UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

SOBRE TEOREMAS LÍMITE EN EL ESPACIO L^2 EN EL CONTEXTO DE LA TEORÍA DE LA PROBABILIDAD

T E S I S

QUE PARA OBTENER EL TíTULO DE:

A C T U A R I O

P R E S E N T A:

ORLANDO RENÉ MARTÍNEZ MEDRANO

DIRECTOR DEL TRABAJO:

M. EN C. RAYBEL ANDRÉS GARCÍA ANCONA

2016

CIUDAD UNIVERSITARIA, CDMX

Datos del jurado

Datos del alumno:
Orlando René Martínez Medrano
(54 85 39 24)
Universidad Nacional Autónoma de México
Facultad de Ciencias
Actuaría
(306163412)

Datos del director de tesis:
M. en C. Raybel Andrés García Ancona

Datos del sinodal 1:
Dr. Fernando Baltazar Larios

Datos del sinodal 2:
Dr. Sergio Iván López Ortega

Datos del sinodal 3:
M. en C. Ricardo Hoyos Argüelles

Datos del sinodal 4:
M. en C. Álvaro Reyes García

Datos del trabajo escrito:
Sobre teoremas límite en el espacio L^2 en el contexto de la teoría de la probabilidad
90 pag.
2016.

Dedicatoria

A mis padres, la causa primera.

Dejo a Sísifo al pie de la montaña. Se vuelve a encontrar siempre su carga. Pero Sísifo enseña la fidelidad superior que niega a los dioses y levanta las rocas. Él también juzga que todo está bien. Este universo en adelante sin amo no le parece estéril ni fútil. Cada uno de los granos de esta piedra, cada trozo mineral de esta montaña llena de oscuridad forma por sí solo un mundo. El esfuerzo mismo para llegar a las cimas basta para llenar un corazón de hombre. Hay que imaginarse a Sísifo dichoso.

Albert Camus. El mito de Sísifo.

Agradecimientos

A mis padres, Carolina y René, por el regalo de la vida. ¿Puede haber gesto más noble que amar algo que aún no existe? Quizás darle vida. No, eso llega después; pero la acción conlleva al fin. ¡Vaya hermosa paradoja! Gracias por ella, que también contiene al amor, o quizá, éste la contiene.

A mis hermanas, Belén y Cyntia, por su enorme apoyo y compañía. Los días son mejores cuando se encuentran cerca.

A Tatiana, por el amor que alimenta el fuego del corazón y vigoriza el alma; por la sabiduría que sólo puede enseñarse con el ejemplo; por la mejora constante a todo mi ser.

A mi familia, por el humor y las risas. Por el ánimo constante.

A mi director de tesis, el M. en C. Raybel García Ancona, por su enorme paciencia y voluntad para la realización de este trabajo. ¿Qué se gana de un amigo que también es un guía? Un maestro.

A mis sinodales: El Dr. Fernando Baltazar Larios, el Dr. Sergio Iván López Ortega, el M. en C. Ricardo Hoyos Argüelles y el M. en C. Álvaro Reyes García, por el tiempo dedicado a la revisión de la tesis; sus valiosas observaciones mejoraron ampliamente este trabajo.

Índice general

Capítulo 1

Introducción

La Teoría de la Probabilidad es una rama de las matemáticas cuyo objetivo es estudiar los fenómenos aleatorios. El estudio de éstos comenzaron a indagarse hace varios siglos atrás pero sin un rigor matemático adecuado. Fue hasta el siglo XIX cuando Andréi Kolmogórov estableció los fundamentos matemáticos de esta área de estudio.

Paul Malliavin (Septiembre 10, 1925 - Junio 3, 2010) fue un matemático francés que contribuyó de manera notable a la Teoría de la Probabilidad al introducir un conjunto de técnicas que llevan su nombre (Cálculo de Malliavin) para analizar ciertos espacios de probabilidad, llamados de Weiner.

Entre una de sus publicaciones destacadas yace un libro titulado "Integration and Probability", véase [7], en el cual expone un camino para abordar la Teoría de la Probabilidad: una interacción entre diversas disciplinas de las matemáticas; varias ramas de ésta trabajando conjuntamente, entre las cuales destacan el Análisis, el Álgebra, la Topología, la Teoría de la Medida, entre otras.

El objetivo de este trabajo es mostrar esta filosofía y atender parte del conocimiento necesario para entender tanto a la Teoría de la Probabilidad en general, como el Cálculo de Malliavin mencionado anteriormente. Comunmente el trato que se le da a la Probabilidad es en cierto modo aislado, cuando ésta puede ser enriquecida o ampliada con el conocimiento de otras áreas, y que dejan bastantes vacíos durante su estudio cuando es abordada de manera "tradicional".Así, el propósito principal es dar un rigor matemático más sólido a la Teoría Probabilística en el espacio L^2, que el usual.

El capítulo 1 tiene por objetivo dar una prueba del Teorema de Representación de Stone, el cual fundamenta la relación entre los eventos de un álgebra Booleana y el cálculo lógico. Se abordan los espacios de probabilidad, se da una demostración del Teorema de Dynkin cuya importancia radica en el nexo que crea entre dos espacios cocientes de probabilidad y las variables aleatorias. Se dan resultados sobre distribuciones, el concepto de esperanza y diferentes tipos de convergencia.

El capítulo 2 versa sobre la esperanza condicional en el espacio L^2, su extensión a L^1, así como una demostración de la desigualdad de Jensen mediante aproximación por $\sigma-$álgebras finitas. Finalmente el estudio de este operador en el espacio L^p.

El capítulo 3 corresponde a la independencia y ortogonalidad de $\sigma-$álgebras y variables aleatorias. El concepto de independencia tiene un rol central en el estudio de la Teoría de la Probabilidad. Éste es abordado en el espacio L^2, así como su extensión al espacio L^1 entre otros.

El capítulo 4 abarca algunas aplicaciones de lo expuesto en los capítulos anteriores; se exiben algunas demostraciones de la ley débil y fuerte de los grandes números entre otros.

El capítulo 5 es un compendio de los resultados usados a lo largo del trabajo.

El capítulo 6 muestra las conclusiones del trabajo.

Capítulo 2

Fundamentos de Teoría de Probabilidad

2.1. Teorema de Stone

Antes de introducir la noción de probabilidad, se presenta una breve descripción del tipo de modelo matemático usado para representar un sistema físico.

Las representaciones pueden estar dadas desde dos puntos de vista distintos: el punto de vista de la esencia o el punto de vista del fenónemo.

El punto de vista de la esencia, generalmente del matemático puro, consiste en pensar que el sistema físico es perfectamente conocido. Se introduce el espacio de todos los posibles estados, donde un estado es un punto en dicho espacio. Este punto de vista es del mecanismo racional.

Por ejemplo el estado de un sistema de n puntos físicos es completamente determinado por un punto en $\mathbb{R}^{6n}$. Para ello, considérese un vector conformado por n puntos (o estados). Supóngase las dimensiones de longitud, altura y profundidad, en adición con la velocidad, la posición y el tiempo. Bajo esta descripción el sistema físico queda completamente descrito como un vector en $\mathbb{R}^{6n}$.

El punto de vista del fenómeno, generalmente del físico experimental, consiste en observar unos cuantos hechos los cuales ocurren en un sistema muy complejo, en el cual el observador, en un principio, nunca podrá entender su estructura básica. Por ejemplo, un físico puede usar herramientas de la Termodinámica para analizar el fenómeno de un gas sin tener que determinar el estado de todas sus moléculas.

El modelo matemático correspondiente a una representación fenomenológica está basado en un cálculo lógico. Se introduce el conjunto $\mathcal{B}$ de todos los eventos posibles que pueden ser observados y estudiados en el sistema físico. $\mathcal{B}$ está dado con la estructura del cálculo lógico en el cual:

- $A + B$ denota la ocurrencia del evento A o la del evento B
- $A \cdot B$ denota la ocurrencia de ambos eventos, es decir, la ocurrencia del evento A y la de B.
- $\mathbf{0}$ denota la imposibilidad del evento y $\mathbf{1}$ el evento seguro.

Conviene definir los siguientes conceptos.

Definición 1. *Sea X un conjunto abstracto no vacío. Una $\sigma-$álgebra en X es una familia $\mathcal{A}$ de subconjuntos de X que satisfacen los siguientes tres axiomas:*

1. $X \in \mathcal{A}$
2. *Si* $A_1 \in \mathcal{A}$ *entonces* ${A_1}^c \in \mathcal{A}$
3. *Si* $A_i \in \mathcal{A}$ *para toda* $i \in \mathbb{N}$ *entonces* $\bigcup_{i \in \mathcal{N}} A_i \in \mathcal{A}$

Un álgebra Booleana en X es una familia $\mathcal{B}$ de subconjuntos de X que satisfacen los axiomas de la definición anterior, únicamente pidiendo que X sea distinta del vacío y en adición verifican la siguiente propiedad:

1. Cada unión finita de conjuntos en $\mathcal{B}$ está en $\mathcal{B}$.

Proposición 2.1.1. *El conjunto $\mathcal{B}$ de todos los eventos forman un álgebra Booleana.*

Demostración. Se quiere demostrar que $\mathcal{B}$ es un álgebra Booleana. Para ello verificamos las propiedades enunciadas anteriormente.

Por definición, $\mathcal{B}$ es el conjunto de todos los eventos, en particular, el elemento $\mathbf{1} \in \mathcal{B}$. Por otra parte, si $A \in \mathcal{B}$ y $\mathcal{B}$ es el conjunto de todos los posibles eventos, implica que $A^c \in \mathcal{B}$, ya que denota la contraparte del evento A, por lo que A^c también es un evento. Además,para $A_1, ..., A_n \in \mathcal{B}$ y $A_{n+k} = \emptyset$ para toda $k \geq 1$, usando la parte 3 de la definición anterior se sigue que

$$\sum_{k=1}^{n} A_k \in \mathcal{B}.$$

Así, se concluye que $\mathcal{B}$ es álgebra Booleana. $\square$

El punto de vista del fenómeno tiene menos alcance que el punto de vista de la esencia; sin embargo es mucho más adaptable para describir ciertas propiedades importantes.

Por ejemplo considérese un sistema descrito por la información que se suscitó en éste hace veinte años por un álgebra Booleana $\mathcal{B}_0$ de eventos, puede ser descrito hoy, después de un análisis más detallado, por un álgebra Booleana $\mathcal{B}_1$. Todos los eventos que aparecieron hace veinte años en $\mathcal{B}_0$ aparecerán en $\mathcal{B}_1$. Entonces hay un mapeo $\mathcal{B}_0 \mapsto \mathcal{B}_1$, el cual conmuta con las operaciones del cálculo lógico y permite identificar a $\mathcal{B}_0$ con un álgebra $\mathcal{B}_1$.

Para ello considérese la inclusión $\iota{:}\mathcal{B}_0 \mapsto \mathcal{B}_1$, entonces $A_1 = A_2$ ya que es inyectiva y de esta forma se preservan las operaciones del cálculo lógico antes descrito.

La manera para describir y entender el sistema está dada por una sucesión de álgebras Booleanas $\mathcal{B}_0 \mapsto \mathcal{B}_1 \mapsto \mathcal{B}_2 \mapsto \cdots$ donde las flechas son homomorfismos inyectivos de álgebras Booleanas.

Obsérvese que $\{\mathcal{B}_i\}_{i\geq 0}$ es una sucesión creciente de álgebras Booleanas, i.e., $\mathcal{B}_i \subseteq \mathcal{B}_j$ para todo $i < j$, $i, j \in I$, I conjunto indizado, $I \neq \emptyset$. Luego, $\iota{:}\mathcal{B}_i \mapsto \mathcal{B}_j$ es inyectiva y preserva el cálculo lógico definido anteriormente; dicho de otro modo, para cualesquiera $A_p, A_q \in \mathcal{B}_1$ y para todo $p, q \in I$, $p \neq q$

$$\iota(A_p + A_q) = A_p + A_q = \iota(A_p) + \iota(A_q)$$

e

$$\iota(A_p \cdot A_q) = A_p \cdot A_q = \iota(A_p) \cdot \iota(A_q),$$

entonces $(\mathcal{B}_i, +, \cdot)\iota_{\hookrightarrow}(\mathcal{B}_j, +, \cdot)$ es un homomorfismo de álgebras Booleanas abstractas para toda $i < j$, $i, j \in I$.

Resulta ser más natural tratar las operaciones antes definidas con las usuales en la Teoría de Conjuntos. Entonces el álgebra Booleana es un conjunto $\mathcal{B}$ junto con dos operaciones conmutativas y asociativas, denotadas por $A \cup A'$ y $A \cap A'$, donde $A, A' \in \mathcal{B}$. Éstas dos operaciones satisfacen una condición de distributividad.

Dentro de este contexto denotaremos a $\mathbf{0}$ como $\emptyset$ y X como el conjunto universal tales que satisfacen $A \cup \emptyset = A$, $A \cap \emptyset = \emptyset$, $A \cup X = X$, $A \cap X = A$. También, un mapeo denominado complementación,$A \mapsto\ A^c$ de $\mathcal{B}$ en $\mathcal{B}$, el cual satisface $A \cup A^c = X$, $A \cap A^c = \emptyset$ y $(A^c)^c = A$, así como las Leyes de De Morgan. Finalmente $X^c = \emptyset$ y $\emptyset^c = X$.

2.1.1. Relación de Orden

A continuación establecemos una relación de orden entre los elementos del álgebra Booleana.

Dada un álgebra Booleana $\mathcal{B}$ y $A, B \in \mathcal{B}$, se dice que A implica B, y se escribe $A \leq B$ si $A \cap B = A$. Se afirma que $(\leq)$ es una relación de orden en $\mathcal{B}$, esto es, $(\leq)$ es una relación reflexiva, transitiva y antisimétrica. Para ver esto, se tiene lo siguiente:

Si $A \in \mathcal{B}$, $A \cap A = A$ y se sigue que $A \leq A$.

Por otra parte, sean $A, B \in \mathcal{B}$ y supongamos que $A \leq B$ y $B \leq A$. Por un lado se tiene que $A \cap B = A$ y por otro $B \cap A = B$, así $A = B$, es decir, la relación es antisimétrica.

Finalmente sean $A, B, C \in \mathcal{B}$ y asumamos que $A \leq B$ y $B \leq C$. Luego $A \cap (B \cap C) = A$. Por la asociatividad de $\cap$ se obtiene que $(A \cap B) \cap C = A$

y en consecuencia $A \cap C = A$, esto es $A \leq C$. Por tanto $(\leq)$ es relación de orden en $\mathcal{B}$.

Con respecto a este orden, X es el elemento más grande y $\emptyset$ es el más pequeño, esto es, para todo $A \in \mathcal{B}$, $\emptyset \leq A \leq X$. En efecto, sea $A \in \mathcal{B}$ y nótese que $\emptyset \cap A = \emptyset$ lo cual implica que $\emptyset \leq A$ y $X \cap A = A$ implica que $A \leq X$, luego $\emptyset \leq A \leq X$.

Usando la conmutatividad de $\cup$ y $\cap$, notamos que $A \cup B$ y $A \cap B$ son respectivamente, una cota superior e inferior de A y B.

De hecho, $A \cup B$ es la mínima cota superior de A y B y $A \cap B$ es la máxima cota inferior de A y B. Para ver esto, sean $A, B, C \in \mathcal{B}$ tales que $A \leq C$ y $B \leq C$. Luego $(A \cup B) \cap C = (A \cap C) \cup (B \cap C) = A \cup B$ por lo que $A \cup B \leq C$, lo que prueba que $A \cup B$ es la mínima cota superior. Además, nótese que si $D \in \mathcal{B}$ tal que $D \leq A$ Y $D \leq B$, entonces

$$(A \cap B) \cap D = A \cap (B \cap D) = A \cap D = D$$

de donde se sigue que $D \leq (A \cap B)$ lo que prueba que $A \cap B$ es la máxima cota inferior.

Una vez establecida la relación de orden procedemos a dar el modelo de representación matemático de estas ideas.

Debido a que la transición del punto de vista de la esencia al del fenómeno es claro, basta con interpretar la información dada por el enfoque matemático para establecer hechos concretos sobre dicho sistema.

Si Ω es el espacio de estados del sistema físico que se ha comenzado a estudiar, asociamos con un evento A de este sistema el siguiente subconjunto de Ω:

$$A^{'} = \{\omega \in \Omega : \omega \quad \text{satisface el evento A}\} \tag{2.1}$$

Denotaremos a $P(\Omega)$ como el conjunto potencia asociado al conjunto Ω.

Proposición 2.1.2. *La terna* $(P(\Omega), \cup, \cap)$ *es un álgebra Booleana.*

Demostración. Obsérvese que por definición, $P(\Omega) = \{A \subseteq \Omega\}$. En particular $\Omega \in P(\Omega)$ según la definición de $P(\Omega)$. Además si $A \in P(\Omega)$, por definición del mismo, entonces $A^c \in P(\Omega)$. Por la definición, se tiene que $\bigcup_{i=1}^{n} A_i \in \mathcal{B}$. Así la colección $P(\Omega)$ es un álgebra Booleana. $\square$

En resumen, los datos de una representación del fenómeno de un sistema físico del cual se conoce el espacio de estados Ω, son equivalentes a los datos de una subálgebra Booleana de $P(\Omega)$.

2.1.2. Teorema de Representación de Stone

El resultado principal de esta sección es el Teorema de Stone, el cual le proporciona una representación identificable a las álgebras Booleanas. Requeriremos de la noción de filtro para dar continuidad a esta discusión.

Definición 2. *Un filtro $\mathcal{F}$ es una familia de $\mathcal{B}$ no vacía tal que:*

1. *Si $A_1, A_2 \in \mathcal{F}$ entonces $A_1 \cap A_2 \in \mathcal{F}$.*
2. *Si $A_1 \in \mathcal{F}$, $A_1 \leq A_2$, entonces $A_2 \in \mathcal{F}$.*
3. $\emptyset \notin \mathcal{F}$.

Definición 3. *Sean $\mathcal{F}_1, \mathcal{F}_2$ filtros. Se dice que $\mathcal{F}_1$ es más fino que $\mathcal{F}_2$ si para todo $A \in \mathcal{F}_2$, $A \in \mathcal{F}_1$.*

Observemos que la relación de orden dada a los elementos en $\mathcal{B}$ es compatible con la de filtros.

Proposición 2.1.3. *La relación de inclusión en la familia $\mathcal{F}$ de $\mathcal{B}$ define una relación de orden en la familia de filtros.*

Demostración. Claramente $\mathcal{F}_1 \leq \mathcal{F}_1$, por lo que la relación es reflexiva. Sean $\mathcal{F}_1, \mathcal{F}_2 \subseteq \mathcal{B}$ y supóngase que $\mathcal{F}_1 \leq \mathcal{F}_2$ y que $\mathcal{F}_2 \leq \mathcal{F}_1$. Si $B \in \mathcal{F}_2$, entonces $B \in \mathcal{F}_1$ y recíprocamente $B \in \mathcal{F}_2$, luego $\mathcal{F}_1 = \mathcal{F}_2$, y por tanto la relación es antisimétrica.

Finalmente sean $\mathcal{F}_1, \mathcal{F}_2 \leq \mathcal{B}$. Asumimos que $\mathcal{F}_1 \leq \mathcal{F}_2$ y que $\mathcal{F}_2 \leq \mathcal{F}_3$. Por un argumento anteriormente expuesto resulta que $\mathcal{F}_1 \leq \mathcal{F}_3$. Por lo cual se concluye el resultado. □

Por otra parte, cada álgebra Booleana puede ser representada como una subálgebra de $\mathcal{P}(\Omega)$ lo cual es probado en en el siguiente teorema fundamental.

Teorema 2.1.4 (de Stone)**.** *Sea $\mathcal{B}$ un álgebra Booleana. Entonces existe un espacio compacto Ω y una representación identificable de $\mathcal{B}$ con un subálgebra Booleana de $\mathcal{P}(\Omega)$ de subconjuntos que son abiertos y cerrados a la vez en Ω.*

Demostración. La prueba del teorema se muestra en dos partes: el caso finito y el caso infinito. La razón de esto es que el caso general absorbe al caso finito, sin embargo se pierden algunos detalles importantes que vale la pena resaltar.

Supóngase que $\mathcal{B}$ es un álgebra finita. Es decir card$(\mathcal{B}) < \infty$. Defínase como eventos atómicos aquellos que son mínimos en $\mathcal{B}$ con respecto a la relación de orden inclusión; sea entonces Ω el conjunto de eventos atómicos.

Por otra parte, sea $f : \mathcal{B} \to \mathcal{P}(\Omega)$ el mapeo definido por $B \longmapsto \bigcup_q H_q$ siendo $\{H_q\}$ los átomos que conforman a B. En otras palabras $B = \bigcup_q H_q$ en virtud de la proposición A.0.17. Probaremos que tal mapeo es primeramente un mapeo Booleano, es decir, un homomorfismo que respeta la estructura de álgebra de Boole. Obsérvese que

i) $f(\emptyset) = \emptyset$.

ii) $f(A \cup B) = \bigcup_q H_q = \bigcup_i H_i \cup \bigcup_j H_j = f(A) \cup f(B)$.

iii) $f(A \cap B) = \bigcap_q H_q = \bigcap_i H_i \cap \bigcap_j H_j = f(A) \cap f(B)$.

iv) $f(A^c) = \bigcup_q H_q = f^c(A)$.

Ahora, resta probar que f es un mapeo biyectivo. Supóngase que

$$f(A) = f(B) \quad A, B \in \mathcal{B}.$$

Entonces $A = B$ según la definición de f. Luego el mapeo es inyectivo. Por otra parte, sea $B \in \mathcal{B}$. Tomando eventos atómicos H_q en $P(\Omega)$, por la proposición A.0.17 se tiene que para algún $B \in \mathcal{B}$, $f(B) = \cup H_q$. Luego f es sobreyectiva y, por tanto, biyectiva.

Para el caso infinito, sea Ω el espacio de ultrafiltros (consúltese 37) en el álgebra Booleana. Sea φ un mapeo de $\mathcal{B}$ en $P(\Omega)$ definido por

$$\varphi(A) = \{\mathcal{U} \in \Omega : A \in \mathcal{U}\}, A \in \mathcal{B}$$

Ahora, sean $A_1, A_2 \in \mathcal{B}$ y $A_2 \in \mathcal{U}$ tales que $A_1 \geq A_2$ donde $\mathcal{U}$ es un ultrafiltro en Ω. Nótese que $\mathcal{U} \in \varphi(A_2)$ y, por definición de filtro, $A_1 \in \mathcal{U}$, de donde se sigue que $\mathcal{U} \in \varphi(A_1)$. En consecuencia, φ es compatible con la relación de orden y por tanto φ define un homomorfismo de álgebras Booleanas.

Demostraremos que φ es inyectiva.

Supóngase que $A \neq B$, $A, B \in \mathcal{U}$, entonces $A \cap B^c \neq \emptyset$ o $A^c \cap B \neq \emptyset$. Verificaremos el caso en el que $A \cap B^c \neq \emptyset$, el otro caso es análogo. Considérese el siguiente conjunto

$$\mathcal{F} = \{X \in \mathcal{B} / X \geq A \cap B^c\}$$

Se afirma que $\mathcal{F}$ es un filtro. Para demostrar esta afirmación debemos mostrar que se satisfacen las condiciones descritas en la definición 2.

- Supóngase que $A_1 \in \mathcal{F}$ y $A_2 \in \mathcal{F}$, entonces $A_1 \geq A \cap B^c$ y $A_2 \geq A \cap B^c$. Por propiedades conjuntistas $A_1 \cap A_2 \geq (A \cap B^c) \cap (A \cap B^c) = A \cap B^c$.
- Supóngase que $A_1 \leq A_2$ tal que $A_1 \in \mathcal{F}$, entonces $A_1 \in \mathcal{B}$ y $A_1 \geq A \cap B^c$, como $A_1 \leq A_2$, se tiene que $A_2 \geq A \cap B^c$, y por transitividad $A_2 \in \mathcal{F}$.
- Si $\emptyset \in \mathcal{F}$, entonces $\emptyset \in \mathcal{B}$ de tal forma que $\emptyset \geq A \cap B^c$ si y sólo si $A \cap B^c = \emptyset$ lo cual es una contradicción. Luego $\emptyset \notin \mathcal{F}$.

