusta bien al concepto de probabilidad condicional más básico; sin embargo, al momento de generalizar este concepto, veremos que es más útil considerar la siguiente medida de probabilidad sobre $(\Omega, \mathcal{F})$:

$$\begin{aligned} P_B\colon \ \mathcal{F} &\longrightarrow [0,1] \\ A &\longmapsto P_B(A) = \frac{P(A \cap B)}{P(B)}. \end{aligned} \tag{3.3}$$

Así como la probabilidad de ocurrencia de un evento A se puede ver modificada por la ocurrencia de otro evento B, lo mismo puede ocurrir con el valor esperado de una variable aleatoria. Esto motiva la siguiente definición:

DEFINICIÓN 3.1 (Esperanza condicional elemental)**.** *Sean X una variable aleatoria integrable y $B \in \mathcal{F}$ tal que $P(B) > 0$. Se define la esperanza condicional de X dado el evento B, como la esperanza de la variable aleatoria X en el espacio $(\Omega, \mathcal{F}, P_B)$ y se denota por*

$$\boldsymbol{E}(X \mid B) = \boldsymbol{E}_B(X) := \int_\Omega X \, dP_B.$$

Podemos ver que, para todo evento $A \in \mathcal{F}$, la esperanza condicional de $\mathbb{1}_A$ está dada por

$$\mathbf{E}(\mathbb{1}_A \mid B) = \int_\Omega \mathbb{1}_A \, dP_B = P_B(A) = \frac{P(A \cap B)}{P(B)} = P(A|B); \tag{3.4}$$

es decir, que la probabilidad condicional elemental es un caso particular de la esperanza condicional elemental.

El siguiente lema nos será de utilidad al momento de calcular esperanzas condicionales:

LEMA 3.1. *Sea X una variable aleatoria tal que $\boldsymbol{E}(X)$ existe. Se tiene que la esperanza de X en el espacio $(\Omega, \mathcal{F}, P_B)$, denotada por $\boldsymbol{E}_B(X)$, existe y está dada por*

$$\boldsymbol{E}_B(X) = \frac{1}{P(B)} \int_B X \, dP = \frac{\boldsymbol{E}(X \mathbb{1}_B)}{P(B)}.$$

Demostración. En esta demostración procederemos de la siguiente manera: primero, probaremos que el resultado es cierto para funciones indicatrices, luego probaremos que también es cierto para funciones simples y, haciendo uso del Teorema de convergencia monótona, extenderemos este resultado a funciones medibles no negativas. Finalmente, probaremos el teorema para funciones medibles en general.

- Si X es una función indicatriz; es decir, existe $A \in \mathcal{F}$ tal que $X = \mathbb{1}_A$, entonces

$$\begin{aligned}
\mathbf{E}_B(X) &= \int_\Omega X \, dP_B \\
&= \int_\Omega \mathbb{1}_A \, dP_B \\
&= P_B(A) \\
&= \frac{P(A \cap B)}{P(B)} \\
&= \frac{1}{P(B)} \int_\Omega \mathbb{1}_{A \cap B} \, dP \\
&= \frac{1}{P(B)} \int_\Omega \mathbb{1}_A \mathbb{1}_B \, dP \\
&= \frac{\mathbf{E}(X \mathbb{1}_B)}{P(B)}. \qquad (3.5)
\end{aligned}$$

- Si X es una función simple; es decir, existe una colección finita $\{A_i\}_{i=1}^n \subseteq \mathcal{F}$ de eventos disjuntos dos a dos y $\{a_i\}_{i=1}^n \subseteq [0, \infty]$ tal que

$$X = \sum_{i=1}^n a_i \mathbb{1}_{A_i},$$

entonces

$$\begin{aligned}
\mathbf{E}_B(X) &= \int_\Omega X \, dP_B \\
&= \int_\Omega \sum_{i=1}^n a_i \mathbb{1}_{A_i} \, dP_B \\
&= \sum_{i=1}^n a_i \int_\Omega \mathbb{1}_{A_i} \, dP_B \\
&= \sum_{i=1}^n a_i \mathbf{E}_B \left(\mathbb{1}_{A_i} \right) \, dP_B,
\end{aligned}$$

y, gracias a (3.5), se sigue que

$$\mathbf{E}_B(X) = \sum_{i=1}^n a_i \frac{\mathbf{E}\left(\mathbb{1}_{A_i} \mathbb{1}_B \right)}{P(B)} \, dP_B$$

$$= \frac{1}{P(B)}\mathbf{E}\left(\sum_{i=1}^{n} a_i \mathbb{1}_{A_i}\mathbb{1}_B\right)$$
$$= \frac{\mathbf{E}\left(X\mathbb{1}_B\right)}{P(B)}. \tag{3.6}$$

- Si X es una variable aleatoria no negativa, entonces existe una sucesión creciente $(X_n)_{n\in\mathbb{N}}$ de variables aleatorias simples tal que

$$X_n \to X.$$

Gracias al Teorema de convergencia monótona, tenemos que

$$\mathbf{E}\left(X_n\right) \to \mathbf{E}\left(X\right) \tag{3.7}$$

y

$$\mathbf{E}_B\left(X_n\right) \to \mathbf{E}_B\left(X\right). \tag{3.8}$$

De (3.6) y (3.7), se tiene que

$$\mathbf{E}_B\left(X_n\right) = \frac{\mathbf{E}\left(X_n\mathbb{1}_B\right)}{P(B)} \to \frac{\mathbf{E}\left(X\mathbb{1}_B\right)}{P(B)},$$

y, por lo tanto, gracias a (3.8),

$$\mathbf{E}_B\left(X\right) = \frac{\mathbf{E}\left(X\mathbb{1}_B\right)}{P(B)}. \tag{3.9}$$

- Por último, supongamos que X es una variable aleatoria cualquiera. Tenemos que $X = X^+ - X^-$, donde X^+ y X^- son no negativas, por lo tanto, gracias a (3.9), se tiene que

$$\mathbf{E}_B\left(X^+\right) = \frac{\mathbf{E}\left(X^+\mathbb{1}_B\right)}{P(B)} \qquad \text{y} \qquad \mathbf{E}_B\left(X^-\right) = \frac{\mathbf{E}\left(X^-\mathbb{1}_B\right)}{P(B)},$$

de donde

$$\begin{aligned}\mathbf{E}_B\left(X\right) &= \frac{\mathbf{E}\left(X^+\mathbb{1}_B\right)}{P(B)} - \frac{\mathbf{E}\left(X^-\mathbb{1}_B\right)}{P(B)}\\ &= \frac{1}{P(B)}\mathbf{E}\left(X^+\mathbb{1}_B - X^-\mathbb{1}_B\right)\\ &= \frac{\mathbf{E}\left(X\mathbb{1}_B\right)}{P(B)},\end{aligned}$$

como buscábamos. □

La esperanza condicional elemental de una variable aleatoria integrable X está condicionada a la ocurrencia de un evento. Es decir, que si el resultado $\omega \in \Omega$ de un experimento aleatorio es un elemento de B, entonces podemos calcular $\mathbf{E}(X \mid B)$, caso contrario, lo que podríamos calcular sería $\mathbf{E}(X \mid B^c)$. Esto nos indica que el valor esperado de X puede cambiar dependiendo del conjunto al que pertenezca ω, es decir que, en realidad, la esperanza condicional es una variable aleatoria. Por otra parte, los objetos matemáticos que nos permite almacenar la información sobre un experimento aleatorio son las σ-álgebras, que en este caso, la σ-álgebra que almacene la información sobre nuestro experimento sería $\mathcal{G} = \{\varnothing, B, B^c, \Omega\}$, por lo tanto, ya no condicionaremos el valor esperado para X con respecto al evento B, sino con respecto a la σ-álgebra $\mathcal{G}$, y, de hecho, lo definiremos por

$$\mathbf{E}(X \mid \mathcal{G}) = \mathbf{E}(X \mid B)\, \mathbb{1}_B + \mathbf{E}(X \mid B^c)\, \mathbb{1}_{B^c}. \tag{3.10}$$

Notemos que

$$\begin{aligned}
\mathbf{E}(\mathbf{E}(X \mid \mathcal{G})\, \mathbb{1}_B) &= \int_\Omega \mathbf{E}(X \mid \mathcal{G})\, \mathbb{1}_B\, dP \\
&= \int_B \mathbf{E}(X \mid \mathcal{G})\, dP \\
&= \int_B (\mathbf{E}(X \mid B)\, \mathbb{1}_B + \mathbf{E}(X \mid B^c)\, \mathbb{1}_{B^c})\, dP \\
&= \int_B \mathbf{E}(X \mid B)\, dP \\
&= \int_B \frac{\mathbf{E}(X\mathbb{1}_B)}{P(B)}\, dP \\
&= \mathbf{E}(X\mathbb{1}_B)\,;
\end{aligned}$$

de manera similar, podemos probar que

$$\mathbf{E}(\mathbf{E}(X \mid \mathcal{G})\, \mathbb{1}_{B^c}) = \mathbf{E}(X\mathbb{1}_{B^c})$$

y, por lo tanto, también

$$\mathbf{E}(\mathbf{E}(X \mid \mathcal{G})) = E(X).$$

Así, hemos probado que, para todo $G \in \mathcal{G}$, se tiene que

$$\mathbf{E}(\mathbf{E}(X \mid \mathcal{G})\, \mathbb{1}_G) = \mathbf{E}(X\mathbb{1}_G).$$

A esta propiedad se la suele llamar *propiedad fundamental* de la esperanza condicional. Además, probamos también que $\mathbf{E}(X \mid \mathcal{G})$ es integrable. Veremos más adelante que estas propiedades se cumplen también en el caso más general.

3.2. El caso general

Ya hemos construido una variable aleatoria, a la que la hemos llamado esperanza condicional, pero en un caso muy simple. La σ-álgebra con respecto a la cual se puede condicionar la esperanza de una variable aleatoria integrable puede ser arbitraria y esto hace que la esperanza condicional no siempre sea fácil de calcular; sin embargo, el siguiente teorema nos garantiza que esta siempre existe:

TEOREMA 3.2. *Sea $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$. Si $\mathcal{G}$ es una sub σ-álgebra de $\mathcal{F}$, entonces existe Y, una variable aleatoria $\mathcal{G}$-medible, tal que*

$$\boldsymbol{E}(Y\mathbb{1}_G) = \boldsymbol{E}(X\mathbb{1}_G),$$

para todo $G \in \mathcal{G}$; o, de manera equivalente,

$$\int_G Y\,dP = \int_G X\,dP,$$

para todo $G \in \mathcal{G}$. La variable aleatoria Y es única casi seguramente.

Este teorema nos permite dar la siguiente definición:

DEFINICIÓN 3.2. *Sea $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$. Para toda sub σ-álgebra $\mathcal{G}$ de $\mathcal{F}$ se define la esperanza condicional de X dada la σ-álgebra $\mathcal{G}$, denotada por $\boldsymbol{E}(X \mid \mathcal{G})$, como una variable aleatoria $\mathcal{G}$-medible tal que, para todo $G \in \mathcal{G}$,*

$$\int_G \boldsymbol{E}(X \mid \mathcal{G})\;dP = \int_G X\,dP.$$

Gracias al Teorema 3.2, podemos garantizar que la esperanza condicional de una variable aleatoria es única casi seguramente.

Antes de pasar a la demostración del Teorema 3.2, probaremos una versión más débil del mismo en la que aprovecharemos el hecho de que $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ es un espacio de Hilbert.

LEMA 3.3. *Sea $X \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$. Si $\mathcal{G}$ es una sub σ-álgebra de $\mathcal{F}$, entonces existe Y, una versión de la esperanza condicional de X.*

Demostración. Tenemos que $\mathcal{L}^2(\Omega, \mathcal{G}, P) \subseteq \mathcal{L}^2(\Omega, \mathcal{F}, P)$ es un espacio de Hilbert y, por lo tanto, completo. Gracias al Teorema 2.25, existe $Y \in \mathcal{L}^2(\Omega, \mathcal{G}, P)$ tal que

$$\mathbf{E}\left((X-Y)^2\right) = \text{ínf}\left\{\mathbf{E}\left((X-W)^2\right) : W \in \mathcal{L}^2(\Omega, \mathcal{G}, P)\right\} \tag{3.11}$$

y

$$\mathbf{E}\left((X - Y)Z\right) = 0, \tag{3.12}$$

para todo $Z \in \mathcal{L}^2(\Omega, \mathcal{G}, P)$; en particular, para todo $G \in \mathcal{G}$, tenemos que

$$\mathbf{E}\left((X - Y)\mathbb{1}_G\right) = 0$$

y, por lo tanto,

$$\mathbf{E}\left(X\mathbb{1}_G\right) = \mathbf{E}\left(Y\mathbb{1}_G\right).$$

Así, $Y = \mathbf{E}\left(X \mid G\right)$, casi seguramente. $\square$

Un par de resultados útiles serán probados previo a la demostración del Teorema 3.2.

Lema 3.4. *Sean $X \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ y $\mathcal{G}$ una sub σ-álgebra de $\mathcal{F}$. Si X es no negativa, entonces $\mathbf{E}\left(X \mid \mathcal{G}\right)$ es no negativa, casi seguramente.*

Demostración. Sea $Y = \mathbf{E}\left(X \mid G\right)$ casi seguramente. Supongamos que $P(Y < 0) > 0$. Definamos la sucesión de eventos $(A_n)_{n\in\mathbb{N}^*}$ de modo que, para cada $n \in \mathbb{N}^*$,

$$A_n = \left\{Y < -\frac{1}{n}\right\}.$$

Podemos ver que $(A_n)_{n\in\mathbb{N}^*}$ es creciente y

$$\bigcup_{n\in\mathbb{N}^*} A_n = \{Y < 0\}.$$

Notemos que existe $n_0 \in \mathbb{N}^*$ tal que $P(A_{n_0}) > 0$, pues de no ser así, tendríamos que $P(Y < 0) = 0$. Además, gracias a que $Y = \mathbf{E}\left(X \mid G\right)$ casi seguramente, tenemos que Y es $\mathcal{G}$-medible, por lo tanto $A_{n_0} \in \mathcal{G}$. Así,

$$\int_{A_{n_0}} X\,dP = \int_{A_{n_0}} Y\,dP < \int_{A_{n_0}} -\frac{1}{n_0}\,dP = -\frac{1}{n_0}P(A_{n_0}) < 0.$$

Por otra parte, gracias a la no negatividad de X, tenemos que

$$0 \leq \int_{A_{n_0}} X\,dP,$$

lo que junto a la última desigualdad, nos lleva a una contradicción. Así, tenemos que $\mathbf{E}\left(X \mid \mathcal{G}\right) \geq 0$ casi seguramente. $\square$

Lema 3.5. *Sean $X, Y \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ y $\mathcal{G}$ una sub σ-álgebra de $\mathcal{F}$. Si X e Y son no negativas y $Y \leq X$, entonces $\mathbf{E}\left(Y \mid \mathcal{G}\right) \leq \mathbf{E}\left(X \mid \mathcal{G}\right)$, casi seguramente.*

Demostración. Tenemos que $X - Y \geq 0$, por lo tanto, gracias al lema anterior,

$$\mathbf{E}\left(X - Y \mid \mathcal{G}\right) \geq 0. \tag{3.13}$$

Notemos que $\mathbf{E}\left(X \mid \mathcal{G}\right) - \mathbf{E}\left(Y \mid \mathcal{G}\right)$ es $\mathcal{G}$-medible, y, para todo $G \in \mathcal{G}$,

$$\begin{aligned}\int_G \left(\mathbf{E}\left(X \mid \mathcal{G}\right) - \mathbf{E}\left(Y \mid \mathcal{G}\right)\right) dP &= \int_G \mathbf{E}\left(X \mid \mathcal{G}\right) \, dP - \int_G \mathbf{E}\left(Y \mid \mathcal{G}\right) \, dP \\ &= \int_G X \, dP - \int_G Y \, dP \\ &= \int_G (X - Y) \, dP,\end{aligned}$$

por lo tanto

$$\mathbf{E}\left(X - Y \mid \mathcal{G}\right) = \mathbf{E}\left(X \mid \mathcal{G}\right) - \mathbf{E}\left(Y \mid \mathcal{G}\right),$$

casi seguramente. Así, gracias a (3.13), se sigue el resultado. □

Con estos resultados, estamos listos para probar el Teorema 3.2.

Demostración del Teorema 3.2. En primer lugar, probaremos la unicidad casi segura de Y. Para ello, supongamos que Y y Y' son dos versiones de $\mathbf{E}\left(X \mid \mathcal{G}\right)$, distintas, casi seguramente. Por una parte, tenemos que

$$\int_G Y \, dP = \int_G X \, dP = \int_G Y' \, dP, \tag{3.14}$$

para todo $G \in \mathcal{G}$; y, por otra, que

$$P(Y' > Y) + P(Y > Y') = P(Y \neq Y') > 0.$$

Sin pérdida de generalidad, podemos suponer que $P(Y > Y') > 0$. Definamos la sucesión de eventos $(A_n)_{n \in \mathbb{N}^*}$ de modo que, para cada $n \in \mathbb{N}^*$,

$$A_n = \left\{Y > Y' + \frac{1}{n}\right\} = \left\{Y - Y' > \frac{1}{n}\right\}.$$

Podemos ver que $(A_n)_{n \in \mathbb{N}^*}$ es creciente y

$$\bigcup_{n \in \mathbb{N}^*} A_n = \{Y > Y'\}.$$

También, tenemos que, para todo $n \in \mathbb{N}^*$, $A_n \in \mathcal{G}$, pues tanto Y como Y' son $\mathcal{G}$-medibles. Además, existe $n_0 \in \mathbb{N}^*$ tal que $P(A_{n_0}) > 0$, pues de no ser así, tendríamos que $P(Y > Y') = 0$. Así, tenemos que

$$\int_{A_{n_0}} Y - Y' \, dP > \int_{A_{n_0}} \frac{1}{n_0} \, dP = \frac{1}{n_0} P(A_{n_0}) > 0,$$

lo cual es absurdo, pues de (3.14) se sigue que

$$\int_{A_{n_0}} Y - Y' \, dP = 0.$$

Luego, $Y = Y'$, casi seguramente.

Probemos ahora que existe una versión de la esperanza condicional de X dada la σ-álgebra $\mathcal{G}$. Notemos que X^+ es no negativa. Sabemos que existe una sucesión creciente $(X_n^+)_{n \in \mathbb{N}}$ de variables aleatorias simples tal que $X_n^+ \to X^+$ y además, para todo $n \in \mathbb{N}$, se tiene que $X_n^+ \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$, por lo tanto, gracias al Lema 3.3, existe $Y_n^+ = \mathbf{E}(X_n^+ \mid \mathcal{G})$. Del Lema 3.4, tenemos que, para todo $n \in \mathbb{N}$, $Y_n^+ \geq 0$, casi seguramente, y, del Lema 3.5, se sigue que la sucesión $(Y_n^+)_{n \in \mathbb{N}}$ es creciente. Definamos

$$Y^+ := \lim_{n \to \infty} Y_n^+.$$

Tenemos, gracias al Teorema de convergencia monótona, que, para todo $G \in \mathcal{G}$,

$$\mathbf{E}\left(X_n^+ \mathbb{1}_G\right) \to \mathbf{E}\left(X^+ \mathbb{1}_G\right) \qquad \text{y} \qquad \mathbf{E}\left(Y_n^+ \mathbb{1}_G\right) \to \mathbf{E}\left(Y^+ \mathbb{1}_G\right),$$

pero, ya que $Y_n^+ = \mathbf{E}(X_n^+ \mid \mathcal{G})$, se tiene que

$$\mathbf{E}\left(X_n^+ \mathbb{1}_G\right) \to \mathbf{E}\left(Y^+ \mathbb{1}_G\right),$$

así

$$\mathbf{E}\left(Y^+ \mathbb{1}_G\right) = \mathbf{E}\left(X^+ \mathbb{1}_G\right),$$

casi seguramente. Por lo tanto, gracias a que Y^+ es $\mathcal{G}$-medible, se sigue que $Y^+ = \mathbf{E}(X^+ \mid \mathcal{G})$, casi seguramente. De la misma manera, podemos probar que existe $Y^- = \mathbf{E}(X^- \mid \mathcal{G})$.

