Índice general

Capítulo 1

Fundamentos de la teoría de probabilidad

En este capítulo presentamos los elementos básicos la de teoría de probabilidad, así como también herramientas del análisis que se requieren para estudiar el concepto de esperanza condicional que se verá en el siguiente capítulo.
Denotaremos por $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ y $\mathbb{R}$ los conjuntos de los números naturales, enteros, racionales y reales, respectivamente.

1.1. Eventos y variables aleatorias

La teoría de probabilidad trabaja con modelos matemáticos de experimentos cuyos resultados dependen del "azar". El conjunto de todos los posibles resultados estarán en un conjunto Ω, llamado *espacio muestral*, donde un elemento de este conjunto lo denotaremos por ω. Si el experimento consiste en lanzar una moneda, entonces $\Omega = \{\text{águila, sol}\}$; para el experimento de lanzar un par de dados (distinguibles entre si) $\Omega = \{(i, j) : 1 \leq i, j \leq 6\}$; para el tiempo de vida de una bombilla, $\Omega = [0, \infty)$; en la observación del nivel del agua del tiempo t_1 al tiempo t_2, el espacio muestral Ω será el conjunto de todas las funciones real valuadas (o quizá continuas) definidas en el intervalo $[t_1, t_2]$. Si A es un subconjunto de Ω, lo indicaremos escribiendo $A \subset \Omega$.

Por otro lado, no todo subconjunto de Ω es un conjunto interesante. Denotaremos por $\mathfrak{F}$ a una colección de subconjuntos de Ω, la cual llamaremos

σ-álgebra (léase "sigma álgebra"), que satisface los siguientes axiomas:

Definición 1.1.1. *Sea $\Omega \neq \emptyset$ un espacio abstracto. Una colección $\mathfrak{F}$ de subconjuntos de Ω es* σ-álgebra *si:*

(i) $\Omega \in \mathfrak{F}$.

(ii) Si $A \in \mathfrak{F}$, entonces $A^c \in \mathfrak{F}$.

(iii) Si $\{A_n\}_{n\in\mathbb{N}}$ es una familia numerable de conjuntos tal que $A_n \in \mathfrak{F}$ para cada $n \in \mathbb{N}$, entonces $\cup_n A_n \in \mathfrak{F}$.

Los elementos de $\mathfrak{F}$ son llamados *conjuntos medibles ó eventos* y al par $(\Omega,\mathfrak{F})$ se le denomina *espacio medible*.

Observación 1.1.1. *Es fácil verificar que si $\{\mathfrak{F}_\alpha\}_{\alpha\in\Lambda}$ es una familia no vacía de σ-álgebras sobre Ω, entonces su intersección $\bigcap_{\alpha\in\Lambda}\mathfrak{F}_\alpha$ es también una σ-álgebra sobre Ω.*

Definición 1.1.2. *Sea $\mathcal{C}$ una familia de subconjuntos de Ω. Denotamos por $\sigma(\mathcal{C})$ la intersección de todas las σ-álgebras que contienen a $\mathcal{C}$, y se le llama* sigma álgebra generada por $\mathcal{C}$, *y es la mínima σ-álgebra que contiene a $\mathcal{C}$. La σ-álgebra $\sigma(\mathcal{C})$ existe por que hay σ-álgebras que contienen a $\mathcal{C}$ (por ejemplo $\mathcal{P}(\Omega)$ el conjunto potencia).*

A continuación definimos el concepto de variable aleatoria la cual denotaremos por v.a., y variables aleatorias lo abreviaremos como vv.aa.

Definición 1.1.3. *Sean $(\Omega,\mathfrak{F})$, $(\Omega',\mathfrak{F}')$, espacios medibles, $X:\Omega \longrightarrow \Omega'$ una aplicación. Diremos que X es una Ω'-valuada* variable aleatoria *si*

$$X^{-1}(A) \in \mathfrak{F} \quad \forall A \in \mathfrak{F}'.$$

En teoría de la medida diremos que X es medible, *o que X es $(\mathfrak{F},\mathfrak{F}')$-medible.*

Ejemplo 1.1.3.1. *Sean $(\Omega,\mathfrak{F})$ un espacio medible, y* la función indicadora *I_A de un subconjunto $A \subset \Omega$ definida por:*

$$I_A(\omega) = \begin{cases} 1, & si \quad \omega \in A \\ 0, & si \quad \omega \notin A. \end{cases}$$

Notemos que I_A es v.a. si y sólo si A es medible, es decir, $A \in \mathfrak{F}$.

Observación 1.1.2. *Sea* $Y : \Omega \longrightarrow \mathbb{R}$. *Definimos*

$$Y^{+} = \text{máx}(Y, 0),\ Y^{-} = -\text{mín}(Y, 0).$$

Notemos que $Y = Y^{+} - Y^{-}$ *y* $|Y| = Y^{+} + Y^{-}$. *Entonces* Y *es v.a. si y solo si* Y^{+}, Y^{-} *son variables aleatorias.*

De nuestras definiciones, se sigue el siguiente resultado de manera inmediata.

Proposición 1.1.1. *Sean* Ω *un conjunto no vacío,* $(\Omega', \mathfrak{F}')$ *un espacio medible,* $X : \Omega \longrightarrow \Omega'$ *una aplicación. Entonces la colección*

$$X^{-1}(\mathfrak{F}') := \{X^{-1}(F') : F' \in \mathfrak{F}'\},$$

es una σ*-álgebra sobre* Ω, *y es la mínima* σ*-álgebra sobre* Ω, *con* $(\Omega', \mathfrak{F}')$ *fijo, que hace de* X *una v.a.*

Definición 1.1.4. *Sean* $(\Omega, \mathfrak{F}), (\Omega', \mathfrak{F}')$, *espacios medibles y* $X : \Omega \longrightarrow \Omega'$ *una v.a. Denotamos por* $\sigma(X)$ *a la* σ*-álgebra inducida (o generada) por* X *como sigue:*

$$\sigma(X) = X^{-1}(\mathfrak{F}'),$$

donde $X^{-1}(\mathfrak{F}') = \{X^{-1}(F') : F' \in \mathfrak{F}'\}$. *Por la proposición 1.1.1, se tiene que* $\sigma(X)$ *es una* σ*-álgebra sobre* Ω *contenida en* $\mathfrak{F}$, *es decir,* $\sigma(X)$ *es una sub* σ*-álgebra de* $\mathfrak{F}$.
En general si Λ *es un subconjunto de índices de tal manera que para cada* $\alpha \in \Lambda$, $X_\alpha : (\Omega, \mathfrak{F}) \longrightarrow (\Omega'_\alpha, \mathfrak{F}'_\alpha)$ *es variable aleatoria, definimos*

$$\sigma\left(X_\alpha, \alpha \in \Lambda\right) = \sigma\left(\bigcup_{\alpha \in \Lambda} \sigma(X_\alpha)\right),$$

y se llama la σ-álgebra generada por la familia $\{X_\alpha\}_{\alpha \in \Lambda}$ *de vv.aa.*

De la proposición 1.1.1 y de la definición precedente, tenemos el siguiente corolario.

Corolario 1.1.1. $X : (\Omega, \mathfrak{F}) \to (\Omega', \mathfrak{F}')$ *es v.a si y sólo si*

$$\sigma(X) \subset \mathfrak{F}.$$

Ejemplo 1.1.4.1. *Sean* $(\Omega, \mathfrak{F}), (\mathbb{R}, \mathcal{B}(\mathbb{R}))$*, espacios medibles,* $X : \Omega \longrightarrow \mathbb{R}$ *tal que* $X(\omega) = 2$ *para toda* ω*, entonces*

$$\begin{aligned} \sigma(X) &= \{X^{-1}(B) : B \in \mathcal{B}\} \\ &= \{\emptyset, \Omega\}. \end{aligned}$$

Ejemplo 1.1.4.2. *Sea* $X : \mathbb{R} \longrightarrow \mathbb{R}$*, consideremos la* σ*-álgebra de Borel* $\mathcal{B}(\mathbb{R})$*, tal que* $X = I_A$ *para algún* $A \in \mathcal{B}(\mathbb{R})$*. Notemos que* X *toma los valores* $\{0, 1\}$ *y*

$$X^{-1}(\{0\}) = A^c, \quad X^{-1}(\{1\}) = A,$$

entonces $\sigma(X) = \{\emptyset, \Omega, A, A^c\}$.

Lema 1.1.1. *Sean* Ω *un conjunto no vacío,* $(\Omega', \mathfrak{F}')$ *un espacio medible,* $X : \Omega \to \Omega'$ *una aplicación. Supongamos que* $\mathcal{C}$ *es una subcolección de* $\mathfrak{F}'$ *tal que* $\mathfrak{F}' = \sigma(\mathcal{C})$*. Entonces*

$$\sigma(X) = \sigma(X^{-1}(\mathcal{C})),$$

es decir,

$$X^{-1}(\sigma(\mathcal{C})) = \sigma(X^{-1}(\mathcal{C})), \quad \textit{(los símbolos } X^{-1} \textit{ y } \sigma \textit{ commutan)}$$

donde $X^{-1}(\mathcal{C}) := \{X^{-1}(C) : C \in \mathcal{C}\}$.

Demostración. Claramente $X^{-1}(\mathcal{C}) \subset X^{-1}(\mathfrak{F}') = \sigma(X)$. De aquí,

$$\sigma(X^{-1}(\mathcal{C})) \subset \sigma(X). \tag{1.1.1}$$

Definamos ahora el conjunto:

$$\mathcal{H} := \{B \in \mathfrak{F}' : X^{-1}(B) \in \sigma(X^{-1}(\mathcal{C}))\}.$$

Claramente $\mathcal{H}$ es una sub-σ-álgebra de $\mathfrak{F}'$ que contiene a la colección $\mathcal{C}$, de donde $\mathfrak{F}' = \sigma(\mathcal{C}) \subset \mathcal{H} \subset \mathfrak{F}'$, es decir, $\mathcal{H} = \mathfrak{F}'$, de donde

$$\sigma(X) = X^{-1}(\mathfrak{F}') \subset \sigma(X^{-1}(\mathcal{C})). \tag{1.1.2}$$

Las relaciones (1.1.1) y (1.1.2) demuestran el lema. $\square$

Proposición 1.1.2. *Sean* $(\Omega, \mathfrak{F})$*,* $(\Omega', \mathfrak{F}')$ *espacios medibles,* $\mathcal{C}'$ *una subcolección de* $\mathfrak{F}'$ *tal que* $\mathfrak{F}' = \sigma(\mathcal{C}')$*,* $X : \Omega \longrightarrow \Omega'$ *una aplicación. Una condición necesaria y suficiente para que* X *sea variable aleatoria es que*

$$X^{-1}(C') \in \mathfrak{F} \quad \forall C' \in \mathcal{C}'.$$

Demostración. Se sigue inmediatamente del corolario 1.1.1 y del lema 1.1.1. En efecto, X es v.a. si y sólo si $\sigma(X) = \sigma(X^{-1}(\mathcal{C})) \subset (F)$ si y sólo si $X^{-1}(\mathcal{C}) \subset \mathfrak{F}$. □

En el caso de que $\Omega' = \mathbb{R}^n$, consideraremos a $\mathbb{R}^n$ dotada de la σ-álgebra de Borel, la cual denotamos por $\mathcal{B}(\mathbb{R}^n)$, y es la σ-álgebra generada por todos los abiertos de $\mathbb{R}^n$. Otras colecciones de subconjuntos de $\mathbb{R}^n$ que generan a $\mathcal{B}(\mathbb{R}^n)$ son los cerrados, los compactos y los intervalos de la forma

$$]-\infty, t_1]\times]-\infty, t_2] \times \cdots \times]-\infty, t_n].$$

Definición 1.1.5. *Los* números reales extendidos *o la* recta acabada *es el conjunto formado por el conjunto de los números reales junto con* $\{+\infty\}$, $\{-\infty\}$, *el cual lo denotaremos por* $\overline{\mathbb{R}}$ *ó* $[-\infty, +\infty]$.

Observación 1.1.3. *Consecuencias de la proposición 1.1.2:*
Sea $(\Omega, \mathfrak{F})$ *un espacio medible*

(1) $X : \Omega \longrightarrow \mathbb{R}^n$ *es v.a. si y sólo si* $X^{-1}(G) \in \mathfrak{F}$ *para cada* G *abierto de* $\mathbb{R}^n$.

(2) $Y : \Omega \longrightarrow \overline{\mathbb{R}}$ *es v.a. si y sólo si* $Y^{-1}((-\infty, t]) = \{\omega \in \Omega : Y(\omega) \leq t\} \in \mathfrak{F}$ *para cada* $t \in \mathbb{R}$. *Es usual denotar por* $[Y \in (-\infty, t]]$ *o* $[Y \leq t]$ *al conjunto* $Y^{-1}((-\infty, t])$. *Si en lugar de* $\overline{\mathbb{R}}$ *fuese* $\mathbb{R}$ *la caracterización es la misma.*

Corolario 1.1.2. *Sea* $(\Omega, \mathfrak{F})$ *un espacio medible. La función*

$$\begin{array}{rcl} X : \Omega & \longrightarrow & \mathbb{R}^n \\ \omega & \longmapsto & (X_1(\omega), \cdots, X_n(\omega)) \end{array}$$

es un vector aleatorio, si y sólo si X_i *es v.a. para cada* $i = 1, \cdots, n$.