Por tanto $\mathcal{F}$ es un filtro como se quería demostrar.

Ahora, sea $\mathcal{U}$ un ultrafiltro más fino que $\mathcal{F}$. Obsérvese que para todo $A \in \mathcal{F}$ resulta cierto que $A \in \mathcal{U}$ pues $\mathcal{F} \leq \mathcal{U}$, luego $\mathcal{U} \in \varphi(A)$, pero $\mathcal{U} \notin \varphi(B)$, de lo contrario, $B \in \mathcal{U}$ y $A \cap B \in \mathcal{U}$, por tanto $A \cap B \cap A \cap B^c = \emptyset \in \mathcal{U}$ lo cual es una contradicción con el hecho de que $\mathcal{U}$ es un filtro. Por tanto $\varphi(A) \neq \varphi(B)$, probando así que φ es inyectiva.

Por otra parte, para dotar a Ω con una topología, considérese el producto infinito de conjuntos de dos elementos indexados por los elementos del álgebra $\mathcal{B}$, el cual denotaremos por $\Omega_1 = 2^{\mathcal{B}}$. Ya que

$$\Omega_1 = \{0, 1\}^{\mathcal{B}}$$

y $\{0, 1\}$ es un conjunto finito el cual es compacto, por el Teorema de Tychonoff (véase [8]) Ω_1 es compacto.

Defínase el mapeo $\Phi : \Omega \longrightarrow \Omega_1$ de la siguiente forma:

$$\Phi(\mathcal{U}) = \{1_{\mathcal{U}}(A)\}_{A \in \mathcal{B}} \tag{2.2}$$

donde definimos a $1_{\mathcal{U}}(A)$ como:

$$1_{\mathcal{U}}(A) = \begin{cases} 1 & \text{si} \quad A \in \mathcal{U} \\ 0 & \quad \text{e.o.c} \end{cases} \tag{2.3}$$

Obsérvese que Φ es inyectiva por la definición de ultrafiltro. Bajo esta observación, Ω puede ser identificado con un subconjunto de Ω_1. Para ver esto, supóngase que $\mathcal{U} \neq \mathcal{V}$, donde $\mathcal{U}, \mathcal{V} \in \Omega$. Así, sea $A \in \mathcal{U}$ y $A \notin \mathcal{V}$. Luego las funciones indicadoras $1_{\mathcal{U}}(A) \neq 1_{\mathcal{V}}(A)$. En consecuencia φ es función inyectiva.

Sea Ω_1 identificado con el conjunto de funciones f definido en $\mathcal{B}$ y con valores en $\{0, 1\}$. Considérese el lema A.0.19 y defínanse los siguientes conjuntos:

$$\mathcal{W}_A = \{f \in \Omega_1 : f(A) + f(A^c) = 1\} \tag{2.4}$$

Entonces $\mathcal{W}_A$ es subconjunto cerrado de Ω_1, y $W = \bigcap_{A \in \mathcal{B}} \mathcal{W}_A$ es un subconjunto cerrado de Ω_1 pues la intersección arbitraria de conjuntos cerrados es cerrada.

De la misma forma, obtenemos que los siguientes conjuntos son subconjuntos cerrados en Ω_1:

$$\mathcal{S}_{\emptyset} = \{f \in \Omega_1 : f(\emptyset) = 0\}$$

$$\mathcal{T}_{A,A'} = \{f \in \Omega_1 : f(A) \leq f(A^{'}), A \leq A^{'}\}$$
$$\mathcal{V}_{A'',A'''} = \{f \in \Omega_1 : f(A^{''} \cap A^{'''}) = 1\}$$

Procediendo de manera similar con los conjuntos anteriores

$$S = \bigcap \mathcal{S}_{\emptyset}, T = \bigcap_{A,A' \in \mathcal{B}} \mathcal{T}_{A,A'}, V = \bigcap_{A'',A''' \in \mathcal{B}} \mathcal{V}_{A'',A'''}$$

son conjuntos cerrados.

En virtud del lema A.0.20 se sigue que $\Phi(\Omega) = S \cap T \cap V \cap W$. De esta forma hemos probado que $\Phi(\Omega)$ es un subconjunto cerrado de Ω_1. Con la topología inducida por Ω_1, $\Phi(\Omega)$ es compacto. Esto se sigue ya que $\Phi(\Omega)$ es un subconjunto cerrado de un conjunto compacto. Luego, haciendo pull back en esta topología se tiene que Ω es un espacio compacto.

Ahora, fijando $A_0 \in \mathcal{B}$ y definiendo $f_0(\mathcal{U}) = 1_{\mathcal{U}}(A_0)$. Entonces

$$\varphi(A_0) = \{\mathcal{U} \in \Omega : f_0(\mathcal{U}) = 1\}$$

Como f_0 es continua, $\varphi(A_0)$ es un subconjunto cerrado de Ω. Pero $(\varphi(A_0))^c = \varphi(A_0^c)$ es también cerrado, entonces $\varphi(A_0)$ es un subconjunto abierto y cerrado de Ω. De esta forma queda demostrado el Teorema de Stone. □

2.2. Espacios de Probabilidad

Comenzaremos esta sección recordando la definición de medida y espacio medible, útiles para abordar y definir conceptos importantes en el campo de la probabilidad.

Definición 4. *El par $(X, \mathcal{A})$ que consiste del conjunto X junto con una σ-álgebra $\mathcal{A}$ de subconjuntos de X es llamado un espacio medible.*

Definición 5. *Sea $(X, \mathcal{A})$ un espacio medible. Una medida en $(X, \mathcal{A})$ es una función conjuntista $\mu : \mathcal{A} \to \overline{\mathbb{R}}$, donde $\overline{\mathbb{R}}$ es el conjunto de números reales extendidos, con las siguientes propiedades:*

- $\mu(\emptyset) = 0$
- $\mu(E) \geq 0$ *para todo* $E \in \mathcal{A}$
- μ *es* σ*-aditiva, es decir, si* $\{E_n\}$ *es una sucesión de elementos disjuntos entre sí de* $\mathcal{A}$*, entonces* $\mu(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \mu(E_n)$

Definición 6. *Un espacio de medida es una terna* $(X, \mathcal{A}, \mu)$ *en la que* $(X, \mathcal{A})$ *es un espacio medible y* $\mu : \mathcal{A} \to \overline{\mathbb{R}}$ *es una medida.*

Un espacio de probabilidad es un espacio medible $(X, \mathcal{A}, \mu)$ para el cual la medida μ verifica que $\mu(X) = 1$. En el campo de la Teoría de la Probabilidad, la simbología habitual $(\Omega, \mathcal{A}, P)$ representa un espacio de probabilidad. Denominamos $A \in \mathcal{A}$ como un evento el cual es un conjunto medible. Tal medida del conjunto A es llamada la probabilidad de A y la denotamos $P(A)$.

Evidentemente $0 \leq P(A) \leq 1$ pues, por hipótesis, $P(\Omega) = 1$ y como

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c) \geq P(A)$$

entonces $0 \leq P(A) \leq 1$. $P(\cdot)$ es llamada la medida de probabilidad.

Definición 7. *Sea* $(X, \mathcal{A}, \mu)$ *un espacio medible. Un subconjunto* Z *de* X *es llamado negligible si existe* $A \in \mathcal{A}$ *tal que* $\mu(A) = 0$ *y* $Z \subset A$.

Definición 8. *Una propiedad* $(\mathcal{P})$ *se dice que es* μ*-verdadera en casi cualquier parte (*μ*-a.e) en el espacio* $(X, \mathcal{A}, \mu)$ *si*

$$\{x : \mathcal{P} \quad \textit{no se verifica en} \quad x\}$$

esta contenido en un conjunto negligible.

En el campo probabilístico una propiedad que es μ-a.e en Ω es llamada casi segura (a.s).

Enseguida enunciaremos algunos conceptos de mapeos medibles importantes en esta sección.

Definición 9. *Dados dos espacios medibles $(X, \mathcal{A})$ y $(X^{'}, \mathcal{A}^{'})$ un mapeo f de X a $X^{'}$ es llamado medible si $f^{-1}(\mathcal{A}) \subset \mathcal{A}^{'}$. Denotaremos al conjunto de mapeos medibles de $(X, \mathcal{A})$ en $(X^{'}, \mathcal{A}^{'})$ por $\mathcal{M}((X, \mathcal{A}); (X^{'}, \mathcal{A}^{'}))$.*

Definición 10. *En $\mathcal{M}((X, \mathcal{A}); (X^{'}, \mathcal{A}^{'})$ se define la relación de equivalencia por $\underline{f} \sim f^{'}$ si $f(x) = f^{'}(x)$ μ-a.e. La clase de equivalencia de f es denotada por $\overline{f}$.*

Verificaremos que efectivamente $\sim$ es de equivalencia.

- Reflexividad: $f(x) = f(x)$ es μ-a.e. pues el conjunto

$$\{x : f(x) = f(x) \quad \text{no se verifica en} \quad x\} = \emptyset,$$

y siendo el conjunto vacío negligible trivialmente se obtiene $f \sim f$.

- Simetría: Supóngase que $f \sim f^{'}$, lo que implica que existe $A \in \mathcal{A}$ conjunto negligible tal que

$$\{x : f(x) = f^{'}(x) \quad \text{no se verifica en} \quad x\} \subset A$$

y evidentemente $f^{'} \sim f$.

- Transitividad: Si $f \sim f^{'}$ y $f^{'} \sim f^{''}$ resulta cierto que existen $A, B \in \mathcal{A}$ conjuntos negligibles tales que

$$\{x : f(x) = f^{'}(x) \quad \text{no se verifica en} \quad x\} \subset A$$

y

$$\{x : f^{'}(x) = f^{''}(x) \quad \text{no se verifica en} \quad x\} \subset B.$$

Entonces

$$\{x : (f(x) = f^{'}(x)) \cap f^{'}(x) = f^{''}(x) \quad \text{no se verifica en} \quad x\} \subset A \cap B.$$

Luego

$$\{x : f(x) = f^{''}(x) \quad \text{no se verifica en} \quad x\} \subset A \cap B.$$

Se sigue que $f \sim f^{''}$.

De los incisos anteriores se concluye que $f \sim f^{'}$ si $f(x) = f^{'}(x)$ μ-a.e es relación de equivalencia.

Para caracterizar la relación de equivalencia anterior en un espacio de probabilidad será indispensable contar con las siguientes definiciones y algunos resultados generales de espacios medibles.

Definición 11. *El espacio cociente de $\mathcal{M}$ para esta relación de equivalencia es denotado por*

$$\mathcal{M}_\mu((X,\mathcal{A});(X^{'},\mathcal{A}^{'})) = \{[f] : f \in \mathcal{M}((\Omega,\mathcal{A});(\Omega^{'},\mathcal{A}^{'}))\}.$$

Definición 12. *Se denotará por $\mathcal{L}^o_{P'}(X^{'},\mathcal{A}^{'})$ al espacio cociente de*

$$\mathcal{L}^o(X^{'},\mathcal{A}^{'}) = \mathcal{M}((X^{'},\mathcal{A}^{'});(\mathbb{R},\mathcal{B}_\mathbb{R})) \tag{2.5}$$

donde $\mathbb{R}$ denota como de costumbre el campo de los números reales y $\mathcal{B}_\mathbb{R}$ el álgebra de Borel definida sobre el conjunto de los números reales.

Ahora, sea $(\Omega,\mathcal{A},P)$ un espacio de probabilidad y sea $(Y,\mathcal{B})$ un espacio medible. Sea $\Phi : \Omega \to Y$ un mapeo medible:

$$\Phi \in \mathcal{M}((\Omega,\mathcal{A});(Y,\mathcal{B})).$$

Entonces una medida de probabilidad P_1 está definida en $(Y,\mathcal{B})$ por

$$P_1(B) = P(\Phi^{-1}(B)).$$

Verificaremos que efectivamente P_1 es una medida de probabilidad.
i) $P_1(\emptyset) = P(\Phi^{-1}(\emptyset)) = P(\emptyset) = 0$ pues por hipótesis, $P(\cdot)$ es medida de probabilidad.
ii) $P_1(A) = P(\Phi^{-1}(A)) = P(B) \geq 0$ para toda $A \in \mathcal{B}$, $\Phi^{-1}(A) = B$.
iii) Sea $\{A_i\}_{i=1}{}^{\infty}$ una sucesión de eventos medibles disjuntos en $\mathcal{B}$, de modo que $\{\Phi^{-1}(A_i)\}_{i=1}^{\infty}$ también lo será. Entonces

$$\begin{aligned} P_1\left(\bigcup_{i=1}^{\infty} A_i\right) &= P\left(\Phi^{-1}\left(\bigcup_{i=1}^{\infty} A_i\right)\right) = P\left(\bigcup_{i=1}^{\infty} \Phi^{-1}(A_i)\right) \\ &= \sum_{i=1}^{\infty} P(\Phi^{-1}(A_i)) = \sum_{i=1}^{\infty} P_1(A_i). \end{aligned}$$

luego entonces $P_1(\cdot)$ es medida. Más aún, $P_1(Y) = P(\Phi^{-1}(Y)) = P(\Omega) = 1$ y por tanto es de probabilidad.

Reescribiremos a P_1 como $P_1 = \Phi_*(P)$ y la nombraremos la imagen directa, o simplemente la imagen de la medida de probabilidad P bajo el mapeo Φ. $\Phi_*(P)$ es llamada en ocasiones como la medida inducida por Φ en Y.

Proposición 2.2.1. *Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, sea $(Y, \mathcal{B})$ un espacio medible, y sea $\Phi, \Phi' \in \mathcal{M}((\Omega, \mathcal{A}); (Y, \mathcal{B}))$. Si $\Phi(\omega) = \Phi'(\omega)$ a.s., entonces $\Phi_* P = {\Phi_*}' P$*

Demostración. Defínase $A_0 = \{\omega \in \Omega : \Phi(\omega) \neq \Phi'(\omega)\}$. Entonces $P(A_0) = 0$ pues por hipótesis, $\Phi(\omega) = \Phi'(\omega)$ es a.s. Obsérvese que

$$P(A \cup A_0^c) = P(A) + P(A_0^c) - P(A \cap A_0^c) = P(A) + 1 - P(A \cap A_0^c).$$

Ya que $A^c \cap A_0 \subseteq A_0$ y por tanto $P(A^c \cap A_0) = 0$, se tiene que

$$\begin{aligned} P(A \cap A_0^c) = P(A) + 1 - P(A \cup A_0^c) &= P(A) + P(\Omega) - P(A \cup A_0^c) \\ &= P(A) + P(\Omega - A \cup A_0^c) \\ &= P(A) + P(A^c \cap A_0) \\ &= P(A). \end{aligned}$$

Así $P(A) = P(A \cap A_0^c)$ para toda $A \in \mathcal{A}$.

En particular se tiene que $P(\Phi^{-1}(B)) = P(\Phi^{-1}(B) \cap {A_0}^c)$ para todo $B \in \mathcal{B}$. Ahora, sea $\omega \in \Phi^{-1}(B) \cap {A_0}^c$, entonces $\omega \in \Phi^{-1}(B)$, luego $\Phi(\omega) \in B$ y como $\Phi(\omega) = \Phi'(\omega) \in B$, por consiguiente $\Phi^{-1}(B) \cap {A_0}^c \subset {\Phi'}^{-1}(B)$. Por monotonía de la probabilidad resulta que

$$P(\Phi^{-1}(B) \cap {A_0}^c) \leq P({\Phi'}^{-1}(B))$$

es decir $\Phi_* P \leq {\Phi_*}' P$. Un argumento simétrico demuestra la desigualdad opuesta: $\Phi_* P \geq {\Phi_*}' P$. Por tanto $\Phi_* P \doteq {\Phi_*}' P$. $\square$

Corolario 2.2.2. *La imagen directa $\Phi_* P$ depende sólo de la clase de equivalencia de Φ en $\mathcal{M}_P((\Omega, A); (Y, \mathcal{B}))$*

Demostración. En virtud de la proposición anterior $\Phi_* P = {\Phi_*}^{'} P$, lo que equivale a decir que la imagen directa depende exclusivamente del mapeo Φ, y como $\Phi(\omega) = \Phi^{'}(\omega)$ a.s. se obtiene el resultado deseado. □

2.3. Morfismos de Probabilidad

A continuación presentaremos dos operaciones que en la Teoría de la Probabilidad juegan un rol importante:
i) Preservar una medida de probabilidad por un mapeo medible; y
ii) restringir una probabilidad a una sub σ-álgebra.

Definición 13. *Sea $(\Omega, \mathcal{A}, P)$ y $(\Omega^{'}, \mathcal{A}^{'}, P^{'})$ espacios de probabilidad y sea*

$$\Phi \in \mathcal{M}_P((\Omega, \mathcal{A}); (\Omega^{'}, \mathcal{A}^{'}))$$

. Si $\Phi_* P = P^{'}$, Φ es llamado un morfismo de espacios de probabilidad y se dice que preserva probabilidades.

Ahora, sea $\Phi \in \mathcal{M}((\Omega, \mathcal{A}); (\Omega^{'}, \mathcal{A}^{'}))$ y sea $(Y, \mathcal{B})$ un espacio medible. Con $u^{'} \in \mathcal{M}((\Omega^{'}, \mathcal{A}^{'}); (Y, \mathcal{B}))$ asociamos $\Phi^* u^{'}$, su imagen inversa bajo Φ, definida por

$$(\Phi^* u^{'})(\omega) = (u^{'} \circ \Phi)(\omega).$$

Entonces

$$(\Phi^* u^{'}) \in \mathcal{M}((\Omega, \mathcal{A}); (Y, \mathcal{B})).$$

Algunas veces $\Phi_* u^{'}$ es llamado el "pullback"de $u^{'}$.

Si asumimos que $(\Omega, \mathcal{A})$ y $(\Omega^{'}, \mathcal{A}^{'})$ están equipados con medidas de probabilidad P y $P^{'}$ y que Φ es un morfismo de espacios de probabilidad, tenemos las siguientes proposiciones que vienen seguidas de sus respectivas demostraciones.

Proposición 2.3.1. *La clase de equivalencia de* $(\Phi^* u^{'})$ *en* $\mathcal{M}_P((\Omega, \mathcal{A}); (Y, \mathcal{B}))$ *depende sólo de la clase de* $u^{'}$ *en* $\mathcal{M}_{P'}((\Omega^{'}, \mathcal{A}^{'}); (Y, \mathcal{B}))$

Demostración. En efecto. Considérese $u^{'}, u^{''} \in \mathcal{M}_{P'}((\Omega^{'}, \mathcal{A}^{'}); (Y, \mathcal{B}))$. En virtud de la relación de equivalencia definida en 10 se tiene que

$$u^{'}(x) = u^{''}(x) \quad P^{'} - \text{a.s.}$$

Por ser Φ morfismo de probabilidades resulta cierto que

$$u^{'}(\Phi) = u^{''}(\Phi) \quad P - \text{a.s.}$$

y con ello la afirmación antes hecha. □

Proposición 2.3.2. *Sean* $u^{'}, {u_1}^{'} \in \mathcal{M}((\Omega^{'}, \mathcal{A}^{'}); (Y, \mathcal{B}))$ *y sean*

$$A = \{\omega : (\Phi^* u^{'})(\omega) \neq (\Phi^* {u_1}^{'})(\omega)\}, A' = \{\omega^{'} : u^{'}(\omega) \neq ({u_1}^{'})(\omega^{'})\}.$$

Entonces $A = \Phi^{-1}(A^{'})$.

Demostración. Para demostrar tal afirmación nótese que $A \subset \Omega$ y $A^{'} \subset \Omega^{'}$. Se demostrará que $A = \Phi^{-1}(A^{'})$.

$\subseteq$) Sea $\omega \in A$. Implica que

$$(\Phi^* u^{'})(\omega) \neq (\Phi^* {u_1}^{'})(\omega).$$

Luego

$$(u^{'} \circ \Phi)(\omega) \neq ({u_1}^{'} \circ \Phi)(\omega),$$

esto es,

$$u^{'}(\Phi(\omega)) \neq {u_1}^{'}(\Phi(\omega)).$$

Como $\omega^{'} = \Phi(\omega) \in \Omega^{'}$ se tiene

$$u^{'}(\omega^{'}) \neq {u_1}^{'}(\omega^{'}),$$

con lo cual $\omega \in \Phi^{-1}(A^{'})$.

$\supseteq$) Sea $\lambda \in \Phi^{-1}(A^{'})$, esto es

$$\begin{aligned} \Phi^{-1}(A^{'}) &= \{\lambda \in \Omega : u^{'}(\Phi(\lambda)) \neq {u_1}^{'}(\Phi(\lambda))\} \\ &= \{\lambda \in \Omega : (u^{'} \circ \Phi)(\lambda) \neq ({u_1}^{'} \circ \Phi)(\lambda)\} \\ &= \{\lambda \in \Omega : (\Phi^{*}u^{'})(\lambda) \neq (\Phi^{*}{u_1}^{'})(\lambda)\}, \end{aligned}$$

entonces $\lambda \in A$. De ambas contenciones se sigue que $A = \Phi^{-1}(A^{'})$.

Obsérvese que si $P^{'} = \Phi_{*}P$ entonces

$$P(A) = P(\Phi^{-1}(A^{'})) = P^{'}(A^{'}) = 0.$$

□

Por abuso del lenguaje, Φ^{*} denotará la imagen inversa del mapeo inducido por Φ entre los espacios $\mathcal{M}_P$ y $\mathcal{M}_{P'}$.

Proposición 2.3.3. *Sea $\Phi, \Phi_1 \in \mathcal{M}((\Omega, \mathcal{A}); (\Omega^{'}, \mathcal{A}^{'}))$. Supóngase $\Phi = \Phi_1$ a.s. Entonces Φ^{*} y ${\Phi_1}^{*}$ define el mismo mapeo de $\mathcal{M}_{P'}((\Omega^{'}, \mathcal{A}^{'}); (Y, \mathcal{B}))$ a $\mathcal{M}_P((\Omega, \mathcal{A}); (Y, \mathcal{B}))$.*

Demostración. Para probar esta afirmación supóngase lo contrario, esto es, que Φ^{*} y ${\Phi_1}^{*}$ no definen el mismo mapeo de $\mathcal{M}_{P'}$ a $\mathcal{M}_P$. Entonces debe de existir $u^{'} \in \mathcal{M}_{P'}$ tal que $A = \{\omega : \Phi^{*}u^{'} \neq {\Phi_1}^{*}u^{'}\}$ y $P(A) > 0$.

Defínase $A_1 = \{\omega \in \Omega : \Phi(\omega) \neq \Phi_1(\omega)\}$. Obsérvese que $A \subset A_1$ puesto que para cualquier $\omega \in A$, implica que $\Phi^{*}u^{'}(\omega) \neq {\Phi_1}^{*}u^{'}(\omega)$, es decir, $(u^{'} \circ \Phi)(\omega) \neq (u^{'} \circ \Phi_1)(\omega)$, luego $\Phi(\omega) \neq \Phi_1(\omega)$, por tanto $\omega \in A_1$ Además, $P(A_1) = 0$ en virtud de la hipótesis $\Phi = \Phi_1$ a.s. Pero $P(A) > 0$, y $P(A) < P(A_1)$ lo cual es un absurdo. En vista de esta contradicción se sigue la veracidad de la proposición. □

Proposición 2.3.4. *Sea $\Phi_3 = \Phi_2 \circ \Phi_1$. Entonces*

$$\Phi_{3*} = \Phi_{2*} \circ \Phi_{1*} \quad y \quad {\Phi_3}^{*} = {\Phi_1}^{*} \circ {\Phi_2}^{*}.$$

Demostración. La razón de tales igualdades se exponen a continuación. Demostraremos primeramente que $\Phi_{3*} = \Phi_{2*} \circ \Phi_{1*}$. Por definición,

$$\begin{aligned}\Phi_{3*}P &= (\Phi_2 \circ \Phi_1)_* P = P(\Phi_1^{-1} \circ \Phi_2^{-1}) = P \circ \Phi_1^{-1} \circ \Phi_2^{-1} \\ &= (P \circ \Phi_1^{-1}) \circ \Phi_2^{-1} = \Phi_{2*}P(\Phi_1^{-1}) = (\Phi_2)_* \circ (\Phi_1)_* P.\end{aligned}$$

Probaremos ahora que ${\Phi_3}^* = {\Phi_1}^* \circ {\Phi_2}^*$.

$${\Phi_3}^* u^{'} = (\Phi_2 \circ \Phi_1)^* u^{'} = u^{'} \circ \Phi_2 \circ \Phi_1 = \Phi_1^*(u^{'} \circ \Phi_2) = ({\Phi_1}^* \circ {\Phi_2}^*)(u^{'}).$$

□

A continuación enunciamos un par de proposiciones fundamentales para los propósitos de esta subsección.