Ahora, notemos que

$$\begin{aligned}
\int_G (\mathbf{E}(X^+ \mid \mathcal{G}) - \mathbf{E}(X^- \mid \mathcal{G})) \, dP &= \int_G \mathbf{E}\left(X^+ \mid \mathcal{G}\right) \, dP - \int_G \mathbf{E}\left(X^- \mid \mathcal{G}\right) \, dP \\
&= \int_G X^+ \, dP - \int_G X^- \, dP \\
&= \int_G (X^+ - X^-) \, dP \\
&= \int_G X \, dP,
\end{aligned}$$

por lo tanto,

$$\mathbf{E}(X \mid \mathcal{G}) = \mathbf{E}\left(X^+ \mid \mathcal{G}\right) - \mathbf{E}\left(X^- \mid \mathcal{G}\right),$$

casi seguramente. $\square$

Este teorema nos garantiza la existencia de la esperanza condicional, dada cualquier sub σ-álgebra de $\mathcal{F}$. De hecho, podemos ver que la función que construimos en la sección anterior y a la que llamamos esperanza condicional, en efecto cumple con las propiedades que el Teorema 3.2 nos dice que debe cumplir. Usualmente, la σ-álgebra con respecto a la cual condicionamos el valor esperado de una variable aleatoria X es la σ-álgebra generada por otra variable aleatoria Y, la cual se define como

$$\sigma(Y) = \{Y^{-1}(B) : B \in \mathcal{B}\}.$$

En ese caso, denotaremos

$$\mathbf{E}(X \mid \sigma(Y)) := \mathbf{E}(X \mid Y).$$

Como dijimos al inicio de esta sección, no siempre es posible calcular explícitamente la esperanza condicional de una variable aleatoria. Además del mostrado en la sección anterior, veremos un caso en el cual es posible hacerlo. De hecho, el siguiente ejemplo no es más que una generalización del ejemplo visto en la sección anterior.

EJEMPLO 2. Supongamos que $\{F_n\}_{n\in\mathbb{N}} \subseteq \mathcal{F}$ es una partición de Ω tal que, para todo $n \in \mathbb{N}$, $P(F_n) > 0$. Si definimos $\mathcal{G} = \sigma(\{F_n\}_{n\in\mathbb{N}})$, entonces, para todo $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$, se tiene que

$$\mathbf{E}(X \mid \mathcal{G}) = \sum_{n\in\mathbb{N}} \mathbf{E}(X \mid F_n) \mathbb{1}_{F_n}.$$

En efecto, podemos ver que el lado derecho de la igualdad es una función $\mathcal{G}$-medible, por lo tanto, resta ver que se cumpla la propiedad característica de la esperanza condicional. Para ello, notemos que todo elemento $G \in \mathcal{G}$ es unión de elementos de $\{F_n\}_{n\in\mathbb{N}}$; es decir, para algún $I \subseteq \mathbb{N}$

$$G = \bigcup_{i\in I} F_i.$$

Además, tenemos que, para todo $i \in \mathbb{N}$,

$$\int_{F_i} \sum_{n\in\mathbb{N}} \mathbf{E}(X \mid F_n) \mathbb{1}_{F_n} \, dP = \sum_{n\in\mathbb{N}} \int_{F_i} \mathbf{E}(X \mid F_n) \mathbb{1}_{F_n} \, dP$$

$$
\begin{aligned}
&= \int_{F_i} \mathbf{E}\left(X \mid F_i\right) \mathbb{1}_{F_i}\, dP \\
&= \mathbf{E}\left(X \mid F_i\right) \int_{F_i} \mathbb{1}_{F_i}\, dP \\
&= \frac{\mathbf{E}\left(X \mathbb{1}_{F_i}\right)}{P(F_i)} P(F_i) \\
&= \mathbf{E}\left(X \mathbb{1}_{F_i}\right),
\end{aligned}
$$

por lo tanto, gracias a que la función

$$
A \mapsto \int_A \sum_{n \in \mathbb{N}} \mathbf{E}\left(X \mid F_n\right) \mathbb{1}_{F_n}\, dP
$$

es una medida con signo (véase [12]), tenemos que

$$
\begin{aligned}
\int_G \sum_{n \in \mathbb{N}} \mathbf{E}\left(X \mid F_n\right) \mathbb{1}_{F_n}\, dP &= \sum_{i \in I} \int_{F_i} \sum_{n \in \mathbb{N}} \mathbf{E}\left(X \mid F_n\right) \mathbb{1}_{F_n}\, dP \\
&= \sum_{i \in I} \mathbf{E}\left(X \mathbb{1}_{F_i}\right) \\
&= \mathbf{E}\left(X \sum_{i \in I} \mathbb{1}_{F_i}\right) \\
&= \mathbf{E}\left(X \mathbb{1}_{G}\right),
\end{aligned}
$$

luego

$$
\mathbf{E}\left(X \mid \mathcal{G}\right) = \sum_{n \in \mathbb{N}} \mathbf{E}\left(X \mid F_n\right) \mathbb{1}_{F_n},
$$

casi seguramente.

El resultado de este ejemplo también se lo conoce como el *Teorema de esperanza total* y nos será de utilidad cuando probemos el Teorema de densidad de Lebesgue.

EJEMPLO 3 (Relación entre la esperanza condicional y la probabilidad condicional elemental)**.** Gracias a la ecuación (3.14) vimos que la probabilidad condicional elemental es un caso particular de la esperanza condicional elemental. Ahora que ya hemos presentado un concepto general de esperanza condicional, este debería generalizar también al de probabilidad condicional elemental.

La probabilidad condicional de un evento $A \in \mathcal{F}$, dada una σ-álgebra se define como

$$
P(A|\mathcal{G}) := \mathbf{E}\left(\mathbb{1}_A \mid \mathcal{G}\right).
$$

Así, tenemos que, para todo $G \in \mathcal{G}$,

$$\begin{aligned}\int_G P(A|\mathcal{G})\,dP &= \int_G \mathbf{E}\left(\mathbb{1}_A \mid \mathcal{G}\right)\,dP\\ &= \int_G \mathbb{1}_A\,dP\\ &= P(A\cap G),\end{aligned}$$

por lo tanto, para $G \in \mathcal{G}$ tal que $P(G) > 0$, tenemos que

$$P(A|G) = \frac{P(A\cap G)}{P(G)} = \frac{1}{P(G)}\int_G P(A|\mathcal{G})\,dP.$$

De hecho, usando una idea similar, podemos probar que, si $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$, entonces

$$\mathbf{E}\left(X \mid G\right) = \frac{1}{P(G)}\int_G \mathbf{E}\left(X \mid \mathcal{G}\right)\,dP.$$

Por último, veamos unas cuantas propiedades de la esperanza condicional cuyas demostraciones se pueden encontrar en [4].

Teorema 3.6. *Sea $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$. Si $\mathcal{G}$ y $\mathcal{H}$ son sub σ-álgebras de $\mathcal{F}$, entonces*

a) *Si Y es una versión de $\boldsymbol{E}\left(X \mid \mathcal{G}\right)$, entonces $\boldsymbol{E}\left(Y\right) = \boldsymbol{E}\left(X\right)$.*

b) *Si X es $\mathcal{G}$-medible, entonces $\boldsymbol{E}\left(X \mid \mathcal{G}\right) = X$, casi seguramente.*

c) ***Linealidad**: Para todo $X_1, X_2 \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ y $a_1, a_2 \in \mathbb{R}$, se tiene que*

$$\boldsymbol{E}\left(a_1X_1 + a_2X_2 \mid \mathcal{G}\right) = a_1\boldsymbol{E}\left(X_1 \mid \mathcal{G}\right) + a_2\boldsymbol{E}\left(X_2 \mid \mathcal{G}\right),$$

casi seguramente.

d) ***No negatividad**: Si $X \geq 0$, entonces $\boldsymbol{E}\left(X \mid \mathcal{G}\right) \geq 0$, casi seguramente.*

e) ***Convergencia monótona**: Si $(X_n)_{n\in\mathbb{N}}$ es una sucesión creciente de variables aleatorias no negativas en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$ tal que $X_n \to X$, entonces*

$$\boldsymbol{E}\left(X_n \mid \mathcal{G}\right) \to \boldsymbol{E}\left(X \mid \mathcal{G}\right),$$

casi seguramente.

f) ***Lema de Fatou**: Si $(X_n)_{n\in\mathbb{N}}$ es una sucesión de variables aleatorias no negativas en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$, entonces*

$$\boldsymbol{E}\left(\liminf X_n \mid \mathcal{G}\right) \leq \liminf \boldsymbol{E}\left(X_n \mid \mathcal{G}\right),$$

casi seguramente.

g) ***Convergencia dominada**: Sea $(X_n)_{n\in\mathbb{N}}$ es una sucesión de variables aleatorias en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$ tal que $X_n \to X$, casi seguramente. Si existe $V \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ tal que, para todo $n \in \mathbb{N}$,*

$$|X_n| \leq V$$

casi seguramente, entonces

$$\boldsymbol{E}(X_n \mid \mathcal{G}) \to \boldsymbol{E}(X \mid \mathcal{G}),$$

casi seguramente.

h) ***Desigualdad de Jensen**: Si $c\colon \mathbb{R} \to \mathbb{R}$ es una función convexa y $\boldsymbol{E}(|c(X)|) < \infty$, entonces*

$$c(\boldsymbol{E}(X \mid \mathcal{G})) \leq \boldsymbol{E}(c(X) \mid \mathcal{G}),$$

casi seguramente.

i) ***Propiedad de torre**: Si $\mathcal{H}$ es una sub σ-álgebra de $\mathcal{G}$, entonces*

$$\boldsymbol{E}(\boldsymbol{E}(X \mid \mathcal{G}) \mid \mathcal{H}) = \boldsymbol{E}(X \mid \mathcal{H}),$$

casi seguramente.

j) *Si Y es $\mathcal{G}$-medible y acotada, entonces*

$$\boldsymbol{E}(XY \mid \mathcal{G}) = Y\boldsymbol{E}(X \mid \mathcal{G}),$$

casi seguramente.

k) *Si X es independiente de $\mathcal{G}$, entonces*

$$\boldsymbol{E}(X \mid \mathcal{G}) = \boldsymbol{E}(X),$$

casi seguramente.

Capítulo 4

Martingalas

En este capítulo utilizaremos las ideas ya concebidas sobre la esperanza condicional para introducir una nueva herramienta: las martingalas. Exploraremos varios resultados concernientes a ellas, de entre los cuales podemos destacar el Teorema de convergencia de submartingalas, mismo que será de utilidad para probar el Teorema de Radon-Nikodym de una manera constructiva.

4.1. Conceptos básicos

Consideremos un espacio de probabilidad $(\Omega, \mathcal{F}, P)$. Un *proceso estocástico* es una familia de variables aleatorias. A una sucesión creciente de sub σ-álgebras de $\mathcal{F}$, $(\mathcal{F}_n)_{n\in\mathbb{N}}$, la llamaremos *filtración*; y a $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{N}}, P)$ lo llamaremos *espacio filtrado* y denotaremos

$$\mathcal{F}_\infty := \sigma\left(\bigcup_{n\in\mathbb{N}} \mathcal{F}_n\right).$$

Se dice que un proceso estocástico $X = (X_n)_{n\in\mathbb{N}}$ es adaptado a $(\mathcal{F}_n)_{n\in\mathbb{N}}$ si, para todo $n \in \mathbb{N}$, X_n es $\mathcal{F}_n$-medible.

A lo largo de este capítulo, trabajaremos sobre un espacio filtrado arbitrario $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{N}}, P)$.

Definición 4.1. *Sea $X = (X_n)_{n\in\mathbb{N}}$ un proceso adaptado a $(\mathcal{F}_n)_{n\in\mathbb{N}}$. Si X cumple que, para todo $n \in \mathbb{N}$,*

a) $\boldsymbol{E}(X_n) < \infty$*; y*

b) $\boldsymbol{E}(X_{n+1} \mid \mathcal{F}_n) = X_n$*,* *se dice que X es una martingala. Si en lugar de b), se cumple que*

b') $\boldsymbol{E}(X_{n+1} \mid \mathcal{F}_n) \leq X_n$ (resp. $\boldsymbol{E}(X_{n+1} \mid \mathcal{F}_n) \geq X_n$),

se dice que X es una supermartingala (resp. submartingala).

Notemos que X es una supermartingala si y sólo si $-X$ es una submartingala. Además, se tiene que X es una martingala si y sólo si es, a la vez, una supermartingala y una submartingala.

Sea X una martingala, tenemos que, para todo $n \in \mathbb{N}$, X_n es $\mathcal{F}_n$-medible, por lo tanto

$$\mathbf{E}(X_n \mid \mathcal{F}_n) = X_n.$$

Así,

$$\begin{aligned}
\mathbf{E}(X_{n+1} \mid \mathcal{F}_n) = X_n &\iff \mathbf{E}(X_{n+1} \mid \mathcal{F}_n) = \mathbf{E}(X_n \mid \mathcal{F}_n) \\
&\iff \mathbf{E}(X_{n+1} \mid \mathcal{F}_n) - \mathbf{E}(X_n \mid \mathcal{F}_n) = 0 \\
&\iff \mathbf{E}(X_{n+1} - X_n \mid \mathcal{F}_n) = 0. \qquad (4.1)
\end{aligned}$$

De manera similar, se tiene que, si X es una supermartingala,

$$\mathbf{E}(X_{n+1} - X_n \mid \mathcal{F}_n) \leq 0. \qquad (4.2)$$

Así, tenemos que (4.1) y (4.2) pueden reemplazar a la segunda propiedad que define martingalas y supermartingalas, respectivamente.

De la definición, tenemos que, para todo $n \in \mathbb{N}$,

$$\mathbf{E}(X_{n+1} \mid \mathcal{F}_n) = X_n,$$

por lo tanto, $\mathbf{E}(X_{n+1}) = \mathbf{E}(X_n)$; y, si X es una submartingala, tendremos que $\mathbf{E}(X_{n+1}) \leq \mathbf{E}(X_n)$. Esto nos indica que, en promedio, una martingala se mantiene constante, mientras que una supermartingala decrece. Gracias a la Propiedad de torre vista en el Teorema 3.6, tenemos que

$$\mathbf{E}(\mathbf{E}(X_{n+1} \mid \mathcal{F}_n) \mid \mathcal{F}_{n-1}) = \mathbf{E}(X_{n+1} \mid \mathcal{F}_{n-1}),$$

por lo tanto

$$\mathbf{E}(X_{n+1} \mid \mathcal{F}_{n-1}) = \mathbf{E}(X_n \mid \mathcal{F}_{n-1}) = X_{n-1}.$$

Si repetimos este procedimiento $(n-1)$-veces, tendremos que

$$\mathbf{E}\left(X_{n+1} \mid \mathcal{F}_0\right) = X_0,$$

y, por lo tanto, $\mathbf{E}\left(X_{n+1}\right) = \mathbf{E}\left(X_0\right)$. Así mismo, en el caso en que X sea una submartingala, tendremos que $\mathbf{E}\left(X_{n+1}\right) \leq \mathbf{E}\left(X_0\right)$.

Ejemplo 4. Supongamos que $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ y definamos, para cada $n \in \mathbb{N}$,

$$M_n := \mathbf{E}\left(X \mid \mathcal{F}_n\right).$$

Notemos que, para todo $n \in \mathbb{N}$,

$$\mathbf{E}\left(M_{n+1} \mid \mathcal{F}_n\right) = \mathbf{E}\left(\mathbf{E}\left(X \mid \mathcal{F}_{n+1}\right) \mid \mathcal{F}_n\right).$$

Ya que $\mathcal{F}_n$ es una sub σ-álgebra de $\mathcal{F}_{n+1}$, por la Propiedad de Torre vista en el Teorema 3.6, tenemos que

$$\mathbf{E}\left(\mathbf{E}\left(X \mid \mathcal{F}_{n+1}\right) \mid \mathcal{F}_n\right) = \mathbf{E}\left(X \mid \mathcal{F}_n\right),$$

por lo tanto

$$\mathbf{E}\left(M_{n+1} \mid \mathcal{F}_n\right) = \mathbf{E}\left(X \mid \mathcal{F}_n\right) = M_n,$$

lo que implica que $\left(M_n\right)_{n \in \mathbb{N}}$ es una martingala.

Este ejemplo de martingala será retomado cuando estudiemos el Teorema de densidad de Lebesgue.

Supongamos que participamos en un juego de apuestas que funciona de la siguiente manera: en cada turno, nos asignarán un número real al azar, y este número se registrará por el proceso estocástico $\left(X_n\right)_{n \in \mathbb{N}}$, donde $X_0 = 0$; para cada $n \in \mathbb{N}^*$, en el n-ésimo turno, apostaremos una cierta cantidad de dinero C_n y el beneficio que obtendremos por cada unidad apostada está dado por la diferencia $X_n - X_{n-1}$; es decir, que nuestras ganancias (o pérdidas) en el n-ésimo turno estarán dadas por $C_n(X_n - X_{n-1})$.

Vamos a diseñar una estrategia para elegir el valor de nuestras apuestas en este juego; es decir, que vamos a determinar la manera en la que escogeremos los valores de C_n. Fijemos dos valores $a, b \in \mathbb{R}$ de modo que $a < b$. Mientras el número que nos asignen sea mayor o igual que a, nos mantendremos sin apostar; es decir, que C_n será igual a cero hasta que $X_n < a$. Una vez que $X_n > a$, inmediatamente empezaremos a hacer apuestas unitarias, es decir, $C_n = 1$, hasta que $X_n > b$. Cuando esto ocurra, pararemos nuestras apuestas ($C_n = 0$) y volveremos a empezar.

En esta estrategia, nos interesa saber cuándo nos asignan un número menor que a o un número mayor que b. Cuando ocurra lo segundo, dejaremos de apostar, es decir, que, de cierta forma, detendremos el juego. A una variable aleatoria que registre los instantes en los que debamos parar el juego se la suele llamar un *tiempo de parada*. Veamos su definición de manera formal:

DEFINICIÓN 4.2. *Sea $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{N}}, P)$ un espacio filtrado. Se dice que una función $T\colon \Omega \to \mathbb{N} \cup \{\infty\}$ es un* tiempo de parada *si, para todo $n \in \mathbb{N} \cup \{\infty\}$,*

$$\{T \leq n\} = \{\omega \in \Omega : T(\omega) \leq n\} \in \mathcal{F}_n;$$

o, de forma equivalente,

$$\{T = n\} = \{\omega \in \Omega : T(\omega) = n\} \in \mathcal{F}_n.$$

La palabra «parada» en la definición anterior no debe tomarse literalmente, pues en la estrategia que acabamos de diseñar, la variable que registra el primer instante en el que nos asignan un número menor que a también es un tiempo de parada aunque después de eso no dejemos de jugar. De hecho, a una variable de este tipo se la conoce como primer tiempo de arribo.