Demostración. Para cada $i = 1, \cdots, n$, sea $\Pi_i : \mathbb{R}^n \to \mathbb{R}$ la proyección en la i-ésima coordenada. Entonces los conjuntos de la forma $C := \cap_{i=1}^n \Pi_i^{-1}(B_i)$ con $B_1, \cdots, B_n \in \mathcal{B}(\mathbb{R})$ borelianos de la recta real, forman una colección de subconjuntos de $\mathbb{R}^n$ que genera a los borelianos de $\mathbb{R}^n$.

Por la proposición 1.1.2 se tiene que $X : \Omega \longrightarrow \mathbb{R}^n$ es vector aleatorio si y sólo si $X^{-1}(\cap_{i=1}^n \Pi_i^{-1}(B_i)) = \cap_{i=1}^n X_i^{-1}(B_i) \in \mathfrak{F}$, para todo $B_1, \cdots, B_n \in \mathcal{B}(\mathbb{R})$, si y sólo si $X_i^{-1}(B) \in \mathfrak{F}$ para todo $B \in \mathcal{B}(\mathbb{R})$ y para todo $i = 1, \cdots, n$, si y sólo si X_i es v.a. para cada $i = 1, \cdots, n$.

□

Proposición 1.1.1. *Sean* $(\Omega, \mathfrak{F})$, $(\Omega', \mathfrak{F}')$, $(\Gamma, \mathfrak{U})$, *espacios medibles. Sean* $X : \Omega \longrightarrow \Omega'$, $Y : \Omega' \longrightarrow \Gamma$ *vv.aa. definidas en sendos espacios medibles. Entonces* $Y \circ X : \Omega \to \Gamma$ *es también una v.a.*

Demostración. En efecto,

$$\sigma(Y \circ X) = (Y \circ X)^{-1}(\mathfrak{U}) = X^{-1}(Y^{-1}(\mathfrak{U})) = X^{-1}(\sigma(Y)) \subset X^{-1}(\mathfrak{F}') = \sigma(X) \subset \mathfrak{F}.$$

□

Corolario 1.1.3. *Sean* $(\Omega, \mathfrak{F})$, *espacio medible,* $X : \Omega \to \mathbb{C}^n$ *un vector aleatorio complejo,* $\varphi : \mathbb{C}^n \to \mathbb{C}^m$ *una aplicación continua. Entonces* $\varphi \circ X : \Omega \to \mathbb{C}^m$ *es también un vector aleatorio. En particular, si* $X_1, \cdots, X_n$ *son vv.aa.* $\Omega \to \mathbb{C}$, $\alpha_1, \cdots, \alpha_n \in \mathbb{C}$, *entonces* $\alpha_1 X_1 + \cdots + \alpha_n X_n$ *es v.a. Una afirmación equivalente se tiene si sustituimos* $\mathbb{C}$ *por* $\mathbb{R}$.

Se tiene el siguiente resultado conocido para variables aleatorias.

Proposición 1.1.3. *Sean* $(\Omega, \mathfrak{F})$, $(\Omega', \mathfrak{F}')$, $(\Gamma, \mathfrak{U})$, *espacios medibles.*

(1) Sean $X : \Omega \longrightarrow \Omega'$, $Y : \Omega \longrightarrow \Gamma$ *variables aleatorias. Si existe* $f : \Omega' \longrightarrow \Gamma$ *aplicación medible (variable aleatoria) tal que* $Y = f \circ X$, *entonces* $\sigma(Y) \subseteq \sigma(X)$.

(2) Sean $X : \Omega \longrightarrow \Omega'$, $Y : \Omega \longrightarrow \mathbb{R}^n$ *variables aleatorias de tal manera que* $\sigma(Y) \subseteq \sigma(X)$. *Entonces existe* $f : \Omega' \to \mathbb{R}^n$ *una función* $\mathfrak{F}'$ *medible, tal que* $Y = f \circ X$.

Demostración.

(1) Note que $\sigma(Y) = Y^{-1}(\mathfrak{U}) = X^{-1}(f^{-1}(\mathfrak{U})) \subseteq X^{-1}(\mathfrak{F}') = \sigma(X)$.

(2) Consideremos el caso $n = 1$. De manera trivial se generaliza a mayor dimensión.

(a) Supongamos que Y es v.a. simple, es decir, existen $\alpha_1, \cdots, \alpha_n \in \mathbb{R}$, $A_1, \cdots, A_n \in \mathfrak{F}$ tal que $Y = \sum_{i=1}^n \alpha_i I_{A_i}$, donde $\alpha_1, \cdots, \alpha_n$ son los valores distintos que toma Y (de aquí, $A_i = [Y = \alpha_i]$). Por lo tanto, $A_i \in \sigma(X)$, de donde existe $B_i \in \mathfrak{F}'$ tal que $A_i = X^{-1}(B_i)$ para todo $i = 1, \cdots, n$. Por lo tanto,

$$Y = \sum_{i=1}^n \alpha_i I_{A_i} = \sum_{i=1}^n \alpha_i I_{X^{-1}(B_i)} = \sum_{i=1}^n \alpha_i I_{B_i} \circ X = f \circ X,$$

donde $f = \sum_{i=1}^{n} \alpha_i I_{B_i} : \Omega' \to \mathbb{R}$ es función $\mathfrak{F}'$-medible. Notemos de aquí que si $Y \geq 0$, entonces también $f \geq 0$.

(b) Supongamos que Y es una v.a. medible positiva. Luego existe una suceción creciente $\{Y_n\}$ de v.a. simples postivas tal que $Y_n(\omega) \uparrow Y(\omega)$ cuando $n \to \infty$, y para todo $\omega \in \Omega$. Por el inciso (a) arriba, para cada $n \in \mathbb{N}$ existe $f_n : \Omega' \to \mathbb{R}$ función $\mathfrak{F}'$-medible positiva tal que $Y_n = f_n \circ X$. Definimos $f : \Omega' \to \mathbb{R}$ como

$$f(\omega') = \begin{cases} \liminf f_n(\omega') & \text{si} \quad \liminf f_n(\omega') < \infty, \\ \\ 0 & \text{si} \quad \liminf f_n(\omega') = \infty. \end{cases}$$

De aquí, $f : \Omega' \to \mathbb{R}$ es función $\mathfrak{F}'$-medible.

Claramente es fácil verificar que $\inf_{k \geq n}\{f_k(X(\omega))\} = \inf_{k \geq n}\{Y_k(\omega)\} = Y_n(\omega)$, de donde $f(X(\omega)) = \liminf_{n \to \infty} f_n(X(\omega)) = \lim_{n \to \infty} Y_n(\omega) = Y(\omega)$ para cada $\omega \in \Omega$. Así, $Y = f \circ X$.

(c) Sea Y es v.a. $\Omega \to \mathbb{R}$ arbitraria. Basta aplicar el inciso (b) a las funciones $\mathfrak{F}$-medibles positivas Y^+ y Y^-, puesto que $Y = Y^+ - Y^-$.

(d) Sea $Y : \Omega \to \mathbb{R}^n$ un vector aleatoria, $Y = (Y_1, \cdots, Y_n)$ sus funciones coordenadas. Entonces cada $Y_i : \Omega \to \mathbb{R}$ es $\sigma(X)$-medible, de donde por los incisos precedente, existen $f_1, \cdots, f_n : \Omega' \to \mathbb{R}$ funciones $\mathfrak{F}'$-medibles tal que $Y_i = f_i \circ X$. Luego $Y = f \circ X$, con $f = (f_1, \cdots, f_n) : \Omega' \to \mathbb{R}^n$ función $\mathfrak{F}'$-medible.

$\square$

1.2. Probabilidad y funciones de distribución

Presentamos a continuación la definición de medida, así como la de función de distribución, y enunciaremos algunas propiedades básicas asociadas a estos conceptos.

Definición 1.2.1. *Sea* $(\Omega, \mathfrak{F})$ *un espacio medible. Una función*

$$\mu : \mathfrak{F} \longrightarrow [0, +\infty]$$

definida sobre $\mathfrak{F}$ *es una* medida *si satisface las siguientes condiciones:*

(i) $0 \leq \mu(A) \leq +\infty$ *para cada* $A \in \mathfrak{F}$,

(ii) $\mu(\emptyset) = 0$,

(iii) si $\{A_n\}_n$ *es una familia numerable disjunta de conjuntos tal que* $A_n \in \mathfrak{F}$ *para cada* $n \in \mathbb{N}$, *entonces*

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

La terna $(\Omega, \mathfrak{F}, \mu)$ se llama *espacio medido*, y $\mu(A)$ es la medida del conjunto medible A. Si $\mu(\Omega) < +\infty$, entonces se dice que μ es finita.

Definición 1.2.2. *Un espacio medido* $(\Omega, \mathfrak{F}, \mu)$ *es llamado espacio de medida* σ*-finita si* Ω *es la unión numerable o finita de conjuntos medibles que satisfacen que* $\mu(A_i) < +\infty$ *para cada* $i \in \mathbb{N}$.

Sea P una medida sobre $(\Omega, \mathfrak{F})$. Decimos que P es una *medida de probabilidad* sobre $(\Omega, \mathfrak{F})$ si $P(\Omega) = 1$. En este caso la terna $(\Omega, \mathfrak{F}, P)$ se llama *un espacio de probabilidad* y los elementos de $\mathfrak{F}$ son llamados *eventos.* La función P asigna a cada evento $A \in \mathfrak{F}$ un número $P(A)$, tal que $0 \leq P(A) \leq 1$.

Además tenemos que:

(a) $P(A) + P(A^c) = P(\Omega) = 1$.

(b) Si $A \subseteq B$, entonces $P(A) \leq P(B)$.

Una de las medidas importantes en el análisis real es la *medida de Lebesgue* denotada por λ sobre los borelianos de $\mathbb{R}^n$. Esta medida asigna a cada intervalo n-dimensional su "volumen":

$$\lambda(\{(x_1, x_2, \cdots, x_n) : a_k \leq x_k \leq b_k\}_{k=1}^n) = \prod_{k=1}^{n}(b_k - a_k)$$

Puesto que $\lambda(\mathbb{R}^n) = +\infty$, la medida de Lebesgue no es finita. Sin embargo, es σ-finita pues $\lambda(\{(x_1, x_2, \cdots, x_n) : -n \leq x_k \leq n, \}_{k=1}^n) = (2n)^n < +\infty$, $n \in \mathbb{N}$, y

$$\mathbb{R}^n = \bigcup_{n=1}^{\infty} \{(x_1, x_2, \cdots, x_n) : -n \leq x_k \leq n, \forall k = 1, \cdots, n\}.$$

Continuamos con espacios de probabilidad.

Definición 1.2.3. *Sean* $(\Omega, \mathfrak{F}, P)$ *un espacio de probabilidad,* A*,* $B \in \mathfrak{F}$*. Decimos que los eventos* A *y* B *son independientes si*

$$P(A \cap B) = P(A)P(B).$$

En general, una familia de eventos $\{A_\alpha : \alpha \in \Lambda\}$ *donde* Λ *denota un conjunto de índices, se dice ser independiente si*

$$P\left(\bigcap_{\alpha \in J} A_\alpha\right) = \prod_{\alpha \in J} P(A_\alpha)$$

para todo subconjunto finito J *de* Λ*.*

Definición 1.2.4. *Sea* Ω *un conjunto arbitrario. Una colección* $\mathcal{A}$ *de subconjuntos de* Ω *es llamada un* álgebra *sobre* Ω *si satisface:*

(1) $\Omega \in \mathcal{A}$*,*

(2) si $A \in \mathcal{A}$*, entonces* $A^c \in \mathcal{A}$*,*

(3) sean $A_1, A_2, \cdots, A_n$ *una familia finita de conjuntos de* $\mathcal{A}$*, entonces*

$$\bigcup_{k=1}^{n} A_k \in \mathcal{A},$$

es decir, $\mathcal{A}$ *es cerrada bajo uniones finitas.*

Definición 1.2.5. *Sea* $(\Omega, \mathfrak{F}, P)$ *un espacio de probabilidad.*

(a) Sean $\mathcal{C}, \mathcal{H}$ *dos sub* σ*-álgebras de* $\mathfrak{F}$*. Decimos que* $\mathcal{C}$ *es* independiente *de* $\mathcal{H}$ *si*

$$P(A \cap B) = P(A)P(B) \quad \forall A \in \mathcal{C} \text{ y } B \in \mathcal{H}.$$

(b) Sea $\{\mathfrak{F}_i\}_{1 \leq i \leq n}$ *una familia de sub* σ*-álgebras de* $\mathfrak{F}$*. Diremos que dicha familia es* independiente *si*

$$P(\cap_{i=1}^{n} B_i) = \prod_{i=1}^{n} P(B_i), \quad \forall B_1 \in \mathfrak{F}_1, \cdots, B_n \in \mathfrak{F}_n.$$