Proposición 2.3.5 (De Inyectividad)**.** *Sea Φ un morfismo de probabilidad del espacio $(\Omega, \mathcal{A}, P)$ en $(\Omega^{'}, \mathcal{A}', P^{'})$ y sea $(Y, \mathcal{B})$ un espacio medible arbitrario. Entonces Φ^* define un mapeo inyectivo*

$$\mathcal{M}_{P'}((\Omega^{'}, \mathcal{A}^{'}); (Y, \mathcal{B})) \to \mathcal{M}_P((\Omega, \mathcal{A}); (Y, \mathcal{B}))$$

Demostración. Sean $u^{'}, {u_1}^{'} \in \mathcal{M}_{P'}((\Omega^{'}, \mathcal{A}^{'}); (Y, \mathcal{B}))$. Defínanse $u = \Phi^* u^{'}$, $u_1 = \Phi^* {u_1}^{'}$ y el par de conjuntos $A = \{\omega : u \neq u_1\}$ y $A^{'} = \{\omega^{'} : u^{'} \neq {u_1}^{'}\}$. Por las igualdad establecida en 2.3.2 se obtiene que $\Phi^{-1}(A^{'}) = A$. Por hipótesis, Φ es morfismo de probabilidad, de donde $P^{'}(A^{'}) > 0$ implica que $P(A) > 0$. De aquí que Φ^* define un mapeo inyectivo de $\mathcal{M}_{P'}$ a $\mathcal{M}_P$. □

Teorema 2.3.6 (de Dynkin: Medibilidad y Dependencia Funcional)**.** *Sea $(\Omega, \mathcal{A}, P)$ y $(\Omega^{'}, \mathcal{A}^{'}, P^{'})$ dos espacios de probabilidad, sea Φ un morfismo del primero en el segundo, y sea $\mathcal{B} = \Phi^{-1}(\mathcal{A}^{'})$. Entonces $u \in \mathcal{L}_p^o(\Omega, \mathcal{A})$ puede ser escrito en la forma*

$$u = u^{'} \circ \Phi, u^{'} \in \mathcal{L}_{p'}^o(\Omega^{'}, \mathcal{A}^{'})$$

si y sólo si la clase de u contiene una función $\mathcal{B}$-medible.

Demostración. $\Rightarrow$) Supóngase que $u \in \mathcal{L}_p^o(\Omega, \mathcal{A})$ puede ser escrito en la forma $u = u^{'} \circ \Phi$ con $u^{'} \in \mathcal{L}_{p'}^o(\Omega^{'}, \mathcal{A}^{'})$. Ya que Φ es morfismo de probabilidades y $u^{'}$ es medible, u es medible. Así,

$$u^{-1}(B) = (u^{'} \circ \Phi)^{-1}(B) = \Phi^{-1}({u^{'}}^{-1}(B)) \in \mathcal{B}.$$

$\Leftarrow$) Supóngase que u es $\mathcal{B}$-medible. Entonces por el corolario A.0.21, existe una sucesión $\{f_n\}_{n\in\mathbb{N}}$ de funciones $\mathcal{B}$-medibles que convergen puntualmente a u. Si $B \in \mathcal{B}$, entonces existe $A^{'} \in \mathcal{A}^{'}$ tal que $B = \Phi^{-1}(A^{'})$, por lo que se tiene $1_B = \Phi^* 1_{A^{'}}$, lo que implica que cada función $\mathcal{B}$-medible simple satisface la ecuación enunciada en el teorema. Entonces $f_n = u_n^{'} \circ \Phi$, con $u_n{}^{'} \in \mathcal{L}^o_{p^{'}}(\Omega^{'}, \mathcal{A}^{'})$.

Mostraremos que $u_n{}^{'}$ converge a.s. en $\Omega^{'}$.

Supóngase lo contrario, que $u_n{}^{'}$ no converge a.s. en $\Omega^{'}$, entonces existe $\epsilon > 0$ y $A^{'} \in \mathcal{A}^{'}$, con $P^{'}(A^{'}) > 0$ tal que

$$\sup_{m,n>p} \mid u_n{}^{'}(\omega) - u_m{}^{'}(\omega) \mid > \epsilon$$

para todo p entero positivo y $\omega \in A^{'}$. Entonces f_n satisface la desigualdad anterior en $\Phi^{-1}(A^{'})$, lo cual contradice la convergencia a.s. de f_n, ya que $P(\Phi^{-1}(A^{'})) = P^{'}(A^{'}) > 0$.

Entonces, sea $u^{'} = \lim_{n\to\infty} u_n{}^{'} \in \mathcal{L}^o_{p^{'}}(\Omega^{'}, \mathcal{A}^{'})$, luego $u = \lim_{n\to\infty} f_n = u^{'} \circ \Phi$. □

Corolario 2.3.7. *Sea Φ un morfismo de espacios de probabilidad de $(\Omega, \mathcal{A}, P)$ a $(\Omega^{'}, \mathcal{A}^{'}, P^{'})$ y sea $\mathcal{B} = \Phi^{-1}(\mathcal{A}^{'})$. Usando Φ^*, se puede identificar $\mathcal{L}^o_{P^{'}}(\Omega^{'}, \mathcal{A}^{'})$ con el subespacio vectorial de $\mathcal{L}^o_P(\Omega, \mathcal{A})$ consistente de funciones $\mathcal{B}$-medibles.*

Demostración. Por la proposición 2.3.5, Φ^* define un mapeo inyectivo medible de $\mathcal{L}^o_{P^{'}}(\Omega^{'}, \mathcal{A}^{'})$ a $\mathcal{L}^o_P(\Omega, \mathcal{A})$. Ya que $u = u^{'} \circ \Phi$, donde $u^{'} \in \mathcal{L}^o_{P^{'}}(\Omega^{'}, \mathcal{A}^{'})$ en virtud del Teorema de Dynkyn 2.3.6, la clase de u contiene una función $\mathcal{B}$-medible para cada u. □

2.4. Variables Aleatorias y Distribuciones de Variables Aleatorias

A continuación introducimos la definición de variable aleatoria, concepto clave y base del cálculo probabilístico.

Definición 14. *Dado un espacio de probabilidad* $(\Omega, \mathcal{A}, P)$, *una variable aleatoria* X *es un elemento de* $\mathcal{L}_P^o(\Omega, \mathcal{A})$. *Abreviaremos v.a.*

Ejemplo 1. *Considérese el espacio muestral* $\Omega = \{-1, 0, 1\}$ *junto con la* $\sigma-$*álgebra* $\mathcal{A} = \{\emptyset, \{0\}, \{-1, 1\}, \Omega\}$. *Sea* $X : \Omega \to \mathbb{R}$ *la función* $X(\omega) = \omega$. *Entonces* $|X|$ *es variable aleatoria pues para cualquier conjunto boreliano* B,

$$|X|^{-1}(B) = \begin{cases} \Omega & si \quad 0, 1 \in B \\ \{-1, 1\} & si \quad 0 \notin B, 1 \in B \\ \{0\} & si \quad 0 \in B, 1 \notin B \\ \emptyset & si \quad 0, 1 \notin B \end{cases}$$

Definición 15. *La distribución de la variable aleatoria* X *es la imagen directa de* P *bajo* X.

Ejemplo 2. *Si se considera el espacio de probabilidad* $(\Omega, \mathcal{A}, P)$ *con* $P(\omega) = 1/3$, *resulta que la función de distribución de la v.a.* $|X|$ *definida en el ejemplo anterior es*

$$P(|X|^{-1}(B)) = \begin{cases} 1 & si \quad 0, 1 \in B \\ 2/3 & si \quad 0 \notin B, 1 \in B \\ 1/3 & si \quad 0 \in B, 1 \notin B \\ 0 & si \quad 0, 1 \notin B \end{cases}$$

Definiremos algunos conceptos concernientes sobre el espacio de medida que usaremos frecuentemente.

Definición 16. *Diremos que un espacio medible* $(X, \mathcal{A}, \mu)$ *es* $\sigma-$*finito si existe una sucesión* $\{A_n\}_{n\in\mathbb{N}}$ *disjunta de elementos de* $\mathcal{A}$ *tal que* $\bigcup_{n\in\mathbb{N}} A_n = X$.

Definición 17. *Un espacio topológico* X *es* $\sigma-$*compacto si es unión númerable de conjuntos compactos.*

Definición 18. *Una medida de Borel* $\sigma-$*finita en un espacio métrico* (X, d) *es localmente finita si cada elemento de* X *tiene una vecindad abierta de medida finita.*

Definición 19. *Una medida* μ, $\sigma-$*finita en un espacio métrico* (X, d) *es llamada una medida de Radon si es localmente finita y si*

$$\mu(B) = \sup\{\mu(K) : K\, compacto, K \subseteq B\}$$

para cada conjunto boreliano B *de* X.

Enunciamos un resultado que será útil en breve.

Proposición 2.4.1. *Una medida $\sigma-$finita localmente finita en un espacio métrico $\sigma-$compacto es una medida de Radon.*

Para ver una demostración de este resultado puede consultarse [13] página 901.

Obsérvese que $(X_*P)(\mathbb{R}) = P(X^{-1}(\mathbb{R})) = P(\Omega) = 1$. Por otra parte el espacio medible $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, X_*P)$ es $\sigma-$finito.

Considérese la sucesión $\{[n, n+1)\}_{n\in\mathbb{Z}} \subset \mathbb{R}$. Nótese que la sucesión es disjunta y es tal que $\bigcup_{n\in\mathbb{Z}}[n, n+1) = \mathbb{R}$, probando así la $\sigma-$ finitud. También es $\sigma-$compacto pues $\bigcup_{n\in\mathbb{N}}[-n, n] = \mathbb{R}$, siendo $[-n, n]$ compacto. Nótese que la medida X_*P es localmente finita pues para cada vecindad abierta V_x de $x \in \mathbb{R}$ resulta que $X_*P(V_x) \leq X_*P(\mathbb{R}) = 1$. Entonces, por la proposición 2.4.1 X_*P define una medida de Radon.

Definición 20. *Dado un conjunto finito $X_1, X_2, \cdots X_k$ de variables aleatorias definidas en el espacio de probabilidad $(X, \mathcal{A}, P)$, su distribución conjunta es la imagen directa de P bajo el mapeo $\Phi : \Omega \to \mathbb{R}^k$ definida por las $X_p(\omega)$, con $1 \leq p \leq k$.*

Obsérvese que por argumentos similares, X_*P define una medida de Radon.

Por otra parte, sea P_1 la proyección de un elemento de $\mathbb{R}^k$ en su primera componente, sea μ la distribución conjunta de $X_1, \cdots, X_k$ y sea μ_1 la distribución de X_1. Entonces $\mu_1 = (P_1)_*\mu$.

En efecto, sea $A \in \mathcal{B}_{\mathbb{R}^k}$, entonces por 2.3.4

$$\begin{aligned}(P_1)_*\mu(A) &= P_{1*}P(\Phi^{-1}(A)) = (P_1 \circ \Phi)_*P \\ &= P((P_1 \circ \Phi)^{-1}(A)) = P(X_1^{-1}(A)) \\ &= X_{1*}P(A) = \mu_1(A).\end{aligned}$$

2.5. Esperanza Matemática y Distribuciones

Enseguida damos algunas características numéricas de las variables aleatorias que sustentan gran parte de la teoría de la Probabilidad: Esperanza Matemática y Funciones de Distribución.

Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y sea $X \in L^1_p(\Omega, \mathcal{A})$. Entonces la esperanza matemática de X es escrita como $E[X]$ y definida por

$$E[X] = \int X(\omega) dP(\omega).$$

Nótese que la medida P, y el espacio Ω están implícitos en la notación anterior. Bajo esta notación, la norma L^q es expresada como

$$(E[|Y|^q])^{1/q} = \| Y \|_{L^q} .$$

A continuación exponemos un resultado importante referente al cambio de variable.

Proposición 2.5.1. *Sea $(\Omega, \mathcal{A}, P)$ y $(\Omega^{'}, \mathcal{A}^{'}, P^{'})$ espacios de probabilidad, sea Φ un morfismo del primero al segundo, y sea $\Phi^* : L^o(\Omega^{'}, \mathcal{A}^{'}) \to L^o(\Omega, \mathcal{A})$. Sea $u^{'} \in L^1_{p'}(\Omega^{'}, \mathcal{A}^{'})$. Entonces $u = (\Phi^* u^{'}) \in L^1_p(\Omega, \mathcal{A})$ y $E[u] = E[u^{'}]$*

Demostración. Se probará la veracidad de la proposición para funciones simples y, a partir de esto, el caso general. Asúmase que $u^{'}$ es una función simple, es decir, existe $\alpha_k \neq 0$, $k \in \{1, \ldots, n\}$ tales que $u^{'} = \sum \alpha_k 1_{A^{'}_k}$. Así $u = \sum \alpha_k 1_{A_k}$, donde $A_k = \Phi^{-1}(A^{'}_k)$. Como Φ es un morfismo de probabilidad entonces

$$P(A_K) = P(\Phi^{-1}(A^{'}_k)) = P^{'}(A^{'}_k).$$

En consecuencia

$$
\begin{aligned}
E[u] &= \int u(\omega)dP(\omega) = \int \sum_{k=1}^{n} \alpha_k 1_{A_k} dP(\omega) \\
&= \sum_{k=1}^{n} \alpha_k \int 1_{A_k} dP(\omega) = P(A_k) = P^{'}(A_k^{'}) \\
&= \sum_{k=1}^{n} \alpha_k \int 1_{A_k^{'}} dP(\omega^{'}) = \int \sum_{k=1}^{n} \alpha_k 1_{A_k^{'}} dP(\omega^{'}) \\
&= \int u^{'}(\omega^{'})dP^{'}(\omega^{'}) = E[u^{'}].
\end{aligned}
$$

Ahora, sea $v^{'} \in L^1_{p^{'}}(\Omega^{'}, \mathcal{A}^{'})$. Entonces, existe $\{u^{'}_n\}_{n\in\mathbb{N}}$ sucesión de funciones simples tales que

$$E[|v^{'} - u^{'}_n|] = \| v^{'} - u^{'}_n \|_{L^1_{p^{'}}} \to 0 \quad \text{cuando} \quad n \to \infty.$$

Sea $u_n = \Phi^* u^{'}_n$. Obsérvese que, para $n, m \in \mathbb{N}$,

$$E[|u_n - u_m|] = E[|u^{'}_n - u^{'}_m|] = \| u^{'}_n - u^{'}_m \|_{L^1_p} \to 0 \quad \text{conforme} \quad n, m \to \infty.$$

En conclusión, $\{u_n\}_{n\in\mathbb{N}}$ es sucesión de Cauchy en L^1_p. Así, sea $v = \lim\limits_{n\to\infty} u_n$, con $v \in L^1_p$. Por otra parte, existe $\{u_{n_k}\}_{k\in\mathbb{N}}$ subsucesión de $\{u_n\}_{n\in\mathbb{N}}$ la cual es convergente a.s. en Ω y también existe $\{u^{'}_{n_k}\}_{k\in\mathbb{N}}$ subsucesión de $\{u^{'}_n\}_{n\in\mathbb{N}}$ la cual es convergenete a.s. en $\Omega^{'}$. Obsérvese además que $\{u_{n_k}\}_{k\in\mathbb{N}}$ converge a v. Por tanto, $u_n = \Phi^* u^{'}_n = u^{'}_n \circ \Phi$ implica que $\lim\limits_{n\to\infty} u_n = \lim\limits_{n\to\infty} (u^{'}_n \circ \Phi)$, es decir, $v = v^{'} \circ \Phi$, esto es, $v = \Phi^* v^{'}$. Más aún, en virtud del Teorema A.0.23,

$$E[v] = \lim_{n\to\infty} E[u_n] \quad \text{y} \quad E[v^{'}] = \lim_{n\to\infty} E[u_{n^{'}}]$$

y por la unicidad del operador límite, se tiene que $E[v] = E[v^{'}]$ ya que $E[u_n] = E[u^{'}_n]$. $\square$

Como consecuencia directa de la proposición anterior, podemos establecer un método para calcular valores medios o esperados de variables aleatorias $X_1, \cdots, X_k$ en el espacio de probabilidad $(\Omega, \mathcal{A}, P)$.

Proposición 2.5.2. *Sea μ la medida de Radon en $\mathbb{R}^k$ la cual es la distribución de $X_1, \cdots, X_k$. Sea $\varphi \in L^1_\mu$ y sea $Y(\omega) = \varphi(X_1(\omega), \cdots, X_k(\omega))$. Entonces*

$$Y \in L^1_p \quad y \quad E[Y] = \int_{\mathbb{R}^k} \varphi d\mu.$$

Demostración. En virtud de la proposición 2.5.1, $Y \in L^1_p$ y

$$E[Y] = E[\varphi(X_1(\omega), \cdots, X_k(\omega))] = \int_{\mathbb{R}^k} \varphi d\mu.$$

□

Ejemplo 3. *Considérese las v.a's X, Y, Z, cada una de las cuales se distribuye exponencial con parámetro λ y tienen por distribución conjunta*

$$F_{X,Y,Z}(x,y,z) = \begin{cases} (1-e^{-\lambda x})(1-e^{-\lambda y})(1-e^{-\lambda z}) & si \quad x,y,z > 0 \\ 0 & e.o.c \end{cases}$$

Considérese la función $\varphi : \mathbb{R}^3 \to \mathbb{R}$ definida por $Y = \varphi(x,y,z) = \max\{x,y,z\}$. Entonces, por la proposición anterior resulta que

$$\begin{aligned} E[Y] &= \int_0^\infty \int_0^\infty \int_0^\infty \varphi(x,y,z) dF_{X,Y,Z} = \int\int\int_{\{x<y<z\}} \max\{x,y,z\} dF_{X,Y,Z} \\ &= 6 \int_0^\infty \int_0^z \int_0^y z\lambda^3 e^{-(x+y+z)} dxdydz = \frac{11}{6}\lambda^3. \end{aligned}$$

2.6. Nociones de Convergencia en Teoría de la Probabilidad

En vista de la definición de valor esperado, será necesario considerar algunos tipos de convergencia que la involucran y algunas formas más débiles que se desprenden de ésta; después de todo resulta natural preguntarse sobre la forma en que convergen las v.a's de un espacio probabilístico determinado si estos son elementos del espacio normado L^p.

Convenientemente damos algunas definiciones previas para el estudio de la convergencia de sucesiones de variables aleatorias $\{X_n\}$ y Y definidas en el espacio de probabilidad $(\Omega, \mathcal{A}, P)$.

Definición 21. *i) X_n converge a Y casi seguramente si X_n converge casi donde sea a Y, esto es, $\lim_{n\to\infty} X_n = Y \mu -$ a.e.. En virtud del espacio probabilístico $(\Omega, \mathcal{A}, P)$, la medida μ es reemplazada por P.*

ii) X_n converge a Y en media si $\| X_n - Y \|_{L^1} \to 0$ o $E[|X_n - Y|] \to 0$.

iii) X_n converge a Y en media cuadrática si $\| X_n - Y \|_{L^2} \to 0$, o $E[|X_n - Y|^2] \to 0$.

iv) X_n converge a Y en probabilidad si X_n converge a Y en medida, esto es,

$$\lim_{n\to\infty} \mu(\{\omega \in \Omega : |X_n(\omega) - Y(\omega)| \geq \epsilon\}) = 0$$

para cada $\epsilon > 0$. En este sentido, la medida μ conveniente es P debido al espacio de estudio $(\Omega, \mathcal{A}, P)$.

Las relaciones entre estos diferentes tipos de convergencia son resultados bien conocidos y pueden ser consultados en [9].

Se define ahora un concepto de alta relevancia en Teoría de la Probabilidad el cual es crucial en lo consecutivo a este estudio: Convergencia en Distribución. Comenzaremos respaldando este concepto con algunas defininciones previas.

Definición 22. *Se denota por $\mathcal{M}^1(X)$ el conjunto de todas las medidas de Radon ν con signo en X tal que $|\nu|$ es finita y define una norma en $M^1(X)$ por*

$$||\nu||_{M^1} = \int d|\nu| = |\nu|(X).$$

Entiéndase una medida con signo como aquella que toma valores reales finitos.

Definición 23. *Dada una sucesión $\{\nu_n\}_{n\in\mathbb{N}}, \nu_n \in M^1(X)$, se dice que converge débilmente a $\nu_0 \in M^1(X)$ si*

$$\int h d\mu_n \to \int h d\nu_0$$

para toda $h \in C_0(X)$, siendo $C_0(X)$ el espacio vectorial de las funciones continuas que se anulan en el infinito. Recordemos que una función f se dice que se anula en el infinito si, para toda $\epsilon > 0$, existe un conjunto compacto K tal que $|f(x)| < \epsilon$ para $x \notin K$.

Ejemplo 4. *Sea $\mu_n = \delta_n$. Mostraremos que converge débilmente a la medida nula* 0*, siendo δ_x, llamada la medida de Dirac, definida por*

$$\delta_x(B) = \begin{cases} 1 & si \quad x \in B \\ 0 & \quad e.o.c \end{cases}$$

Obsérvese que $\int hd\mu_n = \int hd\delta_n = h(n)$ y, por otra parte, $\int h0 = 0$ para toda $h \in C_0(\mathbb{R})$. Por tanto $h(n) \to 0$.

Definición 24. *Dada una sucesión $\{\nu_n\}_{n\in\mathbb{N}}, \nu_n \in M^1(X)$ se dice que converge estrechamente a $\nu_0 \in M^1(X)$ si*

$$\int kd\nu_n \to \int kd\nu_0,$$

para cada $k \in C_b(X)$, siendo $C_b(X)$ el espacio vectorial de funciones continuas acotadas en X.

A continuación damos la definición de convergencia en distribución, la cual está íntimamente relacionada con la convergencia estrecha.

Definición 25. *Sea $(\Omega_n, \mathcal{A}_n, P_n)$ una sucesión de espacios de probabilidad y sea $(\Omega', \mathcal{A}', P')$ otro espacio de probabilidad. Sea $X_n \in L^o(\Omega_n, \mathcal{A}_n, P_n)$ y $Y \in L^o(\Omega', \mathcal{A}', P')$. Se dice que la sucesión de distribuciones de X_n converge a la distribución de Y si, dada la sucesión $(X_n)_* P_n = \mu_n$ y $Y_* P' = \nu$ para las respectivas distribuciones, μ_n converge estrechamente a v. Si μ_n converge estrechamente diremos que las v.a.'s X_n convergen en distribución.*

Ejemplo 5. *Observémos que $\mu_n = \delta_{\frac{1}{n}}$ converge estrechamente a $\mu = \delta_0$, siendo δ_a la medida de Dirac en a. Como $\int kd\delta_{\frac{1}{n}} = k(\frac{1}{n})$ y $\int kd\delta_0 = k(0)$, con $k \in C_b(R)$. Entonces $k(\frac{1}{n}) \to k(0)$.*

Definición 26. *Una sucesión $\{\mu_n\}_{n\in\mathbb{N}}, \mu_n \in M(X)$ se dice que converge vagamente a $\mu_0 \in M(X)$ si*

$$\int fd\mu_n \to \int fd\mu_0,$$

para toda $f \in C_K(X)$, siendo $\mathcal{C}_K(X)$ el espacio vectorial de funciones continuas con soporte compacto K.