EJEMPLO 5 (Tiempos de arribo). Sea $(X_n)_{n\in\mathbb{N}}$ un proceso estocástico adaptado a la filtración $(\mathcal{F}_n)_{n\in\mathbb{N}}$ y $A \in \mathcal{B}$. Se define el *primer tiempo de arribo* del proceso $(X_n)_{n\in\mathbb{N}}$ al conjunto A como

$$T_1 := \text{mín}\{n \geq 1 : X_n \in A\}.$$

Es decir, que T_1 representa el primer momento en el que $(X_n)_{n\in\mathbb{N}}$ toma un valor dentro del conjunto A. Si esto no ocurre, entonces diremos que $T_1 = \infty$.

La variable aleatoria T_1 es, en efecto, un tiempo de parada, pues

$$\{T_1 = n\} = \{X_0 \notin A\} \cap \{X_1 \notin A\} \cap \cdots \cap \{X_{n-1} \notin A\} \cap \{X_n \in A\} \in \mathcal{F}_n.$$

El segundo instante en el que el proceso $(X_n)_{n\in\mathbb{N}}$ tome un valor dentro del conjunto A está dado por

$$T_2 := \text{mín}\{n > T_1 : X_n \in A\}.$$

A esta variable la llamaremos *segundo tiempo de arribo*. Notemos que T_2 también es un tiempo de parada, pues

$$\{T_2 = n\} = \bigcup_{k=0}^{n-1} \{T_1 = k\} \cap \{X_{k+1} \notin A\} \cap \cdots \cap \{X_{n-1} \notin A\} \cap \{X_n \in A\} \in \mathcal{F}_n.$$

De manera similar, podemos definir para cada $n \in \mathbb{N}^*$, el n-ésimo tiempo de arribo y ver que el mismo es un tiempo de parada.

Dado un proceso estocástico $(X_n)_{n\in\mathbb{N}}$ y un tiempo de parada T, al proceso $X^T = (X_{T\wedge n})_{n\in\mathbb{N}}$ se lo conoce como *proceso detenido*. Veamos su relación con las martingalas:

Teorema 4.1. *Sea $X = (X_n)_{n\in\mathbb{N}}$ una martingala (resp. supermartingala) con respecto a una filtración $(\mathcal{F}_n)_{n\in\mathbb{N}}$ y T un tiempo de parada. Se tiene que el proceso detenido $X^T = (X_{T\wedge n})_{n\in\mathbb{N}}$ es una martingala (resp. supermartingala) con respecto a la misma filtración que X.*

Demostración. Sea $n \in \mathbb{N}$. Se tiene que

$$X_{T\wedge n} = \begin{cases} X_T & \text{si } T \leq n \\ X_n & \text{si } T > n, \end{cases}$$

por lo tanto

$$X_{T\wedge n} = \sum_{k=0}^{n} X_k \mathbb{1}_{\{T=k\}} + X_n \mathbb{1}_{\{T>n\}}.$$

Ahora, ya que X es un proceso adaptado, se tiene que, para todo $k \in \{0, \ldots, n\}$, X_k es $\mathcal{F}_k$-medible, y puesto que $\mathcal{F}_k \subseteq \mathcal{F}_n$, se sigue que X_k es $\mathcal{F}_n$-medible. Además, dado que T es un tiempo de parada, tenemos que el conjunto $\{T = k\} \in \mathcal{F}_k \subseteq \mathcal{F}_n$, lo que implica que $\mathbb{1}_{\{T=k\}}$ es $\mathcal{F}_n$-medible. Asimismo, $\{T > n\} = \{T \leq n\}^c \in \mathcal{F}_n$, por lo que $\mathbb{1}_{\{T>n\}}$ es también $\mathcal{F}_n$-medible. Luego, $X_{T\wedge n}$ es $\mathcal{F}_n$-medible y, por ende, X^T es adaptado.

Por otra parte, tenemos que

$$X_{T\wedge(n+1)} = \begin{cases} X_T & \text{si } T < n+1 \\ X_{n+1} & \text{si } T \geq n+1, \end{cases}$$

o, de manera equivalente,

$$X_{T\wedge(n+1)} = \begin{cases} X_T & \text{si } T \leq n \\ X_{n+1} & \text{si } T > n, \end{cases}$$

así

$$X_{T\wedge(n+1)} = \sum_{k=0}^{n} X_k \mathbb{1}_{\{T=k\}} + X_{n+1}\mathbb{1}_{\{T>n\}}.$$

Con esto, tenemos que

$$\begin{aligned} \mathbf{E}\left(X_{T\wedge(n+1)} \,\middle|\, \mathcal{F}_n\right) &= \mathbf{E}\left(\sum_{k=0}^{n} X_k \mathbb{1}_{\{T=k\}} + X_{n+1}\mathbb{1}_{\{T>n\}} \,\middle|\, \mathcal{F}_n\right) \\ &= \sum_{k=0}^{n} \mathbf{E}\left(X_k \mathbb{1}_{\{T=k\}} \,\middle|\, \mathcal{F}_n\right) + \mathbf{E}\left(X_{n+1}\mathbb{1}_{\{T>n\}} \,\middle|\, \mathcal{F}_n\right). \end{aligned}$$

Recordemos que, para todo $k \in \{0, \ldots, n\}$, tenemos que $X_k \mathbb{1}_{\{T=k\}}$ es $\mathcal{F}_n$-medible, por lo tanto,

$$\mathbf{E}\left(X_k \mathbb{1}_{\{T=k\}} \,\middle|\, \mathcal{F}_n\right) = X_k \mathbb{1}_{\{T=k\}};$$

además, $\mathbb{1}_{\{T>n\}}$ es $\mathcal{F}_n$-medible y acotada, entonces

$$\mathbf{E}\left(X_{n+1}\mathbb{1}_{\{T>n\}} \,\middle|\, \mathcal{F}_n\right) = \mathbf{E}\left(X_{n+1} \mid \mathcal{F}_n\right)\mathbb{1}_{\{T>n\}} = X_n \mathbb{1}_{\{T>n\}}.$$

Así, tenemos que

$$\begin{aligned} \mathbf{E}\left(X_{T\wedge(n+1)} \,\middle|\, \mathcal{F}_n\right) &= \sum_{k=0}^{n} X_k \mathbb{1}_{\{T=k\}} + X_n \mathbb{1}_{\{T>n\}} \\ &= X_{T\wedge n}, \end{aligned}$$

es decir, que X^T es una martingala. Para el caso de supermartingalas, la demostración es análoga. □

En la estrategia que habíamos elaborado, nuestra apuesta en el instante $n \in \mathbb{N}^*$ se define según los valores que nos han sido asignados hasta el instante $n-1$; es decir, que la manera en la que elegimos C_n se basará en la información conocida del juego hasta el instante $n-1$. En otras palabras, si la σ-álgebra $\mathcal{F}_{n-1}$ almacena la información del juego hasta el instante $n-1$, entonces la variable aleatoria C_n es $\mathcal{F}_{n-1}$-medible. Esto motiva la siguiente definición:

DEFINICIÓN 4.3. *Sea $C = \{C_n\}_{n\in\mathbb{N}^*}$ un proceso estocástico. Diremos que C previsible, si, para todo $n \in \mathbb{N}^*$, se tiene que C_n es $\mathcal{F}_{n-1}$-medible.*

Consideremos los procesos $X = (X_n)_{n\in\mathbb{N}}$ y $C = (C_n)_{n\in\mathbb{N}}$ definidos en nuestra estrategia de apuestas. Veamos que, en efecto, el proceso C es previsible. Notemos que $C_1 = \mathbb{1}_{\{X_0<a\}}$, por lo tanto C_1 es $\mathcal{F}_0$-medible. Supongamos que para algún natural $k \geq 1$ se tiene que C_k es $\mathcal{F}_{k-1}$-medible, entonces se tiene que C_{k+1} es $\mathcal{F}_k$-medible, ya que

$$C_{k+1} = \mathbb{1}_{\{C_k=1\}}\mathbb{1}_{\{X_n\leq b\}} + \mathbb{1}_{\{C_k=0\}}\mathbb{1}_{\{X_n<a\}}.$$

Por otra parte, si X es una martingala, entonces

$$\mathbf{E}\left(X_{n+1} - X_n \mid \mathcal{F}_n\right) = 0;$$

es decir, que, en promedio, ni ganaremos, ni perderemos dinero. En ese caso, diremos que el juego es *justo*. Si en cambio, X es una supermartingala, tendremos que

$$\mathbf{E}\left(X_{n+1} - X_n \mid \mathcal{F}_n\right) \leq 0,$$

por lo tanto, en promedio, perderemos dinero. En este caso, diremos que el juego es *desfavorable*.

Nuestras ganancias en el n-ésimo turno están dadas por

$$C_n(X_n - X_{n-1}),$$

y por lo tanto, nuestras ganancias acumuladas hasta el tiempo n estarán dadas por el proceso $Y = (Y_n)_{n\in\mathbb{N}}$, definido de la siguiente manera:

$$n \mapsto Y_n = \begin{cases} 0 & \text{si } n = 0, \\ \displaystyle\sum_{k=1}^{n} C_k(X_k - X_{k-1}) & \text{si } n \neq 0. \end{cases} \tag{4.3}$$

Con esto en mente, probemos el siguiente teorema:

TEOREMA 4.2. *Sea C un proceso previsible y acotado en el sentido de que para algún $k \in [0,\infty)$ y para cualesquier $n \in \mathbb{N}$ y $\omega \in \Omega$, se tiene que $|C_n(\omega)| \leq k$. Entonces,*

a) Si C es no negativo y X es una supermartingala (resp. martingala), entonces Y es una supermartingala (resp. martingala) nula en 0.

b) Si X es una martingala, entonces Y es una martingala nula en 0.

Demostración.

a) Haremos la demostración para supermartingalas, pues el caso de martingalas se trata de manera análoga.

Por definición, tenemos que $Y_0 = 0$. Por lo tanto Y es un proceso nulo en 0. Notemos que

$$Y_n - Y_{n-1} = C_n(X_n - X_{n-1}),$$

por lo tanto,

$$\mathbf{E}\left(Y_n - Y_{n-1} \mid \mathcal{F}_{n-1}\right) = \mathbf{E}\left(C_n(X_n - X_{n-1}) \mid \mathcal{F}_{n-1}\right).$$

Además, gracias a que C_n es $\mathcal{F}_{n-1}$- medible y acotada, entonces

$$\mathbf{E}\left(Y_n - Y_{n-1} \mid \mathcal{F}_{n-1}\right) = C_n\mathbf{E}\left(X_n - X_{n-1} \mid \mathcal{F}_{n-1}\right).$$

Por otra parte, gracias a que X es una supermartingala, tenemos que

$$\mathbf{E}\left(X_n - X_{n-1} \mid \mathcal{F}_{n-1}\right) \leq 0.$$

Luego, ya que C es no negativo,

$$C_n\mathbf{E}\left(X_n - X_{n-1} \mid \mathcal{F}_{n-1}\right) = \mathbf{E}\left(Y_n - Y_{n-1} \mid \mathcal{F}_{n-1}\right) \leq 0.$$

b) Nuevamente, por definición, Y es un proceso nulo en 0. De manera análoga a la del literal anterior, tenemos que

$$\begin{aligned}\mathbf{E}\left(Y_n - Y_{n-1} \mid \mathcal{F}_{n-1}\right) &= \mathbf{E}\left(C_n(X_n - X_{n-1}) \mid \mathcal{F}_{n-1}\right)\\ &= C_n\mathbf{E}\left(X_n - X_{n-1} \mid \mathcal{F}_{n-1}\right)\\ &= 0.\end{aligned}$$ □

El resultado de este teorema nos indica que sin importar cuál sea nuestra estrategia de apuestas C, si el juego era desfavorable, seguirá siendo desfavorable.Aunque más allá de esta interpretación, este teorema nos será útil en el camino hacia la demostración del Teorema de convergencia de submartingalas.

4.2. El teorema de convergencia de martingalas

Cuando hablamos de sucesiones, ya sea de números reales, o también de funciones, una de las primeras propiedades que nos interesa estudiar es su convergencia. El caso de las martingalas no será la excepción. Buscaremos probar que, bajo ciertas circunstancias, podemos garantizar la convergencia casi segura de martingalas.

Consideremos los procesos $C = (C_n)_{n\in\mathbb{N}}$ y $X = (X_n)_{n\in\mathbb{N}}$ como en la estrategia de apuestas expuesta en la sección anterior, e Y como en (4.3). Sean a, $b \in \mathbb{R}$ tales que $a < b$.

Notemos que, para cada $n \in \mathbb{N}$, X_n podría tomar valores menores que a o valores menores que b, así como también podría no hacerlo. Si existen n, $m \in \mathbb{N}$ tales que $n < m$, $X_n < a$ y $b < X_m$, diremos que X cruzó el intervalo $[a, b]$. Esto motiva la siguiente definición:

DEFINICIÓN 4.4. *Sean $X = (X_n)_{n\in\mathbb{N}}$ un proceso estocástico y $\omega \in \Omega$. Si $N \in \mathbb{N}$, definimos $U_N[a, b]$ como el número de veces que X cruza el intervalo $[a, b]$ hasta el tiempo N. Es decir, $U_N[a, b](\omega)$ es el mayor número natural k para el cual es posible encontrar un conjunto*

$$\{s_1, s_2, \ldots, s_k, t_1, t_2, \ldots, t_k\} \subseteq \mathbb{N}$$

tal que

$$0 \leq s_1 < t_1 < s_2 < t_2 < \cdots < s_k < t_k \leq N$$

y, para todo $i \in \{1, \ldots, k\}$,

$$X_{s_i} < a \quad y \quad X_{t_i} > b.$$

Notemos que, si X nunca cruza el intervalo $[a, b]$, entonces $U_N[a, b](\omega) = 0$. El siguiente lema nos dará una relación entre nuestras ganancias en el juego de apuestas y el número de veces que X cruza el intervalo $[a, b]$ hasta un determinado tiempo.

LEMA 4.3. *Sea $N \in \mathbb{N}$. Tenemos que*

$$Y_N \geq (b - a)U_N[a, b] - (X_N - a)^-,$$

donde $(X_N - a)^-$ es la parte negativa de $X_N - a$.

Demostración. Supongamos que $N > 0$ y definamos $k := U_N[a, b]$. Si $k = 0$, la desigualdad es trivial, pues $Y_N \geq 0$ y $(X_N - a)^- \geq 0$.

Supongamos que $k > 0$. Entonces existe

$$\{s_1, s_2, \ldots, s_k, t_1, t_2, \ldots, t_k\} \subseteq \mathbb{N}$$

tal que

$$0 \leq s_1 < t_1 < s_2 < t_2 < \cdots < s_k < t_k \leq N$$

y, para todo $i \in \{1, \ldots, k\}$,

$$X_{s_i} < a < b < X_{t_i}.$$

Ya que, para todo $i \in \{1, \ldots, k\}$, $X_{s_i} < a$, debemos apostar en el turno $s_i + 1$, por lo tanto $C_{s_i+1} = 1$. Además, tenemos que seguir apostando hasta que X sobrepase b, por lo tanto, podemos definir, para cada $i \in \{1, \ldots, k\}$, la variable l_i como el primer instante en que X toma un valor mayor a b, después de instante s_i; es decir,

$$l_i := \min\{n \in \mathbb{N} : n > s_i \wedge X_n > b\} \leq t_i,$$

así, si $n \leq t_k$, entonces

$$C_n = \begin{cases} 1 & \text{si } s_i + 1 \leq n \leq l_i, \\ 0 & \text{en caso contrario.} \end{cases}$$

Tenemos que $t_k \leq N$; si $t_k = N$, entonces nuestras ganancias hasta el tiempo N estarán dadas por

$$\begin{aligned} Y_N &= \sum_{j=1}^{N} C_j (X_j - X_{j-1}) \\ &= \sum_{i=1}^{k} \left(\sum_{n=s_i}^{l_i - 1} C_{n+1} (X_{n+1} - X_n) \right) \\ &= \sum_{i=1}^{k} \left(\sum_{n=s_i}^{l_i - 1} (X_{n+1} - X_n) \right) \\ &= \sum_{i=1}^{k} \left(X_{l_i} - X_{s_i} \right). \end{aligned}$$

En cambio, si $t_k < N$, podría ocurrir que hasta el tiempo N, X no vuelva a tomar valores por debajo de a, por lo tanto, para $t_k + 1 \leq n \leq N$, tenemos que $C_n = 0$, es decir que las ganancias hasta el tiempo N serán las mismas que si $t_k = N$, o sea que

$$Y_N = \sum_{i=1}^{k} \left(X_{l_i} - X_{s_i} \right). \tag{4.4}$$

Así, gracias a que, para todo $i \in \{1, \ldots k\}$,

$$X_{l_i} - X_{s_i} > b - a,$$

y a que

$$(X_N - a)^- \geq 0,$$

de (4.4), se sigue que

$$Y_N > k(b-a) \geq k(b-a) - (X_N - a)^-. \tag{4.5}$$

Por otra parte, si X llegara a tomar un valor menor que a antes del tiempo N, podemos definir

$$s_{k+1} := \min\{n \in \mathbb{N} : t_k < n \leq N \wedge X_n < a\},$$

es decir, s_{k+1} es la primera vez, después del último cruce, que X toma un valor menor que a. Después de s_{k+1}, y sin sobrepasar N, X no puede tomar valores mayores a b, pues eso indicaría que $U_N[a,b] > k$, lo cual es absurdo. Así que, si $t_k < n \leq N$, entonces

$$C_n = \begin{cases} 1 & \text{si } \ s_{k+1} + 1 \leq n \leq N, \\ 0 & \text{si } \ t_k + 1 \leq n \leq s_{k+1}. \end{cases}$$

Con esto, tenemos que

$$\begin{aligned} Y_N &= \sum_{j=1}^{N} C_j(X_j - X_{j-1}) \\ &= \sum_{i=1}^{k} \left(\sum_{n=s_i}^{l_i - 1} C_{n+1}(X_{n+1} - X_n) \right) + \sum_{n=s_{k+1}}^{N-1} C_{n+1}(X_{n+1} - X_n) \\ &= \sum_{i=1}^{k} \left(\sum_{n=s_i}^{l_i - 1} (X_{n+1} - X_n) \right) + \sum_{n=s_{k+1}}^{N-1} (X_{n+1} - X_n) \\ &= \sum_{i=1}^{k} \left(X_{l_i} - X_{s_i} \right) + X_N - X_{s_{k+1}}. \end{aligned} \tag{4.6}$$

Notemos que $-X_{s_{k+1}} > -a$, por lo tanto, de (4.6), se sigue que

$$Y_N > k(b-a) + X_N - a = k(b-a) - (a - X_N),$$

Finalmente, si $X_N \geq a$, entonces $(X_N - a)^- = 0$ y

$$Y_N > k(b-a);$$

y, si $X_N < a$, entonces $(X_N - a)^- = a - X_N$, por lo tanto

$$Y_N > k(b-a) - (X_N - a)^-. \tag{4.7}$$

De (4.5) y (4.7), se sigue que

$$Y_N > k(b-a) - (X_N - a)^-. \tag{4.8}$$

Por último, si $N = 0$, entonces $Y_N = 0$ y $k = 0$, por lo tanto

$$Y_N \geq k(b-a) - (X_N - a)^-. \tag{4.9}$$

De (4.8) y (4.9) se sigue el resultado. $\square$

Con este resultado seremos capaces de probar el siguiente lema:

LEMA 4.4 (Lema de cruces)**.** *Sean X una supermartingala y $U_N[a,b]$ el número de veces que X cruzó el intervalo $[a,b]$ hasta el tiempo N. Entonces*

$$(b-a)\boldsymbol{E}\left(U_N[a,b]\right) \leq \boldsymbol{E}\left((X_N - a)^-\right).$$

Demostración. Sabemos que nuestra estrategia de apuestas C es un proceso previsible; tenemos también que C es acotado y no negativo, por lo tanto, por el Teorema 4.2, Y es una supermartingala nula en 0. Así, tenemos que

$$\mathbf{E}\left(Y_N\right) \leq \mathbf{E}\left(Y_0\right) = 0.$$

Con esto, y tomando esperanzas en el resultado del lema anterior, tenemos que

$$\begin{aligned} 0 &\geq \mathbf{E}\left(Y_N\right) \\ &\geq \mathbf{E}\left((b-a)U_N[a,b] - (X_N - a)^-\right) \\ &= (b-a)\mathbf{E}\left(U_N[a,b]\right) - \mathbf{E}\left((X_N - a)^-\right), \end{aligned}$$

de donde se sigue que

$$(b-a)\mathbf{E}\left(U_N[a,b]\right) \leq \mathbf{E}\left((X_N - a)^-\right).$$ $\square$

Este resultado nos dice que, en promedio, el número de veces que una martingala puede cruzar un intervalo hasta cierto tiempo N está dominado por el valor de la martingala en ese mismo tiempo.