(c) *Sea* $\{\mathfrak{F}_\alpha\}_{\alpha\in\Lambda}$ *una familia de sub* σ*-álgebras de* $\mathfrak{F}$. *Diremos que dicha familia es* independiente *si para todo* $J \subset \Lambda$, *subconjunto finito de* Λ *de índices, con* $2 \leq Card(J) < \infty$, *la familia finita de sub* σ*-álgebras de* $\mathfrak{F}$, $\{\mathfrak{F}_\alpha\}_{\alpha\in J}$, *es independiente como en la definición en (b).*

(d) *Sea ahora* $\{X_\alpha\}_{\alpha\in\Lambda}$ *una familia de vv.aa.* $X_\alpha : (\Omega, \mathfrak{F}, P) \longrightarrow (\Omega_\alpha, \mathfrak{F}_\alpha)$. *Diremos que la familia* $\{X_\alpha\}_{\alpha\in\Lambda}$ *es independiente si la familia de sub* σ*-álgebras correspondientes* $\{\sigma(X_\alpha)\}_{\alpha\in J}$ *es una colección de sub* σ*-álgebras de* $\mathfrak{F}$ *independiente.*

Definición 1.2.6. *Sean* $(\Omega, \mathfrak{F}, P)$ *un espacio de probabilidad,* $(\Omega', \mathfrak{F}')$ *espacio medible,* $X : \Omega \longrightarrow \Omega'$ *una variable aleatoria. La* medida de distribución de probabilidad *de* X *es la medida de probabilidad* $\mu_X = P \circ X^{-1}$ *sobre* $(\Omega', \mathfrak{F}')$, *definida por*

$$\mu_X(A) = P(X^{-1}(A)) = P[X \in A] \quad \forall A \in \mathfrak{F}'.$$

Si $\Omega' = \mathbb{R}^n$, *la* función de distribución de X *es una función* $F_X : \mathbb{R}^n \longrightarrow \mathbb{R}$ *definida por:*

$$F_X(x) := \mu_X((-\infty, x]) = P[X \leq x],$$

donde $(-\infty, x] = (-\infty, x_1] \times \cdots \times (-\infty, x_n]$, *con* $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$.

En lo que sigue, nos restringiremos a dar algunas propiedades de las funciones de distribución de variables aleatorias real-valuadas, es decir, $X : \Omega \longrightarrow \mathbb{R}$. Para funciones de distribución de vectores aleatorias $X : \Omega \to \mathbb{R}^n$, se recomienda ver los textos [1, 5, 11, 16].

A continuación enunciamos las propiedades de la función de distribución F_X, donde X es una variable aleatoria real valuada.

Propiedades de $\mathbf{F_X}$

(1) F_X es monótona creciente;

(2) F_X es continua por la derecha y posee límites por la izquierda para cada $x \in \mathbb{R}$;

(3) $\lim_{x\to-\infty} F_X(x) = 0 \qquad \lim_{x\to+\infty} F_X(x) = 1$;

(4) $\mu_X((a,b]) = F_X(b) - F_X(a)$, $\mu_X(\{a\}) = F_X(a) - F_X(a^-)$, para todo $a, b \in \mathbb{R},\ a < b$.

1.3. Teoría de integración, esperanzas

Definición 1.3.1. *Sea X una variable aleatoria positiva ó integrable respecto a una medida de probabilidad P. A la integral*

$$\int_\Omega X \quad dP$$

se le llamará la esperanza de X, *y se denotará por*

$$EX = \int_\Omega X \quad dP.$$

Notemos que $EX = EX^+ - EX^-$.

Teorema 1.3.1. ***(de cambio de variable)***
Sean $(\Omega, \mathfrak{F}, P)$ un espacio de probabilidad, $(\Omega', \mathfrak{F}')$ un espacio medible, y sea $X : \Omega \longrightarrow \Omega'$ función $\mathfrak{F}$-medible. Si $f : \Omega' \longrightarrow \mathbb{R}$ es $\mathfrak{F}'$-medible, entonces $f \circ X$ es $\mathfrak{F}$-medible. Más aún, $f \circ X$ es P-integrable si y solo si f es μ_X-integrable y

$$Ef(X) = \int_\Omega f(X) \quad dP = \int_{\Omega'} f(\omega') \quad \mu_X(d\omega')$$

donde μ_X es la medida de distribución de probabilidad de la v.a. X.

La demostración del resultado anterior la podemos encontrar por ejemplo en [1, 4, 11, 16].

Una consecuencia del Teorema anterior es que si $\Omega' = \mathbb{R}$ y X es v.a. integrable, entonces

$$EX = \int_\Omega X \quad dP = \int_\mathbb{R} x \quad \mu_X(dx).$$

Teorema 1.3.2. ***(Desigualdad de Chebyshev)***
Sea $\varphi : \mathbb{R} \longrightarrow (0, +\infty)$ *una función creciente tal que* $\varphi(x) = \varphi(-x)$*,* X *variable aleatoria sobre* $(\Omega, \mathfrak{F}, P)$ *tal que* $E[\varphi(X)] < +\infty$*, entonces*

$$\forall \epsilon > 0 \quad P[|X| \geq \epsilon] \leq \frac{E\varphi(X)}{\varphi(\epsilon)}.$$

Un caso particular es cuando $\varphi(x) = |x|^p$

$$P[|X| \geq \epsilon] \leq \frac{E|X|^p}{\epsilon^p}.$$

Una demostración del teorema anterior la podemos encontrar por ejemplo en [4, 5].

Notación: Sea $(\Omega, \mathfrak{F}, P)$ un espacio de probabilidad y $T(\omega)$ una propiedad de los puntos $\omega \in \Omega$. Decimos que la propiedad T se cumple *casi seguramente (c.s.) ó con probabilidad 1 (c.p.1.)* si el conjunto de puntos $\omega \in \Omega$ donde dicha propiedad no se cumple es de probabilidad cero.

Algunos resultados importantes que surgen de la teoría de integración respecto a una medida, y en particular, respecto a una medida de probabilidad, son los siguientes:

Teorema 1.3.3. ***(de la convergencia monótona de Beppo-Levi)***
Sea $(\Omega, \mathfrak{F}, P)$ *un espacio de probabilidad,* $\{X_n\}_n$ *una sucesión creciente de variables aleatorias positivas* $X_n \geq 0$*, y supongamos que* $X = \lim_{n\to\infty} X_n$ *casi seguramente, entonces*

$$\lim_{n\to\infty} EX_n = EX.$$

Teorema 1.3.4. ***(de Beppo-Levi para series)***
Sea $(\Omega, \mathfrak{F}, P)$ *un espacio de probabilidad,* $\{X_n\}_n$ *una sucesión de variables aleatorias positivas. Entonces*

$$\int \left(\sum_{n=1}^{\infty} X_n \right) \; dP = \sum_{n=1}^{\infty} \int X_n \; dP.$$

Teorema 1.3.5. ***(Lema de Fatou)***
Sean $(\Omega, \mathfrak{F}, P)$ espacio de probabilidad, $\{X_n\}_n$ sucesión de variables aleatorias positivas, entonces

$$\int \liminf_{n\to\infty} X_n \quad dP \leq \liminf_{n\to\infty} \int X_n \quad dP.$$

Teorema 1.3.6. ***(de la convergencia dominada de Lebesgue)***
Sean $(\Omega, \mathfrak{F}, P)$ espacio de probabilidad, Y una función integrable y $\{X_n\}_n$ una sucesión de variables aleatorias. Supongamos que las relaciones

(1) $X = \lim_{n\to\infty} X_n$

(2) $|X_n| \leq Y \quad \forall n \in \mathbb{N}$

se cumplen casi seguramente. Entonces X, X_n son funciones integrables para cada $n \in \mathbb{N}$. Además

$$EX = \lim_{n\to\infty} EX_n.$$

Las demostraciones de los resultados anteriores podemos encontrarlas por ejemplo en [1, 4, 5, 15].

Ya que estamos trabajando con la integral de Lebesgue consideremos el espacio $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \lambda)$, nosotros escribimos

$$\int_{\mathbb{R}^n} f \quad d\lambda = \int_{\mathbb{R}^n} f(x) \quad dx$$

y para cada $B \in \mathcal{B}(\mathbb{R})^n$

$$\int_B f(x) \quad dx = \int f I_B \quad d\lambda.$$

1.4. Espacios $\mathbf{L}^p$

En esta sección daremos un breve repaso a los espacios $\mathbf{L}^p$ (ver por ejemplo [4, 15], así como de sus principales propiedades ya que la teoría de los capítulos siguientes estará basada principalmente en las propiedades del espacio $\mathbf{L}^1(\Omega, \mathfrak{F}, P)$.

De aquí en adelante denotaremos por $\mathbf{L}^1$ al espacio $\mathbf{L}^1(\Omega, \mathfrak{F}, P)$.

Definición 1.4.1. *Sea $1 \leq p < +\infty$. Definimos al espacio $\mathfrak{L}^p$ como sigue*

$$\mathfrak{L}^p(\Omega, \mathfrak{F}, P) = \{X : X : \Omega \longrightarrow \mathbb{R} \ v.a., E|X|^p < +\infty\}.$$

Definamos para cada p tal que $1 \leq p < +\infty$ la siguiente función

$$\|X\|_p := (E|X|^p)^{\frac{1}{p}}.$$

Para $p = +\infty$ definimos a $\mathfrak{L}^\infty$ como el espacio vectorial de todas las variables aleatorias $X : \Omega \longrightarrow \mathbb{R}$ acotadas casi seguramente ó con probabilidad 1. Así, existe $M > 0$ tal que $|X(\omega)| \leq M$ c.s., es decir, $\{\omega \in \Omega : |X(\omega)| > M\}$ tiene probabilidad cero. Esto nos permite definir la función

$$\|X\|_\infty := \text{ínf}\{M > 0 : \big\{\omega \in \Omega : |X(\omega)| > M\big\} \text{ tiene probabilidad cero}\}.$$

Algunos resultados importantes para los espacios $\mathfrak{L}^p$ son los siguientes:

Teorema 1.4.1. ***(Desigualdad de Hölder)***
Sean $(\Omega, \mathfrak{F}, P)$ un espacio de probabilidad, $1 \leq p, q \leq +\infty$ exponentes conjugados, es decir, $\frac{1}{p} + \frac{1}{q} = 1$. Sean $X \in \mathfrak{L}^p, Y \in \mathfrak{L}^q$. Entonces $XY \in \mathfrak{L}^1$ y

$$|EXY| \leq E|XY| \leq \|X\|_p \|Y\|_q.$$

Notemos que para $p = 2$ se obtiene la desigualdad de Schwarz.

Teorema 1.4.2. ***(Desigualdad de Minkowski)***
Sea $(\Omega, \mathfrak{F}, P)$ un espacio de probabilidad, $p \in \mathbb{R}$ que satisface $1 \leq p \leq +\infty$, $X, Y \in \mathfrak{L}^p$, entonces

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p.$$

Corolario 1.4.3. *Sea $(\Omega, \mathfrak{F}, P)$ un espacio de probabilidad, y sea p tal que $1 \leq p \leq +\infty$. Entonces $\mathfrak{L}^p$ es un espacio vectorial y $\|.\|_p$ es una seminorma.*

Definición 1.4.2. *Sea $N^p = \{X \in \mathfrak{L}^p(\Omega, \mathfrak{F}, P) : \|X\|_p = 0\}$, claramente N^p es un subespacio vectorial de $\mathfrak{L}^p$. El espacio $\mathbf{L}^p$ es definido como el espacio cociente $\mathfrak{L}^p(\Omega, \mathfrak{F}, P)/N^p(\Omega, \mathfrak{F}, P)$, es decir, el conjunto de todas las clases de equivalencia, inducidas por la relación de equivalencia $\sim$, donde $X \sim Y$ si y solo si $X - Y \in N^p(\Omega, \mathfrak{F}, P)$. Sobre $\mathbf{L}^p$ inducimos la norma $\|\big[f\big]\|_p = \|f\|_p$ donde $[f]$ es la clase de equivalencia en $\mathbf{L}^p$ con representante $f \in \mathfrak{L}^p$.*

Sin pérdida de generalidad confundiremos voluntariamente el espacio seminormado $\mathfrak{L}^p$, con el espacio normado $\mathbf{L}^p$.

Teorema 1.4.4. *Sea* $(\Omega, \mathfrak{F}, P)$ *un espacio de probabilidad y sea* p *tal que* $1 \leq p \leq +\infty$. *Entonces* $\mathbf{L}^p$ *es un espacio de Banach con la norma* $\|.\|_p$.

Enunciaremos el siguiente resultado como un lema, pues será de utilidad posteriormente.