Ejemplo 6. *Sea $\mu_n = \mathcal{N}(0,n)$ la distribución normal con media cero y varianza n en $\mathbb{R}$, ν la medida identicamente nula, esto es $\nu(A) = 0$ para toda $A \in \mathcal{B}_{\mathbb{R}}$ y λ la medida de Lebesgue en $\mathbb{R}$. Mostraremos que μ_n converge vagamente a ν.*

Sea $f \in C_K(\mathbb{R})$. De modo que $|f| \leq ||f||_\infty 1_K(x)$ para toda $x \in \mathbb{R}$. Entonces

$$
\begin{aligned}
|\int f d\mu| \leq \int |f| d\mu \quad &\leq \textstyle\int ||f||_\infty 1_K d\mu = \frac{1}{\sqrt{2\pi n}} \int_A f(x) e^{-\frac{x^2}{2n}} dx \\
&\leq \textstyle\frac{1}{\sqrt{2\pi n}} \int_A ||f(x)||_\infty dx \to 0
\end{aligned}
$$

cuando $n \to \infty$, lo cual prueba que μ_n converge vagamente a ν.

Obsérvese que también es posible demostrar que $\sqrt{2\pi n}\mu_n$ converge vagamente a λ. Para ello, nótese que para toda $f \in C_K(\mathbb{R})$ se tiene que $|f(x)(1 - e^{-\frac{x^2}{n}})| \leq ||f(x)||_\infty 1_K(x)$ para toda $x \in \mathbb{R}$. En virtud del teorema A.0.23 se sigue que

$$
\lim_{n\to\infty} |\int f(x)(1 - e^{-\frac{x^2}{2n}}) dx| \leq \int \lim_{n\to\infty} |f(x)(1 - e^{-\frac{x^2}{2n}})| dx \to 0
$$

probando así que $\sqrt{2\pi n}\mu_n$ converge vagamente a λ.

Enseguida se muestra un criterio de convergencia en distribución.

Teorema 2.6.1. *Las v.a.'s X_n convergen en distribución a la distribución de Y si y sólo si $\lim\limits_{n\to\infty} E[\varphi(X_n)] = E[\varphi(Y)]$ para toda $\varphi \in \mathcal{C}_K(\mathbb{R})$.*

Demostración. $\Longrightarrow$) En este caso la proposición es inmediata pues la convergencia estrecha implica la convergencia vaga.

$\Longleftarrow$) En virtud de la proposición 2.6.1

$$
E[\varphi(X_n)] = \int_{\mathbb{R}} \varphi d\mu_n \quad \text{y} \quad E[\varphi(Y)] = \int_{\mathbb{R}} \varphi d\nu.
$$

Luego

$$
\lim_{n\to\infty} E[\varphi(X_n)] = \lim_{n\to\infty} \int_{\mathbb{R}} \varphi d\mu_n = \int_{\mathbb{R}} \varphi d\nu = E[\varphi(Y)]
$$

con $\varphi \in \mathcal{C}_K(\mathbb{R})$, esto es, μ_n converge vagamente a ν.

Obsérvese que $\mu_n(\mathbb{R}) = 1 = \nu(\mathbb{R})$. Por tanto $\lim_{n\to\infty} \mu_n(\mathbb{R}) = \nu(\mathbb{R})$. Por la proposición A.0.24 μ_n converge estrechamente, esto es, en distribución. $\square$

Como se ha observado, la noción de convergencia en distribución univariada alienta la posibilidad de extender este concepto a un modo multivariado, como a continuación se ofrece.

Una m-úpla ordenada de v.a.'s $X_1, \cdots, X_m$ es llamada una v.a. con valores en $\mathbb{R}^m$, es decir, un vector aleatorio. Se denota esto como

$$\overline{X} \in \mathcal{M}_p((\Omega, \mathcal{A}); (\mathbb{R}^m, \mathcal{B}_{\mathbb{R}^m})).$$

En virtud de la definición de vector aleatorio, dada la v.a. $\overline{X}$ con valores en $\mathbb{R}^m$, su distribución es la distribución conjunta de las X^k la cual queda completamente determinada y es una medida de Radón en $\mathbb{R}^m$, como ya se había hecho observar.

Una sucesión de v.a. con valores en $\mathbb{R}^m$, dígase $X_1, \cdots, X_n, \cdots$, se dice que converge en distribución a $\overline{X_0}$ si la sucesión de distribuciones converge estrechamente a $\overline{X_0}$. Así, se tiene un criterio para la convergencia en distribución, el cual es análogo al de la proposición 2.6.1.

Teorema 2.6.2. *La sucesión de v.a.'s $\overline{X_n}$ con valores en $\mathbb{R}^m$ converge a la distribución de $\overline{X_0}$ si y sólo si*

$$\lim_{n\to\infty} E[\varphi(\overline{X_0})] = E[\Phi(\overline{X_0})], \quad \textit{para toda} \quad \varphi \in \mathcal{C}_K(\mathbb{R}^m).$$

Como resultado inmediato del teorema anterior tenemos lo siguiente.

Teorema 2.6.3. *Si $\overline{X_n}$ converge en distribución a $\overline{X_0}$, entonces cada componente X_n^k converge en distribución a X_0^k.*

Demostración. Sea P_k la proyección de la k-ésima v.a. X_n^k. Entonces, como $\overline{X_n} \to \overline{X_0}$ en distribución, por el teorema 2.6.2,

$$\lim_{n\to\infty} E[P_k(\overline{X_n})] = E[P_k(\overline{X_0})],$$

es decir,

$$\lim_{n\to\infty} E[X_n^k] = E[X_0^k]$$

y por el teorema 2.6.1, se sigue que X_n^k converge en distribución a X_0^k, donde $P_k \in \mathcal{C}_K(\mathbb{R}^m)$. □

En la siguiente proposición, establecemos algunas relaciones entre los diferentes tipos de convergencia y la convergencia en distribución.

Proposición 2.6.4. *i) La convergencia casi segura implica convergencia en distribución.*

ii) Convergencia en probabilidad implica convergencia en distribución.

iii) Convergencia en L_p implica convergencia en distribución.

Demostración. i) Supóngase que $\{X_n\}$ converge a.s. a Y. Sea $\varphi \in \mathbb{C}_K(\mathbb{R})$. Luego $\varphi(X_n) \to \varphi(Y)$ a.s. pues φ es continua. Además φ es acotada; por el teorema de convergencia Dominada A.0.23 resulta resulta que

$$\lim_{n\to\infty} E(\varphi(X_n)) = E[\lim_{n\to\infty} \varphi(X_n)] = E[\varphi(Y)]$$

y por el teorema 2.6.1 $\{X_n\}$ converge en distribución a Y.

ii) Asúmase que $\{X_n\}$ converge en probabilidad a Y. Ahora, existe $\{X_{n_k}\}$ subsucesión de $\{X_n\}$ tal que $\{X_{n_k}\}$ converge a.s. según la proposición A.0.25. Sea $\varphi \in \mathcal{C}_K(\mathbb{R})$, por lo que

$$\lim_{k\to\infty} E[\varphi(X_{n_k})] = E[\varphi(Y)]$$

según el inciso anterior. Así, sea $\beta_n = E[\varphi(X_n)]$ y sea $\gamma = E[\varphi(Y)]$. Entonces, toda subsuceión $\{\beta_{n_k}\}$ de $\{\beta_n\}$ converge a γ, y por ende, $\lim_{n\to\infty} \beta_n = \gamma$, y por el teorema 2.6.1, $\{X_n\}$ converge en distribución a Y.

iii) Como convergencia en L_p implica convergencia en probabilidad, por el inciso anterior se sigue el resultado.

□

Capítulo 3

Esperanza Condicional

Retomemos el punto de vista fenomenológico tratado en el capítulo anterior: el conjunto de todas las mediciones que un experimentador puede hacer posiblemente en un sistema físico es representado por un álgebra Booleana $\mathcal{B}$. El físico esta interesado en exhibir las leyes de la naturaleza o el comportamiento del fenómeno en el contexto de $\mathcal{B}$, dadas ciertas medidas, pretende predecir los valores de otros.

Hay dos tipos de predicciones. La primera involucra una dependencia funcional. Por ejemplo, en la ecuación $E = \frac{1}{2}mv^2$, la medición de dos cantidades determina completamente la tercera. La segunda involucra una "correlación" sin necesidad; por ejemplo, una caída substancial de la presión barométrica hace más probable que un ciclón se aproxime.

El experimentador representa la información conocida sobre el sistema físico por una subálgebra $\mathcal{B}'$ de $\mathcal{B}$. Dada una cantidad física X, se pregunta las siguientes cuestiones: a) ¿X está determinada por la información de $\mathcal{B}'$? Esto es, en términos del teorema 2.3.6, ¿es X medible con respecto a la $\sigma -$ álgebra generada por $\mathcal{B}'$?

b) Si no, el experimentador tratará de extraer de la información $\mathcal{B}'$ todo lo que implica sobre X. ¿Cuál es el valor más probable de X? ¿Se arriesga haciendo un error mayor al tomar el valor más probable como el valor de X?

En términos más concretos, el problema en cuestión puede ser planteado de la siguiente forma: Dado un espacio de probabilidad $(\Omega, \mathcal{A}, P)$, una sub $\sigma-$álgebra $\mathcal{A}'$ de $\mathcal{A}$, y $X \in L^o(\Omega, \mathcal{A}, P)$, ¿Puede ser X aproximado por $Y \in L^o(\Omega, \mathcal{A}', P|_{\mathcal{A}'})$, siendo $P|_{\mathcal{A}}$ la restricción de P a $\mathcal{A}'$?

En la siguiente sección, se intentará resolver este problema usando una aproximación que minimice la norma L^2. En forma sintetizada, una proyección ortogonal en L^2.

3.1. Esperanza Condicional como Operador en L^2

Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y sea $\mathcal{B}$ una sub$\sigma-$álgebra de $\mathcal{A}$. El espacio $L^p(\Omega, \mathcal{B}, P)$ será abreviado como L^p.

Lema 3.1.1. *Sea $1 \leq p \leq \infty$. Entonces $L^p(\mathcal{B})$ puede ser identificado con un subespacio vectorial cerrado de $L^p(\mathcal{A})$.*

Demostración. Obsérvese que una función $\mathcal{B}-$medible es $\mathcal{A}-$medible, esto es, $L^o(\Omega, \mathcal{B}) \subset L^o(\Omega, \mathcal{A})$. Así, también la misma contención garantiza que $\mathcal{E}(\mathcal{B}) \subset \mathcal{E}(\mathcal{A},)$ es decir, se sostiene para las funciones simples. Como la medida de probabilidad en $\mathcal{B}$ es la restricción de la de $\mathcal{A}$, entonces la integral de las funciones simples integrables $E^1(\mathcal{B})$ esta dada por la integral definida en $E^1(\mathcal{A})$. Dotando con la norma $\| \cdot \|_{L^p}$ a $E^1(\mathcal{B})$ se tiene que

$$\psi : E^1(\mathcal{B}) \to E^1(\mathcal{A})$$

tal que

$$\| f - g \|_{L^p} = \| \psi(f) - \psi(g) \|_{L^p},$$

donde $f, g \in E^1[\mathcal{B}]$ y siendo ψ la inclusión canónica, ψ es una isometría de $E^1[\mathcal{B}]$ a $E^1[\mathcal{A}]$.

Por otra parte, ya que $E^1[\mathcal{B}]$ es denso en $L^p(\mathcal{B})$ y $L^p(\Omega, \mathcal{A}, P)$ es completo, la isometría ψ se extiende a una isometría de $L^p(\mathcal{B})$ a $L^p(\mathcal{A})$.

Puesto que un espacio completo bajo una isometría es completo; entonces la imagen de $L^p(\mathcal{B})$ es completa y por tanto, cerrada en $L^p(A)$.

□

Recordemos un par de definiciones de utilidad indispensable: espacios de Hilbert y proyección ortogonal.

Definición 27. *Un espacio de Hilbert es un espacio vectorial H con un producto escalar $< \cdot, \cdot >$ que es completo con respecto a la norma inducida.*

Definición 28. *Sea V un subespacio vectorial cerrado de un espacio H. La proyección orotgonal de H sobre V es la función $P_V : H \to V$ que a cada $u \in H$ le asocia el único $P_V u \in V$ tal que $||u - P_V u|| = \inf_{w \in V} ||u - w||$.*

Definición 29. *$E^{\mathcal{B}}$ denotará el operador proyección ortogonal de $L^2(\mathcal{A})$ en $L^2(\mathcal{B})$. Dado $f \in L^2(\mathcal{A})$, $E^{\mathcal{B}}(f)$ es llamada la esperanza condicional de f dado $\mathcal{B}$.*

Daremos algunas propiedades de este operador.

Teorema 3.1.2 (Propiedades de la Esperanza Condicional)**.** *Sea $\mathcal{B}$ y $\mathcal{C}$ sub$\sigma-$álgebras tales que $\mathcal{C} \subset \mathcal{B}$. Entonces:*

i) $E^{\mathcal{B}}(f) \in L^2(\mathcal{B})$.

ii) $\| E^{\mathcal{B}}(f) \|_{L^2} \leq \| f \|_{L^2}$.

iii) $E^{\mathcal{C}} E^{\mathcal{B}} = E^{\mathcal{C}}$.

iv) $E[E^{\mathcal{B}}] = E$.

v) *Sea $\varphi \in L^\infty(\mathcal{B})$. Entonces $E^{\mathcal{B}}(\varphi f) = \varphi E^{\mathcal{B}}(f)$ para toda $f \in L^2(\mathcal{A})$.*

Demostración. i) Por definición

$$E^{\mathcal{B}}(f) : L^2(\mathcal{A}) \to L^2(\mathcal{B}).$$

En consecuencia $E^{\mathcal{B}}(f) \in L^2(\mathcal{B})$.

ii) Puesto que

$$f = E^{\mathcal{B}}(f) + (f - E^{\mathcal{B}}(f))$$

y

$$< E^{\mathcal{B}}(f), f - E^{\mathcal{B}}(f) >= 0$$

pues

$$f - E^{\mathcal{B}}(f) \in (L^2(\mathcal{B}))^{\perp}.$$

Se tiene que

$$\begin{aligned} \| f \|_{L^2}^2 &= \| E^{\mathcal{B}}(f) \|_{L^2}^2 + 2 < E^{\mathcal{B}}(f), f - E^{\mathcal{B}}(f) > + \| f - E^{\mathcal{B}}(f) \|_{L^2}^2 \\ &= \| E^{\mathcal{B}(f)} \|_{L^2}^2 + \| f - E^{\mathcal{B}}(f) \|_{L^2}^2 \ . \end{aligned}$$

Luego

$$\| f \|_{L^2}^2 \geq \| E^{\mathcal{B}}(f) \|_{L^2}^2 \ .$$

Por tanto

$$\| f \|_{L^2} \geq \| E^{\mathcal{B}}(f) \|_{L^2} \ .$$

Nótese que esta desigualdad garantiza la continuidad del operador proyección ortogonal.

iii) Puesto que $L^2(\mathcal{B})$ es cerrado en $L^2(\mathcal{A})$ y en adición es subespacio vectorial,

$$L^2(\mathcal{A}) = L^2(\mathcal{B}) \bigoplus L^2(\mathcal{B})^{\perp}.$$

Por hipótesis, $\mathcal{C} \subset \mathcal{B}$, por lo que $L^2(\mathcal{C}) \subset L^2(\mathcal{B})$. Luego, si $f \in L^2(\mathcal{A})$, entonces existen

$$u \in L^2(\mathcal{B}) \quad \text{y} \quad v \in L^2(\mathcal{B})^{\perp}$$

tales que

$$f = u + v \quad \text{siendo} \quad u = E^{\mathcal{B}}(f).$$

De este modo $u = w + x$, donde $w \in L^2(\mathcal{C})$, $w = E^{\mathcal{C}}(u)$ y $x \in L^2(\mathcal{C})^{\perp}$. En consecuencia $f = w + (x + v)$.

Ya que $L^2(\mathcal{C}) \subset L^2(\mathcal{B})$, entonces $L^2(\mathcal{B})^{\perp} \subset L^2(\mathcal{C})^{\perp}$ y en consecuencia $x + v \in L^2(\mathcal{C})^{\perp}$. Así, por la descomposición de $L^2(\mathcal{A})$, se tiene que $w = E^{\mathcal{C}}(f)$. En consecuencia $E^{\mathcal{C}}(E^{\mathcal{B}}) = E^{\mathcal{C}}$.

iv) Sea $\mathcal{A}_0 = \{\Omega, \emptyset\}$ la $\sigma-$álgebra trivial. Obsérvese que la función Φ es $\mathcal{A}_0-$medible si y sólo si es constante. Para mostrar esto, supóngase que Φ es $\mathcal{A}_0-$medible, entonces, para todo $B \in \mathcal{B}_{\mathbb{R}}$ se tiene que

$$\Phi^{-1}(B) = \Omega \quad \text{y} \quad \Phi^{-1}(B^c) = \emptyset,$$

entonces, si $c \in B, \quad \Phi = c, c$ constante. Recíprocamente, si $\Phi = c$, entonces, para todo $B \in \mathcal{B}_{\mathbb{R}}$, $\Phi^{-1}(B) = \Omega \quad \text{y} \quad \Phi^{-1}(B^c) = \emptyset$. Luego $\Phi^{-1}(B) \in \mathcal{A}_0$.

Ahora obsérvese que

$$(L^2(\mathcal{A}_0))^{\perp} = \{g \in L^2(\mathcal{A}) :< f, g >= 0\} \quad \text{para toda} \quad f \in L^2(\mathcal{A}_0),$$

es decir, la esperanza de los elementos de tal conjunto es cero.

Por otra parte, sea $f \in L^2$ elemento arbitrario, el cual se puede expresar como $f = E(f)1_{\Omega} + h$, donde $E(h) = 0$. Entonces,

$$E^{\mathcal{A}_0}(f) = E(f)1_{\Omega}.$$

En consecuencia, por el inciso iii),

$$E^{\mathcal{A}_0}(E^{\mathcal{B}}) = E^{\mathcal{A}_0} = E.$$

v) Sea $\mathcal{M}_{\varphi}$ el operador acotado definido en $L^2(\mathcal{A})$ por la multiplicación por φ como $\mathcal{M}_{\varphi} : f \mapsto \varphi f$. Como $\varphi \in L^0(\mathcal{B})$ y $L^0(\mathcal{B})$ es álgebra, $\mathcal{M}_{\varphi}(L^2(\mathcal{B})) \subset L^2(\mathcal{B})$. Como

$$E[(\varphi f)|g] = E[f(\varphi g)],$$

para toda $f, g \in L^2(\mathcal{A})$, el operador $\mathcal{M}_{\varphi}$ es hermitiano, esto es,

$$(\mathcal{M}_{\varphi}(f)|g)_{L^2} = (f(|\mathcal{M}_{\varphi}(g)))_{L^2}.$$

Como $L^2(\mathcal{B})$ es invariante bajo $\mathcal{M}_{\varphi}$, su complemento ortogonal también lo es. Entonces, si $f = u + v$, donde $u \in L^2(\mathcal{B})$ y $v \in L^2(\mathcal{B})^{\perp}$ entonces

$$\mathcal{M}_{\varphi} f = \mathcal{M}_{\varphi} u + \mathcal{M}_{\varphi} v,$$

donde

$$\mathcal{M}_{\varphi} u \in L^2(\mathcal{B}) \quad \text{y} \quad \mathcal{M}_{\varphi} v \in L^2(\mathcal{B}) \in L^2(\mathcal{B}).$$

Por lo cual $E^{\mathcal{B}}(\mathcal{M}_{\varphi} f) = \mathcal{M}_{\varphi}(E^{\mathcal{B}})$, esto es, $E^{\mathcal{B}}(\varphi f) = \varphi E^{\mathcal{B}}(f)$.

$\square$

3.2. Esperanza Condicional y Positividad

Proposición 3.2.1. *Sea $f \in L^2(\mathcal{A})$, $f \geq 0$, y sea $\mathcal{B}$ una sub$\sigma-$álgebra de $\mathcal{A}$. Entonces $E^{\mathcal{B}}(f) \geq 0$.*

Demostración. Sea $B \in \mathcal{B}$. Obsérvese que

$$E(E^{\mathcal{B}}(f)1_B) = E(E^{\mathcal{B}}(f1_B)) = E(f1_B)$$

para todo $B \in \mathcal{B}$, de acuerdo con la proposición 3.1.2.

Así, renómbrese $u = E^{\mathcal{B}}(f)$. Se ha probado entonces que $E(u1_B) \geq 0$ pues $E(f1_B) \geq 0$. Defínase

$$B_n = \left\{\omega : u(\omega) < -\frac{1}{n}\right\}.$$

Como $u \in L^o(\mathcal{B})$ con $B_n \in \mathcal{B}$ se tiene que $E(u1_{B_n}) \geq 0$. Más aún,

$$E(u1_{B_n}) < -n^{-1}P(B_n),$$

por lo que

$$-nE(u1_{B_n}) > P(B_n).$$

Luego $P(B_n) = 0$. Así,

$$P(\bigcup_{n\in\mathbb{N}} B_n) = \lim_{n\to\infty} P(B_n) = 0.$$

En consecuencia,

$$P(\bigcap_{n\in\mathbb{N}} B_n^c) = \lim_{n\to\infty} P(B_n^c) = 1.$$

Por tanto $E^{\mathcal{B}}(f) \geq 0$. $\square$

De la proposición anterior se desprende la siguiente consecuencia inmediata.

Corolario 3.2.2. *Sea $f, g \in L^2(\mathcal{A})$. Entonces $f \geq g$ implica que*

$$E^{\mathcal{B}}(f) \geq E^{\mathcal{B}}(g) \quad y \quad |E^{\mathcal{B}}(f)| \leq E^{\mathcal{B}}(|f|).$$

Demostración. Como $f - g \geq 0$ y $f - g \in L^2(\mathcal{A})$, por la proposición 3.2.1,

$$E^{\mathcal{B}}(f - g) \geq 0,$$

esto es

$$E^{\mathcal{B}}(f) \geq E^{\mathcal{B}}(g).$$

Por otra parte,

$$-|f| \leq f \leq |f|.$$

Luego

$$E^{\mathcal{B}}(-|f|) \leq E^{\mathcal{B}}(f) \leq E^{\mathcal{B}}(|f|).$$

De aquí que

$$|E^{\mathcal{B}}(f)| \leq E^{\mathcal{B}}(|f|).$$

□

3.3. Extensión de la Esperanza Condicional a L^1

Hasta este punto, la esperanza condicional ha sido definida en $L^2(\mathcal{A})$. El cuestionamiento natural es saber si ésta puede extenderse a L^1. El siguiente teorema se ocupa de ello.

Teorema 3.3.1. *El operador $E^{\mathcal{B}}$ definido en $L^2(\mathcal{A})$ tiene una extensión continua $\mathcal{E}^{\mathcal{B}}$, definida en $L^1(\mathcal{B})$. Esta extensión tiene las siguientes propiedades:*

i) $\mathcal{E}^{\mathcal{B}}(f) = f$ *para cada* $f \in L^2(\mathcal{B})$

ii) $\| \mathcal{E}^{\mathcal{B}}(f) \|_{L^1} \leq \| f \|_{L^1}$.

iii) Si $\mathcal{C} \subset \mathcal{B}$, *entonces* $\mathcal{E}^{\mathcal{C}}\mathcal{E}^{\mathcal{B}} = \mathcal{E}^{\mathcal{C}}$, *en particular,* $E\mathcal{E}^{\mathcal{B}}E$.

iv) Si $\varphi \in L^\infty(\mathcal{B})$, *entonces* $\mathcal{E}^{\mathcal{B}}(\varphi f) = \varphi\mathcal{E}^{\mathcal{B}}(f)$.