COROLARIO 4.5. *Sea X una supermartingala acotada en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$; es decir,*

$$\sup_{n \in \mathbb{N}} \boldsymbol{E}\left(|X_n|\right) < \infty.$$

Sean a, $b \in \mathbb{R}$ tales que $a < b$. Se tiene que

$$(b-a)\boldsymbol{E}\left(U_\infty[a,b]\right) \leq |a| + \sup_{n \in \mathbb{N}} \boldsymbol{E}\left(|X_n|\right) < \infty,$$

donde

$$U_\infty[a,b] := \lim_{N\to\infty} U_N[a,b].$$

Además

$$P(U_\infty[a,b] = \infty) = 0.$$

Demostración. En primer lugar, gracias a que

$$\mathbf{E}\left(|X_N - a|\right) = \mathbf{E}\left((X_N - a)^+\right) + \mathbf{E}\left((X_N - a)^-\right) \geq \mathbf{E}\left((X_N - a)^-\right)$$

y junto con lema anterior, tenemos que

$$\begin{aligned}(b-a)\mathbf{E}\left(U_N[a,b]\right) &\leq \mathbf{E}\left((X_N - a)^-\right)\\ &\leq \mathbf{E}\left(|X_N - a|\right)\\ &\leq \mathbf{E}\left(|X_N| + |a|\right)\\ &= \mathbf{E}\left(|X_N|\right) + \mathbf{E}\left(|a|\right)\\ &= \mathbf{E}\left(|X_N|\right) + |a|\\ &\leq \sup_{n\in\mathbb{N}} \mathbf{E}\left(|X_n|\right) + |a|.\end{aligned}$$

Por otra parte, tenemos que $(U_N[a,b])_{N\in\mathbb{N}}$ es una sucesión creciente de variables aleatorias no negativas, por lo tanto, por el Teorema de convergencia monótona, tenemos que

$$\mathbf{E}\left(U_\infty[a,b]\right) = \lim_{N\to\infty} \mathbf{E}\left(U_N[a,b]\right).$$

Así, se sigue que

$$(b-a)\mathbf{E}\left(U_\infty[a,b]\right) \leq \sup_{n\in\mathbb{N}} \mathbf{E}\left(|X_n|\right) + |a| < \infty.$$

Por lo tanto, $U_\infty[a,b] \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$, luego $U_\infty[a,b]$ es finita casi seguramente; es decir,

$$P(U_\infty[a,b] = \infty) = 0. \qquad \square$$

Gracias a este corolario, podemos decir que, casi seguramente, una martingala acotada en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$, $(X_n)_{n\in\mathbb{N}}$ cruzará cualquier intervalo $[a,b]$ un número finito de veces, incluso cuando $n \to \infty$.

Habiendo demostrado el Lema 4.4 y su corolario, estamos en capacidad de probar el Teorema que da su nombre a esta sección:

TEOREMA 4.6 (Convergencia de submartingalas). *Sea X una supermartingala aco-*

tada en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$. Entonces el límite

$$X_\infty := \lim_{n\to\infty} X_n$$

existe y es finito casi seguramente. Además, $X_\infty \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$.

Demostración. En primer lugar, definamos el conjunto

$$A := \{w \in \Omega : (X_n(\omega))_{n\in\mathbb{N}} \text{ no converge en } [-\infty, \infty]\}$$

y segundo lugar, para $a, b \in Q$, definamos

$$A_{a,b} := \{w \in \Omega : \liminf X_n(\omega) < a < b < \limsup X_n(\omega)\}.$$

Si $A = \varnothing$, tenemos que $P(A) = 0$, por lo tanto, el límite

$$X_\infty := \lim_{n\to\infty} X_n$$

existe casi seguramente en $[-\infty, \infty]$.

Ahora, supongamos que $A \neq \varnothing$ y tomemos $\omega \in A$. Ya que $(X_n(\omega))_{n\in\mathbb{N}}$ no converge en $[-\infty, \infty]$, tenemos que

$$\liminf X_n(\omega) < \limsup X_n(\omega).$$

Gracias a que $\mathbb{Q}$ es denso en $\mathbb{R}$, existen $a, b \in \mathbb{Q}$ tales que

$$\liminf X_n(\omega) < a < b < \limsup X_n(\omega);$$

por lo tanto,

$$A \subseteq \bigcup_{\substack{a,b\in\mathbb{Q}\\ a<b}} A_{a,b}.$$

En cambio, si existen $a, b \in \mathbb{Q}$ tales que

$$\liminf X_n(\omega) < a < b < \limsup X_n(\omega),$$

entonces la sucesión $(X_n(\omega))_{n\in\mathbb{N}}$ no converge $[-\infty, \infty]$, por lo tanto

$$\bigcup_{\substack{a,b\in\mathbb{Q}\\ a<b}} A_{a,b} \subseteq A.$$

Así, tenemos que

$$A = \bigcup_{\substack{a,b \in \mathbb{Q} \\ a<b}} A_{a,b}.$$

Por otra parte, para $a, b \in \mathbb{Q}$, tenemos que

$$\begin{aligned} \liminf X_n(\omega) < a &\iff \sup_{m \in \mathbb{N}} \inf_{n>m} X_n(\omega) < a \\ &\iff (\forall m \in \mathbb{N})(\exists n > m)(X_n(\omega) < a), \end{aligned}$$

y

$$\begin{aligned} \limsup X_n(\omega) > b &\iff \inf_{m \in \mathbb{N}} \sup_{n>m} X_n(\omega) > b \\ &\iff (\forall m \in \mathbb{N})(\exists n > m)(X_n(\omega) > b); \end{aligned}$$

es decir, que cuando $n \to \infty$, $(X_n(\omega))_{n \in \mathbb{N}}$ habrá cruzado el intervalo $[a, b]$ una cantidad infinita de veces. Dicho de otra forma, tenemos que $U_\infty[a, b](\omega) = \infty$. Por lo tanto,

$$A_{a,b} \subseteq \{\omega \in \Omega : U_\infty[a, b](\omega) = \infty\},$$

y gracias al Corolario 4.5,

$$P(A_{a,b}) \leq P(\{\omega \in \Omega : U_\infty[a, b](\omega) = \infty\}) = 0,$$

lo que implica que $P(A) = 0$. Luego el límite

$$X_\infty := \lim_{n \to \infty} X_n$$

existe casi seguramente en $[-\infty, \infty]$.

Finalmente, gracias al Lema de Fatou, tenemos que

$$\begin{aligned} \mathbf{E}\left(|X_\infty|\right) &= \mathbf{E}\left(\lim_{n \to \infty} |X_n|\right) \\ &= \mathbf{E}\left(\liminf |X_n|\right) \\ &\leq \liminf \mathbf{E}\left(|X_n|\right) \\ &\leq \sup_{n \in \mathbb{N}} \mathbf{E}\left(|X_n|\right) \\ &< \infty, \end{aligned}$$

lo que implica que $X_\infty \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$. $\square$

COROLARIO 4.7. *Si X es una supermartingala no negativa, entonces el límite*

$$X_\infty := \lim_{n\to\infty} X_n$$

existe y es finito casi seguramente.

Demostración. Tenemos que

$$\mathbf{E}\left(|X_n|\right) = \mathbf{E}\left(X_n\right) \leq \mathbf{E}\left(X_0\right) < \infty,$$

por lo tanto X es acotada en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$, y el resultado se sigue inmediatamente. □

Notemos que los resultado obtenidos son aplicables también a martingalas, pues estas son también supermartingalas.

4.3. Integrabilidad uniforme

En la sección anterior, garantizamos la convergencia casi segura de martingalas bajo ciertas condiciones; en esta sección, buscaremos mejorar ese resultado garantizando su convergencia en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$. Una de las propiedades que nos permitirá garantizar dicha convergencia es la de integrabilidad uniforme, pero antes de estudiarla, veamos un par de propiedades útiles de las variables aleatorias integrables.

LEMA 4.8. *Sea $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$. Para todo $\varepsilon > 0$, existe $\delta > 0$ tal que, para todo $F \in \mathcal{F}$,*

$$P(F) < \delta \implies \int_F |X|\, dP < \varepsilon.$$

Demostración. Por reducción al absurdo, supongamos que existe $\varepsilon > 0$ tal que, para cada $n \in \mathbb{N}$, existe $F_n \in \mathcal{F}$ tal que

$$P(F_n) < \frac{1}{2^n} \qquad \text{y} \qquad \int_{F_n} |X|\, dP \geq \varepsilon.$$

Así, tenemos que, por un parte,

$$\limsup \int_{F_n} |X|\, dP \geq \varepsilon,$$

y por otra

$$\sum_{n\in\mathbb{N}} P(F_n) < 2 < \infty.$$

De esta última relación, gracias al Lema de Borel Cantelli, tenemos que

$$P(\limsup F_n) = 0.$$

Consideremos la sucesión de funciones $(f_n)_{n\in\mathbb{N}}$ definida por

$$\begin{aligned} f_n \colon \ \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto f_n(\omega) = |X(\omega)|\mathbb{1}_{F_n}(\omega). \end{aligned}$$

Notemos que $0 \leq f_n \leq |X|$, para todo $n \in \mathbb{N}$. Por lo tanto, gracias al Lema de Fatou inverso (Teorema 2.18), tenemos que

$$\mathbf{E}\left(\limsup f_n\right) \geq \limsup \mathbf{E}\left(f_n\right).$$

Ahora, ya que

$$\limsup f_n = |X|\mathbb{1}_{\limsup F_n},$$

se sigue que

$$\mathbf{E}\left(|X|\mathbb{1}_{\limsup F_n}\right) \geq \limsup \mathbf{E}\left(f_n\right),$$

y en consecuencia

$$0 = \int_{\limsup F_n} |X| \geq \limsup \mathbf{E}\left(f_n\right) \geq \varepsilon,$$

lo cual es absurdo. □

Esta propiedad nos dice que la integral de una variable aleatoria integrable es, de cierta forma, continua en conjuntos de probabilidad nula.

Supongamos que $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$. Tenemos que X es finita, casi seguramente, por lo tanto, la familia

$$\{|X| > k\}_{k>0}$$

es decreciente y tiene como límite el evento de probabilidad nula $\{|X| = \infty\}$. Entonces, a medida que tomemos valores más grandes para k, esperaríamos que la integral de $|X|$ sobre el evento $\{|X| > k\}$ time valores cercanos a cero. En efecto, esto ocurre y es un corolario del lema anterior.

COROLARIO 4.9. *Sea $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$. Para todo $\varepsilon > 0$, existe $k \in [0, \infty)$ tal que*

$$\int_{\{|X|>k\}} |X| \, dP < \varepsilon.$$

Demostración. Sea $\varepsilon > 0$. Gracias al lema anterior, existe $\delta > 0$ tal que, para todo

$F \in \mathcal{F}$,

$$P(F) < \delta \implies \int_F |X|\, dP < \varepsilon.$$

Ya que, para todo $k \in [0, \infty)$,

$$\begin{aligned}
\mathbf{E}\left(|X|\right) &= \int_{\{|X|>k\}} |X|\, dP + \int_{\{|X|\leq k\}} |X|\, dP \\
&\geq \int_{\{|X|>k\}} |X|\, dP \\
&> \int_{\{|X|>k\}} k\, dP \\
&= kP(|X| > k),
\end{aligned}$$

podemos tomar k lo suficientemente grande, de modo que

$$P(|X| > k) \leq \frac{\mathbf{E}\left(|X|\right)}{k} < \delta,$$

y, por lo tanto,

$$\int_{\{|X|>k\}} |X|\, dP < \varepsilon. \qquad \square$$

Esta última propiedad nos permitirá definir lo que es una familia de variables aleatorias uniformemente integrable.

DEFINICIÓN 4.5. *Una colección $\mathcal{C}$ de variables aleatorias se dice* uniformemente integrable *si para todo $\varepsilon > 0$, existe $k \in [0, \infty)$ tal que, para cualquier $X \in \mathcal{C}$,*

$$\int_{\{|X|>k\}} |X|\, dP < \varepsilon.$$

Notemos que, si $\mathcal{C}$ es uniformemente integrable, entonces existe $k \in [0, \infty)$ tal que, para cualquier $X \in \mathcal{C}$,

$$\int_{\{|X|>k\}} |X|\, dP < 1.$$

Por lo tanto,

$$\begin{aligned}
\mathbf{E}\left(|X|\right) &= \int_{\{|X|>k\}} |X|\, dP + \int_{\{|X|\leq k\}} |X|\, dP \\
&< 1 + \int_{\{|X|\leq k\}} |X|\, dP \\
&\leq 1 + \int_{\{|X|\leq k\}} k\, dP \\
&\leq 1 + kP(|X| \leq k) \\
&\leq 1 + k;
\end{aligned}$$

así, toda familia uniformemente integrable es acotada en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$.

Hasta ahora hemos trabajado con dos tipos de convergencia de variables aleatorias: la convergencia casi segura y la convergencia en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$, o conocida también como *convergencia en media*. A continuación presentaremos un nuevo modo de convergencia.

DEFINICIÓN 4.6. *Sea X una variable aleatoria y $(X_n)_{n\in\mathbb{N}}$ una sucesión de variables aleatorias. Diremos que X_n converge en probabilidad a X, y lo denotaremos por*

$$X_n \xrightarrow{p} X,$$

si, para todo $\varepsilon > 0$,

$$P(|X_n - X| > \varepsilon) \to 0,$$

cuando $n \to \infty$.

La convergencia casi segura es un concepto más fuerte que el de convergencia en probabilidad, en el sentido de que la primera implica la segunda. Para una demostración de este hecho, además de una introducción a los distintos tipos de convergencia, se puede consultar [13].

En el primer capítulo, vimos que una de las hipótesis del Teorema de convergencia acotada es que $X_n \to X$ casi seguramente. Ahora veremos que este teorema seguirá siendo válido cuando relajemos dicha hipótesis y la reemplacemos por la de convergencia en probabilidad.

TEOREMA 4.10. *Sean X una variable aleatoria y $(X_n)_{n\in\mathbb{N}}$ una sucesión de variables aleatorias. Supongamos que $X_n \xrightarrow{p} X$ y que existe $k \in (0, \infty)$ tal que, para todo $n \in \mathbb{N}$ y $\omega \in \Omega$,*

$$|X_n(\omega)| \leq k.$$

Entonces

$$\boldsymbol{E}(|X_n - X|) \to 0.$$

Demostración. Sea $m \in \mathbb{N}^*$. Notemos que, para todo $n \in \mathbb{N}$,

$$\begin{aligned} k + \frac{1}{m} < |X| &\implies k + \frac{1}{m} < |X - X_n| + |X_n| \\ &\implies k + \frac{1}{m} < |X - X_n| + k \\ &\implies \frac{1}{m} < |X - X_n|, \end{aligned}$$

por lo tanto
$$\left\{|X| > k + \frac{1}{m}\right\} \subseteq \left\{|X - X_n| > \frac{1}{m}\right\}.$$
Así,
$$P\left(|X| > k + \frac{1}{m}\right) \leq P\left(|X - X_n| > \frac{1}{m}\right).$$
Además, gracias a que $X_n \xrightarrow{p} X$, tenemos que
$$P\left(|X - X_n| > \frac{1}{m}\right) \to 0,$$
cuando $n \to \infty$. Por lo tanto,
$$P\left(|X| > k + \frac{1}{m}\right) = 0.$$
Por otro lado, ya que
$$\{|X| > k\} = \bigcup_{m \in \mathbb{N}} \left\{|X| > k + \frac{1}{m}\right\},$$
tenemos que
$$\begin{aligned}
P(\{|X| > k\}) &= P\left(\bigcup_{m \in \mathbb{N}} \left\{|X| > k + \frac{1}{m}\right\}\right) \\
&\leq \sum_{m \in \mathbb{N}} P\left(\left\{|X| > k + \frac{1}{m}\right\}\right) \\
&= 0,
\end{aligned}$$
por lo tanto, $P(\{|X| \leq k\}) = 1$; es decir, $|X| \leq k$ casi seguramente.

Sea $\varepsilon > 0$. Tenemos que
$$P\left(|X - X_n| > \frac{\varepsilon}{3}\right) \to 0,$$
cuando $n \to \infty$, por lo tanto, existe $N \in \mathbb{N}$ tal que, si $n \geq N$, entonces
$$P\left(|X - X_n| > \frac{\varepsilon}{3}\right) < \frac{\varepsilon}{3k}.$$
Además,
$$\begin{aligned}
\mathbf{E}\left(|X_n - X|\right) &= \int_{\{|X_n - X| > \frac{\varepsilon}{3}\}} |X_n - X|\, dP + \int_{\{|X_n - X| \leq \frac{\varepsilon}{3}\}} |X_n - X|\, dP \\
&\leq \int_{\{|X_n - X| > \frac{\varepsilon}{3}\}} |X_n - X|\, dP + \int_{\{|X_n - X| \leq \frac{\varepsilon}{3}\}} \frac{\varepsilon}{3}\, dP
\end{aligned}$$

$$
\begin{aligned}
&\leq \int_{\{|X_n - X| > \frac{\varepsilon}{3}\}} |X_n - X| \, dP + \mathbf{E}\left(\frac{\varepsilon}{3}\right) \\
&\leq \int_{\{|X_n - X| > \frac{\varepsilon}{3}\}} |X_n - X| \, dP + \frac{\varepsilon}{3} \\
&\leq \int_{\{|X_n - X| > \frac{\varepsilon}{3}\}} |X_n| \, dP + \int_{\{|X_n - X| > \frac{\varepsilon}{3}\}} |X| \, dP + \frac{\varepsilon}{3} \\
&\leq \int_{\{|X_n - X| > \frac{\varepsilon}{3}\}} k \, dP + \int_{\{|X_n - X| > \frac{\varepsilon}{3}\}} k \, dP + \frac{\varepsilon}{3} \\
&= 2kP\left(|X_n - X| > \frac{\varepsilon}{3}\right) + \frac{\varepsilon}{3} \\
&< 2k\frac{\varepsilon}{3k} + \frac{\varepsilon}{3} \\
&= \varepsilon,
\end{aligned}
$$

por lo tanto

$$\mathbf{E}\left(|X_n - X|\right) \to 0. \qquad \square$$

Antes de pasar a dar el resultado más importante de esta sección, probemos el siguiente lema:

LEMA 4.11. *Sea $k \in [0, \infty)$. Definamos, para cada $k \in [0, \infty)$ la función $\varphi_k \colon \mathbb{R} \to [-k, k]$ por*

$$\varphi_k(x) = \begin{cases} k & \text{si } x > k \\ x & \text{si } |x| \leq k \\ -k & \text{si } x < -k. \end{cases}$$

Se tiene que, φ_k es Lipschitz continua.