Lema 1.4.1. *Sea* $(\Omega, \mathfrak{F}, P)$ *un espacio de probabilidad, sea* $\mathfrak{A}_0$ *un álgebra sobre* Ω *tal que* $\mathfrak{F} = \sigma(\mathfrak{A}_0)$. *Entonces* $\mathfrak{A}_0$ *es denso en* $\mathfrak{F}$*, es decir, para cada* $A \in \mathfrak{F}$ *y para cada* $\epsilon > 0$ *existe* $A_0 \in \mathfrak{A}_0$ *tal que* $P(A\Delta A_0) < \epsilon$*, donde* $A\Delta A_0 = (A - A_0) \cup (A_0 - A)$ *es la diferencia simétrica de* A *y* A_0.

Las demostraciones de estos resultados las podemos encontrar por ejemplo en [4, 15].

Observación 1.4.1. *Si* $X \in \mathbf{L}^2$*, entonces* $X \in \mathbf{L}^1$.

En general, si $1 \leq p < q$, $X \in \mathbf{L}^q$, entonces $X \in \mathbf{L}^p$, y $\|X\|_p \leq \|X\|_q$, en otras palabras, la inmersión $\mathbf{L}^q \to \mathbf{L}^p$, $X \mapsto X$, es continua.

Definición 1.4.3. *Sea* X *una variable aleatoria,* $a \in \mathbb{R}$*,* $r \in \mathbb{N}$. *Si* $E|X-a|^r < +\infty$*, entonces* $E(X-a)^r$ *es el* r*-ésimo momento de* X *centrado en* a.
Si $a = 0, r = 1 \quad \mu = EX$ *es el valor esperado o esperanza.*
Si $a = \mu, r = 2 \quad VarX = E(X-\mu)^2 = EX^2 - (EX)^2$ *es la varianza.*
Si X, Y *son variables aleatorias tal que* XY *es integrable, entonces*

$$Cov(X,Y) = E(X - EX)(Y - EY).$$

1.5. Modos de convergencia

En esta sección enunciaremos algunas modos de convergencia que usaremos posteriormente (ver referencias [1, 4, 5, 15]).

Sea $(\Omega, \mathfrak{F}, P)$ un espacio de probabilidad, $X_1, X_2, \ldots, X_n, \ldots, X : \Omega \longrightarrow \mathbb{R}$

(1) *Convergencia casi segura.*
Decimos que $\{X_n\}_n$ converge *casi seguramente* (c.s.) a X si existe $N \in \mathfrak{F}$ con $P(N) = 0$, tal que, para cada $\omega \notin N$

$$X(\omega) = \lim_{n\to\infty} X_n(\omega).$$

Este tipo de convergencia puede ser expresada como:

$$X_n \xrightarrow[n\to\infty]{c.s.} X.$$

(2) *Convergencia en probabilidad.*
Diremos que $\{X_n\}_n$ *converge en probabilidad* a X cuando $n \to \infty$ si para cada $\epsilon > 0$

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0.$$

(3) Sea $\{X_n\}_n$ una sucesión de variables aleatorias en $\mathbf{L}^p$, $X \in \mathbf{L}^p$ tal que $E|X_n - X|^p \to 0$ cuando $n \to \infty$, entonces se dice que $\{X_n\}_n$ converge a X en *p-promedio* ó en la norma de $\mathbf{L}^p$. Para $p = 1$ decimos que converge en la media. Para $p = 2$ decimos que converge en media cuadrática.

Observación 1.5.1. *Convergencia en p-promedio equivale a la convergencia en la norma de* $\mathbf{L}^p$.

Relación entre convergencia casi segura y convergencia en probabilidad.

Teorema 1.5.1. *Sean* $(\Omega, \mathfrak{F}, P)$ *un espacio de probabilidad y*

$$X_1, \cdots, X_n, \cdots, X : \Omega \longrightarrow \mathbb{R} \textit{ variables aleatorias.}$$

Supongamos que $\{X_n\}_n$ *converge a* X *casi seguramente, entonces* $\{X_n\}_n$ *converge a* X *en probabilidad.*

Teorema 1.5.2. *Sean* $(\Omega, \mathfrak{F}, P)$ *un espacio de probabilidad,*

$$X_1, \cdots, X_n, \cdots, X : \Omega \longrightarrow \mathbb{R} \textit{ variables aleatorias.}$$

Supongamos que $\{X_n\}_n$ *converge a* X *en probabilidad. Entonces existe una subsucesión* $\{X_{n_k}\}_k$ *de* $\{X_n\}_n$ *la cual converge a* X *casi seguramente.*

Sabemos que en los cursos de teoría de la medida, convergencia uniforme implica convergencia puntual. Sin embargo el recíproco no es cierto en general, pero en un espacio medido finito, el teorema de Egoroff nos proporciona el recípro bajo ciertas hipótesis.

Teorema 1.5.3. ***(Teorema de Egoroff)***
Sean $(\Omega, \mathfrak{F}, P)$ *un espacio de probabilidad,* $X_1, \cdots, X_n, \cdots, X : \Omega \longrightarrow \mathbb{R}$ *variables aleatorias. Si* $X_n \xrightarrow[n\to\infty]{c.s.} X$*, entonces para cada* $\epsilon > 0$ *existe* $B \in \mathfrak{F}$ *tal que* $P(B^c) < \epsilon$*, y* $\{X_n\}_n$ *converge uniformemente a* X *sobre* B*.*

Teorema 1.5.4. *Sean* $(\Omega, \mathfrak{F}, P)$ *un espacio de probabilidad y*

$$X_1, \cdots, X_n, \cdots, X : \Omega \longrightarrow \mathbb{R}$$

funciones que estan en $\mathbf{L}^1$*. Si* $\{X_n\}_n$ *converge a* X *en la media,* $\{X_n\}_n$ *converge a* X *en probabilidad.*

Una generalización del teorema de la convergencia dominada de Lebesgue es

Teorema 1.5.5. *Sean* $(\Omega, \mathfrak{F}, P)$ *un espacio de probabilidad y*

$$X_1, \cdots, X_n, \cdots, X : \Omega \longrightarrow \mathbb{R}$$

funciones que estan en $\mathbf{L}^1$*. Si* $\{X_n\}_n$ *converge a* X *c.s. ó en probabilidad, y si existe* $g : \Omega \longrightarrow [0, +\infty]$ *una función integrable tal que*

(1) $|X_n| \leq g$ *para cada* n*,* $|X| \leq g$ *c.s.,*

entonces $\{X_n\}_n$ *converge a* X *en la media, es decir, converge en la norma de* $\mathbf{L}^1$*:*

$$\lim_{n\to\infty} E|X_n - X| = 0.$$

$$\begin{tikzcd} \boxed{\text{Convergencia casi segura}} \arrow[r, Rightarrow] & \boxed{\text{Convergencia en probabilidad}} \\ & \boxed{\text{Convergencia en la media}} \arrow[u, Rightarrow] \end{tikzcd}$$

Figura 1.1: Relación de los tipos de convergencia.

1.6. Procesos estocásticos

Consideremos un sistema que puede modelar algún fenómeno y que se caracteriza por estar en cualquier conjunto de estados previamente especificado. Supongamos que el sistema evoluciona o cambia de un estado a otro a lo largo del tiempo de acuerdo con una cierta ley de movimiento. Sea X_t el estado del sistema en el tiempo t. Si pensamos que la forma en que evoluciona este sistema es de manera azarosa (donde el azar se mide a través de una medida de probabilidad), podemos considerar a X_t como una variable aleatoria que toma valores para cada t. Esta es una definición intuitiva de un proceso estocástico.

En esta sección daremos elementos básicos de la teoría de procesos estocásticos para posteriormente en capítulos siguientes presentar una clase de ellos llamados *martingalas*.

Definición 1.6.1. *Un proceso estocástico es una familia $\mathcal{X} = \{X_\alpha : \alpha \in \Lambda\}$ de variables aleatorias definidas en un espacio de probabilidad $(\Omega, \mathfrak{F}, P)$, indexadas por un conjunto Λ que usualmente es $\Lambda \subset \mathbb{R}$, llamado el conjunto de parámetros. Las variables aleatorias toman valores en un conjunto S llamado* espacio de estados. *El espacio de estados S, por lo general, es un subconjunto de $\mathbb{R}$ ó $\overline{\mathbb{R}}$ ó $\mathbb{R}^n$.*

Cuando el conjunto de parámetros Λ es numerable, se dice que $\mathcal{X}$ es un proceso con parámetro (o tiempo) discreto. En este caso, y sin pérdida de generalidad, tomamos $\Lambda = \mathbb{N}$ de tal manera que el proceso estocástico $\mathcal{X}$ toma la forma $\{X_1, X_2, ...\}$.

En estas notas consideraremos inicialmente procesos estocásticos a tiempo discreto.

1.7. Clases de Dynkin

Esta breve sección está dedicada a un repaso sobre las clases de Dynkin y a una técnica que es usada para verificar la igualdad en medidas.

Definición 1.7.1. *Sea D un conjunto no vacío. Una colección Π de subconjuntos de D es llamada π sistema si es cerrado bajo intersecciones finitas.*

Definición 1.7.2. *Sea D un conjunto no vacío. Una colección $\mathfrak{D}$ de subconjuntos de D es llamada clase de Dynkin de D si:*

(i) $D \in \mathfrak{D}$.

(ii) Si $A, B \in \mathfrak{D}$ y $B \subset A$, entonces $A - B \in \mathfrak{D}$.

(iii) Si $\{A_n\}_n$ es una sucesión creciente de conjuntos en $\mathfrak{D}$, entonces

$$\bigcup_n A_n \in \mathfrak{D}.$$

Ejemplo 1.7.2.1. *Toda σ-álgebra es una clase de Dynkin.*

Ejemplo 1.7.2.2. *Sean μ, ν dos medidas finitas sobre el espacio medible $(\Omega, \mathfrak{F})$ tales que $\mu(\Omega) = \nu(\Omega)$, entonces $\mathfrak{D} = \{A \in \mathfrak{F} : \mu(A) = \nu(A)\}$ es una clase de Dynkin.*

Sea $\{\mathfrak{D}_\alpha\}_{\alpha\in\Lambda}$ una familia de clases de Dynkin sobre un conjunto D. Entonces la intersección de dicha familia, $\bigcap_{\alpha\in\Lambda} \mathfrak{D}_\alpha$ es también una clase de Dynkin sobre D.

Sea $\mathcal{C}$ una colección de subconjuntos de D. Denotamos por $\mathfrak{D}(\mathcal{C})$ la intersección de todas las clases de Dynkin que contienen a $\mathcal{C}$. Así, $\mathfrak{D}(\mathcal{C})$ es la mínima clase de Dynkin que contiene a $\mathcal{C}$ y se llama la *clase de Dynkin generada por $\mathcal{C}$*.

Teorema 1.7.1 (Teorema de las clases de Dynkin). *Sea D un conjunto no vacío, sea Π un π-sistema sobre D. Entonces la σ-álgebra generada por Π coincide con la clase de Dynkin generada por Π, es decir, $\sigma(\Pi) = \mathfrak{D}(\Pi)$.*

Las demostraciones de los resultados anteriores podemos encontrarlas por ejemplo en [4].

Capítulo 2

Esperanza condicional

En este capítulo estudiaremos el concepto de *esperanza condicional* y sus propiedades más importantes. Este concepto extiende el de *probabilidad condicional*, siendo una herramienta teórica importante en el desarrollo de la teoría de probabilidades. En nuestro caso será útil en la definición de martingala.

Sean $(\Omega, \mathfrak{F}, P)$ un espacio de probabilidad, $C \in \mathfrak{F}$, con $P(C) > 0$. En los cursos básicos de probabilidad se define la *probabilidad condicional respecto o relativa a* C como la medida de probabilidad $P(\cdot|C) : \mathfrak{F} \longrightarrow [0, 1]$ tal que

$$P(A|C) = \frac{P(A \cap C)}{P(C)} \quad \forall A \in \mathfrak{F}. \tag{2.0.1}$$

La integral respecto a esta medida de probabilidad es llamada la *"esperanza condicional relativa a C"* y la denotamos por $E[\cdot|C]$. Así, si X es v.a. integrable respeto a $P(\cdot|C)$, entonces

$$E[X|C] = \int_\Omega X(\omega) P(d\omega|C).$$

Observación 2.0.1. *Notemos que*

$$\begin{aligned} P(A|C) &= \frac{1}{P(C)} \int_C I_A \quad dP \quad \forall A \in \mathfrak{F}, \\ E[X|C] &= \frac{1}{P(C)} \int_C X \quad dP. \end{aligned}$$

Si X es P-integrable, entonces X es $P(\cdot|C)$-integrable.
En caso de que $P(C) = 0$ convenimos en poner $E[X|C] \equiv 0$.

Ahora bien, sea $\mathcal{C}$ la sub σ-álgebra de $\mathfrak{F}$ generada por una partición numerable $\{C_n\}_n$ de conjuntos medibles de Ω (por ejemplo, si Y es v.a. discreta con valores $\{c_n\}_n$, $C_n := [Y = c_n]$ es una partición de Ω de conjuntos medibles y $\mathcal{C} = \sigma(Y)$).