Demostración. Sea $f \in L^2$. Por el corolario 3.2.2, $|E^{\mathcal{B}}(f)| \leq E^{\mathcal{B}}(|f|)$. Aplicando el operador esperanza,

$$E(|E^{\mathcal{B}}(f))| \leq E(E^{\mathcal{B}}(|f|)),$$

y por las propiedades de la esperanza condicional de 3.1.2 se tiene que

$$E(E^{\mathcal{B}}(|f|)) = E(|f|) = \| f \|_{L^1} .$$

En consecuencia

$$\| E^{\mathcal{B}}(f) \|_{L^1} \leq \| f \|_{L^1}$$

para toda $f \in L^2(\mathcal{A})$. Entonces $E^{\mathcal{B}}$ es un operador acotado cuando $L^2(\mathcal{A})$ es equipado con la norma L^1. En virtud de que $L^1(\mathcal{B})$ es completo y de que $L^2(\mathcal{A})$ es denso en $L^1(\mathcal{A})$ $E^{\mathcal{B}}$ puede ser extendido a un operador

$$\mathcal{E}^{\mathcal{B}} : L^1(\mathcal{A}) \to L^1(\mathcal{B}).$$

Como $E^{\mathcal{A}}(f) = f$ si $f \in L^2(\mathcal{B})$ y ya que $L^2(\mathcal{B})$ es denso en $L^1(\mathcal{B})$, el operador extendido por continuidad tiene la misma propiedad, es decir, $\mathcal{E}^{\mathcal{A}}(f) = f$ si $f \in L^1(\mathcal{B})$.

El inciso ii) ya ha sido probado y iii) y iv) se siguen por la extensión de la continuidad y en virtud de 3.1.2. □

De ahora en adelante se usará la misma notación $E^{\mathcal{B}}$, para $\mathcal{E}^{\mathcal{B}}$ y $E^{\mathcal{B}}$.

Ahora nos ocupamos de calcular el operador $E^{\mathcal{B}}$ cuando $\mathcal{B}$ es una $\sigma-$álgebra finita.

Proposición 3.3.2. *Sea $\mathcal{B}$ una sub$\sigma-$álgebra finita de $\mathcal{A}$ y sean $e_1, \cdots, e_n$ los átomos de $\mathcal{B}$ con estricta probabilidad positiva. Entonces $E^{\mathcal{B}}(f) = \sum \alpha_k 1_{e_k}$, donde $\alpha_k = \frac{1}{P(e_k)} E(f 1_{e_k})$.*

Demostración. Como $E^{\mathcal{B}}(f) \in L^o(\mathcal{B})$, será suficiente con probar que $f - E^{\mathcal{B}}(f)$ es ortogonal a $L^o(\mathcal{B})$. Ya que las $\{1_{e_k}\}_{k=1}^n$ forman una base de $L^0(\mathcal{B})$ se tiene que

$$E^{\mathcal{B}}(f) = \sum_{i=1}^{n} \alpha_i 1_{e_i}.$$

Por lo que bastará probar que

$$E((f - E^{\mathcal{B}}(f))1_{e_i}) = 0, i \in \{1, \cdots, n\}.$$

No obstante, se tiene que

$$E^{\mathcal{B}}(f1_{e_i}) = \alpha_i 1_{e_i} \quad \text{para algún} \quad i \in \{1, \cdots, n\}.$$

Luego

$$E(E^{\mathcal{B}}(f1_{e_i})) = \alpha_i E(f1_{e_i}) = \alpha_i P(e_i) \quad \text{para algún} \quad i \in \{1, \cdots, n\},$$

esto es,

$$E(f1_{e_i}) = \alpha_i P(e_i) \quad \text{para algún} \quad i \in \{1, \cdots, n\}.$$

De aquí que

$$\alpha_i = \frac{1}{P(e_i)} E(f1_{e_i}), i \in \{1, \cdots, n\}.$$

Además,

$$\begin{aligned} E((f - E^{\mathcal{B}}(f)1_{e_i})) &= E(f1_{e_i} - E^{\mathcal{B}(f)}1_{e_i}) = E(f1_{e_i}) - E(E^{\mathcal{B}}(f1_{e_i})) \\ &= E(f1_{e_i}) - E(f1_{e_i}) = 0. \end{aligned}$$

En conclusión,

$$E^{\mathcal{B}}(f) = \sum_{i=1}^{n} \frac{1}{P(e_i)} E(f1_{e_i})1_{e_i}.$$

□

En virtud de la proposición anterior, es posible definir la siguiente medida en $\mathcal{A}$.

Definición 30. *Sea μ_k una medida definida en $\mathcal{A}$ por $\mu_k = \frac{1}{P(e_k)} P(A \cap e_k)$.*

Nótese que efectivamente, μ_k es medida en $\mathcal{A}$:

i) $\mu_k(\emptyset) = \frac{P(\emptyset \cap e_k)}{P(e_k)} = \frac{P(\emptyset)}{P(e_k)} = 0.$

ii) Sea $\{A_i\}_{i\in\mathcal{N}} \in \mathcal{A}$ tales que $A_i \cap A_j = \emptyset$ para toda $i, j \in \mathbb{N}, i \neq j$. Entonces

$$\begin{aligned}\mu_k(\bigcup_{n\in\mathbb{N}} A_i) &= \frac{P(\bigcup_{n\in\mathbb{N}} A_i \cap e_k)}{P(e_k)} = \frac{P(\bigcup_{n\in\mathbb{N}}(A_i \cap e_k))}{P(e_k)} \\ &= \frac{\sum_{i\in\mathbb{N}} P(A_i \cap e_k)}{P(e_k)} = \sum_{i\in\mathbb{N}} \mu_k(A_i).\end{aligned}$$

iii) Más aún, $\mu_k(\Omega) = \frac{P(\Omega\cap e_k)}{P(e_k)} = \frac{P(e_k)}{P(e_k)} = 1$.

μ_k es llamada la probabilidad condicional dado el átomo e_k. Esto es

$$\alpha_k = \int f d\mu_k.$$

Proposición 3.3.3. *Sea $\mathcal{B}$ una $\sigma-$álgebra finita de $\mathcal{A}$, sea φ una función convexa, y sea $f \in L^1(\mathcal{A}), f \geq 0$. Entonces $\varphi(E^{\mathcal{B}}(f)) \leq E^{\mathcal{B}}(\varphi(f))$*

Demostración. Por la proposición 3.3.2,

$$E^{\mathcal{B}}(\varphi(f)) = \sum_{k=1}^{n} \beta_k 1_{e_k} \quad \text{donde} \quad \beta_k = \int \varphi(f) d\mu_k$$

con respecto al átomo e_k. Por otra parte,

$$\varphi(E^{\mathcal{B}}(f)) = \varphi(\sum_{k=1}^{n} \alpha_k 1_{e_k}) = \sum_{k=1}^{n} \varphi(\alpha_k) 1_{e_k}.$$

Como $\mu_k(\Omega) = 1$ y φ es convexa por hipótesis, por la desigualdad de Jensen A.0.26, $\varphi(\alpha_k) \leq \beta_k$. Por lo que $\varphi(E^{\mathcal{B}}(f)) \leq E^{\mathcal{B}}(\varphi(f))$. □

En la siguiente sección mostramos un teorema de existencia sobre $\sigma-$álgebras finitas para aproximar esperanzas condicionales; útil para la generalización de la desigualdad de Jensen.

3.4. Aproximación por $\sigma-$álgebras finitas

Proposición 3.4.1. *Sea $f_1, \cdots, f_n \in L^1(\mathcal{A})$. Entonces existe una sucesión creciente $\mathcal{B}_1 \subset \mathcal{B}_2 \subset \cdots \subset \mathcal{B}_k \subset \cdots \subset \mathcal{B}$ de $\sigma-$álgebras finitas tales que*

$$\| E^{\mathcal{B}_k} f_j - E^{\mathcal{B}} f_j \|_{L^1} \to 0 \quad \textit{cuando} \quad k \to \infty, j \in \{i, \cdots, n\}.$$

Demostración. Por inducción matemática sobre n.

i) Base de inducción. Si $n = 1$, sea $u = E^{\mathcal{B}}(f_1)$, luego $u \in L^1(\mathcal{B})$; así, existe una sucesión de funciones simples $\{u_k\} \in L^1(\mathcal{B})$ tales que $\lim\limits_{n\to\infty} u_k = u$. Sea $\mathcal{B}_k$ la $\sigma-$álgebra generada por las u's, $s \leq k$. En consecuencia $\mathcal{B}_x \subset \mathcal{B}$. Más aún, u_k es B_k-medible. Como $E^{\mathcal{B}}_k(u_k) = u_k$ y

$$\| E^{\mathcal{B}_k}(u_k) - E^{\mathcal{B}_k}(u) \|_{L^1} = \| E^{\mathcal{B}_k}(u_k - u) \|_{L^1} \leq \| u_k - u \|_{L^1}$$

se tiene que

$$\begin{aligned}
\| E^{\mathcal{B}_k}(u) - u \|_{L^1} &= \| E^{\mathcal{B}_k}(u) - E^{\mathcal{B}_k}(u_k) + E^{\mathcal{B}_k}(u_k) - u \|_{L^1} \\
&\leq \| E^{\mathcal{B}_k}(u) - E^{\mathcal{B}_k}(u_k) \|_{L^1} + \| E^{\mathcal{B}_k}(u_k) - u \|_{L^1} \\
&\leq \| u_k - u \|_{L^1} + \| u_k - u \|_{L^1} = 2 \| u_k - u \|_{L^1} .
\end{aligned}$$

Además,

$$E^{\mathcal{B}_k}(u) = E^{\mathcal{B}_k}(E^{\mathcal{B}}(f_1)) = E^{\mathcal{B}_k}(f_1).$$

Por consiguiente

$$\| E^{\mathcal{B}_k}(f_1) - E^{\mathcal{B}}(f_1) \|_{L^1} \leq 2 \| u - u_k \|_{L^1} \to 0 \quad \text{cuando} \quad k \to \infty.$$

ii) Hipótesis de Inducción: Supóngase que existe una sucesión creciente $\mathcal{B}_1 \subset \cdots \subset \mathcal{B}_k \subset \cdots \mathcal{B}$ de $\sigma-$álgebras finitas tales que

$$\| E^{\mathcal{B}_k}(f_j) - E^{\mathcal{B}}(f_j) \|_{L^1} \to 0 \quad \text{cuando} \quad k \to \infty, j \in \{1, \cdots, n\}.$$

Por demostrar válido para $n + 1$.

iii) Paso de Inducción: Sea $\{\mathcal{B}_k'\}$ una sucesión de $\sigma-$álgebras finitas adaptadas a $f_1, \cdots, f_n$. Sea $\mathcal{B}_k''$ la $\sigma-$álgebra generada por $\mathcal{B}_k'$ y $\mathcal{B}_k$, la $\sigma-$álgebra generada por la sucesión de funciones simples $u_s, s \leq k$ tal que $\{u_k\}$ converge a u=$E^{\mathcal{B}_k}(f_{n+1}) \in L^1(\mathcal{B}), u_k \in \mathcal{B}_k$. En virtud de la base de inducción,

$$\| E^{\mathcal{B}_k''}(f_{n+1}) - E^{\mathcal{B}}(f_{n+1}) \|_{L^1} \to 0 \quad \text{cuando} \quad k \to \infty.$$

Por el principio de inducción, la proposición es válida para toda $n \in \mathbb{N}$.

□

Corolario 3.4.2 (Desigualdad de Jensen Generalizada)**.** *Sea $\mathcal{B}$ una sub$\sigma-$álgebra de $\mathcal{A}$, sea $f \in L^1(\mathcal{A}), f \geq 0$, sea φ una función convexa no negativa tal que*
$E(|\varphi(f)|) < \infty$. Entonces $\varphi(E^{\mathcal{B}}(f)) \leq E^{\mathcal{B}}(\varphi(f))$.

Demostración. Por la proposición 3.4.1, existe $\{\mathcal{B}_k\}$ sucesión de $\sigma-$álgebras finitas tales que

$$\| E^{\mathcal{B}_k}(f) - E^{\mathcal{B}}(f) \|_{L^1} \to 0 \quad \text{y} \quad \| E^{\mathcal{B}_k}(\varphi(f)) - E^{\mathcal{B}}(\varphi(f)) \|_{L^1} \to 0.$$

Por la proposición 3.3.3,

$$\varphi(E^{\mathcal{B}_k}(f)) \leq E^{\mathcal{B}_k}(\varphi(f)),$$

esto es

$$E^{\mathcal{B}_k}(\varphi(f)) - \varphi(E^{\mathcal{B}_k}(f)) \geq 0.$$

Por tanto $E^{\mathcal{B}}(\varphi(f)) \leq \varphi(E^{\mathcal{B}}(f))$. □

3.5. Esperanza Condicional en Espacios L^p

Se termina este capítulo dando algunas propiedades del operador $E^{\mathcal{B}}$ en los espacios L_p.

Obsérvese que dado $p \in (1, \infty]$, $L^p \subset L^1$ según A.0.28. Entonces el operador esperanza condicional $E^{\mathcal{B}}$ puede ser definido en L^p.

Proposición 3.5.1. *Sea* $1 \leq p \leq \infty$. *Si* $f \in L^p(\mathcal{A})$, *entonces* $E^{\mathcal{B}}(f) \in L^p(\mathcal{B})$ *y*

i) $\| E^{\mathcal{B}}(f) \|_{L^p} \leq \| f \|_{L^p}$.

ii) *Sean* p *y* q *exponentes conjugados. Entonces* $E^{\mathcal{B}}(fg) = gE^{\mathcal{B}}(f)$, *para toda* $f \in L^p(\mathcal{A}), g \in L^q(\mathcal{A})$.

iii) $E(E^{\mathcal{B}}g(E^{\mathcal{B}}f)) = E(gE^{\mathcal{B}}(f)) = E(fE^{\mathcal{B}}(g))$, *para toda* $f \in L^p(\mathcal{A})$ *y* $g \in L^q(\mathcal{B})$.

Demostración. i) Supóngase que $1 \leq p < \infty$. Considérese la función $\varphi(t) = t^p, t \geq 0$, la cual es convexa. Así, por la desigualdad de Jensen generalizada 3.4.2, $(E^{\mathcal{B}}(f))^p \leq E^{\mathcal{B}}(f^p)$. Luego

$$\left(\int (E^{\mathcal{B}}(f))^p\right)^{1/p} \leq \left(\int (E^{\mathcal{B}}(f^p))\right)^{1/p},$$

es decir,

$$\| E^{\mathcal{B}}(f) \|_{L^p} \leq \| f \|_{L^p} .$$

Ahora, asúmase que $p = \infty$. Obsérvese que

$$\| E^{\mathcal{B}_k}(f) \|_{L^\infty} = \sup |\int f d\mu_k| \leq \sup \int |f| d\mu_k = \| f \|_{L^\infty} \quad f \in L^\infty.$$

Por la proposición 3.4.1 existe una sucesión de $\sigma-$álgebras finitas crecientes $\{\mathcal{B}_k\}$ tales que

$$\| E_k^{\mathcal{B}}(f) - E^{\mathcal{B}}(f) \|_{L^1} \to 0.$$

Por la proposición A.0.25 existe $\{u_{k_s}\}$ subsucesión de $u_k = E^{\mathcal{B}_k}(f)$ tal que

$$u_{k_s} \to E^{\mathcal{B}}(f) \quad \text{a.s.}$$

Por lo que

$$|u_{k_s}| \leq \| f \|_{L^\infty} .$$

En consecuencia

$$\| E^{\mathcal{B}}(f) \|_{L^\infty} \leq \| f \|_{L^\infty} .$$

ii) En virtud del teorema 3.1.2, $E^{\mathcal{B}}(fg) = gE^{\mathcal{B}}(f)$ para funciones acotadas. Entonces existen $\{f_n\} \in L^\infty(\mathcal{A})$ y $\{g_n\} \in L^\infty(\mathcal{B})$ tales que

$$\| f_n - f \|_{L^p} \to 0 \quad \text{y} \quad \| g_n - g \|_{L^p} \to 0.$$

De aquí que

$$\| E^{\mathcal{B}}(f_n) - E^{\mathcal{B}}(f) \|_{L^p} \leq \| f_n - f \|_{L^p} \to 0.$$

Así por 3.1.2 $E^{\mathcal{B}}(g_n f_n) = g_n E^{\mathcal{B}}(f_n)$, y como

$$\begin{aligned} \| f_n g_n - fg \|_{L^1} &= \| f_n g_n - fg_n + fg_n - fg \|_{L^1} \\ &= \| g_n(f_n - f) + f(g_n - g) \|_{L^1} \\ &\leq \| g_n(f_n - f) \|_{L^1} + \| f(g_n - g) \|_{L^1} \\ &\leq \| g_n \|_{L^p} \| f_n - f \|_{L^p} + \| f \|_{L^q} \| g_n - g \|_{L^q} \end{aligned}$$

según la desigualdad de Hölder A.0.27. Entonces se tiene que

$$\| f_n g_n - fg \|_{L^1} \to 0.$$

Por consiguiente

$$\| E^{\mathcal{B}}(f_n g_n) - E^{\mathcal{B}}(fg) \|_{L^1} \to 0.$$

Es decir,

$$g_n E^{\mathcal{B}}(f_n) \to gE^{\mathcal{B}}(f) \quad \text{en} \quad L^1.$$

De manera análoga para $f_n E^{\mathcal{B}}(g)$.

iii) Supóngase que $f, g \in L^2(\mathcal{A})$. Defínase $< f, g >= E(fg)$ un producto interior. Por las propiedades de la proyección ortogonal, se tiene que

$$< E^{\mathcal{B}}(f), g >_{L^2} = < f, E^{\mathcal{B}}(g) >_{L^2} = < E^{\mathcal{B}}(f), E^{\mathcal{B}}(g) >_{L^2}$$

ya que $L^\infty \subset L^2$, se prueba la proposición para $f, g \in L^\infty(\mathcal{A})$. No obstante, nuevamente, existen $\{f_n\} \in L^\infty(\mathcal{A})$ y $\{g_n\} \in L^\infty(\mathcal{B})$ tales que

$$\| f_n - f \|_{L^p} \to 0 \quad \text{y} \quad \| g_n - g \|_{L^q} \to 0.$$

Luego $\| f_n g_n - fg \|_{L^1} \to 0$, de modo que

$$E^{\mathcal{B}}(g_n) \to E^{\mathcal{B}}(g) \quad \text{y} \quad E^{\mathcal{B}}(f_n) \to E^{\mathcal{B}}(f) \quad \text{en} \quad L^1,$$

entonces,

$$< E^{\mathcal{B}}(f), g > = < f, E^{\mathcal{B}}(g) > = < E^{\mathcal{B}}(f), E^{\mathcal{B}}(g) > \quad \text{en} \quad f, g \in L^1.$$

Y ya que $L^\infty \subset L^2 \subset L^1$, se tiene la proposición para toda $f \in L^p(\mathcal{A})$ y $g \in L^q(\mathcal{B})$.

$\square$

Capítulo 4

Independencia y Ortogonalidad

La noción de ortogonalidad es un concepto bien conocido por su característica geométrica en espacios usuales como $\mathbb{R}^n$. Esta condición es extendida a espacios vectoriales más generales a los que se les dota con una norma y un producto interior. Así, el espacio de funciones continuas, integrables, de sucesiones convergentes, entre otros, poseen esta propiedad fundamental. En el estudio de los espacios L^p esta noción es útil y de gran importancia pues los vectores de dichas estructuras que verifican la propiedad de ortogonalidad facilitan el estudio de tales espacios y, en gran medida, son entendibles y abordados como espacios más simples tales como $\mathbb{R}^2$ y $\mathbb{R}^3$. En esta sección, veremos que la ortogonalidad de los elementos en L^p da como resultado, de manera natural, la independencia de estructuras o conjuntos, a saber $\sigma-$álgebras y, posteriormente, a variables aleatorias, mediante el operador esperanza.

4.1. Independencia de dos sub$-\sigma-$álgebras.

Definición 31. *Sea $\mathcal{B}$ y $\mathcal{C}$ dos sub$-\sigma-$álgebras de espacios de probabilidad $(\Omega, \mathcal{A}, P)$. Decimos que $\mathcal{B}$ y $\mathcal{C}$ son independientes (relativas a P) si $L^2(\mathcal{B})$ y $L^2(\mathcal{C})$ son ortogonales en las funciones constantes. Dicho de otra forma,*

$$f \in L^2(\mathcal{B}), g \in L^2(\mathcal{C}) \quad y \quad E(f) = E(g) = 0 \quad implica \quad E(fg) = 0.$$

Obsérvese que la noción de independencia involucra la norma L^2 y por tanto a la medida de probabilidad P. Para ser precisos, debemos hablar de independencia relativa a P. Ya que hemos considerado a P dado y común para todos, diremos simplemente independiente.

Ya que $L^2(\mathcal{B})$ y $L^2(\mathcal{C})$ contienen a la función 1_Ω, estos no pueden ser ortogonales pues $E(1_\Omega) = P(\Omega) = 1$; la independencia corresponde a la noción más fuerte de ortogonalidad que puede ser esperada.

Ahora, considérese el subespacio vectorial $\mathcal{H}$ compuesto de funciones ortogonales a las funciones constantes:

$$\mathcal{H} = \{f \in L^2(\mathcal{A}) : E(f) = 0\}.$$

La relación $E(E^{\mathcal{B}}) = E$ implica que $E^{\mathcal{B}}(\mathcal{H}) \subset \mathcal{H}$. En efecto. Sea $f \in \mathcal{H}$. Entonces $E(E^{\mathcal{B}}(f)) = E(f) = 0$.

Así, la definición 31 es equivalente a que

$$\mathcal{H} \cap L^2(\mathcal{B}) \quad \text{es ortogonal a} \quad \mathcal{H} \cap L^2(\mathcal{C}).$$

Obsérvese que de la misma defición $L^2(\mathcal{B}) \cap L^2(\mathcal{C})$ se reduce a las funciones constantes.

Ya que $L^2(\mathcal{B}) \cap L^2(\mathcal{C}) = L^2(\mathcal{B} \cap \mathcal{C})$, se concluye que si $\mathcal{B}$ y $\mathcal{C}$ son $\sigma-$álgebras independientes, entonces $\mathcal{B} \cap \mathcal{C}$ se reduce a los conjuntos de probabilidad cero y sus complementos.

Establecemos ahora la definición de independencia mutua entre un conjunto finito de sub$-\sigma-$álgebras.

Definición 32. *Sean $\mathcal{B}_1, \cdots, \mathcal{B}_n$ sub$-\sigma-$álgebras de $\mathcal{A}$, sea H un subconjunto de $[0,1]$, y sea $\mathcal{B}_H$ la $\sigma-$álgebra generada por $\{\mathcal{B}_i : i \in H\}$. Entonces $\mathcal{B}_1, \cdots, \mathcal{B}_n$ se dicen mutuamente independientes si $\mathcal{B}_H$ y $\mathcal{B}_{H^c}$ son $\sigma-$álgebras independientes para cada $H \in \mathcal{P}([0,1])$.*

Utilizaremos las definiciones previas para dar lugar al concepto de independencia de variables aleatorias.

4.2. Independencia de Variables Aleatorias y de $\sigma-$álgebras.

Obsérvese lo siguiente: sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y sea $\mathcal{B}$ y $\mathcal{C}$ dos sub$-\sigma-$álgebras que son independientes en este espacio. Sean $\mathcal{B}'$ y $\mathcal{C}'$ otras dos sub$\sigma-$álgebras tales que $\mathcal{B}' \subset \mathcal{B}$ y $\mathcal{C}' \subset \mathcal{C}$. Entonces $\mathcal{B}'$ y $\mathcal{C}'$ son independientes. De hecho,

$$L^2(\mathcal{B}') \cap \mathcal{H} \subset L^2(\mathcal{B}) \cap \mathcal{H} \quad \text{y} \quad L^2(\mathcal{C}') \cap \mathcal{H} \subset L^2(\mathcal{C}) \cap \mathcal{H}.$$

Entonces la ortogonalidad de los primeros pares de subespacios implica la ortogonalidad del segundo par.