Demostración. Vamos a probar que, para todo $x, y \in \mathbb{R}$, se tiene que

$$|\varphi_k(x) - \varphi_k(y)| \leq |x - y|.$$

Sean $x, y \in \mathbb{R}$. Teniendo en cuenta que φ_k está definida por partes, analizaremos los siguientes casos:

- Si $x > k$, $y > k$, entonces

$$|\varphi_k(x) - \varphi_k(y)| = |k - k| = 0 \leq |x - y|.,$$

- Si $x > k$ y $y < -k$, entonces

$$|\varphi_k(x) - \varphi_k(y)| = |k - (-k)| \leq |x - y|,$$

- Si $x > k$ y $-k \leq y \leq k$, entonces

$$|\varphi_k(x) - \varphi_k(y)| = |k - y| \leq |x - y|.$$

Los casos en los que $x < -k$ y $-k \leq x \leq k$ se tratan de manera similar a los ya mostrados, así que los omitiremos. $\square$

TEOREMA 4.12. *Sean $(X_n)_{n\in\mathbb{N}}$ una sucesión en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$ y $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$. Se tiene que X_n converge a X en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$ si y sólo si se cumplen las siguientes condiciones:*

a) X_n converge a X en probabilidad,

b) $(X_n)_{n\in\mathbb{N}}$ es uniformemente integrable.

Demostración. Supongamos que a) y b) se cumplen y tomemos $\varepsilon > 0$. Gracias a la integrabilidad uniforme de $(X_n)_{n\in\mathbb{N}}$, existe $k_1 \in [0, \infty)$ tal que, para todo $n \in \mathbb{N}$,

$$\int_{\{|X_n| > k_1\}} |X_n| \, dP < \frac{\varepsilon}{3};$$

así mismo, usando el Corolario 4.9, existe $k_2 \in [0, \infty)$ tal que

$$\int_{\{|X| > k_2\}} |X| \, dP < \frac{\varepsilon}{3}.$$

Tomando $k = \text{máx}\{k_1, k_2\}$, tenemos que, para todo $n \in \mathbb{N}$,

$$\begin{aligned}
\mathbf{E}\left(|\varphi_k(X_n) - X_n|\right) &= \int_{\{\varphi_k(X_n) - X_n > 0\}} |\varphi_k(X_n) - X_n| \; dP \\
&\quad + \int_{\{\varphi_k(X_n) - X_n < 0\}} |\varphi_k(X_n) - X_n| \; dP \\
&= \int_{\{\varphi_k(X_n) > X_n\}} \varphi_k(X_n) - X_n \, dP \\
&\quad + \int_{\{\varphi_k(X_n) < X_n\}} |X_n - \varphi_k(X_n)| \; dP \\
&= \int_{\{X_n < -k\}} (-k - X_n) \, dP + \int_{\{X_n > k\}} (X_n - k) \, dP \\
&= \int_{\{X_n < -k\}} (-X_n) \, dP + \int_{\{X_n > k\}} X_n \, dP \\
&\quad - k(P(X_n < -k) + P(X_n > k)) \\
&= \int_{\{X_n < -k\}} |X_n| \, dP + \int_{\{X_n > k\}} |X_n| \, dP - kP(|X_n| > k)
\end{aligned}$$

$$
\begin{aligned}
&= \int_{\{|X_n|>k\}} |X_n|\, dP - kP(|X_n| > k) \\
&\leq \int_{\{|X_n|>k\}} |X_n|\, dP \\
&\leq \int_{\{|X_n|>k_1\}} |X_n|\, dP \\
&< \frac{\varepsilon}{3};
\end{aligned}
$$

por lo tanto

$$\mathbf{E}\left(|\varphi_k(X_n) - X_n|\right) < \frac{\varepsilon}{3}.$$

De manera análoga, se tiene que

$$\mathbf{E}\left(|\varphi_k(X) - X|\right) < \frac{\varepsilon}{3}.$$

Ahora, gracias a que φ_k es Lipschitz continua, se sigue que, para todo $n \in \mathbb{N}$,

$$\{|\varphi_k(X_n) - \varphi_k(X)| > \varepsilon\} \subseteq \{|X_n - X| > \varepsilon\},$$

por lo tanto

$$P(|\varphi_k(X_n) - \varphi_k(X)| > \varepsilon) \leq P(|X_n - X| > \varepsilon).$$

Así, y sabiendo que X_n converge en probabilidad a X, se tiene que, si $n \to \infty$,

$$P(|\varphi_k(X_n) - \varphi_k(X)| > \varepsilon) \to 0;$$

es decir, que $\varphi_k(X_n)$ converge en probabilidad a $\varphi_k(X)$. Luego, gracias a que, para todo $n \in \mathbb{N}$ y todo $\omega \in \Omega$,

$$|\varphi_k(X_n(\omega)| \leq k,$$

y junto con el Teorema 4.12, tenemos que

$$\mathbf{E}\left(|\varphi_k(X_n) - \varphi_k(X)|\right) \to 0,$$

cuando $n \to \infty$. Luego existe $N \in \mathbb{N}$ tal que, si $n \geq N$,

$$\mathbf{E}\left(|\varphi_k(X_n) - \varphi_k(X)|\right) < \frac{\varepsilon}{3}.$$

Finalmente, tenemos que, si $n \geq N$,

$$
\begin{aligned}
\mathbf{E}\left(|X_n - X|\right) &\leq \mathbf{E}\left(|X_n - \varphi_k(X_n)|\right) + \mathbf{E}\left(|\varphi_k(X_n) - \varphi_k(X)|\right) + \mathbf{E}\left(|\varphi_k(X) - X|\right) \\
&< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3}
\end{aligned}
$$

$$= \varepsilon.$$

Es decir, X_n converge a X en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$.

Recíprocamente, supongamos que $X_n \to X$ en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$. Sea $\varepsilon > 0$. Tenemos que existe $N \in \mathbb{N}$ tal que, para todo $n \geq N$,

$$\mathbf{E}\left(|X_n - X|\right) < \frac{\varepsilon}{2}.$$

Por el Lema 4.8, para cada $n \in \{0, \ldots, N\}$, existe $\delta_n > 0$ tal que, para todo $F \in \mathcal{F}$

$$P(F) < \delta_n \implies \int_F |X_n| \, dP < \varepsilon;$$

además, existe $\delta' > 0$ tal que, para todo $F \in \mathcal{F}$

$$P(F) < \delta' \implies \int_F |X| \, dP < \frac{\varepsilon}{2}.$$

Tomando $\delta = \text{mín}\{\delta', \delta_0, \ldots, \delta_N\}$, tenemos que, para todo $F \in \mathcal{F}$

$$P(F) < \delta \quad \implies \quad \int_F |X_n| \, dP < \varepsilon \wedge \int_F |X| \, dP < \frac{\varepsilon}{2}.$$

Gracias a que $(X_n)_{n \in \mathbb{N}}$ es acotada en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$, tenemos que

$$\sup_{n \in \mathbb{N}} \mathbf{E}\left(|X_n|\right) < \infty,$$

por lo tanto, existe $k \in \mathbb{N}^*$ tal que

$$\frac{1}{k} \sup_{n \in \mathbb{N}} \mathbf{E}\left(|X_n|\right) < \delta.$$

Con esto, tenemos que, para todo $n \in \mathbb{N}$,

$$\begin{aligned} P(|X_n| > k) &= \int_{\{|X_n| > k\}} dP \\ &= \frac{1}{k} \int_{\{|X_n| > k\}} k \, dP \\ &< \frac{1}{k} \int_{\{|X_n| > k\}} |X_n| \, dP \\ &\leq \frac{1}{k} \mathbf{E}\left(|X_n|\right) \\ &\leq \frac{1}{k} \sup_{n \in \mathbb{N}} \mathbf{E}\left(|X_n|\right), \end{aligned}$$

por ende,

$$P(|X_n| > k) < \delta,$$

así, para $n \in \{0, \ldots, N\}$,

$$\int_{\{|X_n|>k\}} |X_n| \, dP < \varepsilon$$

En cambio, si $n \geq N$, tenemos que

$$\begin{aligned}
\int_{\{|X_n|>k\}} |X_n| \, dP &\leq \int_{\{|X_n|>k\}} |X_n - X| \, dP + \int_{\{|X_n|>k\}} |X| \, dP \\
&\leq \mathbf{E}\left(|X_n - X|\right) + \int_{\{|X_n|>k\}} |X| \, dP \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\
&= \varepsilon.
\end{aligned}$$

Luego, $(X_n)_{n\in\mathbb{N}}$ es uniformemente integrable.

Finalmente, notemos que

$$\begin{aligned}
P(|X_n - X| > \varepsilon) &= \int_{\{|X_n-X|>\varepsilon\}} dP \\
&= \frac{1}{\varepsilon} \int_{\{|X_n-X|>\varepsilon\}} \varepsilon \, dP \\
&< \frac{1}{\varepsilon} \int_{\{|X_n-X|>\varepsilon\}} |X_n - X| \, dP \\
&\leq \frac{1}{\varepsilon} \mathbf{E}\left(|X_n - X|\right),
\end{aligned}$$

por lo tanto,

$$P(|X_n - X| > \varepsilon) \to 0,$$

cuando $n \to \infty$. Es decir, X_n converge en probabilidad a X. □

En el Ejemplo 4, vimos que, para toda variable aleatoria integrable X, se verifica que el proceso

$$(M_n)_{n\in\mathbb{N}} := \left(\mathbf{E}\left(X \mid \mathcal{F}_n\right)\right)_{n\in\mathbb{N}}$$

es una martingala, y, como habíamos dicho, este ejemplo será muy útil al momento de probar el Teorema de densidad de Lebesgue, por lo tanto, veremos un par de propiedades interesantes de las que goza este proceso. La primera tiene que ver con integrabilidad uniforme:

PROPOSICIÓN 4.13. *Sea $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$. Se tiene que la familia*

$$\{\boldsymbol{E}(X \mid \mathcal{G}) : \mathcal{G} \text{ es una sub } \sigma\text{-álgebra de } \mathcal{F}\}$$

es uniformemente integrable.

Demostración. Sea $\varepsilon > 0$. Gracias al Lema 4.8, existe $\delta > 0$ tal que, para todo $F \in \mathcal{F}$,

$$P(F) < \delta \implies \int_F |X| \, dP < \varepsilon. \tag{4.10}$$

Tomemos $k \in \mathbb{N}^*$ tal que

$$\frac{1}{k} \mathbf{E}\left(|X|\right) < \delta.$$

Gracias a la Desigualdad de Jensen vista en el Teorema 3.6, tenemos que, si $\mathcal{G}$ es una sub σ-álgebra de $\mathcal{F}$, entonces

$$|\mathbf{E}\left(X \mid \mathcal{G}\right)| \leq \mathbf{E}\left(|X| \mid \mathcal{G}\right), \tag{4.11}$$

casi seguramente. Así,

$$\begin{aligned}
P(|\mathbf{E}\left(X \mid \mathcal{G}\right)| > k) &= \int_{\{|\mathbf{E}(X \mid \mathcal{G})| > k\}} dP \\
&= \frac{1}{k} \int_{\{|\mathbf{E}(X \mid \mathcal{G})| > k\}} k \, dP \\
&< \frac{1}{k} \int_{\{|\mathbf{E}(X \mid \mathcal{G})| > k\}} |\mathbf{E}\left(X \mid \mathcal{G}\right)| \, dP \\
&\leq \frac{1}{k} \int_{\{|\mathbf{E}(X \mid \mathcal{G})| > k\}} \mathbf{E}\left(|X| \mid \mathcal{G}\right) \, dP \\
&\leq \frac{1}{k} \mathbf{E}\left(\mathbf{E}\left(|X| \mid \mathcal{G}\right)\right) \\
&= \frac{1}{k} \mathbf{E}\left(|X|\right),
\end{aligned}$$

por lo tanto

$$P(|\mathbf{E}\left(X \mid \mathcal{G}\right)| > k) < \delta.$$

Con esto, y gracias a (4.10) y (4.11), tenemos que

$$\int_{\{|\mathbf{E}(X \mid \mathcal{G})| > k\}} |\mathbf{E}\left(X \mid \mathcal{G}\right)| \, dP \leq \int_{\{|\mathbf{E}(X \mid \mathcal{G})| > k\}} |X| \, dP < \varepsilon,$$

luego, la familia

$$\{\mathbf{E}\left(X \mid \mathcal{G}\right) : \mathcal{G} \text{ es una sub } \sigma\text{-álgebra de } \mathcal{F}\}$$

es uniformemente integrable. □

Gracias a este resultado, tenemos que la martingala $(M_n)_{n \in \mathbb{N}}$ es uniformemente integrable, por lo tanto, tenemos además que es convergente casi seguramente y también en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$. El siguiente resultado nos permite saber cuál es su límite:

TEOREMA 4.14 (Lévy). *Sea $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$. Se tiene que el proceso*

$$M = (M_n)_{n\in\mathbb{N}} := (\boldsymbol{E}(X \mid \mathcal{F}_n))_{n\in\mathbb{N}}$$

es una martingala uniformemente integrable y converge a $\boldsymbol{E}(X \mid \mathcal{F}_\infty)$, casi seguramente y también en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$.

Demostración. Sin pérdida de generalidad, supongamos que $X \geq 0$. Gracias a la proposición anterior, tenemos que M es uniformemente integrable y, por lo tanto, converge casi seguramente y también en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$; así, sólo nos resta probar que

$$M_\infty := \lim_{n\to\infty} M_n = \mathbf{E}\left(X \mid \mathcal{F}_\infty\right),$$

casi seguramente.

Sean $n \in \mathbb{N}$ y $F \in \mathcal{F}_n$. Gracias a la Propiedad de torre vista en el Teorema 3.6, tenemos que

$$\mathbf{E}\left(\mathbf{E}\left(X \mid \mathcal{F}_\infty\right) \mid \mathbb{F}_n\right) = \mathbf{E}\left(X \mid \mathcal{F}_n\right),$$

casi seguramente, de donde se sigue que

$$\int_F \mathbf{E}\left(X \mid \mathcal{F}_\infty\right)\, dP = \int_F M_n\, dP. \tag{4.12}$$

Supongamos que $r \geq n$. Ya que M es una martingala, tenemos que

$$\mathbf{E}\left(M_r \mid \mathcal{F}_n\right) = M_n,$$

casi seguramente, por lo tanto,

$$\int_F M_r\, dP = \int_F M_n\, dP. \tag{4.13}$$

Por otra parte, tenemos que

$$\left| \int_F M_r\, dP - \int_F M_\infty\, dP \right| \leq \int_F |M_r - M_\infty|\, dP \leq \mathbf{E}\left(|M_r - M_\infty|\right). \tag{4.14}$$

Tomando el límite cuando $r \to \infty$ en (4.13) y (4.14), y gracias a que M_n converge a M_∞ en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$, tenemos que

$$\lim_{r\to\infty} \int_F M_r\, dP = \int_F M_n\, dP = \int_F M_\infty\, dP. \tag{4.15}$$

De (4.12) y (4.15), se sigue que

$$\int_F \mathbf{E}\left(X \mid \mathcal{F}_\infty\right)\, dP = \int_F M_\infty\, dP.$$

Con esto, tenemos que las medidas

$$\begin{aligned} Q_1 \colon \mathcal{F}_\infty &\longrightarrow [0, \infty) \\ F &\longmapsto Q_1(F) = \int_F \mathbf{E}\left(X \mid \mathcal{F}_\infty\right) dP \end{aligned}$$

y

$$\begin{aligned} Q_1 \colon \mathcal{F}_\infty &\longrightarrow [0, \infty) \\ F &\longmapsto Q_1(F) = \int_F M_\infty \, dP. \end{aligned}$$

coinciden en el π-sistema $\bigcup_{n \in \mathbb{N}} \mathcal{F}_n$; más aún, gracias al Lema 2.7, coinciden en $\mathcal{F}_\infty$. Es decir que, para todo $F \in \mathcal{F}_\infty$,

$$\int_F M_\infty \, dP = \int_F \mathbf{E}\left(X \mid \mathcal{F}_\infty\right) dP,$$

y gracias a que M_∞ es $\mathcal{F}_\infty$-medible, tenemos que

$$M_\infty = \mathbf{E}\left(X \mid \mathcal{F}_\infty\right),$$

casi seguramente. □

4.4. El Teorema de Radon-Nikodym

Las herramientas y resultados que hemos construido a lo largo de este capítulo nos permitirá mostrar que existe cierta relación entre el Teorema de Radon-Nikodym y el Teorema de convergencia de submartingalas. De hecho, como habíamos indicado al inicio de este capítulo, daremos una demostración constructiva del Teorema de Radon-Nikodym; así, probaremos que este último es, en realidad, un teorema dentro de la Teoría de probabilidades. Para ello necesitaremos de la siguiente definición:

DEFINICIÓN 4.7. *Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad. Se dice que una medida finita Q, definida sobre sobre $(\Omega, \mathcal{F})$, es absolutamente continua con respecto a P si para todo $F \in \mathcal{F}$ se tiene que*

$$P(F) = 0 \implies Q(F) = 0.$$

La siguiente proposición le dará sentido a la palabra «continuidad» en la definición anterior:

PROPOSICIÓN 4.15. *Sea Q una medida absolutamente continua con respecto a P.*

Entonces, para todo $\varepsilon > 0$, existe $\delta > 0$ tal que, para todo $F \in \mathcal{F}$,

$$P(F) < \delta \implies Q(F) < \varepsilon.$$

Demostración. Por reducción al absurdo, supongamos que existe $\varepsilon > 0$ tal que para todo $\delta > 0$ existe $F \in \mathcal{F}$ que cumple

$$P(F) < \delta \qquad \text{y} \qquad Q(F) \geq \varepsilon.$$

Así, para cada $n \in \mathbb{N}$ existe $F_n \in \mathcal{F}$ tal que

$$P(F_n) < \frac{1}{2^n} \qquad \text{y} \qquad Q(F_n) \geq \varepsilon,$$

por lo tanto

$$\sum_{n \in \mathbb{N}} P(F_n) < 2 < \infty.$$

Gracias al Lema de Borel-Cantelli, se sigue que

$$P(\operatorname{lím\,sup} F_n) = 0,$$

y, ya que Q es absolutamente continua con respecto a P, tenemos que

$$Q(\operatorname{lím\,sup} F_n) = 0.$$

Por otra parte, notemos que

$$\operatorname{lím\,sup} Q(F_n) \geq \varepsilon;$$

así que, por el Lema de Fatou invertido para eventos, tenemos que

$$0 = Q(\operatorname{lím\,sup} F_n) \geq \operatorname{lím\,sup} Q(F_n) \geq \varepsilon,$$

lo cual es contradictorio. $\square$

Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad. Si Q es una medida absolutamente continua con respecto a P y X es una variable aleatoria integrable y no negativa tal que, para todo $F \in \mathcal{F}$,

$$Q(F) = \int_F X \, dP,$$

diremos que X es la *derivada de Radon-Nikodym* de Q en $(\Omega, \mathcal{F})$, y la denotaremos por

$$X = \frac{dQ}{dP},$$

en $(\Omega, \mathcal{F})$.

El Teorema de Radon-Nikodym garantiza la existencia de la derivada de Radon-Nikodym y también su unicidad casi segura. En su versión general, el Teorema de Radon-Nikodym considera un espacio de probabilidad cualquiera $(\Omega, \mathcal{F}, P)$ y una medida absolutamente continua con respecto a P, pero para fines prácticos, probaremos primero una versión más débil en la que supondremos que $\mathcal{F}$ es una σ-álgebra separable en el sentido de que existe una sucesión de eventos $(F_n)_{n\in\mathbb{N}}$ tal que

$$\mathcal{F} = \sigma(\{F_n\}_{n\in\mathbb{N}}).$$

La σ-álgebra de Borel en $(0,1)$ es un ejemplo de una σálgebra separable. En efecto, como podemos ver en [12], se tiene que

$$\mathcal{B}(0,1) = \sigma(\{(a,b) \subseteq (0,1) : a, b \in \mathbb{Q}\}).$$

LEMA 4.16. *Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad. Sea Q una medida absolutamente continua con respecto a P. Si $\mathcal{F}$ es separable, existe una variable aleatoria no negativa $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ tal que*

$$X = \frac{dQ}{dP},$$

en $(\Omega, \mathcal{F})$. La variable X es única, casi seguramente.