Dada X una v.a. P-integrable, definimos la v.a. $E[X|\mathcal{C}] : \Omega \longrightarrow \mathbb{R}$ por:

$$\begin{aligned} E[X|\mathcal{C}](\omega) &= \sum_{i=1}^{\infty} E[X|C_i] I_{C_i}(\omega) \\ &= \sum_{i=1}^{\infty} \left(\frac{1}{P(C_i)} \int_{C_i} X \ \ dP \right) I_{C_i}(\omega). \end{aligned}$$

Aquí estamos suponiendo que $P(C_i) > 0$ para cada i. Por otro lado, puesto que X es P-integrable, se tiene que $E[X|\mathcal{C}]$ es también P-integrable, más aún

$$\begin{aligned} \int_{\Omega} E[X|\mathcal{C}] \ \ dP &= \sum_{i=1}^{\infty} \left(\frac{1}{P(C_i)} \int_{C_i} X \ \ dP \right) \int_{\Omega} I_{C_i} \ \ dP \\ &= \int_{\Omega} X \ \ dP. \end{aligned}$$

es decir, $E[E[X|\mathcal{C}]] = EX$.
Más aún, si $C \in \mathcal{C}$, entonces C es la unión numerable de una subfamilia de la partición $\{C_n\}_n$, $C = \cup_{i \in \Lambda} C_i$, luego

$$\begin{aligned}
\int_C E[X|\mathcal{C}] \ \ dP &= \sum_{i\in\Lambda} E[X|C_i] \int_C I_{C_i}(\omega) \ \ dP \\
&= \sum_{i\in\Lambda} E[X|C_i] P(C_i) \\
\text{por la observación 2.0.1} \quad &= \sum_{i\in\Lambda} \left(\frac{1}{P(C_i)} \int_{C_i} X \ \ dP \right) P(C_i) \\
&= \sum_{i\in\Lambda} \int_{C_i} X \ \ dP = \int_{\bigcup_{i\in\Lambda} C_i} X \ \ dP \\
&= \int_C X \ \ dP.
\end{aligned}$$

Resumiendo todo lo anterior, tenemos que la v.a. $E[X|\mathcal{C}]$ satisface

(a) $E[X|\mathcal{C}]$ es $\mathcal{C}$-medible.

(b) $\int_C E[X|\mathcal{C}] \ \ dP = \int_C X \ \ dP \quad \forall C \in \mathcal{C}$.

El inciso (b) establece que el promedio ó valor esperado de la v.a. X sobre los eventos $C \in \mathcal{C}$ se puede obtener como el valor esperado de la v.a. $\mathcal{C}$-medible, $E[X|\mathcal{C}]$ sobre los eventos $C \in \mathcal{C}$, lo que es de suma importancia, puesto que el comportamiento de X sobre $\mathcal{C}$ puede ser obtenido a través de una v.a. "minimal" que contenga la mínima información requerida, es decir, que sea $\mathcal{C}$-medible.

Observación 2.0.2. *Si Y es otra v.a. $\mathcal{C}$-medible integrable tal que*

$$\int_C Y \ \ dP = \int_C X \ \ dP \quad \forall C \in \mathcal{C}$$

entonces $Y = E[X|\mathcal{C}]$ c.p.1. (ó c.s.), así pues $E[X|\mathcal{C}]$ es única salvo un conjunto de probabilidad cero.

En lo que sigue extenderemos este concepto a un marco general de tal forma que tenga sentido $E[X|\mathcal{C}]$, donde $\mathcal{C}$ sea cualquier sub σ-álgebra de $\mathfrak{F}$. Para ello necesitamos del teorema de Radon-Nikodym.

2.1. Teorema de Radon-Nikodym

En esta sección enunciaremos el teorema de Radon-Nikodym que establece condiciones bajo las cuales se puede tener la existencia de funciones de densidad para medidas absolutamente continuas respecto a una medida de probabilidad dada.

Extendemos el concepto de esperanza o promedio de una magnitud aleatoria respecto a un conjunto de condiciones o magnitudes aleatorias, haciendo uso del teorema de Radon-Nikodym, permitiendo formalizar y poner bajo un marco teórico preciso el concepto de esperanza condicional.

Definición 2.1.1. *Sean* $(\Omega, \mathfrak{F})$ *un espacio medible,* $\nu : \mathfrak{F} \longrightarrow \mathbb{R}$ *una función. Decimos que* ν *es una* medida signada finita *sobre* $(\Omega, \mathfrak{F})$ *si*

(i) $|\nu(\Omega)| < +\infty$,

(ii) $\nu(\emptyset) = 0$,

(iii) ν *es aditivamente numerable, es decir, si* $\{A_n\}_n$ *es una familia numerable disjunta de conjuntos, elementos de la* σ*-álgebra de* $\mathfrak{F}$*, entonces*

$$\nu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \nu(A_n).$$

Ejemplo 2.1.2. *Sean* $(\Omega, \mathfrak{F}, \mu)$ *espacio medido,* $X : \Omega \longrightarrow \mathbb{R}$ *una función integrable respecto a* μ*. Definimos* $\nu : \mathfrak{F} \longrightarrow \mathbb{R}$ *como*

$$\nu(F) = \int_F X \quad d\mu = \int_F X^+ \quad d\mu - \int_F X^- d\mu \quad \forall F \in \mathfrak{F}.$$

Entonces ν *es una medida signada finita.*

En efecto, aplicando el teorema de Beppo-Levi para series de funciones medibles positivas (ver teorema 1.3.4), a los integrandos X^+ *y* X^-*, vemos que* ν *es aditivamente numerable. También es claro que* $\nu(\emptyset) = 0$*. Así,* ν *es una medida signada finita sobre* $(\Omega, \mathfrak{F})$*.*

Definición 2.1.3. *Sea* $(\Omega, \mathfrak{F}, \mu)$ *un espacio medido,* ν *una medida signada finita sobre* $(\Omega, \mathfrak{F})$*. Decimos que* ν es absolutamente continua a la medida μ*, hecho que denotamos por*

$$\nu << \mu,$$

si vale la implicación

$$A \in \mathfrak{F} \ , \quad \mu(A) = 0 \Longrightarrow \nu(A) = 0.$$

Ejemplo 2.1.4. *En el ejemplo 2.1.2, la medida signada ν es absolutamente continua respecto a μ. Así, la medida signada finita del ejemplo 2.1.2 es una medida signada absolutamente continua respecto a μ.*

Se tiene el siguiente resultado que caracteriza la continuidad absoluta, y cuya demostración puede verse en [4, 15]:

Proposición 2.1.5. *Una condición necesaria y suficiente para que $\nu << \mu$ es que para cada $\epsilon > 0$ existe $\delta > 0$ tal que para cada $A \in \mathfrak{F}$ satisfaciendo la condición $\mu(A) < \delta$, entonces necesariamente $\nu(A) < \epsilon$.*

De la proposición 2.1.5 y el ejemplo 2.1.2 se desprende el siguiente corolario:

Corolario 2.1.6. ***(continuidad absoluta o uniforme de la integral)*** *Si X es integrable respecto a μ, entonces para cada $\epsilon > 0$ existe $\delta > 0$ tal que*

$$\mu(A) < \delta \Longrightarrow \int_A |X| \ \ d\mu < \epsilon.$$

Tenemos una especie de recíproco al ejemplo 2.1.4 anterior en el siguiente teorema conocido como el *teorema de Radon-Nikodym*.

Teorema 2.1.7. ***(de Radon-Nikodym)***
Sea $(\Omega, \mathfrak{F}, \mu)$ un espacio de medida finita o σ-finita, ν una medida signada finita sobre $(\Omega, \mathcal{C})$, (donde $\mathcal{C}$ es una sub σ-álgebra de $\mathfrak{F}$) tal que $\nu << \mu$. Entonces existe una v.a. $g : \Omega \longrightarrow \mathbb{R}$, $\mathcal{C}$-medible, de esperanza finita tal que

$$\nu(F) = \int_F g \ \ d\mu \quad \forall F \in \mathcal{C}.$$

La función g es única salvo igualdad casi en todas partes y se llama la derivada de Radon-Nikodym *de ν respecto a μ, denotándola por $\frac{d\nu}{d\mu}$.*

Las demostración del teorema 2.1.7 se encuentra por ejemplo en [1, 4, 5, 15].

Gracias a este resultado, cobra sentido la definición dada en cursos precedentes sobre variables aleatorias *absolutamente continuas*, el cual volvemos a recoger en la siguiente definición:

Definición 2.1.8. *Sean* $(\Omega, \mathfrak{F}, P)$ *un espacio de probabilidad, una variable aleatoria* $X : \Omega \longrightarrow \mathbb{R}^n$ *se dice ser absolutamente continua si la distribución de probabilidad* μ_X *es absolutamente continua respecto a la medida de Lebesgue de* $\mathbb{R}^n$*, es decir, si existe* $f_X : \mathbb{R}^n \longrightarrow [0, +\infty)$ *función Borel-medible tal que* $\int_{\mathbb{R}^n} f_X(u_1, \cdots, u_n) \ \ du_1 \cdots du_n = 1$ *y*

$$\mu_X(B) = P(X \in B) = \int_B f_X(u_1, \cdots, u_n) \ \ du_1 \cdots du_n \quad \forall B \in \mathcal{B}(\mathbb{R}^n).$$

Notemos que la función de distribución asociada a X *viene dada por*

$$F_X(x_1, \cdots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_X(u_1, \cdots, u_n) \ \ du_1 \cdots du_n.$$

A f_X *se le llama función de densidad de* X*.*

2.2. Esperanzas condicionales

La definición de esperanza condicional puede ser desarrollada como sigue.

Sean $(\Omega, \mathfrak{F}, P)$ un espacio de probabilidad, $\mathcal{C}$ una sub σ-álgebra de $\mathfrak{F}$, $X \in \mathbf{L}^1$ (no necesariamente $\mathcal{C}$-medible). Definimos

$$\nu(C) = \int_C X \ \ dP = E[XI_C] \quad \forall C \in \mathcal{C}. \tag{2.2.1}$$

Claramente ν es una medida signada sobre $(\Omega, \mathfrak{F})$ la cual es absolutamente continua respecto a P. Por el teorema de Radon-Nikodym (ver teorema 2.1.7) existe Y función $\mathcal{C}$-medible tal que

$$\int_C Y \ \ dP = \int_C X \ \ dP \quad \forall C \in \mathcal{C}.$$

Dicha función es única salvo un conjunto de probabilidad cero. A Y la denotaremos por $E[X|\mathcal{C}]$ y es llamada *la esperanza condicional de* X *bajo la sub* σ*-álgebra* $\mathcal{C}$ de tal forma que

$$\nu(C) = \int_C E[X|\mathcal{C}] \ \ dP \quad \forall C \in \mathcal{C}.$$

En resumen, la esperanza condicional de X dado $\mathcal{C}$, $E[X|\mathcal{C}]$, es una variable aleatoria $\mathcal{C}$-medible y está únicamente determinada salvo un conjunto de probabilidad cero, satisfaciendo

$$\int_C X \;\; dP = \int_C E[X|\mathcal{C}] \;\; dP \quad \forall C \in \mathcal{C}.$$

En lo que sigue estableceremos otras propiedades importantes de esperanza condicional.

Teorema 2.2.1. *Sean X una v.a. sobre el espacio medido $(\Omega, \mathfrak{F}, P)$, $\mathcal{C}$ una sub σ-álgebra de $\mathfrak{F}$. Las siguientes propiedades son válidas c.p.1. (o casi seguramente):*

(a) *Si $\mathcal{C} = \{\emptyset, \Omega\}$, entonces $E[X|\mathcal{C}] = EX$.*

(b) *Si $X \geq 0$, entonces $E[X|\mathcal{C}] \geq 0$.*

(c) *Si X es $\mathcal{C}$-medible, entonces $E[X|\mathcal{C}] = X$.*

(d) *Si X =constante= a, entonces $E[X|\mathcal{C}] = a$.*

(e) *Si $X, Y \in \mathbf{L}^1$, entonces $E[aX + bY|\mathcal{C}] = aE[X|\mathcal{C}] + bE[Y|\mathcal{C}]$ $a, b \in \mathbb{R}$.*

(f) *Si $X \leq Y$, entonces $E[X|\mathcal{C}] \leq E[Y|\mathcal{C}]$.*

(g) *Sean $X, Y, XY \in \mathbf{L}^1$, X variable aleatoria $\mathcal{C}$-medible, entonces*

$$E[XY|\mathcal{C}] = XE[Y|\mathcal{C}].$$

En particular $E[E[X|\mathcal{C}]Y|\mathcal{C}] = E[X|\mathcal{C}]E[Y|\mathcal{C}]$.

(h) *Si $\mathcal{H}$ es independiente de las σ-álgebras $\sigma(X)$, $\mathcal{C}$, entonces*

$$E[X|\sigma(\mathcal{C}, \mathcal{H})] = E[X|\mathcal{C}].$$

En particular si $X, \mathcal{H}$ son independientes, entonces $E[X|\mathcal{H}] = EX$.