Definición 33. *Sea $X_1, \cdots, X_n$ v.a's y sea $\mathcal{B}_k = X_K{}^{-1}(\mathcal{B}_{\mathbb{R}})$. Entonces $X_1, \cdots, X_n$ se dicen mutuamente independientes si las $\mathcal{B}_k$ son $\sigma-$álgebras mutuamente independientes.*

De la observación y definición anterior, se desprende el siguiente resultado:

Proposición 4.2.1. *Sea $\mathcal{D}_1, \cdots, \mathcal{D}_n$ sub$-\sigma-$álgebras mutuamente independientes del espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Sea X_k una v.a. $\mathcal{D}_k-$medible definida en $(\Omega, \mathcal{A}, P)$. Entonces las v.a.'s X_k son independientes.*

Demostración. Como X_k es v.a. $\mathcal{D}_k-$medible, entonces $X_k{}^{-1}(\mathcal{B}_{\mathbb{R}}) \subset \mathcal{D}_k$, para toda $k \in \{1, \ldots, n\}$. Así, por lo observado con anterioridad, $\mathcal{B}_1, \ldots, \mathcal{B}_n$ son mutuamente independientes pues, por hipótesis, $\mathcal{D}_1, \ldots, \mathcal{D}_n$ son mutuamente independientes. En consecuencia $X_1, \ldots, X_n$ son v.a's independientes. □

La independencia de las variables aleatorias es preservada bajo un cambio de variable como lo muestra la siguiente proposición.

Proposición 4.2.2. *Sea $X_1, \ldots, X_n$ v.a.'s independientes, sea $\varphi_1, \ldots, \varphi_n$ funciones Borelianas de $\mathbb{R}$ a $\mathbb{R}$, y sea $Y_k = \varphi_k(X_k)$ para toda $k \in \{1, \ldots, n\}$. Entonces las Y_k's son v.a.'s mutuamente independientes.*

Demostración. Puesto que

$$Y_k{}^{-1}(\mathcal{B}_{\mathbb{R}}) \subset X_k{}^{-1}(\varphi_k{}^{-1}(\mathcal{B}_{\mathbb{R}})),$$

y como φ es mapeo medible, entonces

$$X_k{}^{-1}(\varphi_k{}^{-1}(\mathcal{B}_{\mathbb{R}})) \subset X_k{}^{-1}(\mathcal{B}_{\mathbb{R}}).$$

Por tanto

$$Y_k{}^{-1}(\mathcal{B}_{\mathbb{R}}) \subset X_k{}^{-1}(\mathcal{B}_{\mathbb{R}}).$$

Por hipótesis $\{X_i\}_{i=1}^n$ son v.a's independientes. En consecuencia las $X_k{}^{-1}(\mathcal{B}_{\mathbb{R}})$ son mutuamente independientes y, por tanto, $Y_k{}^{-1}(\mathcal{B}_{\mathbb{R}})$ también lo son. $\square$

4.3. Esperanza de un producto de v.a's independientes

Presentamos ahora una equivalencia con la independencia de dos $sub-\sigma-$álgebras.

Teorema 4.3.1. *Sea $\mathcal{B}$ y $\mathcal{C}$ dos sub$-\sigma-$álgebras del espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Entonces los siguientes enunciados son equivalentes:*

i) $\mathcal{B}$ y $\mathcal{C}$ son independientes.

ii) $E(fg) = E(f)E(g)$ para toda $f \in L^2(\mathcal{B}), g \in L^2(\mathcal{C})$.

Demostración. $\Rightarrow$) Supóngase que $\mathcal{B}$ y $\mathcal{C}$ son independientes. Sean $f \in L^2(\mathcal{B})$ y $g \in L^2(\mathcal{C})$. Descomponiendo a f y g como

$$f = u + E(f)1_\Omega \quad \text{tal que} \quad E(u) = 0$$

y

$$g = v + E(g)1_\Omega \quad \text{tal que} \quad E(v) = 0$$

se tiene que

$$\begin{aligned} E(fg) &= E(uv + uE(g)1_\Omega + vE(f)1_\Omega + E(f)E(g)1_\Omega^2) \\ &= E(uv) + E(g)E(u1_\Omega) + E(f)E(v1_\Omega) + E(f)E(g) \end{aligned}$$

Obsérvese que

$$u \in \mathcal{H} \cap L^2(\mathcal{B}) \quad \text{y} \quad v \in \mathcal{H} \cap L^2(\mathcal{C}).$$

También

$$E(u1_\Omega) = 0 = E(v1_\Omega).$$

En consecuencia se tiene que

$$E(fg) = E(uv) + E(f)E(g).$$

Por lo observado anteriomente $E(uv) = 0$, obteniéndose así el resultado deseado.

$\Leftarrow$) Supóngase ahora que

$$E(fg) = E(f)E(g) \quad \text{para toda} \quad f \in L^2(\mathcal{B}), g \in L^2(\mathcal{C})$$

y además que

$$E(f) = 0 = E(g).$$

Por tanto $E(fg) = 0$, probando así que $\mathcal{B}$ y $\mathcal{C}$ son independientes. □

A continuación establecemos un resultado referente a independencia de sub$-\sigma-$álgebras y funciones en L^∞.

Proposición 4.3.2. *Sea $\mathcal{B}_1, \cdot, \mathcal{B}_n$ sub$-\sigma-$álgebras mutuamente independientes del espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Si $f_i \in L^\infty(\mathcal{B}_i), i \in \{1, \cdots, n\}$, entonces*

$$E\left(\prod_{i=1}^{n} f_i\right) = \prod_{i=1}^{n} E(f_i).$$

Demostración. Por inducción matemática sobre n.

Si $n = 1$ se obtiene una identidad trivial. Si $n = 2$ entonces $\mathcal{B}_1$ y $\mathcal{B}_2$ son independientes. Sea $\mathcal{B}_H$ la $\sigma-$álgebra generada por $f_2^{-1}(\mathcal{B}_\mathbb{R})$. De modo que $f_1 f_2 \in L^\infty(\mathcal{B}_H)$. Como $\mathcal{B}_H$ y $f_1^{-1}(\mathcal{B}_\mathbb{R})$ son $\sigma-$álgebras independientes en virtud de la ecuación (4.2.2). En consecuencia $E(f_1 f_2) = E(f_1)E(f_2)$.

Supóngase cierta la proposición para n, esto es,

$$E\left(\prod_{i=1}^{n} f_i\right) = \prod_{i=1}^{n} E(f_i).$$

Por demostrar válida para $n+1$.

Sea

$$h = \prod_{i=2}^{n+1} E(f_i).$$

Sea $\mathcal{B}_H$ la $\sigma-$álgebra generada por

$$\{f_i^{-1}(\mathcal{B}_{\mathbb{R}}) : 2 \leq i \leq n\}.$$

De esta forma $h \in L^\infty(\mathcal{B}_H)$. En virtud de la definición 32

$$f_1^{-1}(\mathcal{B}_{\mathbb{R}}) \quad \text{y} \quad \mathcal{B}_H \quad \text{son idependientes.}$$

Por la base de inducción,

$$E(f_1 h) = E(f_1)E(h).$$

Así, por hipótesis de inducción,

$$E\left(\prod_{i=1}^{n+1} f_i\right) = E(f_1)\prod_{i=2}^{n+1} E(f_i) = \prod_{i=1}^{n+1} E(f_i).$$

Por inducción, la proposición es válida para toda $n \in \mathbb{N}$. $\square$

Corolario 4.3.3. *Sean $f_1, \ldots, f_n \in L^1(\Omega, \mathcal{A}, P)$ y sea $h = \prod_{i=1}^{n} f_i$. Si las v.a.'s f_i's son independientes, entonces*

$$h \in L^1(\Omega, \mathcal{A}, P) \quad y \quad E(h) = \prod_{i=1}^{n} E(f_i).$$

Demostración. Supóngase que $f_i \geq 0, i \in \{1, \cdots, n\}$. Sea T_q el operador truncamiento definido en la ecuación (A.2). Por la proposición 4.2.2 las $T_q(f_i)$'s son independientes; por la proposición 4.3.2,

$$E\left(\prod_{i} T_q(f_i)\right) = \prod_{i=1}^{n} E(T_q(f_i)) \leq \prod_{i=1}^{n} E(f_i) = M.$$

Sea $U_q = \prod_i T_q(f_i)$. Obsérvese que $\{U_q\}_{q\geq 1}$ es una sucesión creciente. También se tiene que $E(U_q) \leq M$. Entonces por el teorema A.0.29 resulta que

$$\lim_{q\to\infty} U_q = h \in L^1 \quad \text{y} \quad \lim_{q\to\infty} E(U_q) = E(h).$$

Para el caso general, sea $f_i = f_i^0 - f_i^1$ donde

$$f_i^0 = f_i^+ = \sup(f_i, 0) \quad \text{y} \quad f_i^1 = f_i^- = \sup(-f_i, 0).$$

Como las $f_i^{\alpha_i}$ son no negativas

$$\prod_{i=1}^n f_i = \sum_\alpha (-1)^{\alpha_1+\ldots+\alpha_n} \prod_{i=1}^n E(f_i^{\alpha_i}) = \prod_{i=1}^n E(f_i).$$

De este modo, el producto de las f_i's puede verse como suma de funciones en $\mathcal{L}^1$. Por tanto $\prod_{i=1}^n f_i \in L^1$. □

Enseguida se presenta el caso de la independencia en el operador esperanza condicional.

4.4. Esperanza Condicional e Independencia

Teorema 4.4.1. *Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y sea $\mathcal{B}$ y $\mathcal{C}$ dos sub $-\sigma-$álgebras. Entonces las siguientes proposiciones son equivalentes:*

i) Las $\sigma-$álgebras $\mathcal{B}$ y $\mathcal{C}$ son independientes.

ii) $E^{\mathcal{B}}(f) = E(f)$ para toda $f \in L^1(\mathcal{C})$.

Demostración. Supóngase que $\mathcal{B}$ y $\mathcal{C}$ son independientes. Defínase

$$\bar{f} = f - E(f)1_\Omega \quad \text{donde} \quad f \in L^2(\mathcal{C}).$$

Obsérvese que $\bar{f} \in \mathcal{H}$ pues

$$(f - E(f)1_\Omega) \in L^2(\mathcal{C}) \quad \text{y} \quad E(\bar{f}) = E(f) - E(f) = 0.$$

Entonces

$$E^{\mathcal{B}}(\bar{f}) = E^{\mathcal{B}}(f) - E^{\mathcal{B}}(E(f)1_\Omega) = f - E(f)1_\Omega = \bar{f}.$$

Luego $\bar{f} \in L^2(\mathcal{C}) \cap \mathcal{H}$.

Así, $\bar{f}$ es ortogonal a $L^2(\mathcal{B})$. Por tanto $E^{\mathcal{B}}(\bar{f}) = 0$, es decir,

$$E^{\mathcal{B}}(f) = E(f)1_\Omega.$$

Cuando $f \in L^1$, el operador truncación definido en la ecuación (A.2) es usado y se pasa al límite, produciéndose el resultado deseado.

Ahora, supóngase que $f \in L^2(\mathcal{C}) \cap \mathcal{H}$. Entonces, $E^{\mathcal{B}}(f) = 0 = E(f)$, por lo que cada $f \in L^2(\mathcal{C}) \cap \mathcal{H}$ es ortogonal a $L^2(\mathcal{B})$. Se concluye que $\mathcal{B}$ y $\mathcal{C}$ son independientes. □

El teorema anterior tiene una interpretación concreta. Si $\mathcal{B}$ y $\mathcal{C}$ son independientes, entonces "la información de los eventos en la $\sigma -' algebra\ \mathcal{B}$"no altera el "el valor promedio"de una v.a. $\mathcal{C} - medible$.

Corolario 4.4.2. *Sean $\mathcal{B}$ y $\mathcal{C}$ $\sigma-$álgebras independientes del espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Entonces $E^{\mathcal{B}}E^{\mathcal{C}} = E$.*

Demostración. Sea $f \in L^1(\Omega, \mathcal{A}, P)$. Entonces $E^{\mathcal{C}}(f) \in L^1(\mathcal{C})$. Sea $u = E^{\mathcal{C}}(f)$, entonces

$$E(u) = E(E^{\mathcal{C}}(f)) = E(f).$$

Ya que $u \in L^1(\mathcal{C})$, por el teorema 4.4.1 $E^{\mathcal{B}}(u) = E(u) = E(f)$, es decir, $E^{\mathcal{B}}E^{\mathcal{C}} = E$. □

4.5. Independencia y distribuciones (caso de dos variables aleatorias.)

Proposición 4.5.1. *Sea X_1 y X_2 dos v.a.'s definidas en el espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Sea μ_1 y μ_2 que denotan la distribución de X_1 y X_2, respectivamente, y sea μ que denota su distribución conjunta. Entonces los siguientes enunciados son equivalentes:*

i) *X_1 y X_2 son v.a.'s independientes.*

ii) *Para toda función Boreliana acotada φ_1, φ_2, definidas en $\mathbb{R}$,*

$$E(\varphi_1(X_1)\varphi_2(X_2)) = E(\varphi_1(X_1))E(\varphi_2(X_2))$$

iii) $\mu = \mu_1 \otimes \mu_2$

Demostración. i)$\Leftrightarrow$ ii). Sea $\mathcal{B}_i = X_i^{-1}(\mathcal{B}_{\mathbb{R}})$. Entonces $\mathcal{B}_i$ son independientes pues las v.a.'s X_1 y X_2 lo son. Sea $f_i \in L^2(\mathcal{B}_i)$. Por el teorema de Dynkin 2.3.6 existen funciones Borelianas $\psi_i : \mathbb{R} \to \mathbb{R}$ tales que $\psi_i(X_i) = f_i, i \in \{1, 2\}$. Entonces, i) es equivalente a $E(\varphi_1(X_1)\varphi_2(X_2)) = E(\varphi_1(X_1))E(\varphi_2(X_2))$ por la proposición 4.3.1 para toda función Boreliana ψ_i tal que $\psi_i(X_i) \in L^2$. Obsérvese que, usando el operador truncación, se muestra que la condición antes expuesta es equivalente pues siempre es posible aproximar funciones acotadas mediante este operador. El recíproco se sigue inmediatamente del teorema 4.3.1.

ii)$\Rightarrow$ *iii)*

Sea $C, D \in \mathcal{B}_{\mathbb{R}}$. Defínase $\varphi = 1_C$ y $\varphi = 1_D$. Entonces,

$$\int_{\mathbb{R}^2} 1_{C\times D} d\mu = E(1_C(X_1)1_D(X_2)).$$

Pero $\int_{\mathbb{R}} 1_C d\mu_1 = E(1_C(X_1))$. Luego, por ii) $\mu(C \times D) = \mu(C)\mu(D)$. Como μ es medida de Borel en $\mathbb{R}^2$ y, por la definición A.0.31, $\mathcal{B}_{\mathbb{R}^2} = \mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}}$, se tiene que $\mu = \mu_1 \otimes \mu_2$.

iii)$\Rightarrow$ ii).

$$
\begin{aligned}
E(\varphi_1(X_1)\varphi_2(X_2)) &= \int_{\mathbb{R}^2} \varphi_1(\xi_1)\varphi_2(\xi_2)d\mu_1(\xi_1)d\mu_2(\xi_2) \\
&= \int_{\mathbb{R}} \varphi_1(\xi_1)d\mu_1(\xi_1)\left(\int_{\mathbb{R}} \varphi_2(\xi_2)d\mu_2(\xi_2)\right)
\end{aligned}
$$

según el teorema A.0.30. □

4.6. Un espacio de funciones en la $\sigma-$álgebra generada por dos $\sigma-$álgebras.

A continuación se presenta un espacio de funciones para sigmas álgebras generadas por otras dos. Enunciamos este hecho como un teorema cuya prueba esta dada en una versión debilitada pues queda fuera de los límites del presente trabajo. La nombramos debilitada debido a la hipótesis adicional sobre la existencia de ciertos mapeos que son fundamentales para la prueba que se dará en breve. El caso general puede ser revisado en [7] págs. 218-228.

Teorema 4.6.1. *Sea $\mathcal{B}$ y $\mathcal{C}$ dos sub$\sigma-$álgebras del espacio de probabilidad $(\Omega, \mathcal{A}, P)$ y sea $\mathcal{D}$ que denota la $\sigma-$álgebra que ellos generan. Sea V el subespacio vectorial de $L^\infty(\mathcal{A})$ definido por*

$$
V = \left\{ h \in L^\infty(\mathcal{A}) : h = \sum_{i=1}^{n} f_i g_i, f_i \in L^\infty(\mathcal{B}) \quad y \quad g_i \in L^\infty(\mathcal{C}) \right\}.
$$

Entonces $V \subset L^2(\mathcal{D})$ y V es denso en $L^2(\mathcal{D})$.

Demostración. Supóngase que existen dos mapeos

$$
u : \Omega \to \mathbb{R}^n \quad \text{y} \quad v : \Omega \to \mathbb{R}^n
$$

tales que

$$
u^{-1}(\mathcal{B}_{\mathbb{R}^n}) = \mathcal{B} \quad \text{y} \quad v^{-1}(\mathcal{B}_{\mathbb{R}^p}) = \mathcal{C}.
$$

Sea $W : \Omega \to \mathcal{R}^{n+p}$ definido por $W(\omega) = (u(\omega), v(\omega))$. Entonces $W^{-1}(\mathcal{B}_{\mathbb{R}^{n+p}})$ es una $\sigma-$álgebra que contiene a $\mathcal{B}$ y $\mathcal{C}$. Más aún, por A.0.31

$\mathcal{B}_{\mathbb{R}^{n+p}} = \mathcal{B}_{\mathbb{R}^n} \otimes \mathcal{B}_{\mathbb{R}^p}$. Entonces $\mathcal{B}_{\mathbb{R}^{n+p}}$ es generada por los rectángulos $R = X \times Y$, donde $X \in \mathcal{B}_{\mathbb{R}^n}, Y \in \mathcal{B}_{\mathbb{R}^p}$. Así

$$W^{-1}(\mathbb{R}) = \{\omega : u(\omega) \in X \quad \text{y} \quad v(\omega) \in Y\} = u^{-1}(X) \cap v^{-1}(Y).$$

Luego $W^{-1}(R) = \mathcal{D}$.

Ahora se demostrará que $\mathcal{D} = W^{-1}(\mathbb{R}^{n+p})$.

Sea ρ la distribución de W Y W^* el mapeo imágen inversa. Por la proposición 2.5.2

$$W^* : L^2(\mathcal{B}_{\mathbb{R}^{n+p}}, \rho) \to L^2(\mathcal{D}),$$

el cual es una simetría suprayectiva. Cabe mencionar que las funciones continuas con soporte compacto,$C_K(\mathbb{R}^{n+p})$, son densas en $L^2(\mathcal{B}_{\mathbb{R}^{n+p}}, \rho)$. Puede consularse [7] págs. 78-79 para una demostración de este hecho.

Sea $\varphi \in C_K(\mathbb{R}^{n+p})$. Entonces, por el teorema de Stone-Weiestrass (véase [4]) existe una sucesión de polinomios P_r que convergen uniformemente a φ en un conjunto compacto $K_1 \times K_2$ el cual contiene el soporte de φ. Sea $g_r = P_r 1_{K_1 \times K_2}$. Entonces $\| g_r - \varphi \|_{L^2(\rho)} \to 0$.

Ahora se probará que $W^*(g_r) \in V$. Como P_r es la suma de monomios de la forma

$$(u^1)^{m_1} \dots (u^n)^{m_n} (v^1)^{q_1} \dots (v^p)^{q_p}$$

y siendo

$$f = 1_{K_1}(u^1)^{m_1} \dots (u^n)^{m_n} \quad \text{y} \quad g = 1_{K_2}(v^1)^{q_1} \dots (v^p)^{q_p},$$

$W^*(g_r)$ es combinación lineal de funciones de la forma fg. De modo que $W^*(C_k(\mathbb{R}^{n+p}))$ es denso en $L^2(\mathcal{D})$, y la convergencia de g_r a φ en $L^2(\mathcal{C})$ corresponde una convergencia en $L^2(\mathcal{D})$. En conclusión $\mathcal{D} = W^{-1}(\mathbb{R}^{n+p})$. $\square$

Corolario 4.6.2. *Sea $\mathcal{B}_1 \dots \mathcal{B}_n$ una colección finita de sub$\sigma-$álgebras del espacio de probabilidad $(\Omega, \mathcal{A}, P)$ y sea $\mathcal{D}$ la $\sigma-$álgebra generada por $\mathcal{B}_1 \dots \mathcal{B}_n$. Sea*

$$W_n = \left\{ \sum_{p=1}^{q} f_p^1 f_p^2 \cdots f_p^n, f_p^i \in L^\infty(\mathcal{B}_i) \right\}.$$

Entonces $W_n \subset L^2(\mathcal{D})$ y W_n es denso en $L^2(\mathcal{D})$

Demostración. Por inducción sobre n.

Si $n = 2$, el teorema 4.6.1 prueba este caso.

Supóngase que W_n es denso en $L^2(\mathcal{C})$, siendo $\mathcal{C}$ la $\sigma-$álgebra generada por $\mathcal{B}_2 \dots \mathcal{B}_{n+1}$. Por demostrar válido para $n+1$.

La $\sigma-$álgebra generada por $\mathcal{B}_1 \dots \mathcal{B}_{n+1}$ es la $\sigma-$álgebra generada por $\mathcal{B}_1$ y $\mathcal{C}$. Sea

$$V = \left\{ h : h = \sum f_i g_i, f_i \in L^\infty(\mathcal{B}_i), g_i \in L^\infty(\mathcal{C}) \right\}.$$

Entonces por la base de inducción, es decir, por el teorema 4.6.1, V es denso en $L^2(\mathcal{D})$. Sea

$$V^{'} = \left\{ h : h = \sum f_i g_i, f_i \in L^\infty(\mathcal{B}_1), g_i \in L^2(\mathcal{C}) \right\}.$$

Entonces $V^{'} \subset L^2(\mathcal{D})$ y $V^{'}$ es denso en $L^2(\mathcal{D})$ puesto que $V \subset V^{'}$. Por hipótesis de inducción, cada g_i puede ser aproximada por elementos de W_n. Entonces existe una sucesión $K_i^s \in W_n$ tal que $\| K_i^s - g_i \|_{L^2} \rightarrow 0$ y

$$\| \sum f_i g_i - \sum f_i K_i^s \|_{L^2} \leq \sum \| f_i \|_{L^\infty} \| g_i - K_i^s \|_{L^2} \rightarrow 0.$$

Además, $\sum f_i K_i^s \in W_n$. Por tanto, $W_{n+1} \subset L^2(\mathcal{D})$ y W_{n+1} es denso en $L^2(\mathcal{D})$. □

4.7. Independencia y Distribución (caso de n variables aleatorias).

Teorema 4.7.1. *Sea $\mathcal{B}_1, \dots \mathcal{B}_n$ $\sigma-$álgebras del espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Entonces las siguientes afirmaciones son equivalentes:*

i) $\mathcal{B}_1, \dots \mathcal{B}_n$ *son mutuamente independientes.*

ii) $E\left(\prod_{i=1}^{n} f_i\right) = \prod_{i=1}^{n} E(f_i)$ *para toda* $f_i \in L^\infty(\mathcal{B}_i)$.

Demostración. La implicación i)⇒ii) fue probada en el teorema 4.5.1. Demostraremos la otra implicación. Sea H un subconjunto de $\{1, \ldots, n\}$. Sea $H^{'}$ el complemento de H, y sea $\mathcal{C}$ y $\mathcal{C}^{'}$ las $\sigma-$álgebras generadas por $\{\mathcal{B}_i : i \in H\}$ y $\{\mathcal{B}_j : j \in H^{'}\}$ respectivamente. Se demostrará la independencia de $\mathcal{C}$ y $\mathcal{C}^{'}$. Por le teorema 4.3.1, se tendrá que $E(gg^{'}) = E(g)E(g^{'})$, para toda $g \in L^2(\mathcal{C}), g^{'} \in L^2(\mathcal{C}^{'})$. Entonces, por el espacio de funciones construido en 4.6.1 en la $\sigma-$álgebra generada por $\mathcal{C}$ y $\mathcal{C}^{'}$ y la bilinealidad del proucto, se tiene que

$$E\left(\prod_{i\in H} f_i \prod_{j\in H'} f_j\right) = \prod_{i\in H} E(f_i) \prod_{j\in H'} E(f_j).$$

Entonces, si $f_i = 1$ si $i \in H^{'}$, se tiene que el primer término en la ecuación anterior del lado derecho es $E(g)$ y similarmente el segundo $E(g^{'})$. En consecuencia $E(gg^{'}) = E(g)E(g^{'})$. □

Como consecuencia inmediata de este teorema, se tiene el siguiente resultado que ha sido probado en 4.3.1 para $n = 2$ y cuya demostración general es idéntica a éste. Lo enunciamos en el siguiente teorema.