Demostración. Vamos a construir una martingala convergente, cuyo límite será la variable que buscamos. Para ello, trabajaremos con sub σ-álgebras de $\mathcal{F}$ que serán definidas a continuación.

Por la separabilidad de $\mathcal{F}$, existe una sucesión $(F_n)_{n\in\mathbb{N}}$ de eventos tal que $\mathcal{F} = \sigma(\{F_n\}_{n\in\mathbb{N}})$. Con esto, definimos, para cada $n \in \mathbb{N}$, la σ-álgebra

$$\mathcal{F}_n = \sigma(\{F_0, \ldots F_n\}),$$

y con ello formamos la filtración $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$.

Sea $n \in \mathbb{N}$. Tenemos que $\mathcal{F}_n$ está formada por uniones de intersecciones finitas de los elementos de $\{F_0, \ldots F_n\}$ o de sus complementos; es decir que los elementos de $\mathcal{F}_n$ son uniones de eventos de la forma

$$H_0 \cap \cdots \cap H_n, \tag{4.16}$$

donde, para cada $i \in \{0, \ldots, n\}$, tenemos que H_i es F_i o F_i^c; estos eventos son disjuntos dos a dos, y, de hecho, gracias a que $\{F_0, \ldots F_n\}$ es una colección finita, podemos formar una cantidad finita de ellos, digamos $r(n)$. Si denotamos por

$$A := \{A_{n,1}, A_{n,2}, \dots A_{n,r(n)}\}$$

a la colección de eventos de la forma (4.16), tenemos que

$$\mathcal{F}_n = \sigma(A). \tag{4.17}$$

Más aún, por la manera en la que están definidos los elementos de A, estos son los eventos más pequeños que podemos construir en $\mathcal{F}_n$, salvo por el conjunto vacío; es decir, A es la colección de átomos de la σ-álgebra $\mathcal{F}$. La relación (4.17) nos permitirá definir la martingala que buscamos.

De (4.17), tenemos que, para cada $F \in \mathcal{F}_n$ existe $I_F \subseteq \{0, \dots, r(n)\}$ tal que

$$F = \bigcup_{i \in I_F} A_{n,i}.$$

Si definimos $I_F^* := \{i \in I_F : P(A_{n,i}) = 0\}$, tenemos además que

$$F = \left(\bigcup_{i \in I_F \setminus I_F^*} A_{n,i} \right) \cup \left(\bigcup_{i \in I_F^*} A_{n,i} \right).$$

Definamos la función

$$\begin{aligned} X_n \colon \; \Omega &\longrightarrow [0, \infty) \\ \omega &\longmapsto X_n(\omega) = \sum_{i \in I_\Omega \setminus I_\Omega^*} \mathbb{1}_{A_{n,i}}(\omega) \frac{Q(A_{n,i})}{P(A_{n,i})} + \sum_{i \in I_\Omega^*} \mathbb{1}_{A_{n,i}}(\omega). \end{aligned}$$

Por la manera en la que está definida, tenemos que X_n es $\mathcal{F}_n$-medible. Además, tenemos que

$$\begin{aligned} \mathbf{E}\left(X_n\right) &= \mathbf{E}\left(\sum_{i \in I_\Omega \setminus I_\Omega^*} \mathbb{1}_{A_{n,i}} \frac{Q(A_{n,i})}{P(A_{n,i})} + \sum_{i \in I_\Omega^*} \mathbb{1}_{A_{n,i}} \right) \\ &= \sum_{i \in I_\Omega \setminus I_\Omega^*} \mathbf{E}\left(\mathbb{1}_{A_{n,i}} \frac{Q(A_{n,i})}{P(A_{n,i})} \right) + \sum_{i \in I_\Omega^*} \mathbf{E}\left(\mathbb{1}_{A_{n,i}} \right) \\ &= \sum_{i \in I_\Omega \setminus I_\Omega^*} \frac{Q(A_{n,i})}{P(A_{n,i})} \mathbf{E}\left(\mathbb{1}_{A_{n,i}} \right) + \sum_{i \in I_\Omega^*} P(A_{n,i}). \end{aligned} \tag{4.18}$$

Recordemos que$Q$ es absolutamente continua con respecto a P, por lo tanto, para todo $i \in I^*_\Omega$, tenemos que $P(A_{n,i}) = Q(A_{n,i}) = 0$; así, de (4.18), se tiene que

$$\begin{aligned}
\mathbf{E}\left(X_n\right) &= \sum_{i \in I_\Omega \setminus I^*_\Omega} \frac{Q(A_{n,i})}{P(A_{n,i})} P(A_{n,i}) + \sum_{i \in I^*_\Omega} Q(A_{n,i}) \\
&= \sum_{i \in I_\Omega \setminus I^*_\Omega} Q(A_{n,i}) + \sum_{i \in I^*_\Omega} Q(A_{n,i}) \\
&= \sum_{i \in I_\Omega} Q(A_{n,i}) \\
&= Q\left(\bigcup_{i \in I_\Omega} A_{n,i} \right) \\
&= Q(\Omega),
\end{aligned}$$

con lo cual obtenemos que

$$\mathbf{E}\left(X_n\right) = \int_\omega X_n \, dP = Q(\Omega) < \infty;$$

es decir, $X_n \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$. De la misma manera, podemos probar que

$$\int_F X_n \, dP = Q(F), \tag{4.19}$$

para todo $F \in \mathcal{F}_n$. Luego, para todo $n \in \mathbb{N}$, $X_n = \dfrac{dQ}{dP}$ en $(\Omega, \mathcal{F}_n)$.

Probemos ahora que $(X_n)_{n\in\mathbb{N}}$ es una martingala. Sean $n \in \mathbb{N}$ y $F \in \mathcal{F}_n \subseteq \mathcal{F}_{n+1}$. Gracias a que X_n y X_{n+1} son las derivadas de Radon-Nikodym de Q en $(\Omega, \mathcal{F}_n)$ y $(\Omega, \mathcal{F}_{n+1})$, respectivamente, tenemos que

$$\int_F X_n \, dP = Q(F) = \int_F X_{n+1} \, dP.$$

Por otra parte, por definición de esperanza condicional, tenemos que

$$\int_F \mathbf{E}\left(X_{n+1} \mid \mathcal{F}_n\right) \, dP = \int_F X_{n+1} \, dP;$$

así,

$$\int_F \mathbf{E}\left(X_{n+1} \mid \mathcal{F}_n\right) \, dP = \int_F X_n \, dP.$$

Luego, gracias a que X_n es $\mathcal{F}_n$-medible, se sigue que

$$X_n = \mathbf{E}\left(X_{n+1} \mid \mathcal{F}_n\right),$$

casi seguramente. Con esto, obtenemos que $(X_n)_{n\in\mathbb{N}}$ es una martingala con respecto a la filtración $(\mathcal{F}_n)_{n\in\mathbb{N}}$; y al ser no negativa, por el Corolario 4.7, se tiene que el límite

$$X_\infty := \lim_{n\to\infty} X_n$$

existe casi seguramente y además $X_\infty \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$.

La variable X_∞ es nuestra candidata a ser $\dfrac{dQ}{dP}$ en $(\Omega, \mathcal{F})$. Al ser límite de variables aleatorias no negativas, es también no negativa; y, como ya vimos, es integrable. Por lo tanto, nos resta probar que, para todo $F \in \mathcal{F}$,

$$\int_F X\, dP = Q(F).$$

Para ello, probaremos que X_n converge a X en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$.

Sea $\varepsilon > 0$. Gracias a que Q es absolutamente continua con respecto a P, de la Proposición 4.15, tenemos que existe $\delta > 0$ tal que, para todo $F \in \mathcal{F}$,

$$P(F) < \delta \implies Q(F) < \varepsilon. \tag{4.20}$$

Tomemos $k \in \mathbb{N}^*$ tal que

$$\frac{1}{k} Q(\Omega) < \delta.$$

Notemos que, para todo $n \in \mathbb{N}$,

$$\begin{aligned} P(X_n > k) &= \int_{\{X_n > k\}} dP \\ &= \frac{1}{k} \int_{\{X_n > k\}} k\, dP \\ &< \frac{1}{k} \int_{\{X_n > k\}} X_n\, dP \\ &\leq \frac{1}{k} \mathbf{E}\left(X_n\right), \end{aligned}$$

de donde, gracias a (4.19), tenemos que

$$P(X_n > k) < \frac{1}{k} Q(\Omega) < \delta;$$

así, de (4.20), se sigue que

$$\int_{\{X_n > k\}} X_n\, dP = Q(X_n > k) < \varepsilon.$$

Luego, X es uniformemente integrable. Ahora, gracias a que X_n converge casi seguramente a X_∞, tenemos que X_n converge a X en probabilidad, por lo tanto, el

Teorema 4.12, implica que X_n converge a X en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$; es decir,

$$\lim_{n\to\infty} \mathbf{E}\left(|X_n - X_\infty|\right) = 0. \tag{4.21}$$

De esta última igualdad, se puede ver que si $F \in \mathcal{F}$, entonces

$$\lim_{n\to\infty} \mathbf{E}\left(|X_n \mathbb{1}_F - X_\infty \mathbb{1}_F|\right) = 0, \tag{4.22}$$

lo que implica que

$$\lim_{n\to\infty} \int_F X_n \, dP = \int_F X_\infty \, dP. \tag{4.23}$$

Teniendo este resultado, lo único que nos resta probar es que

$$\lim_{n\to\infty} \int_F X_n \, dP = Q(F),$$

y lo haremos utilizando el Lema 2.7, que nos garantiza la unicidad de extensiones de medidas finitas.

Si $F \in \bigcup_{n\in\mathbb{N}} \mathcal{F}_n$, existe $n_1 \in \mathbb{N}$ tal que, para todo $n \geq n_1$, $F \in \mathcal{F}_n$ y por lo tanto, gracias a que, $X_n = \dfrac{dQ}{dP}$ en $(\Omega, \mathcal{F}_n)$, tenemos que

$$\int_F X_n \, dP = Q(F).$$

Así,

$$\lim_{n\to\infty} \int_F X_n \, dP = Q(F).$$

Luego, para todo $F \in \bigcup_{n\in\mathbb{N}} \mathcal{F}_n$

$$\int_F X_\infty \, dP = Q(F).$$

Esta última igualdad, nos indica que las medidas

$$F \mapsto Q(F) \qquad \text{y} \qquad F \mapsto \int_F X_\infty \, dP$$

coinciden en $\bigcup_{n\in\mathbb{N}} \mathcal{F}_n$. Con esto, y gracias a que $\bigcup_{n\in\mathbb{N}} \mathcal{F}_n$ es un π-sistema que genera a $\mathcal{F}$, del Lema 2.7, se sigue que

$$Q(F) = \int_F X_\infty \, dP,$$

para todo $F \in \mathcal{F}$. Luego $X_\infty = \dfrac{dQ}{dP}$ en $(\Omega, \mathcal{F})$.

Finalmente, supongamos que existe otra variable aleatoria no negativa

$Y \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ tal que

$$\int_F Y\, dP = Q(F).$$

Definamos

$$F := \{\omega \in \Omega : X_\infty(\omega) < Y(\omega)\}$$

y supongamos que $P(F) > 0$. Tenemos que

$$Q(F) = \int_F X_\infty\, dP > \int_F Y\, dP = Q(F),$$

lo cual es absurdo. Por lo tanto $\dfrac{dQ}{dP}$ es única casi seguramente. $\square$

Con este resultado, podemos garantizar la existencia de la derivada de la derivada de Radon-Nikodym en espacios de probabilidad dotados de σ-álgebras separables; sin embargo, este no es el único caso que nos interesa, pues existen σ-álgebras de interés que no son separables; de hecho, como se puede ver en [12], la σ-álgebra de Lebesgue en $(0, 1)$ no es separable.

A partir de ahora, consideraremos un espacio de probabilidad cualquiera $(\Omega, \mathcal{F}, P)$. Vamos a probar que el resultado del Lema 4.16 también se cumple si eliminamos la hipótesis de separabilidad sobre la σ-álgebra $\mathcal{F}$. Para ello daremos una noción de convergencia para familias de variables aleatorias, así como una propiedad muy similar a la propiedad de Cauchy definida sucesiones.

Sea Q una medida absolutamente continua con respecto a P. Definamos

$$\mathrm{Sep}(\Omega, \mathcal{F}) := \{\mathcal{G} \subseteq \mathcal{F} : \mathcal{G} \text{ es una } \sigma\text{-álgebra separable}\}.$$

Gracias al Lema 4.16, tenemos que, para cada $\mathcal{G} \in \mathrm{Sep}(\Omega, \mathcal{F})$, existe una variable aleatoria $X_\mathcal{G} \in \mathcal{L}^1(\Omega, \mathcal{G}, P)$, tal que

$$X_\mathcal{G} = \frac{dQ}{dP},$$

en $(\Omega, \mathcal{G})$. Diremos que la familia de variables aleatorias $(X_\mathcal{G})_{\mathcal{G} \in \mathrm{Sep}(\Omega, \mathcal{F})}$ es *convergente* en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$ si existe $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ tal que, para todo $\varepsilon > 0$, existe $\mathcal{K} \in \mathrm{Sep}(\Omega, \mathcal{F})$ tal que para $\mathcal{G} \in \mathrm{Sep}(\Omega, \mathcal{F})$ se verifica que

$$\mathcal{K} \subseteq \mathcal{G} \implies \|X_\mathcal{G} - X\|_1 < \varepsilon.$$

Así mismo, diremos que la familia $(X_\mathcal{G})_{\mathcal{G} \in \mathrm{Sep}(\Omega, \mathcal{F})}$ es *de Cauchy* en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$ si, para todo $\varepsilon > 0$, existe $\mathcal{K} \in \mathrm{Sep}(\Omega, \mathcal{F})$ tal que, para $\mathcal{G}_1, \mathcal{G}_2 \in \mathrm{Sep}(\Omega, \mathcal{F})$, se verifica que

$$\mathcal{K} \subseteq \mathcal{G}_1 \wedge \mathcal{K} \subseteq \mathcal{G}_2 \implies \|X_{\mathcal{G}_1} - X_{\mathcal{G}_2}\|_1 < \varepsilon.$$

Esta noción de convergencia para una familia indexada por σ-álgebras es la que corresponde a la de convergencia de Moore-Smith para redes considerando la clase parcialmente ordenada $(\mathrm{Sep}(\Omega, \mathcal{F}), \subseteq)$ (véase [5]).

Gracias a que el espacio $\mathcal{L}^1(\Omega, \mathcal{F}, P)$ es completo, esperaríamos que, si la familia $(X_{\mathcal{G}})_{\mathcal{G} \in \mathrm{Sep}(\Omega, \mathcal{F})}$ es de Cauchy, entonces sea también convergente, en los sentidos antes expuestos. En efecto, tenemos el siguiente lema:

LEMA 4.17. *Si $(X_{\mathcal{G}})_{\mathcal{G} \in \mathrm{Sep}(\Omega, \mathcal{F})}$ es de Cauchy en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$, entonces es convergente en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$.*

Demostración. Supongamos que $(X_{\mathcal{G}})_{\mathcal{G} \in \mathrm{Sep}(\Omega, \mathcal{F})}$ es de Cauchy en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$. Tenemos que, para cada $n \in \mathbb{N}^*$, existe $\mathcal{K}_n \in \mathrm{Sep}(\Omega, \mathcal{F})$ tal que para $\mathcal{G}_1, \mathcal{G}_2 \in \mathrm{Sep}(\Omega, \mathcal{F})$ se verifica que

$$\mathcal{K}_n \subseteq \mathcal{G}_1 \wedge \mathcal{K}_n \subseteq \mathcal{G}_2 \implies \|X_{\mathcal{G}_1} - X_{\mathcal{G}_2}\|_1 < \frac{1}{2n}.$$

Si definimos $\mathcal{H}_n := \sigma(\mathcal{K}_1, \ldots, \mathcal{K}_n) \in \mathrm{Sep}(\Omega, \mathcal{F})$, entonces la sucesión $(X_{H_n})_{n \in \mathbb{N}^*}$ es de Cauchy en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$. En efecto, gracias a que, para todo $n \in \mathbb{N}^*$ y todo $m \geq n$

$$\mathcal{K}_n \subseteq \mathcal{H}_n \subseteq \mathcal{H}_m,$$

se sigue que

$$\|X_{\mathcal{H}_n} - X_{\mathcal{H}_m}\|_1 < \frac{1}{2n} \to 0,$$

cuando $n, m \to \infty$. Gracias a que $\mathcal{L}^1(\Omega, \mathcal{F}, P)$ es completo, existe $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ tal que

$$\|X_{\mathcal{H}_n} - X\|_1 \to 0.$$

cuando $n \to \infty$.

Probemos que la familia $(X_{\mathcal{G}})_{\mathcal{G} \in \mathrm{Sep}(\Omega, \mathcal{F})}$ es convergente en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$ y que su límite es X.