(i) *Sean $\mathcal{C}_1 \subset \mathcal{C}_2 \subset \mathfrak{F}$ subsigma álgebras de $\mathfrak{F}$, entonces $E[E[X|\mathcal{C}_2]|\mathcal{C}_1] = E[E[X|\mathcal{C}_1]|\mathcal{C}_2] = E[X|\mathcal{C}_1]$.*

Demostración. (a) Sea $h : \Omega \longrightarrow \mathbb{R}$ función constante tal que $h(\omega) = E[X]$ para cada $\omega \in \Omega$. Veamos que h es $\mathcal{C}$-medible; en efecto:

$$\{\omega \in \Omega : h(\omega) \geq r\} = \begin{cases} \Omega & , si \quad E[X] \geq r \\ \emptyset & , si \quad E[X] < r \end{cases}$$

$\forall r \in \mathbb{R}$, y $\{\Omega, \emptyset\} \subset \mathcal{C}$. Como h es $\mathcal{C}$-medible, resta probar que

$$\int_C X\ dP = \int_C h\ dP \quad \forall C \in \mathcal{C} = \{\emptyset, \Omega\}.$$

En efecto , si $C = \Omega$

$$E[X] = \int_\Omega X\ dP = \int_\Omega h\ dP = E[X]\, P(\Omega) = E[X]$$

Si $C = \emptyset$

$$\int_\emptyset X\ dP = 0 = \int_\emptyset h\ dP$$

Por lo tanto

$$\int_C X\ dP = \int_C h\ dP \text{ para cada } C \in \mathcal{C}.$$

Así por unicidad de la esperanza condicional $E[X|\mathcal{C}] = E[X]$.

(b) Sea $h = E[X|\mathcal{C}]$, como $X \geq 0$ entonces

$$\int_C h\ dP = \int_C X\ dP \geq 0 \quad \forall C \in \mathcal{C}.$$

Queremos ver que $h \geq 0$, en efecto, tomemos $C_1 = \{\omega \in \Omega : h(\omega) < 0\} \in \mathcal{C}$. Entonces

$$0 \leq \int_{C_1} h\ dP \leq 0 \quad \Longrightarrow \quad \int_{C_1} h\ dP = 0.$$

Note que

$$hI_{C_1} = hI_{\{\omega\in\Omega: h(\omega)<0\}} = -h^-$$

entonces

$$\int_{C_1} h\ dP = \int_{C_1} -h^- = 0.$$

Como $h^- \geq 0$, necesariamente se tiene que $h^- = 0$ c. s. (c.p.1.), se sigue que $-h^- = 0$ c. s. y $h = h^+ - h^- = h^+$ de aquí que $h \geq 0$.

(c) Este inciso se sigue de la unicidad casi seguramente de la esperanza condicional.

(d) Note que $\int_C a\ dP = \int_C X\ dP = \int_C E[X|\mathcal{C}]\ dP$ y sabemos que la función constante es una función $\mathcal{C}$-medible, para toda $\mathcal{C}$ σ-álgebra; por unicidad de la esperanza condicional, $a = E[X|\mathcal{C}]$.

(e) Note que

$$\begin{aligned}
\int_C E[\alpha X + \beta Y|\mathcal{C}]\ dP &= \int_C (\alpha X + \beta Y)\ dP \\
&= \int_C \alpha X\ dP + \int_C \beta Y\ dP \\
&= \alpha \int_C E[X|\mathcal{C}]\ dP + \beta \int_C E[Y|\mathcal{C}]\ dP \\
&= \int_C (\alpha E[X|\mathcal{C}] + \beta E[Y|\mathcal{C}])\ dP.
\end{aligned}$$

La funciones $\alpha E[X|\mathcal{C}] + \beta E[Y|\mathcal{C}]$ y $E[\alpha X + \beta Y|\mathcal{C}]$ son $\mathcal{C}$-medibles y por unicidad de la esperanza condicional

$$E[\alpha X + \beta Y|\mathcal{C}] = \alpha E[X|\mathcal{C}] + \beta E[Y|\mathcal{C}] \text{ c.s. ó c.p.1.}$$

(f) Definamos $f := Y - X$, $f \geq 0$. Usando (b) y (e) se sigue inmediatamente el resultado.

(g) Sea $\nu(C) = \int_C Y\ dP = \int_C E[Y|\mathcal{C}]\ dP \quad \forall C \in \mathcal{C}$. Siendo X $\mathcal{C}$-medible, por un teorema de cambio de variable, así como la definición de esperanza condicional, tenemos

$$\int_C E[XY|\mathcal{C}]\ dP = \int_C XY\ dP = \int_C X\ d\nu = \int_C XE[Y|\mathcal{C}]\ dP \quad \forall C \in \mathcal{C}.$$

Por ser $X, E[X|\mathcal{C}]$ funciones $\mathcal{C}$-medibles, se tiene que $E[XY|\mathcal{C}] = XE[Y|\mathcal{C}]$ con probabilidad uno.

(h) Supongamos primero que $X \geq 0$ y $E[X] < +\infty$.
Notemos que $\sigma(\mathcal{C}, \mathcal{H})$ es generada por el π-sistema formado por los conjuntos de la forma $C \cap H$, $C \in \mathcal{C}$, $H \in \mathcal{H}$. Usaremos el Teorema

de las clases de Dynkin 1.7.1. Por inciso (b) arriba $E[X|\sigma(\mathcal{C},\mathcal{H})] \geq 0$ y $E[X|\mathcal{C}] \geq 0$. Definamos las medidas finitas positivas sobre el subespacio medible $(\Omega, \sigma(\mathcal{C},\mathcal{H}))$

$$\begin{aligned} \mu(D) &:= \int_D E[X|\sigma(\mathcal{C},\mathcal{H})]\ dP \\ \nu(D) &:= \int_D E[X|\mathcal{C}]\ dP \quad \forall D \in \sigma(\mathcal{C},\mathcal{H}) \end{aligned}$$

claramente

$$\mu(\Omega) = \nu(\Omega) = \int_\Omega X dP = E[X].$$

Por independencia de $\mathcal{H}$ respecto a $\sigma(X)$ y $\mathcal{C}$, tenemos que para cada $C \in \mathcal{C}$, $H \in \mathcal{H}$

$$\begin{aligned} \mu(C\cap H) &= \int_{C\cap H} E[X|\sigma(\mathcal{C},\mathcal{H})]\ dP \\ &= \int_{C\cap H} X\ dP = \left(\int_C X\ dP\right)\left(\int_H\ dP\right) \\ &= \left(\int_C E[X|\mathcal{C}]\ dP\right)\int_H\ dP \\ &= \int_{C\cap H} E[X|\mathcal{C}]\ dP = \nu(C\cap H). \end{aligned}$$

Ahora bien, si definimos por $\mathfrak{D}$ la colección de todos los eventos $A \in \sigma(\mathcal{C},\mathcal{H})$ tal que $\mu(A) = \nu(A)$, se tiene que $\mathfrak{D}$ es una clase de Dynkin (ver ejemplo 1.7.2.2), la cuel contiene al π-sistema formado por los conjuntos de la forma $C\cap H$, $C \in \mathcal{C}$, $H \in \mathcal{H}$. Como consecuencia del Teorema de las clases de Dynkin (ver teorema 1.7.1), la clase de Dynkin $\mathfrak{D}$ coincide con la sub σ-álgebra $\sigma(\mathcal{C},\mathcal{H})$, o sea, se sigue que

$$\int_D E[X|\sigma(\mathcal{C},\mathcal{H})]\ dP = \int_D E[X|\mathcal{C}]\ dP \quad \forall D \in \sigma(\mathcal{C},\mathcal{H}).$$

De esta igualdad y del hecho de que $E[X|\mathcal{C}]$ es $\mathcal{C}$-medible y
$\mathcal{C} \subset \sigma(\mathcal{H},\mathcal{C})$ luego $E[X|\mathcal{C}]$ es $\sigma(\mathcal{C},\mathcal{H})$-medible, por univocidad de la esperanza condicional se sigue

$$E[X|\sigma(\mathcal{C},\mathcal{H})] = E[X|\mathcal{C}].$$

Para el caso general de X una v.a. arbitraria tal que $E[X] < +\infty$ basta poner $X = X^+ - X^-$, considerar el inciso (e) precedente, y lo hecho para el caso de v.a. positivas:

$$\begin{aligned} E[X|\sigma(\mathcal{C},\mathcal{H})] &= E[X^+ - X^-|\sigma(\mathcal{C},\mathcal{H})] \\ &= E[X^+|\sigma(\mathcal{C},\mathcal{H})] - E[X^-|\sigma(\mathcal{C},\mathcal{H})] \\ &= E[X^+|\mathcal{C}] - E[X^-|\mathcal{C}] \\ &= E[X|\mathcal{C}]. \end{aligned}$$

(i) Sean $Y = E[X|\mathcal{C}_2]$, $C_1 \in \mathcal{C}_1$. Note que

$$\begin{aligned} \int_{C_1} E[Y|\mathcal{C}_1]\ dP &= \int_{C_1} Y\ dP \\ &= \int_{C_1} E[X|\mathcal{C}_2]\ dP \\ &= \int_{C_1} X\ dP \text{ (ya que } C_1 \in \mathcal{C}_2) \\ &= \int_{C_1} E[X|\mathcal{C}_1]\ dP. \end{aligned}$$

Por unicidad de esperanza condicional, dado que $E[Y|\mathcal{C}_1], E[X|\mathcal{C}_1]$ son $\mathcal{C}_1$-medibles, entonces $E[E[X|\mathcal{C}_2]|\mathcal{C}_1] = E[X|\mathcal{C}_1]$.
La igualdad $E[E[X|\mathcal{C}_1]|\mathcal{C}_2] = E[X|\mathcal{C}_1]$ se sigue del inciso (c), pues $E[X|\mathcal{C}_1]$ es $\mathcal{C}_2$ medible ya que $\mathcal{C}_1 \subset \mathcal{C}_2$. ■

□

Para una discusión más profunda sobre las propiedades de la esperanza condicional, se pueden consultar [1, 5, 16].

Observación 2.2.2. *Sean $X \in \mathbf{L}_1(\Omega, \mathfrak{F}, P)$, $\mathfrak{C}$ una sub σ-álgebra de $\mathfrak{F}$. Si $E[X|\mathfrak{C}] \leq E[X]$ c.p.1., entonces $E[X|\mathfrak{C}] = E[X]$ c.p.1.*
En efecto:

Sea $Y = E[X] - E[X|\mathfrak{C}] \geq 0$ *c.s., notemos que*

$$\begin{aligned}\int_\Omega Y\ dP &= \int_\Omega E[X]\ dP - \int_\Omega E[X|\mathfrak{C}]\ dP \\ &= E[X]P(\Omega) - \int_\Omega X\ dP \\ &= 0\ c.s.\end{aligned}$$

Siendo $Y \geq 0$ *c.p.1., se sigue sin más que* $Y = 0$ *c.s., es decir,* $E[X|\mathfrak{C}] = E[X]$ *c.p.1.*

A continuación veremos un teorema que será útil para variables aleatorias discretas.

Teorema 2.2.3. *Sean* $\mathfrak{C}$ *una sub* σ*-álgebra de* $\mathfrak{F}$ *generada por eventos disjuntos* $\{B_1, B_2, ..\}$*, donde* $P(B_n) > 0$ *para cada* $n \in \mathbb{N}$*, tal que* $\Omega = \cup_{n=1}^{\infty} B_n$*, entonces para cada* n *y para cada* $\omega \in B_n$*,* $E[X|\mathfrak{C}](\omega) = \frac{E[XI_{B_n}]}{P(B_n)}$.

Dem. Sea $W = \displaystyle\sum_{j=1}^{\infty} \frac{E[XI_{B_i}]}{P(B_i)} I_{B_i}$. Sea $C \in \mathfrak{C}$ arbitrario, entonces existe una subsucesión de índices $\{n_j\}$ tal que $C = \bigcup_j B_{n_j}$. De aquí

$$\begin{aligned}\int_C W\ dP &= \sum_j \int_{B_{n_j}} W dP = \sum_j \int I_{B_{n_j}} W\ dP \\ &= \sum_j \frac{E[XI_{B_{n_j}}]}{P(B_{n_j})} \int I_{B_{n_j}}\ dP = \sum_j E[XI_{B_{n_j}}] = E[XI_C] = \int_C X\ dP.\end{aligned}$$

Puesto que W es $\mathfrak{C}$-medible entonces $W = E[X|\mathfrak{C}]$ c.p.1. ■

Como una aplicación inmediata de la esperanza condicional, podemos extender la noción de probabilidad condicional respecto a un evento, a la probabilidad condicional respecto a una σ-álgebra como sigue:

Definición 2.2.4. *Sean* $\mathcal{C} \subset \mathcal{F}$ *una sub* σ*-álgebra de* $\mathfrak{F}$*,* A *un evento en* $\mathfrak{F}$*. Definimos la probabilidad condicional de* A *respecto a la sub* σ*-álgebra* $\mathcal{C}$ *de* $\mathfrak{F}$*,* $P(A|\mathcal{C})$*, como sigue*

$$P(A|\mathcal{C}) := E[I_A|\mathcal{C}].$$

Así, la probabilidad condicional es una v.a. $\mathcal{C}$-medible.