Teorema 4.7.2. *Sea $X_1, \ldots, X_n$ v.a's en el espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Los siguientes enunciados son equivalentes:*

i) Las v.a.'s X_k son mutuamente independientes.

ii) Para toda función Boreliana acotad φ_k en $\mathbb{R}$,

$$E\left(\prod_{k=1}^{n} \varphi_k(X_k)\right) = \prod_{k=1}^{n} E(\varphi_k(X_k)).$$

iii) Sea μ la disribución conjunta de $X_1, \ldots, X_n$ y sea μ_i la distribución de X_i. Entonces

$$\mu(A_1 \times A_2 \times \cdots \times A_n) = \prod_{i=1}^{n} \mu_i(A_i)$$

para toda $A_i \in \mathcal{B}_{\mathbb{R}}$. En otras palabras, $\mu_{\otimes}\mu_i$.

Capítulo 5

Aplicaciones

El objetivo de este capítulo es mostrar la aplicación de la teoría expuesta con anterioridad en ciertos teoremas límite que son de gran relevancia en teoría de la probabilidad y en estadística, así como algunos otros ejemplos.

La siguiente proposición muestra la aproximación a la distribución binomial por la distribución poisson. Ésta da una cota superior del error cometido por reemplazar una distribución binomial por una distribución poisson, también trata el caso de experimentos que son independientes pero no idénticamente distribuidos. Requeriremos de algunos resultados previos para su demostración y de la definición de la operación convolución.

Definición 34. *La función $z \mapsto \int_{-\infty}^{\infty} f_X(x)f_Y(y)(z-x)dx$ es llamada la convolución de f_X y f_Y y se denota por $f_X * f_Y$.*

Obsérvese que si X y Y son independientes, $f_X * f_Y$ es una función de densidad de $X+Y$.

Lema 5.0.3. *Supóngase que δ_a es la medida de Dirac en a, $p \in (0,1)$ y $\lambda > 0$. Considérese las siguientes dos medidas de probabilidad en $\mathbb{N}$:*

$$\nu_p = (1-p)\delta_0 + p\delta_1 \quad (\textit{Distribución Bernoulli con parámetro} \quad p)$$

$$\mu_\lambda = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}\delta_k \quad (\textit{Distribución Poisson con parámetro} \quad \lambda)$$

Entonces el límite vago de la sucesión $\{\nu^{*n}_{\frac{\lambda}{n}}\}_{n>\lambda}$ *es* μ_λ*, donde* $*$ *denota la operación convolución.*

Demostración. Se probará que que la sucesión $\{\nu^{*n}_{\frac{\lambda}{n}}\}_{n>\lambda}$ converge en distribución a μ_λ. Sea $p \in (0,1)$ y sea $q = 1-p$. Obsérvese que

$$\nu_p^{*n} = (q\delta_0 + p\delta_1)^{*n} = \sum_{k=1}^{n} \binom{n}{k} p^k q^{n-k} \delta_k,$$

que es la distribución de la v.a. binomial. Entonces

$$\begin{aligned} \nu^n_{\frac{\lambda}{n}}(k) &= \binom{n}{k} \frac{\lambda^k}{n^k}(1-\frac{\lambda}{n})^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!}\frac{\lambda^k}{n^k}(1-\frac{\lambda}{n})^{n-k} \\ &= \frac{\lambda^k}{k!}(1-\frac{1}{n})\cdots(1-\frac{k+1}{n})(1-\frac{\lambda}{n})^{n-k} \to \frac{\lambda^k}{k!}e^{-\lambda} = \mu_\lambda(k) \end{aligned}$$

cuando $n \to \infty$, esto es, $\{\nu^{*n}_{\frac{\lambda}{n}}\}_{n>\lambda}$ converge en distribución a μ_λ. Como la convergencia en distribución implica la convergencia vaga, se obtiene el resultado deseado. □

Lema 5.0.4. *Sea* $0 < p < 1$*. Considérese la medida* m_p *en* $\mathbb{N}^2$ *concentrada en los puntos* $(0,0), (0,1), (1,1)$ *y* $(k,0)$ *con* $k \geq 2$ *tales que* X *tiene distribución* μ_p *y* Y *tiene la distribución* ν_p *si* (x,y) *tiene distribución* m_p*. Entonces* $P(X \neq Y) \leq 2p^2$.

Demostración. Comenzaremos calculando m_p.

Puesto que $P(X=1) = pe^{-p}$, entonces $m_p(1,1) = pe^{-p}$ debido a la ausencia del punto $(1,0)$. Por otra parte,

$$P(Y=1) = p = m_p(0,1) + m_p(1,1),$$

luego

$$m_p(0,1) = p - m_p(1,1) = p - pe^{-p} = p(1-e^{-p}).$$

De aquí que

$$P(X=0) = e^{-p} = m_p(0,0) + m_p(0,1)$$

implica que

$$m_p(0,0) = e^{-p} - m_p(0,1) = e^{-p} - p(1-e^{-p}).$$

De hecho $m_p(k,0) = e^{-p}\frac{p^k}{k!}$. Luego

$$\begin{aligned} P(X = Y) &= m_p(0,0) + m_p(1,1) = e^{-p} - p(1-e^{-p}) + pe^{-p} \\ &= e^{-p} + 2pe^{-p} - p = e^{-p}(1+2p) - p \\ &\geq (1-p)(1+2p) - p = 1 - 2p^2. \end{aligned}$$

Por tanto $1 - P(X \neq Y) \geq 1 - 2p^2$ y por consiguiente $P(X \neq Y) \leq 2p^2$. □

Lema 5.0.5. *Si (X,Y) es una v.a. arbitraria en $\mathbb{N}^2$ y $A \subset \mathbb{N}$, se tiene que*

$$|P(X \in A) - P(Y \in A)| \leq P(X \neq Y).$$

Demostración. Obsérvese que para una v.a. Y fija se tiene que

$$\{X \in A\} = \{X = Y \in A\} \cup \{X \neq Y \in A\} \subset \{Y \in A\} \cup \{X \neq Y\}.$$

Por monotonía de la medida de probabilidad

$$P(X \in A) \leq P(Y \in A) + P(X \neq Y).$$

De manera análoga se obtiene

$$P(Y \in A) \leq P(X \in A) + P(X \neq Y).$$

En conclusión

$$|P(X \in A) - P(Y \in A)| \leq P(X \neq Y).$$

□

Proposición 5.0.6. *Sean $(X_1,Y_1),\ldots,(X_n,Y_n)$ v.a.'s independientes con valores en $\mathbb{N}^2$ y con distribuciones $m_{p_1},\ldots,m_{p_n}$. Sea $A \subset \mathbb{N}$. Entonces*

$$|P((X_1 + \cdots + X_n) \in A) - P((Y_1 + \cdots + Y_n) \in A)| \leq 2\sum_{j=1}^{n} p_j^2.$$

Demostración. Sean $X = X_1 + \cdots X_n$ y $Y = Y_1 + \cdots + Y_n$. Como

$$\{X \neq Y\} \subset \bigcup_{j=1}^{n} \{X_j \neq Y_j\},$$

por el lema anterior

$$|P(X \in A) - P(Y \in A)| \leq P(X \neq Y) \leq \sum_{j=1}^{n} P(X_j \neq Y_j) \leq \sum_{j=1}^{n} 2p_j^2.$$

□

Proposición 5.0.7. *Si $n > \lambda$ y $A \subset \mathbb{N}$, $|\nu_{\frac{\lambda}{n}}^{*n}(A) - \mu_\lambda(A)| \leq \frac{2\lambda^2}{n}$.*

Demostración. Haciendo $p_j = \frac{\lambda}{n}$ para toda $j \in \{1, \ldots, n\}$, aplicamos la proposición anterior y se obtiene la desigualdad buscada puesto que la distribución de $X_1 + \cdots + X_n$ es μ_λ y la de $Y_1 + \cdots + Y_n$ es $\nu_{\frac{\lambda}{n}}^*$ segun el primer lema. □

A continuación damos pruebas breves de la ley débil y fuerte de los grandes números. Comenzaremos enunciando y demostrando dos desigualdades importantes.

Lema 5.0.8 (Desigualdad de Chebyshev)**.** *Sea Y una v.a. positiva. Para toda $y > 0$,*

$$P(Y \geq y) \leq \frac{1}{y} E(Y)$$

Demostración. Ya que

$$Y1_{\{Y \geq y\}} + Y1_{\{Y < y\}} \leq y1_{\{Y \geq y\}}$$

entonces

$$E(Y) \geq E(y1_{\{Y \geq y\}}) = yP(Y \geq y).$$

Por tanto

$$\frac{1}{y} E(Y) \geq P(Y \geq y).$$

□

Lema 5.0.9 (Desigualdad de Beinamé)**.** *Sea X v.a. tal que $E(X^2) < \infty$. Si $m = E(X)$, para todo $t > 0$,*

$$P(|X - m| \geq t) \leq \frac{1}{t^2}(E((X - m)^2)).$$

Demostración. Sea $Y = (X - m)^2$ y $y = t^2$. Por el inciso anterior se tiene que

$$P(|X - m| \geq t) \leq \frac{1}{t^2}(E((X - m)^2).$$

□

Proposición 5.0.10 (Ley Débil de los Grandes Números)**.** *Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de v.a.'s reales independientes con la misma distribución y tal que $E(X_1^2) < \infty$. Si $m = E(X_1)$. Entonces para toda $\epsilon > 0$ y para toda $\alpha \in [0, \frac{1}{2})$*

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - m\right| \geq \frac{\epsilon}{n^\alpha}\right) \to 0 \quad \textit{cuando} \quad n \to \infty.$$

Demostración. Sea $X = \frac{1}{n}(X_1 + \cdots + X_n)$. Entonces

$$E(X) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}nE(X_1) = E(X_1).$$

También

$$\begin{aligned} E((X - m)^2) &= Var(X) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} E((X - m)^2) = \frac{n}{n^2}E((X_1 - m)^2). \end{aligned}$$

Luego, dada $\frac{\epsilon}{n^\alpha} > 0$, por la desigualdad de Beinamé

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - m\right| \geq \frac{\epsilon}{n^\alpha}\right) \leq \frac{n^{2\alpha}}{\epsilon^2}E((X-m)^2) = \frac{n^{2\alpha}}{n\epsilon^2}E((X_1-m)^2) \to 0$$

cuando $n \to \infty$. □

Lema 5.0.11. *Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de v.a.'s independientes e identicamente distribuidas para las cuales existe $k > 0$ tal que $E(e^{k|X_1|}) < \infty$. Si $m = E(X_1)$, para cada $\epsilon > 0$ existe $q \in (0,1)$ tal que*

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - m\right| \geq \epsilon\right) \leq 2q^n.$$

Demostración. Defínase

$$A_n = \{\frac{1}{n}(X_1 + \cdots + X_n) - m \geq \epsilon\} \quad \text{y} \quad B_n = \{\frac{1}{n}(X_1 + \cdots + X_n) - m \leq \epsilon\}.$$

Entonces, por la desigualdad de Chebyshev

$$P(A_n) = P(e^{s(X_1+\cdots+X_n)} \geq e^{sn(\epsilon+m)}) \leq e^{-sn(\epsilon+m)}E(e^{s(X_1+\cdots+X_n)})$$

con $0 \leq s < \alpha$. Defínase

$$\varphi(s) = e^{-s(\epsilon+m)}E(e^{sX_1}).$$

Entonces

$$P(A_n) \leq \varphi(s)^n.$$

Nótese que

$$\varphi^{'}(s) = -(\epsilon + m)e^{-s(\epsilon+m)}E(e^{sX_1}) + e^{-s(\epsilon+m)}E(X_1 e^{sX_1}).$$

Luego $\varphi^{'}(0) = -\epsilon - m + m = -\epsilon < 0$. Además $\varphi(0) = 1$. Luego existe $s_1 \in (0, \alpha]$ tal que $q_1 = \varphi(s_1) \in (0,1)$. Por otra parte, aplicando el resultado anterior a $(-X_n)_{n\in\mathbb{N}}$ implica que existe $q_2 \in (0,1)$ tal que $P(B_n) \leq q_2^n$. Sea $q = \text{máx}\{q_1, q_2\}$. Luego

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - m\right| \geq \epsilon\right) \leq 2q^n.$$

□

Proposición 5.0.12 (Ley Fuerte de los Grandes Números)**.** *Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de v.a.'s independientes e identicamente distribuidas para las cuales existe $k > 0$ tal que $E(e^{k|X_1|}) < \infty$. Si $m = E(X_1)$,*

$$\frac{1}{n}(X_1 + \cdots + X_n) \to m \quad \textit{a.s. cuando} \quad n \to \infty.$$

Demostración. Del lema 5.0.11 se sigue que

$$\begin{aligned}\sum_{n=1}^{\infty} P\left(\left|\frac{X_1+\cdots+X_n}{n}-m\right|\geq\epsilon\right) &\leq \sum_{n=1}^{\infty} 2q^n = 2\sum_{n=0}^{\infty} q^n - 1 \\ &= 2(\frac{1}{1-q}-1) = 2(\frac{q}{1-q}) < \infty.\end{aligned}$$

Haciendo uso del lema de Borell-Cantelli A.0.32:

$$P(\limsup_{n\to\infty} |\frac{1}{n}(X_1+\cdots+X_n)-m|\geq\epsilon) = 0.$$

Entonces

$$P(\liminf_{n\to\infty} |\frac{1}{n}(X_1+\cdots+X_n)-m|\leq\epsilon) = 1,$$

es decir,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to m \quad \text{a.s. cuando} \quad n\to\infty.$$

□

La siguiente proposición hace uso de un bien conocido teorema llamado del Límite Central, el cual enunciamos en la sección de preliminares, así como de la ley débil de los grandes números 5.0.10.

Proposición 5.0.13. *Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de v.a's independientes reales con la misma distribución y tales que $E(X_1)=0$ y $E(X_1^2)<\infty$. Sea $S_n = X_1+\cdots+X_n$. Entonces*

$$\lim_{n\to\infty} P(S_n\geq 0) = \frac{1}{2} \quad y \quad \lim_{n\to\infty}\left(E(e^{-S_n 1_{S_n\geq 0}})\right)^{1/n} = 1.$$

Demostración. En virtud del Teorema del Límite Central A.0.33, la distribución μ_n de $\frac{S_n}{\sqrt{n}\sigma}$ converge en distribución a $\nu = e^{-\frac{x^2}{2}}\frac{dx}{\sqrt{2\pi}}$. Así

$$\mu_n([0,\infty)) \to \nu([0,\infty)).$$

Como

$$\nu([0,\infty)) = \int_0^{\infty} e^{-\frac{x^2}{2}}\frac{dx}{\sqrt{n}\sigma} = \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\pi}{2}} = \frac{1}{2}$$

se obtiene el resultado deseado. Considérese la función $f(x) = x^{n^{-3/4}}$ con $x > 0$. Nótese que tal función es cóncava, de modo que la desigualdad de Jensen se satisface pero en orden opuesto. Por tanto se obtiene

$$(E(e^{-S_n}1_{\{S_n \geq 0\}}))^{n^{-3/4}} \geq E(e^{-n^{-3/4}S_n}1_{\{S_n \geq 0\}}).$$

Por otra parte, en virtud de la ley débil de los grandes números (Proposición (5.0.12)), se tiene que

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n}\right| \geq \frac{\epsilon}{n^{1/4}}\right) = \lim_{n\to\infty} P\left(\left|\frac{S_n}{n^{3/4}}\right| \geq \epsilon\right) = 0,$$

asegurándose así que

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n^{3/4}}\right| \leq \epsilon\right) = 1$$

para toda $\epsilon > 0$. Como consecuencia de esto, se precisa que

$$e^{-n^{-3/4}S_n} \geq 1 - n^{-3/4}S_n \geq 1 - \epsilon.$$

De aquí que

$$\begin{aligned} E(e^{-n^{-3/4}S_n}1_{\{S_n \geq 0\}}) &\geq (1-\epsilon)P(S_n \geq 0) \\ &\geq (1-\epsilon)P(\{S_n \geq 0\} \cap \{e^{-n^{-3/4}S_n} \geq 1-\epsilon\}). \end{aligned}$$

Obsérvese que

$$\begin{aligned} 1 &\geq P(\{S_n \geq 0\} \cup \{e^{-n^{-3/4}S_n} \geq 1-\epsilon\}) \\ &= P(\{S_n \geq 0\}) + P(\{e^{-n^{-3/4}S_n} \geq 1-\epsilon\}) \\ &- P(\{S_n \geq 0\} \cap \{e^{-n^{-3/4}S_n} \geq 1-\epsilon\}). \end{aligned}$$

De aquí que

$$(1-\epsilon)(P(\{S_n \geq 0\} \cap \{e^{-n^{-3/4}S_n} \geq 1-\epsilon\})$$

$$\geq (1-\epsilon)(P(\{S_n \geq 0\}) - P(\{e^{-n^{-3/4}S_n} < 1-\epsilon\})).$$

En resumen,

$$(E(e^{-S_n}1_{\{S_n\geq 0\}}))^{n^{-3/4}} \geq (1-\epsilon)(P(\{S_n \geq 0\}) - P(\{e^{-n^{-3/4}S_n} < 1-\epsilon\})).$$

De donde se sigue que

$$\liminf_{n\to\infty}(E(e^{-S_n}1_{\{S_n\geq 0\}}))^{n^{-3/4}} \geq (1-\epsilon)\frac{1}{2} > 0$$

y por tanto

$$\liminf_{n\to\infty}(E(e^{-S_n}1_{\{S_n\geq 0\}}))^{1/n} \geq 1.$$

Ya que

$$e^{-S_n}1_{\{S_n\geq 0\}} \leq 1,$$

se obtiene el resultado buscado. □

Proposición 5.0.14. *Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de v.a's reales independientes con la misma distribución. Supóngase que $\varphi(t) = E(e^{tX_1})$ existe para toda t en un intervalo abierto I conteniendo al cero y un número real fijo $s > E(X)$ tal que $t \mapsto e^{-ts}\varphi(t)$ alcanza su mínimo $\alpha(s)$ en un punto τ de I. Defínase $S_n = X_1 + \cdots + X_n$. Entonces*

$$\alpha(s) = \lim_{n\to\infty}\left[P\left(\frac{S_n}{n} \geq s\right)\right]^{\frac{1}{n}}.$$

Demostración. La demostración requerirá de varios pasos previos. Primeramente mostraremos que

$$\left[P\left(\frac{S_n}{n} \geq s\right)\right]^{\frac{1}{n}} \leq \alpha(s).$$

Para ello se probará que $\log(\varphi(t))$ es convexa en I y que $\tau > 0$. Nótese que $\varphi^n(t) = E(X_1^n e^{tX_1})$. Probaremos la convexidad de la función $\log(\varphi(t))$ haciendo uso de la desigualdad de Hölder. Sea $\theta \in (0,1)$. Defínase las v.a.'s $Y = e^{(1-\theta)t_1X_1}$ y $Z = e^{\theta t_2 X_1}$. Considérese $p = \frac{1}{1-\theta}$ y $q = \frac{1}{\theta}$. Nótese que $\frac{1}{p} + \frac{1}{q} = 1$, es decir, p y q son exponentes conjugados. Entonces, por la desigualdad de Hölder A.0.27 se tiene que

$$E(|YZ|) \leq (E(|Y|^p))^{1/p}(E(|Z|^q))^{1/q}.$$

Dicho de otro modo

$$E(e^{(1-\theta)t_1X_1+\theta t_2X_1}) \leq (E(e^{t_1X_1}))^{1-\theta}(E(e^{t_2X_1}))^{\theta}.$$

Por lo tanto

$$\log(E(e^{(1-\theta)t_1X_1+\theta t_2X_1})) \leq \theta\log(E(e^{t_1X_1})) + (1-\theta)E(e^{t_2X_1}).$$

Cuando $\theta = 0,1$ la inecuación se convierte en igualdad. En conclusión, $\log(\varphi(t))$ es convexa, lo cual implica que $(\log(\varphi(t)))^{''} \geq 0$. En consecuencia $\frac{\varphi^{'}(t)}{\varphi(t)}$ es creciente. Por otra parte, ya que en τ se alcanza el mínimo de $g(t) = e^{-st}\varphi(t)$, resulta cierto que

$$g^{'}(\tau) = -se^{-s\tau}\varphi(\tau) + e^{-s\tau}E(X_1e^{\tau X_1}) = 0.$$

Luego

$$E(X_1e^{\tau X_1}) = s\varphi(\tau)$$

y finalmente

$$\frac{\varphi^{'}(\tau)}{\varphi(\tau)} = s.$$

Obsérvese que si $\tau \leq 0$ se tendría que

$$s = \frac{\varphi^{'}(\tau)}{\varphi(\tau)} \leq \frac{\varphi^{'}(0)}{\varphi(0)} = E(X_1)$$

lo cual contradice la hipótesis de que $s > E(X_1)$. Luego $\tau > 0$. Finalmente, haciendo uso de la desigualdad de Chebyshev obtenemos

$$P\left(\frac{S_n}{n} \geq s\right) = P(e^{\tau S_n} \geq e^{\tau sn}) \leq (e^{-\tau s}\varphi(\tau))^n = (\alpha(s))^n.$$

Por otra parte sea μ_1 la distribución de $X_1 - s$ y sea $\nu(dx) = \frac{e^{\tau x}}{\alpha(s)}\mu_1(dx)$. Probaremos que ν es una medida de probabilidad, que $\int x\nu(dx) = 0$, y también que $\int x^2\nu(dx) < \infty$. Obsérvese que

$$\int \nu(dx) = \frac{1}{\alpha(s)}E(e^{\tau(X_1-s)}) = \frac{E(e^{\tau(X_1-s)})}{E(e^{\tau(X_1-s)})} = 1.$$

Ya que τ pertenece a I,

$$\int |x|^n \nu(dx) = \frac{1}{\alpha(s)} E(|X_1 - s|^n e^{\tau(X_1 - s)}) < \infty$$

para toda $n \in \mathbb{N}$. Además, como $s = \frac{\varphi'(\tau)}{\varphi(\tau)}$ resulta

$$\begin{aligned}
\int x\nu(dx) &= \frac{e^{-\tau s}}{\alpha(s)} E((X_1 - s)e^{\tau X_1}) = \frac{E(X_1 e^{\tau X_1}) - sE(e^{\tau X_1})}{E(e^{\tau X_1})} \\
&= \frac{E(X_1 e^{\tau X_1}) - \frac{\varphi'(\tau)}{\varphi(\tau)}\varphi(\tau)}{E(e^{\tau X_1})} = \frac{E(X_1 e^{\tau X_1}) - E(X_1 e^{\tau X_1})}{E(e^{\tau X_1})} = 0.
\end{aligned}$$

Finalmente, se afirma que para $\{Z_n\}_{n=1}^{\infty}$ una sucesión de v.a's independientes con misma distribución ν,

$$P\left(\frac{S_n}{n} \geq s\right) = (\alpha(s))^n E(e^{-\tau(Z_1 + \cdots + Z_n)} 1_{\{Z_1 + \cdots + Z_n \geq 0\}}).$$

Damos razón de esto.