Sea $\varepsilon > 0$ y fijemos $n_1 \in \mathbb{N}^*$ tal que $\dfrac{1}{n_1} < \varepsilon$. Tomemos $\mathcal{G} \in \mathrm{Sep}(\Omega, \mathcal{F})$ tal que $\mathcal{H}_{n_1} \subseteq \mathcal{G}$, entonces

$$\|X_{\mathcal{H}_{n_1}} - X_{\mathcal{G}}\|_1 < \frac{1}{2n_1},$$

y, de hecho, si $\mathcal{H}_n \subseteq \mathcal{G}$, entonces, para todo $n \geq n_1$,

$$\|X_{\mathcal{H}_n} - X_{\mathcal{G}}\|_1 < \frac{1}{2n_1}.$$

Por otra parte, existe $N_1 \in \mathbb{N}$ tal que si $n \geq N_1$, entonces

$$\|X_{\mathcal{H}_n} - X\|_1 < \frac{1}{2n_1}.$$

Así, si $N = \max\{n_1, N_1\}$, tenemos que, para todo $n \geq N$, y todo $\mathcal{G} \in \operatorname{Sep}(\Omega, \mathcal{F})$,

$$\mathcal{H}_n \subseteq \mathcal{G} \implies \|X_{\mathcal{H}_n} - X_{\mathcal{G}}\|_1 < \frac{1}{2n_1} \quad \wedge \quad \|X_{\mathcal{H}_n} - X\|_1 < \frac{1}{2n_1}.$$

En particular, si $\mathcal{G} \in \operatorname{Sep}(\Omega, \mathcal{F})$ es tal que $\mathcal{H}_N \subseteq \mathcal{G}$, tenemos que

$$\|X_{\mathcal{H}_N} - X_{\mathcal{G}}\|_1 < \frac{1}{2n_1} \qquad \text{y} \qquad \|X_{\mathcal{H}_N} - X\|_1 < \frac{1}{2n_1}.$$

Por lo tanto

$$\|X - X_{\mathcal{G}}\|_1 \leq \|X - X_{\mathcal{H}_N}\|_1 + \|X_{\mathcal{H}_N} - X_{\mathcal{G}}\|_1 < \frac{1}{n_1} < \varepsilon,$$

es decir,

$$X_{\mathcal{G}} \to X,$$

en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$. La no negatividad de X se sigue inmediatamente de la no negatividad de $(X_{H_n})_{n \in \mathbb{N}^*}$. $\square$

Teniendo este resultado, ya sólo nos resta probar que la familia $(X_{\mathcal{G}})_{\mathcal{G} \in \operatorname{Sep}(\Omega, \mathcal{F})}$ es de Cauchy en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$. En efecto, se tiene el siguiente lema:

LEMA 4.18. *Se tiene que $(X_{\mathcal{G}})_{\mathcal{G} \in Sep(\Omega, \mathcal{F})}$ es de Cauchy en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$.*

Demostración. Supongamos que no es así. Es decir, supongamos que existe $\varepsilon > 0$ tal que, para todo $\mathcal{K} \in \operatorname{Sep}(\Omega, \mathcal{F})$, existen $\mathcal{G}_1, \mathcal{G}_2 \in \operatorname{Sep}(\Omega, \mathcal{F})$ de modo que

$$(\mathcal{K} \subseteq \mathcal{G}_1 \wedge \mathcal{K} \subseteq \mathcal{G}_2) \wedge \|X_{\mathcal{G}_1} - X_{\mathcal{G}_2}\|_1 \geq \varepsilon.$$

Sea $\mathcal{K}_0 \in \operatorname{Sep}(\Omega, \mathcal{F})$, existen $\mathcal{G}_{0,1}, \mathcal{G}_{0,2} \in \operatorname{Sep}(\Omega, \mathcal{F})$ tales que $\mathcal{K}_0 \subseteq \mathcal{G}_{0,1} \subseteq \mathcal{G}_{0,2}$ y cumplen que

$$\|X_{\mathcal{G}_{0,1}} - X_{\mathcal{G}_{0,2}}\|_1 \geq \varepsilon.$$

Definamos $\mathcal{K}_1 := \mathcal{G}_{0,1}$ y $\mathcal{K}_2 := \mathcal{G}_{0,2}$. Tenemos que

$$\|X_{\mathcal{K}_1} - X_{\mathcal{K}_2}\|_1 \geq \varepsilon.$$

Ahora, para $\mathcal{K}_2 \in \text{Sep}(\Omega, \mathcal{F})$, existen $\mathcal{G}_{2,1}, \mathcal{G}_{2,2} \in \text{Sep}(\Omega, \mathcal{F})$ tales que $\mathcal{K}_2 \subseteq \mathcal{G}_{2,1} \subseteq \mathcal{G}_{2,2}$ y además

$$\|X_{\mathcal{G}_{2,1}} - X_{\mathcal{G}_{2,2}}\|_1 \geq \varepsilon.$$

Definamos $\mathcal{K}_3 := \mathcal{G}_{2,1}$ y $\mathcal{K}_4 := \mathcal{G}_{2,2}$. Así, tenemos que

$$\|X_{\mathcal{K}_3} - X_{\mathcal{K}_4}\|_1 \geq \varepsilon.$$

Procediendo inductivamente, siguiendo este procedimiento, habremos construido una sucesión creciente de σ-álgebras separables $(\mathcal{K}_n)_{n\in\mathbb{N}^*}$ y una sucesión de variables aleatorias $(X_{\mathcal{K}_n})_{n\in\mathbb{N}^*}$ tal que, para todo $n \in \mathbb{N}^*$,

$$\|X_{\mathcal{K}_{2n-1}} - X_{\mathcal{K}_{2n}}\|_1 \geq \varepsilon. \tag{4.24}$$

Sean $n \in \mathbb{N}^*$ y $F \in \mathcal{K}_{n-1}$. Por definición de esperanza condicional, tenemos que

$$\int_F \mathbf{E}\left(X_{\mathcal{K}_n} \mid \mathcal{K}_{n-1}\right) dP = \int_F X_{\mathcal{K}_n} \, dP.$$

Por otra parte, ya que $\mathcal{K}_{n-1} \subseteq \mathcal{K}_n$, y gracias a que

$$X_{\mathcal{K}_{n-1}} = \frac{dQ}{dP} \qquad \text{y} \qquad X_{\mathcal{K}_n} = \frac{dQ}{dP}$$

en $(\Omega, \mathcal{K}_{n-1})$ y $(\Omega, \mathcal{K}_n)$, respectivamente tenemos que

$$\int_F X_{\mathcal{K}_n} \, dP = Q(F) = \int_F X_{\mathcal{K}_{n-1}} \, dP,$$

así

$$\int_F \mathbf{E}\left(X_{\mathcal{K}_n} \mid \mathcal{K}_{n-1}\right) dP = \int_F X_{\mathcal{K}_{n-1}} \, dP.$$

Puesto que $F \in \mathcal{K}_{n-1}$ es arbitrario y $X_{\mathcal{K}_{n-1}}$ es $\mathcal{K}_{n-1}$-medible, se sigue que

$$\mathbf{E}\left(X_{\mathcal{K}_n} \mid \mathcal{K}_{n-1}\right) = X_{\mathcal{K}_{n-1}},$$

casi seguramente. Con esto, tenemos que $(X_{\mathcal{K}_n})_{n\in\mathbb{N}^*}$ es una martingala. Además, procediendo como en el Lema 4.16, tenemos que $(X_{\mathcal{K}_n})_{n\in\mathbb{N}^*}$ es uniformemente integrable y, por ende, convergente en $\mathcal{L}^1(\Omega, \mathcal{F}, P)$, lo que contradice (4.24). □

Ya tenemos las herramientas necesarias para probar el teorema más importante de esta sección:

TEOREMA 4.19 (Radon-Nikodym). *Sean $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y Q una medida absolutamente continua con respecto a P. Existe una variable aleatoria* *no negativa* $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ *tal que*

$$X = \frac{dQ}{dP},$$

en $(\Omega, \mathcal{F})$*. La variable* X *es única casi seguramente.*

Demostración. Gracias al Lema 4.18, y al Lema 4.17, existe una variable aleatoria no negativa $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ tal que, si fijamos $\varepsilon > 0$, existe $\mathcal{K} \in \mathrm{Sep}(\Omega, \mathcal{F})$ para la cual

$$\mathcal{K} \subseteq \mathcal{G} \implies \|X_{\mathcal{G}} - X\|_1 < \varepsilon.$$

Sea $F \in \mathcal{F}$. Tenemos que $\mathcal{K} \subseteq \sigma(\{\mathcal{K}, F\}) \in \mathrm{Sep}(\Omega, \mathcal{F})$ y $F \in \sigma(\{\mathcal{K}, F\})$. Por lo tanto, gracias a que

$$X_{\sigma(\{\mathcal{K},F\})} = \frac{dQ}{dP},$$

en $(\Omega, \sigma(\{\mathcal{K}, F\})$, tenemos que

$$\begin{aligned}
\left| \int_F X \, dP - Q(F) \right| &= \left| \int_F X \, dP - \int_F X_{\sigma(\{\mathcal{K},F\})} \, dP \right| \\
&= \left| \int_F X - X_{\sigma(\{\mathcal{K},F\})} \, dP \right| \\
&\leq \int_F \left| X - X_{\sigma(\{\mathcal{K},F\})} \right| \, dP \\
&\leq \mathbf{E} \left(|X - X_{\sigma(\{\mathcal{K},F\})}| \right) \\
&= \|X - X_{\sigma(\{\mathcal{K},F\})}\|_1 \\
&< \varepsilon.
\end{aligned}$$

Luego, gracias a que ε y F son arbitrarios, se sigue que, para todo $F \in \mathcal{F}$,

$$\int_F X \, dP = Q(F).$$

Así,

$$X = \frac{dQ}{dP},$$

en $(\Omega, \mathcal{F})$. La unicidad de X se demuestra tal y como en el Lema 4.16. $\square$

Capítulo 5

El Teorema de densidad de Lebesgue

En este capítulo, mostraremos la relación que guardan el Teorema de diferenciación de Lebesgue y el Teorema de densidad de Lebesgue. Daremos además una demostración probabilista del Teorema de densidad de Lebesgue haciendo uso de una martingala adecuada y de los resultados obtenidos en el capítulo anterior.

5.1. El Teorema de diferenciación de Lebesgue

En 1910, Henri Lebesgue [10], considerando el espacio medido $(\mathbb{R}^n, \mathrm{Leb}, \lambda)$, probó el siguiente resultado conocido como el Teorema de diferenciación de Lebesgue:

TEOREMA 5.1. *Si $f\colon \mathbb{R}^n \to \mathbb{R}$ es integrable, entonces, para casi todo punto $x \in \mathbb{R}^n$, se tiene que*

$$\lim_{\delta\to 0} \frac{1}{\lambda(B(x,\delta))} \int_{B(x,\delta)} f(y)\, d\lambda(y) = f(x). \tag{5.1}$$

En [9], se prueba que, bajo las hipótesis del teorema anterior, para casi todo punto $x \in \mathbb{R}^n$,

$$\lim_{\delta\to 0} \frac{1}{\lambda(B(x,\delta))} \int_{B(x,\delta)} |f(y) - f(x)|\, d\lambda(y) = 0,$$

de donde se deduce fácilmente (5.1). De hecho, otra versión del Teorema de diferenciación de Lebesgue, y con las que trabajaremos en este capítulo, es la siguiente:

TEOREMA 5.2. *Si $f \in \mathcal{L}^1(\mathbb{R}^n, \mathrm{Leb}, \lambda)$, entonces, para casi todo punto $x \in (0,1)^n$, se tiene que*

$$\lim_{\delta\to 0} \frac{1}{\lambda(B(x,\delta))} \int_{B(x,\delta)} |f(s) - f(x)|\, d\lambda(s) = 0.$$

Notemos que el Teorema 5.1 es consecuencia del Teorema 5.2. En efecto, supongamos que, si $f \in \mathcal{L}^1(\mathbb{R}^n, \text{Leb}, \lambda)$, entonces, para casi todo punto $x \in (0,1)^n$, se tiene que

$$\lim_{\delta \to 0} \frac{1}{\lambda(B(x,\delta))} \int_{B(x,\delta)} |f(s) - f(x)| \, d\lambda(s) = 0.$$

Vamos a probar que si $g \in \mathcal{L}^1(\mathbb{R}^n, \text{Leb}, \lambda)$, entonces, para casi todo punto $x \in \mathbb{R}^n$, se tiene que

$$\lim_{\delta \to 0} \frac{1}{\lambda(B(x,\delta))} \int_{B(x,\delta)} |f(y) - f(x)| \, d\lambda(y) = 0.$$

Supongamos que $g \in \mathcal{L}^1(\mathbb{R}^n, \text{Leb}, \lambda)$ y sea $x \in \mathbb{R}^n$. Definamos

$$\begin{aligned} f\colon \mathbb{R}^n &\longrightarrow \mathbb{R} \\ s &\longmapsto g(h^{-1}(s)), \end{aligned}$$

donde $h\colon \mathbb{R}^n \to \mathbb{R}^n$ es una traslación que envía a x en un elemento de $(0,1)^n$; es decir, $h(x) \in (0,1)^n$. Dado que $f \in \mathcal{L}^1(\mathbb{R}^n, \text{Leb}, \lambda)$ y $h(x) \in (0,1)^n$, tenemos que

$$\lim_{\delta \to 0} \frac{1}{\lambda(B(h(x),\delta))} \int_{B(h(x),\delta)} |f(s) - f(h(x))| \, d\lambda(s) = 0.$$

Gracias a que h es una traslación, tenemos que

$$B(h(x),\delta) \overset{h^{-1}}{\longmapsto} B(x,\delta)$$

y, además,

$$\lambda(B(h(x),\delta)) = \lambda(B(x,\delta)).$$

Por lo tanto,

$$\lim_{\delta \to 0} \frac{1}{\lambda(B(x,\delta))} \int_{B(h(x),\delta)} |g(h^{-1}(s)) - g(x)| \, d\lambda(s) = 0. \tag{5.2}$$

Así, tomando el cambio de variable $y = h^{-1}(s)$ en (5.2), tenemos que

$$\lim_{\delta \to 0} \frac{1}{\lambda(B(x,\delta))} \int_{B(x,\delta)} |g(y) - g(x)| \, d\lambda(s) = 0. \tag{5.3}$$

Gracias a este resultado, en lo que sigue, no referiremos al Teorema 5.2 como el Teorema de diferenciación de Lebesgue.

5.2. El Teorema de densidad de Lebesgue

El Teorema de densidad de Lebesgue establece que:

TEOREMA 5.3. *Para todo* $A \in \mathrm{Leb}(0,1)$*, se tiene que, para casi todo punto* $x \in A$*,*

$$\lim_{\delta \to 0} \frac{\lambda(A \cap (x-\delta, x+\delta))}{2\delta} = 1,$$

donde λ *es la medida de Lebesgue.*

Este resultado se enuncia, por lo general, como un corolario del Teorema de diferenciación de Lebesgue. En efecto, supongamos que $A \in \mathrm{Leb}(0,1)$. Tenemos que $\mathbb{1}_A$ es integrable, y si $\delta > 0$ entonces, para casi todo punto $x \in \mathbb{R}$,

$$\left| \frac{1}{2\delta} \int_{(x-\delta,x+\delta)} \mathbb{1}_A(y) - \mathbb{1}_A(x)\, d\lambda(y) \right| \le \frac{1}{2\delta} \int_{(x-\delta,x+\delta)} |\mathbb{1}_A(y) - \mathbb{1}_A(x)|\, d\lambda(y).$$

Así, gracias al Teorema 5.2, tenemos que, para casi todo punto $x \in \mathbb{R}$,

$$\lim_{\delta \to 0} \left| \frac{1}{2\delta} \int_{(x-\delta,x+\delta)} \mathbb{1}_A(y) - \mathbb{1}_A(x)\, d\lambda(y) \right| = 0. \tag{5.4}$$

Por otra parte, podemos ver que

$$\begin{aligned} \frac{1}{2\delta} \int_{(x-\delta,x+\delta)} \mathbb{1}_A(y) - \mathbb{1}_A(x)\, d\lambda(y) &= \frac{1}{2\delta} \left(\int_{(x-\delta,x+\delta)} \mathbb{1}_A(y)\, d\lambda(y) \right. \\ &\qquad \left. - \int_{(x-\delta,x+\delta)} \mathbb{1}_A(x)\, d\lambda(y) \right) \\ &= \frac{1}{2\delta}(\lambda(A \cap (x-\delta, x+\delta)) - 2\delta \mathbb{1}_A(x)) \\ &= \frac{\lambda(A \cap (x-\delta, x+\delta))}{2\delta} - \mathbb{1}_A(x), \end{aligned}$$

por lo tanto, por (5.4), tenemos que, para casi todo punto $x \in \mathbb{R}$,

$$\lim_{\delta \to 0} \frac{\lambda(A \cap (x-\delta, x+\delta))}{2\delta} = \mathbb{1}_A(x).$$

Luego, para casi todo punto $x \in A$ se tiene que

$$\lim_{\delta \to 0} \frac{\lambda(A \cap (x-\delta, x+\delta))}{2\delta} = 1.$$

La demostración clásica del Teorema de diferenciación de Lebesgue, y por ende, la del Teorema de densidad de Lebesgue, se realiza mediante recubrimientos de Vitali (véase [9]). En cambio, si suponemos que $A \in \mathrm{Leb}(0,1)$ y $x \in A$, entonces, para todo $\delta > 0$, se tiene que

$$\frac{\lambda(A \cap (x-\delta, x+\delta))}{2\delta} = \frac{\lambda(A \cap (x-\delta, x+\delta))}{\lambda(x-\delta, x+\delta)} = \lambda(A | (x-\delta, x+\delta)),$$

por lo tanto, el límite en el Teorema 5.3 es un límite de probabilidades condicionales, lo cual se traduce en un límite de esperanzas condicionales. Esto motiva el enfoque que proponemos para tratar este teorema.

Antes de pasar la demostración del Teorema de densidad de Lebesgue usando martingalas, probaremos unos cuantos lemas y proposiciones que serán de mucha utilidad.

PROPOSICIÓN 5.4. *Para todo $n \in \mathbb{N}$, se tiene que las familias*

$$A_n = \left\{ \left(\frac{k}{2^n}, \frac{k+1}{2^n} \right] : k \in \mathbb{Z} \right\}$$

y

$$A_n' = \left\{ \left(\frac{k}{2^n} + \frac{1}{3}, \frac{k+1}{2^n} + \frac{1}{3} \right] : k \in \mathbb{Z} \right\}$$

son particiones de $\mathbb{R}$. Además, las sucesiónes $(\sigma(A_n))_{n\in\mathbb{N}}$ y $(\sigma(A_n'))_{n\in\mathbb{N}}$ son filtraciones sobre $\mathbb{R}$.

Demostración. Sean $n \in \mathbb{N}$ y $x \in \mathbb{R}$. Por definición de la función techo, tenemos que

$$\lceil x2^n \rceil - 1 < x2^n \leq \lceil x2^n \rceil;$$

es decir,

$$x \in \left(\frac{\lceil x2^n \rceil - 1}{2^n}, \frac{\lceil x2^n \rceil}{2^n} \right] \in A_n.$$

Así, $\mathbb{R} \subseteq \bigcup_{n\in\mathbb{N}} A_n$. Por otra parte, podemos ver que, si $k \in \mathbb{Z}$, entonces

$$\left(\frac{k-1}{2^n}, \frac{k}{2^n} \right] \cap \left(\frac{k}{2^n}, \frac{k+1}{2^n} \right] = \varnothing,$$

lo que implica que, para todo $k, j \in \mathbb{Z}$ tales que $k < j$, se tiene que

$$\left(\frac{k}{2^n}, \frac{k+1}{2^n} \right] \cap \left(\frac{j}{2^n}, \frac{j+1}{2^n} \right] = \varnothing.$$

Con esto, tenemos que A_n es una partición de $\mathbb{R}$; y, de una manera similar, se tiene que A_n' también lo es.

Por último, sean $n \in \mathbb{N}$ y $k \in \mathbb{Z}$. Tenemos que

$$\left(\frac{k}{2^n}, \frac{k+1}{2^n} \right] = \left(\frac{2k}{2^{n+1}}, \frac{2k+2}{2^{n+1}} \right] = \left(\frac{2k}{2^{n+1}}, \frac{2k+1}{2^{n+1}} \right] \cup \left(\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}} \right] \in \sigma(A_{n+1}),$$

por lo tanto

$$\sigma(A_n) \subseteq \sigma(A_{n+1}). \qquad \square$$

Así, tenemos que tanto $(\sigma(A_n))_{n\in\mathbb{N}}$ como $(\sigma(A'_n))_{n\in\mathbb{N}}$ son filtraciones sobre $\mathbb{R}$.