En particular si $\mathcal{C}$ es numerablemente generada por una familia $\{A_n\}_n$ de conjuntos disjuntos a pares tal que $P(A_n) > 0$ para toda n, tenemos que

$$P(A|\mathcal{C})(\omega) = \sum_{n=1}^{\infty} \frac{P(A \cap A_n)}{P(A_n)} I_{A_n}(\omega)$$

Notemos que estamos extendiendo la noción de probabilidad condicional dada en (2.0.1) a través de una variable aleatoria $\mathcal{C}$-medible.

En seguida mencionaremos algunas propiedades de la probabilidad condicional, cuyas demostraciones son evidentes. Para más propiedades de la probabilidad condicional se puede consultar las referencias ya citadas [1, 5, 16].

Propiedades de la probabilidad condicional

(a) $P(A|\mathcal{C}) \geq 0$ para todo $A \in \mathfrak{F}$.

(b) $P(\Omega|\mathcal{C}) = 1$, $P(\emptyset|\mathcal{C}) = 0$.

(c) Sean $A_1, A_2, ...$ eventos disjuntos elementos de $\mathfrak{F}$, entonces

$$P(\bigcup_n A_n|\mathcal{C}) = \sum_n P(A_n|\mathcal{C}).$$

2.2.1. Formas condicionales de los teoremas de la convergencia monótona, convergencia dominada y lema de Fatou

En esta parte se establecerán las versiones análogas del teorema de la convergencia monótona (teorema 1.3.3), lema de Fatou (teorema 1.3.5), teorema de la convergencia dominada de Lebesgue (teorema 1.3.6) extendidas para esperanzas condicionales; herramientas que son necesarias para la teoría que se estudiará en el siguiente capítulo.

Teorema 2.2.5. ***(Forma condicional del teorema de la convergencia monótona.)*** *Sea $\{X_n\}_n$ una sucesión creciente de variables aleatorias positivas tal que $X_n \xrightarrow[n\to\infty]{c.s.} X$, con $X, X_n \in \mathbf{L}^1$ para cada n. Entonces para cualquier $\mathcal{C}$ sub σ-álgebra de $\mathfrak{F}$*

$$E[X_n|\mathcal{C}] \xrightarrow[n\to\infty]{c.s.} E[X|\mathcal{C}].$$

Demostración. Como $0 \leq X_n \leq X_{n+1} \leq X$ para cada $n \in \mathbb{N}$, y X con esperanza finita, entonces X_n tiene esperanza finita para cada $n \in \mathbb{N}$. Por el inciso (f) del teorema 2.2.1, tenemos

$$0 \leq E[X_n|\mathcal{C}] \leq E[X_{n+1}|\mathcal{C}] \leq E[X|\mathcal{C}] \quad \forall n \text{ c. s.}$$

entonces $\lim_{n\to\infty} E[X_n|\mathcal{C}]$ existe c.s., y además es $\mathcal{C}$-medible. Resta demostrar que para cada $C \in \mathcal{C}$

$$\int_C (\lim_{n\to\infty} E[X_n|\mathcal{C}])\ dP = \int_C E[X|\mathcal{C}]\ dP.$$

Sean $Y := \lim_{n\to\infty} E[X_n|\mathcal{C}]$, y $Y_n := E[X_n|\mathcal{C}]$, por el teorema de la convergencia monótona de Beppo-Levi (teorema 1.3.3)

$$\begin{aligned}
\int_C Y\ dP &= \lim_{n\to\infty} \int_C Y_n\ dP = \lim_{n\to\infty} \int_C E[X_n|\mathcal{C}]\ dP \\
&= \lim_{n\to\infty} \int_C X_n\ dP \text{ (por definición)} \\
&= \int_C X\ dP \text{ (Teorema de Beppo-Levy)} \\
&= \int_C E[X|\mathcal{C}]\ dP
\end{aligned}$$

Por unicidad de la esperanza condicional se sigue que $\lim_{n\to\infty} E[X_n|\mathcal{C}] = E[X|\mathcal{C}]$. □

Teorema 2.2.6. ***(Forma condicional del lema de Fatou.)*** *Sea $\{X_n\}_n$ una sucesión de variables aleatorias positivas en $\mathbf{L}^1$. Si $\mathcal{C}$ es una sub σ-álgebra de $\mathfrak{F}$, tal que $E(\liminf_{n\to\infty} X_n) < +\infty$, entonces*

$$E[\liminf_{n\to\infty} X_n|\mathcal{C}] \leq \liminf_{n\to\infty} E[X_n|\mathcal{C}] \quad \textit{c.p.1.}$$

Demostración. Sea $Z_n = \text{ínf}_{k\geq n} X_k$. Así Z_n es una sucesión creciente de variables aleatorias positivas, tales que

$$\liminf_n X_n = \lim_n Z_n \quad \text{c.p.1.}$$

Por el teorema precedente (teorema 2.2.5) se tiene

$$E[Z_n|\mathcal{C}] \xrightarrow[n\to\infty]{c.s.} E[\liminf_n X_n|\mathcal{C}]. \tag{2.2.2}$$

Por otro lado, note que $Z_n \leq X_k$ para cada $k \geq n$. Por el inciso (f) del teorema 2.2.1

$$E[Z_n|\mathcal{C}] \leq E[X_k|\mathcal{C}].$$

Luego

$$E[Z_n|\mathcal{C}] \leq \inf_{k\geq n} E[X_k|\mathcal{C}]. \tag{2.2.3}$$

De las relaciones (2.2.2) y (2.2.3) se sigue lo que se quiere demostrar. □

Teorema 2.2.7. ***(Forma condicional del teorema de la convergencia dominada de Lebesgue.)*** *Sean $Y, \{X_n\}_n$ una sucesión de variables aleatorias, satisfaciendo $|X_n| \leq Y$ c.p.1. para cada n, con $Y, X_n \in \mathbf{L}^1$ para cada n. Supongamos además que $X_n \xrightarrow[n\to\infty]{c.s.} X$. Luego si $\mathcal{C}$ es una sub σ-álgebra de $\mathfrak{F}$, entonces*

$$E[X_n|\mathcal{C}] \xrightarrow[n\to\infty]{c.s.} E[X|\mathcal{C}].$$

Demostración. Supongamos que las hipótesis anteriores se cumplen para toda $\omega \in \Omega$, y también que $Y \geq 0$. Entonces $\{Y + X_n\}_n$ es una sucesión de funciones medibles positivas tal que

$$\lim_{n\to\infty} (Y + X_n) = Y + X.$$

Entonces por el teorema anterior (teorema 2.2.6)

$$\begin{aligned} E[Y + X|\mathcal{C}] &= E[\liminf_{n\to\infty} (Y + X_n)|\mathcal{C}] \\ &\leq \liminf_{n\to\infty} E[Y + X_n|\mathcal{C}] = E[Y|\mathcal{C}] + \liminf_{n\to\infty} E[X_n|\mathcal{C}] \text{ c. s.} \end{aligned}$$

De aquí se sigue que

$$E[X|\mathcal{C}] \leq \liminf_{n\to\infty} E[X_n|\mathcal{C}] \text{ c. s.} \tag{2.2.4}$$

De manera similar para la sucesión $\{Y - X_n\}_n$ se demuestra que

$$E[Y - X|\mathcal{C}] = [\liminf_{n\to\infty} (Y - X_n)|\mathcal{C}] \leq \liminf_{n\to\infty} E[Y - X_n|\mathcal{C}] \text{ c. s.}$$

lo que implica que

$$E[X|\mathcal{C}] = E[\limsup_{n\to\infty} X_n|\mathcal{C}] \geq \limsup_{n\to\infty} E[X_n|\mathcal{C}] \text{ c. s.} \tag{2.2.5}$$

De las relaciones (2.2.4) y (2.2.5) se tiene el resultado. □

2.2.2. Forma condicional de la desigualdad de Jensen

En este apartado, daremos un breve repaso sobre las funciones convexas, para después dar lugar a la desigualdad de Jensen para esperanzas condicionales.

Definición 2.2.8. *Sea I un intervalo y $g : I \to \mathbb{R}$ una función. Se dice que g es una* función convexa *si para cada $x, y \in I$, $\lambda \in (0, 1)$ se satisface la desigualdad*

$$g(\lambda x + (1 - \lambda)y) \leq \lambda g(x) + (1 - \lambda)g(y).$$

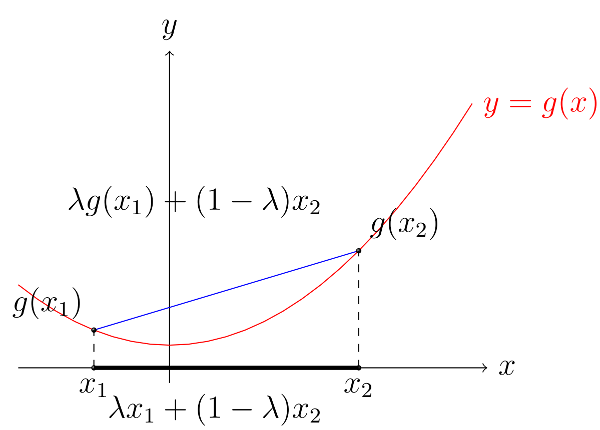

Figura 2.1: Representación geométrica de una función convexa.

Enunciamos algunas propiedades de las funciones convexas, estudiadas en los cursos de cálculo y análisis. Un estudio sistemático de las funciones convexas se tiene en el texto de Royden [15].

Propiedades de las funciones Convexas

- Una función g es cóncava si satisface la desigualdad contraria de la definición; es decir, si g es convexa entonces $-g$ es cóncava.
- Toda combinación lineal con coeficientes positivos de funciones convexas es una función convexa.
- Toda función convexa es continua.
- Toda función convexa $g : (a,b) \longrightarrow \mathbb{R}$ es monótona sobre $\mathbb{R}$, ó bien existe un $x_0 \in (a,b)$ tal que g es decreciente sobre $(a, x_0]$ y creciente sobre $[x_0, b)$.
- Si $g : (a,b) \longrightarrow \mathbb{R}$ es una función convexa, entonces para cada $x_0 \in (a,b)$ existe una recta $y = m(x - x_0) + g(x_0)$ que siempre está por debajo de la gráfica de g; es decir,

  $$m(x - x_0) + g(x_0) \leq g(x) \quad \forall x \in (a,b).$$

  Dicha recta la llamaremos *recta de soporte en el punto* x_0 y siempre estará por debajo de la gráfica de g.

Teorema 2.2.9. ***(Forma condicional de la desigualdad de Jensen)*** *Sea $g : \mathbb{R} \to \mathbb{R}$ una función convexa, y sea X una variable aleatoria tal que $X, g(X) \in \mathbf{L}^1$. Si $\mathcal{C}$ es una sub σ-álgebra de $\mathfrak{F}$, entonces*

$$g(E[X|\mathcal{C}]) \leq E[g(X)|\mathcal{C}] \quad c.s.$$

La demostración de este teorema podemos encontrarla por ejemplo en [5, 15, 16]. Sin embargo haremos un bosquejo de dicha prueba.

Demostración. Considere la recta de soporte en el punto x_0 que debe estar por debajo de la gráfica de g de tal forma que:

$$g(x_0) + m(x - x_0) \leq g(x) \quad \forall x \in \mathbb{R} \tag{2.2.6}$$

donde m es la pendiente de dicha recta de soporte que pasa por el punto $(x_0, g(x_0))$.

Consideremos particularmente $x_0 := E[X|\mathcal{C}]$ y $x = X$, sustituyendo en (2.2.6) obtenemos

$$g\left(E[X|\mathcal{C}]\right) + m(X - E[X|\mathcal{C}]) \leq g(X). \tag{2.2.7}$$

Tomando la esperanza condicional respecto a $\mathcal{C}$ de ambos lados de la desigualdad (2.2.7), para el lado izquierdo se tiene

$$g\left(E[X|\mathcal{C}]\right) + mE[(X - E[X|\mathcal{C}])\,|\mathcal{C}] = g\left(E[X|\mathcal{C}]\right) \tag{2.2.8}$$

ya que $E[(X - E[X|\mathcal{C}])|\mathcal{C}] = 0$. Por otro lado al calcular la esperanza condicional del lado derecho de la desigualdad (2.2.7) se tiene $E[g(X)|\mathcal{C}]$ y por lo tanto de las ecuaciones (2.2.7) y (2.2.8) se tiene que

$$g\left(E[X|\mathcal{C}]\right) \leq E[g(X)|\mathcal{C}].$$

□

Consecuencia: $\big|E[X|\mathcal{C}]\big| \leq E\big[|X|\big|\mathcal{C}\big]$.

Notación: Sean X, Y variables aleatorias, X con esperanza finita. Denotamos $E[X|Y]$ a la esperanza condicional $E[X|\sigma(Y)]$. En general, si Y_1,Y_2,...,Y_n son variables aleatorias, $E[X|Y_1, \cdots, Y_n]$ denotará a $E[X|\sigma(Y_1, \cdots, Y_n)]$. Si $\{Y_\alpha\}_{\alpha\in\Lambda}$ es una familia de variables aleatorias, $E[X|\{Y_\alpha\}_{\alpha\in\Lambda}]$ denotará a $E[X|\sigma(\{Y_\alpha\}_{\alpha\in\Lambda})]$.

Concluimos este apartado considerando un caso particular para la existencia de *densidades condicionales*.

Sean $X \in \mathbf{L}^1$, Y una variable aleatoria. Para cada $B \in \mathcal{B}(\mathbb{R})$, definamos a φ por

$$\varphi(B) = \int_{[Y\in B]} X \ \ dP \tag{2.2.9}$$

Es claro que φ es una medida signada finita boreliana. Consideremos la medida de distribución de Y, μ_Y, es decir,

$$\mu_Y(B) = P[Y \in B] \quad \text{para cada } B \in \mathcal{B}(\mathbb{R}).$$

Claramente φ es una medida absolutamente continua respecto a μ_Y por lo que podemos aplicar el Teorema de Radon-Nikodym (teorema 2.1.7). Definimos $E[X|Y = y]$ como la función Borel medible que está únicamente determinada excepto sobre un conjunto A con $\mu_Y(A) = 0$ por

$$\varphi(B) = \int_B E[X|Y = y] \ \ d\mu_Y(y). \tag{2.2.10}$$

Notemos que $E[X|Y = y]$ es la derivada de Radon-Nikodym de φ con respecto a μ_Y, es decir,

$$E[X|Y = y] = \frac{d\varphi}{d\mu_Y}(y).$$

A $E[X|Y = y]$ se llamará la *densidad condicional* de X dada la condición $Y = y$.

Ahora mostraremos la conección entre esta definición de esperanza condicional y la precedente.

Teorema 2.2.10. *Sea $X \in \mathbf{L}^1$ y sea Y una variable aleatoria. Si g es la función definida en (2.2.10), es decir, $g(y) = E[X|Y = y]$, entonces $g(Y) = E[X|Y]$ c.p.1.*

Demostración. Como g es una función Borel medible y además tenemos que g, $E[X|Y]$ son $\sigma(Y)$-medibles. Necesitamos demostrar

$$\int_C g(Y) \; dP = \int_C E[X|Y] \; dP \quad \forall C \in \sigma(Y).$$

De la definición de $E[X|Y]$, únicamente necesitamos demostrar que

$$\int_C g(Y) \; dP = \int_C X \; dP \quad \forall C \in \sigma(Y).$$

Para cada $C \in \sigma(Y)$ existe un $B \in \mathcal{B}(\mathbb{R})$ tal que

$$C = [Y \in B].$$

Entonces por (2.2.9) y (2.2.10)

$$\begin{aligned} \int_C g(Y) \; dP &= \int_{[Y\in B]} g(Y) \; dP \\ &= \int_B g(y) \; d\mu_Y(y) \\ &= \int_C X \; dP. \end{aligned}$$

Por lo tanto $g(Y) = E[X|Y]$ c.p.1. □

De manera semejante podemos generalizar el resultado anterior estableciendo:

Teorema 2.2.11. *Sean* $(\Omega, \mathfrak{F}, P)$ *un espacio de probabilidad,*

$$Y_1, Y_2, ..., Y_n, X : \Omega \longrightarrow \mathbb{R}$$

variables aleatorias. Supongamos que $E|X| < +\infty$*, entonces existe*

$$g : \mathbb{R}^n \longrightarrow \mathbb{R}$$

una función Borel-medible tal que

$$E[X|Y_1, ..., Y_n] = g(Y_1, .., Y_n).$$

Dicha g *es única* μ_Y*-casi seguramente, donde* Y *es el vector aleatorio definido como* $Y = (Y_1, \cdots, Y_n)$*,* $\mu_Y = P \circ Y^{-1}$*.*
Dicha g *se denotará por* $g(y_1, .., y_n) = E[X|Y_1 = y_1, ..., Y_n = y_n]$*.*

Demostración. Recuerde que $\sigma(Y) = Y^{-1}(\mathcal{B}(\mathbb{R}^n))$. Sea $B \in \mathcal{B}(\mathbb{R}^n)$ y sea

$$\nu(B) := \int_{Y^{-1}(B)} X \; dP \text{ para cada } B \in \mathcal{B}(\mathbb{R}^n)$$

ν es medida signada sobre $\mathcal{B}(\mathbb{R}^n)$ y es finita pues $E[|X|] < \infty$, veamos que $\nu << \mu_Y$, en efecto, sea $B_0 \in \mathcal{B}(\mathbb{R}^n)$ tal que $\mu_Y(B_0) = 0$ lo que implica que

$$\begin{aligned} 0 = \mu_Y(B_0) &= P(Y^{-1}(B_0)) \\ &\Longrightarrow Y^{-1}(B_0) \text{ es } P \text{ despreciable} \\ &\Longrightarrow \nu(B_0) = 0 \end{aligned}$$

por lo tanto, $\nu << \mu_Y$; por el teorema de Radon Nikodym (teorema 2.1.7) existe una única $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ tal que

$$\begin{aligned} \nu(B) &= \int_B g(y_1, .., y_n) \; \mu_Y d(y_1, .., y_n) \\ &= \int g(y_1, .., y_n) I_B(y_1, .., y_n) \; dP \circ Y^{-1} \\ &= \int g(Y) I_B(Y) \; dP \\ &= \int g(Y) I_{Y^{-1}(B)} \; dP \\ &= \int_{Y^{-1}(B)} g(Y) \; dP \end{aligned}$$

por lo tanto $\int_{Y^{-1}(B)} X\ dP = \int_{Y^{-1}(B)} g(Y)\ dP$. De aquí se sigue que

$$\int_{Y^{-1}(B)} E[X|Y_1, ..., Y_n]\ dP = \int_{Y^{-1}(B)} g(Y)\ dP$$

es decir,

$$\int_C E[X|Y_1, ..., Y_n]\ dP = \int_C g(Y)\ dP \text{ para cada } C \in \sigma(Y).$$

Como $E[X|Y_1, ..., Y_n], g(Y)$ son $\sigma(Y)$-medibles, entonces

$$E[X|Y_1, ..., Y_n] = g(Y) = g(Y_1, ..., Y_n).$$

□

Ejemplo 2.2.12. *Sean $X, Y : \Omega \longrightarrow \mathbb{R}$ variables aleatorias, X función integrable, $(X, Y) : \Omega \longrightarrow \mathbb{R}^2$ absolutamente continua, con funciones de densidad $f_Y, f_{X,Y}$ respectivamente. Mediante un cálculo elemental usando la relación (2.2.10) obtenemos que*

$$E[X|Y = y] = \frac{\int_{-\infty}^{+\infty} x f_{X,Y}(x, y)\ \ dx}{f_Y(y)}$$

para aquellas $y \in \mathbb{R}$ tal que $f_Y(y) > 0$.

Demostración. Sea $B \in \mathcal{B}(\mathbb{R})$ un boreliano de $\mathbb{R}$. Luego

$$\begin{aligned}
\int_B E[X|Y = y]\ d\mu_Y(y) &= \int_{Y^{-1}(B)} E[X|Y]\ dP = \int_{Y^{-1}(B)} X\ dP \\
&= \int_\Omega X I_B(Y)\ dP = \iint_{\mathbb{R}^2} x I_B(y)\ dP \circ (X, Y)^{-1}(x, y) \\
&= \iint_{\mathbb{R}^2} x I_B(y) f_{X,Y}(x, y)\ dx\ dy \\
&= \int_B \left\{ \int_{-\infty}^{\infty} x f_{X,Y}(x, y)\ dx \right\}\ dy.
\end{aligned}$$

Pero la integral de la derecha, al considerar que Y tiene densidad de probabilidad f_Y

$$\int_B E[X|Y = y]\ d\mu_Y(y) = \int_B E[X|Y = y] f_Y(y)\ dy.$$

Comparando ambas relaciones:

$$\int_B E[X|Y=y]f_Y(y)\ dy = \int_B \left\{ \int_{-\infty}^{\infty} x f_{X,Y}(x,y)\ dx \right\}\ dy \quad \forall B \in \mathcal{B}(\mathbb{R}),$$

de donde

$E[X|Y=y]f_Y(y) = \displaystyle\int_{-\infty}^{\infty} x f_{X,Y}(x,y)\, dx \quad y-c.s.$ respecto a la medida de Lebesgue en $\mathbb{R}$.

$\square$

2.3. Ejercicios

1. Sea $X \in \mathbf{L}_1(\Omega, \mathfrak{F}, P)$ v.a. integrable, y sea $\mathfrak{B}$ una sub-sigma álgebra de $\mathfrak{F}$. Definamos

$$\varphi(B) := \int_B X \; dP, \qquad \text{para todo } B \in \mathfrak{B}.$$

Pruebe que φ es una medida signada finita sobre $(\Omega, \mathfrak{B}, P)$, la cual es absolutamente continua con respecto a la restricción de P sobre $(\Omega, \mathfrak{B})$.

2. Sean $X, Y \in \mathbf{L}_1(\Omega, \mathfrak{B}, P)$. Pruebe que $X = Y$ c.s. si y sólo si

$$\int_B X \; dP = \int_B Y \; dP, \qquad \text{para todo } B \in \mathfrak{B}.$$

3. Pruebe que toda función convexa es continua.

4. Pruebe que si $g : \mathbb{R} \to \mathbb{R}$ es una función convexa, entonces ó bien g es monótona, ó existe un punto $x_0 \in \mathbb{R}$ tal que g es decreciente en el intervalo $(-\infty, x_0]$, y creciente en el intervalo $[x_0, \infty)$.

5. Sea $(\Omega, \mathfrak{A}, P)$ el espacio de probabilidad del intervalo unitario, y sea $\mathfrak{B}$ la sub-sigma álgebra generada por los intervalos $\{[0, \frac{1}{4}], (\frac{1}{4}, \frac{2}{3}], (\frac{2}{3}, 1]\}$, y sea $X(\omega) = \omega^2$ para todo $\omega \in \Omega = [0, 1]$. Demuestre que $E[X|\mathfrak{B}]$ se puede expresar como

$$E[X|\mathfrak{B}] = a_1 I_{[0,\frac{1}{4}]} + a_2 I_{(\frac{1}{4},\frac{2}{3}]} + a_3 I_{(\frac{2}{3},1]}.$$

Calcule las constantes a_1, a_2, a_3.

6. Si Y es una v.a. discreta tal que $P(Y = y_n) > 0$ para cada $n = 1, 2, \cdots$, y $\sum_{n=1}^{\infty} P(Y = y_n) = 1$. Dado $X \in \mathbf{L}_1(\Omega, \mathfrak{F}, P)$, definamos $\varphi : \mathbb{R} \to \mathbb{R}$ como

$$\varphi(y) = \begin{cases} \int_\Omega X \; dP(\cdot|[Y = y_n]) & \text{si} \quad y = y_n; \\ \\ 0 & \text{si} \quad y \notin \{y_n\}. \end{cases}$$

Pruebe que $\varphi(y) = E[X|Y = y]$ μ_Y-c.s.

7. Pruebe las afirmaciones hechas en el ejemplo 2.2.12.

Bibliografía

[1] Ash, R., Doléans-Dade C., *Probability & Measure Theory*, Academic Press, Reimpreso 2008.

[2] Bergström, H., *Weak Convergence of Measures*, Academic Press, 2014.

[3] Billingsley, P., *Convergence of Probability Measures*, Wiley, New York, 1999.

[4] Cohn, D. L., *Measure Theory*, Birkhäuser, segunda edición, 2010.

[5] Dudley, R. M., *Real Analysis and Probability*, Cambridge University Press, segunda edición, 2011.

[6] Durrett, R., *Probability. Theory and examples*, Cambridge University Press; cuarta edición, 2010.

[7] Ethier, S., Kurtz, T., *Markov Processes: Characterization and Convergence*, Wiley-Interscience; 2nd edition, 2005.

[8] Föllmer, H., Schied, A., *Stochastic Finance: An Introduction in Discrete Time*, de Gruyter; Edición: 4th Rev., ed. 2016.

[9] Gnedenko, B.V., Kolmogorov, A.N., *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley, 1968.

[10] Iglehart, D. L. *Weak convergence in applied probability*, Stoch. Proc. Appl. 2, pp. $211 - 241$, 1974.

[11] Kallenberg, O., *Foundations of Modern Probability*, Springer, Probability and its Applications, 2001.

[12] KALLENBERG, O., *Random Measures, Theory and Applications*, Springer, Probability Theory and Stochastic Modelling, 2017.

[13] KUSHNER, H., *Approximation and Weak Convergence Methods with Applications to Stochastic Systems Theory*, MIT Press, 2008.

[14] KUSHNER, H., *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer, serie: Stochastic Modelling and Applied Probability (Book 24), 2000.

[15] ROYDEN, H. L., *Real Analysis*, Prentice Hall College, tercera edición 1988.

[16] TUCKER, H. G., *A graduate Course in Probability*, Dover Publications, Inc., 2014.