La colecciones A_n y A'_n serán muy importantes al momento de probar el Teorema de densidad de Lebesgue, pues, entre otras cosas, verifican los siguientes resultados:

PROPOSICIÓN 5.5. *Si*

$$A_\infty := \sigma\left(\bigcup_{n\in\mathbb{N}} A_n\right) \qquad y \qquad A'_\infty := \sigma\left(\bigcup_{n\in\mathbb{N}} A'_n\right),$$

entonces

$$A_\infty = A'_\infty = \mathcal{B}.$$

Demostración. Sea $x \in \mathbb{R}$. Vamos a probar que $(-\infty, x) \in A_\infty$. Primeramente, notemos que, para todo $k \in \mathbb{Z}$ y $n \in \mathbb{N}$,

$$\left(-\infty, \frac{k}{2^n}\right] = \bigcup_{\substack{m\in\mathbb{Z}\\ m\leq k}} \left(\frac{m-1}{2^n}, \frac{m}{2^n}\right] \in A_\infty. \tag{5.5}$$

Sea $n \in \mathbb{N}$. Tenemos que

$$\lfloor x2^n \rfloor \leq x2^n = \frac{x2^{n+1}}{2},$$

por lo tanto,

$$2\lfloor x2^n \rfloor \leq x2^{n+1}.$$

Ya que $2\lfloor x2^n \rfloor \in \mathbb{Z}$, se tiene que

$$2\lfloor x2^n \rfloor \leq \lfloor x2^{n+1} \rfloor \leq x2^{n+1},$$

y así

$$\frac{\lfloor x2^n \rfloor}{2^n} \leq \frac{x2^{n+1}}{2^{n+1}};$$

es decir, que la sucesión $\left(\frac{\lfloor x2^n \rfloor}{2^n}\right)_{n\in\mathbb{N}}$ es creciente. Por otra parte, gracias a que $|\lfloor x2^n \rfloor - x2^n| < 1$, tenemos que

$$\begin{aligned}\left|\frac{\lfloor x2^n \rfloor}{2^n} - x\right| &= \left|\frac{\lfloor x2^n \rfloor - x2^n}{2^n}\right| \\ &= \frac{1}{2^n}\left|\lfloor x2^n \rfloor - x2^n\right| \\ &< \frac{1}{2^n}.\end{aligned}$$

Así, si $n \to \infty$, entonces $\dfrac{\lfloor x2^n \rfloor}{2^n} \to x$.

Gracias a (5.5), tenemos que

$$(-\infty, x) = \bigcup_{n\in\mathbb{N}} \left(-\infty, \frac{\lfloor x2^n \rfloor}{2^n}\right] \in A_\infty,$$

por lo tanto

$$\sigma(\{(-\infty, x) : x \in \mathbb{R}\}) = \mathcal{B} \subseteq A_\infty.$$

La contenencia

$$A_\infty \subseteq \mathcal{B}$$

es trivial, pues A_∞ está generada por una colección de borelianos. Así, se sigue que $A_\infty = \mathcal{B}$. De manera análoga, tenemos que $A'_\infty = \mathcal{B}$. $\square$

LEMA 5.6. *Sean $n \in \mathbb{N}$ y $x \in \mathbb{R}$. Definimos $A_n(x)$ como el único elemento de A_n tal que $x \in A_n(x)$; y definimos $A'_n(x)$ como el único elemento de A'_n tal que $x \in A'_n(x)$. Se tiene que, para todo $\delta \in \left(0, \frac{1}{3\cdot 2^n}\right)$,*

$$(x-\delta, x+\delta) \subseteq A_m(x) \cup A'_m(x).$$

Demostración. Por la definición de A_n, existe $k \in \mathbb{Z}$ tal que

$$A_n(x) = \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right].$$

Notemos que, para todo $j \in \mathbb{Z}$, se tiene que

$$\frac{j}{2^n} + \frac{1}{3} \neq \frac{k}{2^n};$$

pues, de no ser así, existiría $j \in \mathbb{Z}$ tal que $2^n = 3(k-j)$, lo cual contradice el hecho de que 2^n y 3 son coprimos. Con esto, y gracias a que A'_n es una partición de $\mathbb{R}$, existe $j \in \mathbb{Z}$ tal que

$$\frac{j-1}{2^n} + \frac{1}{3} < \frac{k}{2^n} < \frac{j}{2^n} + \frac{1}{3} < \frac{k+1}{2^n} < \frac{j+1}{2^n} + \frac{1}{3}. \tag{5.6}$$

Así, si $x \leq \dfrac{j}{2^n} + \dfrac{1}{3}$, tenemos que

$$A'_n(x) = \left(\frac{j-1}{2^n} + \frac{1}{3}, \frac{j}{2^n} + \frac{1}{3}\right]; \tag{5.7}$$

y si $x > \dfrac{j}{2^n} + \dfrac{1}{3}$, entonces

$$A'_n(x) = \left(\frac{j}{2^n} + \frac{1}{3}, \frac{j+1}{2^n} + \frac{1}{3}\right]. \tag{5.8}$$

De (5.6), podemos ver que

$$\begin{aligned}
\frac{k}{2^n} < \frac{j}{2^n} + \frac{1}{3} < \frac{k+1}{2^n} &\iff k < j + \frac{2^n}{3} < k+1 \\
&\iff 3k < 3j + 2^n < 3k+3 \\
&\iff 3j + 2^n \in \{3k+1, 3k+2\} \\
&\iff \frac{j}{2^n} + \frac{1}{3} \in \left\{\frac{k}{2^n} + \frac{1}{3\cdot 2^n}, \frac{k}{2^n} + \frac{2}{3\cdot 2^n}\right\}.
\end{aligned}$$

Por lo tanto, hay dos valores posibles para $\frac{j}{2^n} + \frac{1}{3}$; sin embargo, en cualquier caso, la demostración subsiguiente es análoga, así que vamos a suponer que

$$\frac{j}{2^n} + \frac{1}{3} = \frac{k}{2^n} + \frac{1}{3\cdot 2^n}. \tag{5.9}$$

Así,

$$\frac{j-1}{2^n} + \frac{1}{3} = \frac{k}{2^n} - \frac{2}{3\cdot 2^n} \qquad \text{y} \qquad \frac{j+1}{2^n} + \frac{1}{3} = \frac{k+1}{2^n} + \frac{1}{3\cdot 2^n},$$

y, por lo tanto, gracias a (5.7) y (5.8), se tiene que

$$A'_m(x) = \begin{cases} \left(\dfrac{k}{2^n} - \dfrac{2}{3\cdot 2^n}, \dfrac{k}{2^n} + \dfrac{1}{3\cdot 2^n}\right] & \text{si } x \le \dfrac{k}{2^n} + \dfrac{1}{3\cdot 2^n} \\ \left(\dfrac{k}{2^n} + \dfrac{1}{3\cdot 2^n}, \dfrac{k+1}{2^n} + \dfrac{1}{3\cdot 2^n}\right] & \text{si } x > \dfrac{k}{2^n} + \dfrac{1}{3\cdot 2^n} \end{cases} \tag{5.10}$$

Ahora, sea $\delta \in \left(0, \frac{1}{3\cdot 2^n}\right)$. Supongamos que $x \le \dfrac{k}{2^n} + \dfrac{1}{3\cdot 2^n}$. Se tiene que

$$\frac{k}{2^n} < x \le \frac{k}{2^n} + \frac{1}{3\cdot 2^n},$$

por lo tanto

$$\frac{k}{2^n} - \frac{1}{3\cdot 2^n} < x - \delta < x + \delta \le \frac{k}{2^n} + \frac{2}{3\cdot 2^n}.$$

Con esto, y gracias a que (5.10) implica que

$$A_n(x) \cup A'_n(x) = \left(\frac{k}{2^n} - \frac{2}{3\cdot 2^n}, \frac{k+1}{2^n}\right],$$

se sigue que $(x-\delta, x+\delta) \subseteq A_n(x) \cup A'_n(x)$. De una manera similar, si $x > \dfrac{k}{2^n} +$

$\frac{1}{3\cdot 2^n}$, se tiene que

$$\frac{k}{2^n} < x-\delta < x+\delta \leq \frac{k+1}{2^n} + \frac{1}{3\cdot 2^n}.$$

y

$$A_n(x) \cup A'_n(x) = \left(\frac{k}{2^n}, \frac{k+1}{2^n} + \frac{1}{3\cdot 2^n}\right],$$

por lo tanto, se sigue nuevamente que $(x-\delta, x+\delta) \subseteq A_n(x) \cup A'_n(x)$. □

LEMA 5.7. *Si $\delta > 0$, existe un único $n \in \mathbb{Z}$ tal que*

$$\frac{1}{3\cdot 2^{n+1}} \leq \delta < \frac{1}{3\cdot 2^n}.$$

Demostración. Sea $n = -\lfloor \log_2(3\delta)+1 \rfloor$. Entonces

$$\begin{aligned}
n = -\lfloor \log_2(3\delta)+1\rfloor &\iff -n \leq \log_2(3\delta)+1 < -n+1\\
&\iff -n-1 \leq \log_2(3\delta) < -n\\
&\iff -n-1 \leq \log_2(\delta) - \log_2\left(\frac{1}{3}\right) < -n\\
&\iff \log_2\left(\frac{1}{2^{n+1}}\right) \leq \log_2(\delta) - \log_2\left(\frac{1}{3}\right) < \log_2\left(\frac{1}{2^n}\right)\\
&\iff \log_2\left(\frac{1}{3\cdot 2^{n+1}}\right) \leq \log_2(\delta) < \log_2\left(\frac{1}{3\cdot 2^n}\right)\\
&\iff \frac{1}{3\cdot 2^{n+1}} \leq \delta < \frac{1}{3\cdot 2^n}.
\end{aligned}$$

□

De este último resultado, se sigue que, si $\delta \to 0$, entonces $n \to \infty$. Por lo tanto, si δ es suficientemente pequeño, entonces $n \in \mathbb{N}$.

Teniendo en mente estos lemas y proposiciones, estamos listos para probar el Teorema de densidad de Lebesgue:

Demostración del Teorema 5.3. Sea $A \in \text{Leb}(0,1)$ y procedamos por casos considerando la medida de A.

Si $\lambda(A) = 0$, entonces, para todo $x \in A$ y $\delta > 0$,

$$\lambda(A \cap (x-\delta, x+\delta)) \leq \lambda(A) = 0,$$

por lo tanto

$$\lim_{\delta\to 0} \frac{\lambda(A\cap(x-\delta,x+\delta))}{2\delta} = 0.$$

Así,

$$\left\{x \in A : \lim_{\delta \to 0} \frac{\lambda(A \cap (x-\delta, x+\delta))}{2\delta} \neq 1\right\} = A,$$

de donde se sigue que

$$\lambda\left(\left\{x \in A : \lim_{\delta \to 0} \frac{\lambda(A \cap (x-\delta, x+\delta))}{2\delta} \neq 1\right\}\right) = 0,$$

y, por ende, también el resultado.

Si, ahora, $\lambda(A) > 0$, entonces $A \in \mathcal{B}(0,1)$. Consideremos la filtración $(\mathcal{F}_n)_{n\in\mathbb{N}}$ definida sobre el espacio de probabilidad $((0,1), \mathcal{B}(0,1), \lambda)$, donde

$$\mathcal{F}_n = \sigma(A_n) \cap (0,1);$$

y definamos el proceso

$$\left(\mathbf{E}\left(|\mathbb{1}_A - 1| \mid \mathcal{F}_n\right)\right)_{n\in\mathbb{N}}.$$

Gracias a lo visto en el Ejemplo 4 del capítulo anterior, este proceso es una martingala, y gracias al Teorema 4.14, tenemos que

$$\mathbf{E}\left(|\mathbb{1}_A - 1| \mid \mathcal{F}_n\right)(x) \to \mathbf{E}\left(|\mathbb{1}_A - 1| \mid \mathcal{F}_\infty\right)(x),$$

casi seguramente en $(0,1)$; además, gracias a la Proposición 5.5, se sigue que

$$\mathbf{E}\left(|\mathbb{1}_A - 1| \mid \mathcal{F}_n\right)(x) \to \mathbf{E}\left(|\mathbb{1}_A - 1| \mid \mathcal{B}(0,1)\right)(x). \tag{5.11}$$

Ahora, ya que $A \in \mathcal{B}(0,1)$, se tiene que $|\mathbb{1}_A - 1|$ es $\mathcal{B}(0,1)$-medible, y por lo tanto

$$\mathbf{E}\left(|\mathbb{1}_A - 1| \mid \mathcal{B}(0,1)\right)(x) = |\mathbb{1}_A - 1|(x) = |\mathbb{1}_A(x) - 1|;$$

luego, de (5.11), se tiene que

$$\mathbf{E}\left(|\mathbb{1}_A - 1| \mid \mathcal{F}_n\right)(x) \to |\mathbb{1}_A(x) - 1|.$$

Con esto, tenemos que

$$\mathbf{E}\left(|\mathbb{1}_A - 1| \mid \mathcal{F}_n\right)(x) \to 0, \tag{5.12}$$

casi seguramente en A.

Notemos que, para cada $x \in (0,1)$ existe $N \in \mathbb{N}$ tal que, si $n \geq N$, entonces $A_n(x) \subseteq (0,1)$. Por lo tanto, gracias a que $\mathcal{F}_n$ está definida a partir de una partición finita de $(0,1)$, y al Teorema de esperanza total visto en el Ejemplo 2, se sigue que, para $n \geq N$,

$$\mathbf{E}\left(|\mathbb{1}_A - 1| \mid \mathcal{F}_n\right)(x) = \frac{1}{\lambda(A_n(x))} \int_{A_n(x)} |\mathbb{1}_A(t) - 1| \, d\lambda(t).$$

Así, gracias a (5.12), tenemos que

$$\frac{1}{\lambda(A_n(x))} \int_{A_n(x)} |\mathbb{1}_A(t) - 1| \, d\lambda(t) \to 0, \tag{5.13}$$

casi seguramente en A. De la misma manera, se tiene que

$$\frac{1}{\lambda(A_n'(x))} \int_{A_n'(x)} |\mathbb{1}_A(t) - 1| \, d\lambda(t) \to 0, \tag{5.14}$$

casi seguramente en A.

Sea $\delta > 0$. Gracias al Lema 5.7, si δ es suficientemente pequeño, existe $n \in \mathbb{N}$ tal que

$$\frac{1}{3 \cdot 2^{n+1}} \leq \delta < \frac{1}{3 \cdot 2^n}.$$

Por lo tanto, por el Lema 5.6,

$$(x - \delta, x + \delta) \subseteq A_n(x) \cup A_n'(x),$$

y además

$$\frac{1}{2\delta} \leq \frac{1}{\frac{1}{3 \cdot 2^n}} = \frac{1}{\lambda(A_n(x))} = \frac{1}{\lambda(A_n'(x))}.$$

Así,

$$\begin{aligned}
\left| \frac{1}{2\delta} \int_{(x-\delta,x+\delta)} \mathbb{1}_A(t) - 1 \, d\lambda(t) \right| &\leq \frac{1}{2\delta} \int_{(x-\delta,x+\delta)} |\mathbb{1}_A(t) - 1| \, d\lambda(t) \\
&\leq \frac{1}{2\delta} \int_{A_n(x) \cup A_n'(x)} |\mathbb{1}_A(t) - 1| \, d\lambda(t) \\
&\leq \frac{1}{\lambda(A_n(x))} \int_{A_n(x)} |\mathbb{1}_A(t) - 1| \, d\lambda(t) \\
&\quad + \frac{1}{\lambda(A_n'(x))} \int_{A_n'(x)} |\mathbb{1}_A(t) - 1| \, d\lambda(t).
\end{aligned}$$

Ahora, si $\delta \to 0$, entonces $n \to \infty$, por lo tanto, de (5.13) y (5.14), tenemos que

$$\lim_{\delta \to 0} \left| \frac{1}{2\delta} \int_{(x-\delta,x+\delta)} \mathbb{1}_A(t) - 1 \, d\lambda(t) \right| = 0. \tag{5.15}$$

Finalmente, tenemos que

$$\frac{1}{2\delta} \int_{(x-\delta,x+\delta)} \mathbb{1}_A(t) - 1 \, d\lambda(t) = \frac{1}{2\delta} \lambda(A \cap (x - \delta, x + \delta)) - 1,$$

y gracias a (5.15), se sigue que

$$\lim_{\delta \to 0} \frac{\lambda(A \cap (x - \delta, x + \delta))}{2\delta} = 1,$$

casi seguramente en A. $\square$

Conclusiones y Reomendaciones

Conclusiones

a) Gracias al Teorema 2.25, es posible demostrar la existencia de la esperanza condicional sin hacer uso del Teorema de Radon-Nikodym.

b) Si $X \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$, entonces $\mathbf{E}(X \mid \mathcal{G})$ es una versión de la proyección ortogonal de X sobre $\mathcal{L}^2(\Omega, \mathcal{G}, P)$, con lo cual tenemos que $\mathbf{E}(X \mid \mathcal{G})$ es la mejor aproximación $\mathcal{G}$-medible de X por mínimos cuadrados, pues

$$\mathbf{E}(X - \mathbf{E}(X \mid \mathcal{G})) = \text{mín}\{\mathbf{E}\left((X - Z)^2\right) : Z \in \mathcal{L}^2(\Omega, \mathcal{G}, P)\}.$$

c) Del Teorema 4.2, tenemos que, si un juego de apuestas se puede modelar mediante una supermartingala, es decir, que el juego es desfavorable, entonces, sin importar la estrategia que diseñemos para apostar, el juego seguirá siendo desfavorable.

d) Es posible, haciendo uso del Teorema de convergencia de Martingalas, probar el Teorema de Radon-Nikodym, probando así que este último es un teorema dentro de la Teoría de probabilidades.

e) Gracias al Teorema de Lévy, una consecuencia del Teorema de convergencia de martingalas, es posible demostrar que el Teorema de densidad de Lebesgue es un teorema dentro de la Teoría de probabilidades.

f) Inicialmente, las martingalas fueron una herramienta diseñada para estudiar juegos de apuestas. El Teorema 4.2 es sólo un ejemplo de su aplicación a este tipo de problemas; sin embargo, hemos probado también que son un herramienta poderosa al momento de buscar resultados teóricos como los Teoremas de Radon-Nikodym y de densidad de Lebesgue.

Recomendaciones

Como una continuación de este trabajo, se podría plantear los siguientes problemas:

a) Extender el Teorema de densidad de Lebesgue a $\mathbb{R}^n$ considerando, además de la medida de Lebesgue, otras medidas de probabilidad.

b) Estudiar el Teorema de densidad de Lebesgue (o algún análogo) en espacios medidos abstractos.

c) Estudiar el Teorema de densidad de Lebesgue (o algún análogo) en el espacio $\mathbb{R}^\infty$.

Bibliografía

[1] A. ANDRETA AND R. CAMERLO, *The descriptive set theory of the Lebesgue Density Theorem*, Advances in Mathematics, 234 (2013), pp. 1–42.

[2] S. BANACH AND A. TARSKI, *Sur la décomposition des ensembles de points en parties respectivement congruentes*, Fundamenta Mathematicae, 6 (1924), pp. 244–277.

[3] T. BEDFORD AND A. FISHER, *Analogues of the Lebesgue Density Theorem for Fractal Sets of Reals and Integers*, Proceedings of the London Mathematical Society, s3-64(1) (1992), pp. 95–124.

[4] A. BOBROWSKI, *Functional Analysis for Probability and Stochastic Processes*, Cambridge University Press, Estados Unidos, 2005.

[5] G. BREDON, *Topology and Geometry*, Springer-Verlag, Estados Unidos, 1993.

[6] M. BRUCKNER, *Differentiation of Integrals*, The American Mathematical Monthly, 78 (1971), pp. 1–51.

[7] V. CAZAUBON, *In search of a Lebesgue density theorem for* $\mathbb{R}^\infty$, University of Ottawa, (2008).

[8] K. CHUNG, *A Course in Probability Theory*, Academic Press, Estados Unidos, 2001.

[9] G. FOLLAND, *Real analysis: modern techniques and their applications*, Wiley, Estados Unidos, 1999.

[10] H. LEBESGUE, *Sur l'intégration des fonctions discontinues*, Annales scientifiques de l'École Normale Supérieure, 27 (1910), pp. 361–450.

[11] M. MORAYNE AND S. SOLECKI, *Martingale proof of the existence of Lebesgue Points*, Real Analysis Exchange, 15 (1989-1990), pp. 401–406.

[12] I. RANA, *An Introduction to Measure and Integration*, American Mathematical Society, Estados Unidos, 2002.

[13] K. ROSS AND E. PEKÖS, *A Second Course in Probability*, www.ProbabilityBookstore.com, Estados Unidos, 2007.

[14] D. WILLIAMS, *Probability with Martingales*, Cambridge University Press, Estados Unidos, 1991.