Defínase $Z_i = X_i - s$ para toda $i \in \{1, \ldots, n\}$. De modo que

$$\begin{aligned}
P\left(\frac{S_n}{n} \geq s\right) &= \int_{Z_1 + \cdots + Z_n \geq 0} \mu_1(dz_1) \ldots \mu_n(dz_n) \\
&= (\alpha(s))^n \int_{Z_1 + \cdots + Z_n \geq 0} e^{-\tau(Z_1 + \cdots + Z_n)} \nu(dz_1) \ldots \nu(dz_n) \\
&= (\alpha(s))^n E(e^{-\tau(Z_1 + \cdots + Z_n)} 1_{\{Z_1 + \cdots + Z_n \geq 0\}}).
\end{aligned}$$

Por la proposición anterior concluimos que, para las v.a.'s $\tau Z_1, \ldots, \tau Z_n$,

$$\lim_{n \to \infty} P\left(\frac{S_n}{n} \geq s\right)^{1/n} = \alpha(s).$$

□

Como ejemplos muy particulares tenemos que, para la distribución normal, $\varphi(t) = e^{t^2/2}$ y por tanto $\alpha(s) = e^{-\frac{s^2}{2}}$. Mientras que para la distribución gamma con $t < 1, \beta > 0$, $\varphi(t) = (1-t)^{-\beta}$, y por tanto $\alpha(s) = e^{-s+\beta}(\frac{s}{\beta})^{\beta}$.

La caracterización de la esperanza condicional que es abordada en la siguiente proposición es casi siempre tomada como una definición en la literatura matemática.

Proposición 5.0.15. *Sea* $(\Omega, \mathcal{A}, P)$ *un espacio de probabilidad y sea* $\mathcal{B}$ *una* $sub-\sigma algebra$ *de* $\mathcal{A}$. *Si* $X \in L^1(\mathcal{A})$, *entonces*

$$\int_B X dP = \int_B E(X|\mathcal{B}) dP \quad \textit{para toda} \quad B \in \mathcal{B}, \tag{5.1}$$

la cual caracteriza a $E(X|\mathcal{B})$.

Demostración. Empezamos mostrando que 5.1 se verifica si $X \in L^2(\mathcal{A})$.

Por definición, $X - E(X|\mathcal{B})$ es ortogonal a $L^2(\mathcal{B})$ y por tanto, a la función 1_B para cada $B \in \mathcal{B}$. Esto es

$$E((X - E(X|\mathcal{B}))1_B) = E(X1_B) - E(E(X|\mathcal{B})1_B) = 0.$$

Lo cual es equivalente a que

$$\int_B X dP = \int_B E(X|\mathcal{B}) dP.$$

Extendemos lo probado hasta ahora.

Afirmamos lo siguiente: si $X > 0$, entonces $L(X) = \lim_{n\to\infty} E(\min(X, n)|\mathcal{B})$. Si $X \in L^1(\mathcal{A})$, definimos $L(X) = L(X^+) - L(X^-)$, donde $X^+ = \max(X, 0)$ y $X^- = \max(-X, 0)$. Mostraremos que $L(X) \in L^1(\mathcal{B})$ y también que

$$\int_B (X - L(X)) dP = 0 \quad \text{para toda} \quad B \in \mathcal{B}.$$

Ya que $X > 0$, entonces $\min(X, n) \in L^2(\mathcal{B})$. Así, si $B \in \mathcal{B}$, por el teorema de convergencia monótona A.0.22

$$\int_B X dP = \int_B \lim_{n\to\infty} \min(X, n) dP = \lim_{n\to\infty} \int_B \min(X, n) dP.$$

En virtud de lo anterior resulta que

$$\lim_{n\to\infty} \int_B \min(X, n) dP = \lim_{n\to\infty} \int_B E(\min(X, n)|\mathcal{B}) dP.$$

Obsérvese que la función $n \mapsto E(\text{mín}(X,n)|\mathcal{B})$ es una sucesión creciente según 3.2.2. Nuevamente, por el teorema de convergencia monótona A.0.22 tenemos que

$$\lim_{n\to\infty} \int_B E(\text{mín}(X,n)|\mathcal{B})dP = \int_B L(X)dP.$$

En resumen, hemos probado que

$$\int_B XdP = \int_B L(X)dP.$$

Al igual que el límite de funciones $\mathcal{B} - medibles$ positivas, $L(X)$ es también positivo y $\mathcal{B} - medible$. Ya que $\int_\Omega XdP = \int_\Omega L(X)dP$, se tiene que $L(X)$ es integrable. El caso cuando X no necesariamente es positivo se sigue de la descomposición de $L(X) = L(X^+) - L(X^-)$.

Por otra parte, obsérvese que si $f, g \in L^1(\mathcal{B})$ son tales que $\int_B (f-g)dP = 0$ para cada $B \in \mathcal{B}$, entonces $f = g$. Damos un argumento para esto.

Supóngase que $f - g > 0$. Sea $\epsilon > 0$. Defínase $B_\epsilon = \{f - g \geq \epsilon\}$. Por hipótesis se tiene

$$0 = \int_{B_\epsilon} (f-g)dP \geq \epsilon P(B_\epsilon).$$

Con lo cual se sigue que $P(B_\epsilon) = 0$. Por tanto

$$P(f-g>0) = P(\cap_{n=1}^{\infty} B_{\frac{1}{n}}) \leq \sum_{n=1}^{\infty} P(B_{\frac{1}{n}}) = 0,$$

y en consecuencia $P(f - g > 0) = 0$. El caso cuando $f - g < 0$ se reduce al caso anterior y por tanto $P(f - g < 0) = 0$. En consecuencia $P(f = g) = 1$.

Finalmente afirmamos que $L(X)$ es un operador lineal acotado de $L^1(\mathcal{A})$ a $L^1(\mathcal{B})$ e inferimos que $L(X) = E(X|B)$. Damos demostración de esto.

Nótese que $L(X)$ es el único elemento de $L^1(\mathcal{B})$ que satisface la condición $\int_B (X - L(X))dP = 0$ para cada $b \in \mathcal{B}$. Para ver esto, supóngase otro elemento $S(X)$ que satisface la misma propiedad. Entonces

$$\int_B (X - L(X))dP = \int_B (X - S(X))dP,$$

de donde se obtiene que

$$\int_B (S(X) - L(X))dP = 0.$$

Por lo expuesto hasta ahora tenemos que

$$S(X) = L(X).$$

De este modo tenemos que para $X \in L^2(\mathcal{A})$ $L(X) = E(X|\mathcal{B})$. Por otra parte, L es lineal en L^1 pues para cualsequiera escalares $\lambda, \mu \in \mathbb{R}$ se tiene que

$$\int_B (\lambda X + \mu Y - \lambda L(X) - \mu L(Y))dP = 0$$

para toda $B \in \mathcal{B}$. Además, por lo anterior se tiene que

$$\begin{aligned} E(|L(X)|) &\leq E(L^+(X) + L^-(X)) = E(L^+(X)) + E(L^-(X)) \\ &= E(X^+) + E(X^-) = E(|X|), \end{aligned}$$

lo que prueba que es acotado. Ya que la extensión de $E(X|\mathcal{B})$ de $L^2(\mathcal{A})$ a $L^1(\mathcal{A})$ es única según 3.3.1 se sigue que $L(X) = E(X|\mathcal{B})$. □

En la siguiente proposición requeriremos la noción de clase monótona, la cual definimos a continuación.

Definición 35. *Una clase monótona es una familia $\mathcal{M}$ de subconjuntos de X tal que si $\{A_n\}_{n\in\mathbb{N}}$ es una sucesión monótona para la cual $A_n \in \mathcal{M}$ para cada n, entonces su límite está en $\mathcal{M}$.*

Proposición 5.0.16. *Supóngase que $(\Omega, \mathcal{A}, P)$ es un espacio de probabilidad, $\mathcal{B}$ es una sub $-\sigma-$álgebra de $\mathcal{A}$, Y es una v.a. $\mathcal{B}-$medible, y X es una v.a. independiente de $\mathcal{B}$. Considérese $f : \mathbb{R}^2 \to \mathbb{R}$ tal que $f(X, Y)$ es integrable. Si μ es la distribución de X, entonces*

$$E(f(X, Y)|\mathcal{B}) = \int_{-\infty}^{\infty} f(x, Y)\mu(dx). \tag{5.2}$$

Demostración. La demostración de esta proposición la damos en varios pasos.

Comenzamos probando que 5.2 se satisface si $f(x,y) = 1_I(x)1_J(y)$, donde I, J son subconjuntos borelianos de $\mathbb{R}$. Sea $\varphi = 1_J(Y)$. Aplicando el teorema 3.3.1 inciso iv) obtenemos el resultado deseado.

Sea $\mathcal{P}$ el álgebra Booleana en $\mathbb{R}^2$ que consiste de todos los conjuntos de la forma $E = \cup_{p=1}^{q} I_p \times J_p$, donde I_p, J_p son subconjuntos borelianos de $\mathbb{R}$. Mostramos que 5.2 se satisface si $f(x,y) = 1_E(x,y)$ con $E \in \mathcal{P}$.

Supondremos sin perdida de generalidad que $I_p \times J_p \cap I_q \times J_q = \emptyset$ para toda $p \neq q$, pues si algún rectángulo se traslapa con otro, su intersección pertenece a $\mathcal{P}$ y por tanto, tales rectángulos pueden ser puestos como unión de conjuntos disjuntos. De este modo 5.2 se sigue de lo anterior por linealidad.

Construimos ahora una clase especial de conjuntos útil para la prueba.

Sea $\mathcal{M}$ la familia de subconjuntos M de $\mathbb{R}^2$ tales que $f(x,y) = 1_M(x,y)$ satisface 5.2. Afirmamos que $\mathcal{M}$ es una clase monótona. Supóngase que $M_n \in \mathcal{M}$, con $\{M_n\}_{n\in\mathbb{N}}$ sucesión monótona creciente y $M = \cup_{n=1}^{\infty}$. Obsérvese que $1_{M_n}(X,Y) \to 1_M(X,Y)$ en $L^2(\mathcal{A})$. Entonces, por el teorema de convergencia monótona se tiene que

$$
\begin{aligned}
E(1_M(X,Y)|\mathcal{B}) &= \lim_{n\to\infty} E(1_{M_n}(X,Y)|\mathcal{B}) \\
&= \lim_{n\to\infty} \int_{-\infty}^{\infty} 1_{M_n}(X,Y)\mu(dx) \\
&= \int_{-\infty}^{\infty} (x,Y)\mu(dx).
\end{aligned}
$$

La demostración cuando $\{M_n\}_{n\in\mathbb{N}}$ es una sucesión monótona decreciente es idéntica a la ya expuesta. En conclusión, $\mathcal{M}$ es clase monótona.

Terminamos esta demostración probando los siguientes casos:

a) f es una función simple en $\mathbb{R}^2$.

   Haciendo uso del teorema de las clases monótonas A.0.34 en $\mathcal{M}$ y $\mathcal{P}$ resulta que $\mathcal{M}$ es el álgebra de Borel en $\mathbb{R}^2$. Por tanto 5.2 se satisface para combinaciones lineales finitas de funciones indicadoras de conjuntos borelianos en $\mathbb{R}^2$.

b) f es una función medible positiva con $f(X, Y)$ integrable.

Asumiendo que $f \geq 0$, entonces existe una sucesión creciente de funciones simples borelianas no negativas $\{f_n\}_{n \in \mathbb{N}}$ tal que $f_n \to f$. De modo que 5.2 se sigue por un argumento análogo al expuesto en el de la clase monótona construida.

c) El caso general. Éste se reduce al de b) puesto que $f = f^+ - f^-$, con $f^+ = \text{máx}(0, f)$ y $f^- = \text{máx}(0, -f)$.

$\square$

Capítulo 6

Conclusiones

Con base en el libro de Paul Malliavin, "Integration and Probability" [7], se han exhibido resultados importantes en el estudio de la Teoría de la Probabilidad: el Teorema de Representación de Stone, proposición base para fundamentar la relación entre el cálculo proposicional y el cálculo probabilístico mediante operaciones conjuntistas; el teorema de Dependencia Funcional y Medibilidad de Dynkin que permite la conexión entre espacios de probabilidad y variables aleatorias; nociones de convergencia para variables aleatorias; el estudio del operador esperanza condicional en el espacio natural L^2, sus posibles extensiones y una demostración de la desigualdad de Jensen condicionada usando aproximación mediante $\sigma-$álgebras; demostración poco común en la literatura probabilística; finalmente la noción de ortogonalidad e independencia en tal espacio. La interacción entre diversas disciplinas matemáticas queda clarificada.

Como posibles extensiones de este trabajo, queda por supuesto la conexión entre las variables aleatorias y el Análisis Complejo mediante el uso de funciones características; el uso de polinomios de Hermite y teoremas de convergencia de martingalas, herramientas claves para la comprensión de las bases de la Teoría de la Probabilidad y que da entrada al estudio de los espacios de probabilidad Gaussianos, los cuales son fundamentales en el Cálculo Estocástico de Variaciones, mejor conocido como Cálculo de Malliavin.

Apéndice A

Conceptos y resultados varios

La finalidad de este apéndice es dar los fundamentos y herramientas necesarias para abordar los capítulos anteriores. Para ello damos las definiciones y teoremas previos a la teoría expuesta.

Comenzamos dando la definición de átomo, término básico en diversas ramas de la matemática como la Probabilidad, la Topología y la Teoría de Conjuntos.

Definición 36. *Sea $\mathcal{A}$ una $\sigma-$álgebra finita. Sea $B \in \mathcal{A}$ tal que*

$$B^{'} \subset B, \quad B^{'} \in \mathcal{A} \quad \textit{implica que} \quad B^{'} = B \quad \textit{o} \quad B^{'} = \emptyset.$$

A los conjuntos que verifican esta propiedad se les denomina átomos.

Obsérvese que los átomos son los elementos mínimos con respecto a la relación de inclusión en una $\sigma-$álgebra. Si B y $\tilde{B}$ son átomos distintos, entonces $B \cap \tilde{B} = \emptyset$. Esto implica el siguiente resultado.

Proposición A.0.17. *Sea $\mathcal{A}$ una $\sigma-$álgebra finita en un conjunto Y. Entonces cada subconjunto no vacío en $\mathcal{A}$ es la unión de los átomos que contiene.*

A continuación damos un lema importante para abordar un resultado fundamental en la sección (2.1): el Teorema de Representación de Stone.

Lema A.0.18. *Sea $\mathcal{F}$ un filtro en $\mathcal{B}$ y sea $A_0 \in \mathcal{B}$ tal que $A \cap A_0 \neq \emptyset$ para todo $A \in \mathcal{F}$. La familia*

$$\mathcal{F}_{A_0} = \{Z \in \mathcal{B} : \textit{ Z contiene un conjunto de la forma } A \cap A_0 \textit{ con } A \in \mathcal{F}\} \textit{ es un filtro.} \tag{A.1}$$

Definición 37. *Un ultrafiltro es un filtro $\mathcal{U}$ de $\mathcal{B}$ tal que $\mathcal{F} = \mathcal{U}$ para cada filtro más fino que $\mathcal{U}$.*

El lema de Zorn muestra que, dado un filtro $\mathcal{F}_0$ siempre existe un ultrafiltro $\mathcal{U}$ más fino que $\mathcal{F}_0$. Para mayores detalles véase [2].

El siguiente resultado facilita la identificación de ultrafiltros.

Lema A.0.19. *Una condición necesaria y suficiente para que un filtro $\mathcal{U}$ sea un ultrafiltro en $\mathcal{B}$ es que, para toda $A_0 \in \mathcal{B}$, $A_0 \in \mathcal{U}$ o $A_0^c \in \mathcal{U}$.*

Una demostración de este resultado puede consultarse en [8].

Ahora bien, sea Φ un mapeo tal que $\Phi : \Omega \longrightarrow \Omega_1$ definido como en la ecuación (2.2), se tiene el siguiente resultado.

Lema A.0.20. $f \in \Phi(\Omega)$ *si y sólo si se satisfacen para todo* $A', A'', A''' \in \mathcal{B}$:

- $f(\emptyset) = 0$
- $f(A) \leq f(A')$ *si* $A \leq A'$
- $f(A'' \cap A''') = \inf\{f(A''), f(A''')\}$
- $f(A) + f(A^c) = 1$

Demostración. Supóngase que $f \in \Phi(\Omega)$, entonces $f(A) = 1_{\mathcal{U}_1}(A)$, $A \in \mathcal{B}$. Por propiedades de la función indicadora se tiene que

- $f(\emptyset) = 1_{\mathcal{U}}(\emptyset) = 0$.
- Si $A \leq A'$ entonces $f(A) = 1_{\mathcal{U}}(A) \leq 1_{\mathcal{U}}(A') = f(A')$.

- $f(A'' \cap A''') = 1_{\mathcal{U}}(A'' \cap A''') = \text{ínf}(1_{\mathcal{U}}(A''), 1_{\mathcal{U}}(A''')) = \text{ínf}(f(A''), f(A'''))$.
- Por el lema A.0.19 si $A \in \mathcal{U}$ entonces A^c no pertenece a $\mathcal{U}$, luego se tiene que $f(A) + f(A^c) = 1$.

Ahora supóngase que se satisfacen las cuatro propiedades de la proposición. En virtud de la cuarta propiedad y de la definición de ultrafiltro $f(A)$ es función característica. Las restantes propiedades se satisfacen para funciones características y, por tanto, $f \in \Phi(\Omega)$.

$\square$

Corolario A.0.21. *Sea $f \in \mathcal{L}^o(X, \mathcal{A})$. Entonces existe una sucesión φ_n de funciones simples convergiendo puntualmente a f, donde $\mathcal{L}^o(X, \mathcal{A})$ se define como en (2.5).*

Definición 38. *Se denota por $\mathcal{E}^1_\mu(X, \mathcal{A})$ el conjunto de todas las funciones simples integrables.*

Teorema A.0.22 (De Convergencia Monótona). *Si $\{f_n\}$ es una sucesión de funciones medibles en $\mathcal{M}^+(X, \mathcal{A})$, la cual es creciente (decreciente) y convergente a f, entonces $\displaystyle\int f d\mu = \lim_{n\to\infty} \int f_n d\mu$.*

Teorema A.0.23 (De Convergencia Dominada). *Sea $\{f_n\}$ una sucesión de funciones integrables la cual converge c.s. a una función medible de variable real f. Si existe una función integrable g tal que $|f_n| \leq g$ para toda $n \in \mathbb{N}$, entonces f es integrable y $\int f d\mu = \displaystyle\lim_{n\to\infty} \int f_n d\mu$.*

Proposición A.0.24. *Si X es un espacio métrico localmente compacto, $\sigma-$compacto y $\{\mu_n\} \subset M^1(X)$ una sucesión, entonces*

a) *si $\mu_n \to \mu$ vagamente en $M^1(X)$ y para cada $\epsilon > 0$ existe un conjunto compacto $K \subseteq X$, tal que $|\mu_n|(K^c) < \epsilon$ para toda $n \geq n_0$, entonces $\mu_n \to \mu$ estrechamente en $M^1(X)$.*

b) *si $\mu_n \to \mu$ vagamente en $M^1(X)$ y $\mu_n(X) \to \mu(X)$, entonces $\mu_n \to \mu$ estrechamente en $M^1(X)$.*

La demostración de este resultado puede consultarse en [14].

Proposición A.0.25. *Supóngase que* $(X, \mathcal{A}, \mu)$ *es un espacio de medida completo,* Y *es un espacio métrico,* $f_0 \in M_\mu((X, \mathcal{A}); (Y, \mathcal{B}_Y))$, *y* $\{f_n\}_{n\in\mathbb{N}}$ *es una sucesión en* $M_\mu((X, \mathcal{A}); (Y, \mathcal{B}_Y))$.

Si $d_\mu(f_n, f_0) \to 0$, *entonces existe una subsucesión* $\{f_{n_k}\}_{k\in\mathbb{N}}$ *de* $\{f_n\}_{n\in\mathbb{N}}$ *tales que* $f_{n_k} \to f_0 \quad \mu - a.e.$

La prueba de éste resultado puede ser consultada en [7] pág. 23.

Proposición A.0.26 (Desigualdad de Jensen)**.** *Sea* φ *una función convexa en* $[a, b] \subset \mathbb{R}$. *Sea* $\alpha_k (1 \leq k \leq n)$ *números positivos tales que* $\displaystyle\sum_{k=1}^{n} \alpha_k = 1$. *Entonces*

$$\varphi\left(\sum_{k=1}^{n} \alpha_k t_k\right) \leq \sum_{k=1}^{n} \alpha_k \varphi(t_k)$$

para toda $t_k \in [a, b]$.

Proposición A.0.27 (Desigualdad de Hölder)**.** *Sean* $p, q \in (1, \infty)$ *tales que* $\frac{1}{p} + \frac{1}{q} = 1$. *Si* $f \in L^p(\Omega)$ *y* $g \in L^q(\Omega)$, *entonces*

$$fg \in L^1(\Omega) \quad y \parallel fg \parallel_1 \leq \parallel f \parallel_p \parallel g \parallel_q .$$

Si $f \in L^\infty(\Omega)$ *y* $g \in L^1(\Omega)$, *entonces*

$$fg \in L^1(\Omega) \quad y \parallel fg \parallel_1 \leq \parallel f \parallel_\infty \parallel g \parallel_1 .$$

Proposición A.0.28. *Si* $\mu(\Omega) < \infty$ *y* $1 \leq p < s \leq \infty$, *entonces*

$$L^s(\Omega) \subset L^p(\Omega)$$

y esta inclusión es continua.

Definición 39. *Se definen los siguientes espacios de funciones acotadas medibles:*

$$\mathcal{L}^\infty(X, \mathcal{A}) = \{f \in \mathcal{L}^0(X, \mathcal{A}) : existe \quad M < \infty, |f(x)| < \infty\}.$$

$$\mathcal{L}_\mu^{\infty,1}(X, \mathcal{A}) = \{f \in \mathcal{L}^\infty(X, \mathcal{A}) : \mu(\{x : f(x) \neq 0\}) < \infty\}.$$

Definición 40. *Se define el operador truncamiento como*

$$T_n(f) = \begin{cases} f(x) & si \quad |f(x)| \leq n \\ n & si \quad f(x) > n \\ -n & si \quad f(x) < -n \end{cases} \tag{A.2}$$

Donde $f \in \mathcal{L}_{\mu}^{\infty,1}(X, \mathcal{A})$.

Teorema A.0.29 (De Fatou-Beppo Levi). *Sea* $\{f_n\}$ *una sucesión creciente de funciones integrables tales que* $\int f_n \leq C$, *donde* C *es una constante independiente de* n. *Entonces*

i) $\lim f_n = f_\infty$ *existe y es finito* $\mu-$*c.d.*

ii) $f_\infty \in L_\mu^1$, *y*

iii) $\| f_n - f_\infty \|_{L^1} \rightarrow 0$.

Definición 41. *Sea* $(X_1, \mathcal{A}_1, \mu_1)$ *y* $(X_2, \mathcal{A}_2, \mu_2)$ *espacios medibles, sea* $X = X_1 \times X_2$ *el espacio producto, y sea* $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2$ *la* $\sigma-$*álgebra producto. La medida producto es una medida* μ *en el espacio medible* $(X, \mathcal{A})$ *y satisface*

$$\mu(A_1 \times A_2) = \mu(A_1)\mu(A_2) \quad si \quad \mu_i(A_1) < \infty, i \in \{1, 2\}.$$

Teorema A.0.30 (De Fubini). *Sea* $(X, \mathcal{B}_X, \mu_X)$ *y* $(Y, \mathcal{B}_Y, \mu_Y)$ *espacios medibles* $\sigma-$*finitos completos y sea* $f : X \times Y \rightarrow \mathbb{C}$ *absolutamente integrable con respecto a la completación* $\mathcal{B}_X \bar{\times} \mathcal{B}_Y$. *Entonces:*

i) Para μ_X-*c.d.* $x \in X$, *la función* $y \longmapsto f(x, y)$ *es absolutamente integrable con respecto a* μ_Y, *y en particular* $\displaystyle\int_Y f(x, y) d\mu_Y(y)$ *existe. Más aún, el mapeo definido* μ_X-*c.d.* $x \longmapsto \displaystyle\int_Y f(x, y) d\mu_Y(y)$ *es absolutamente integrable con respecto a* μ_X.

ii) Para μ_Y-*c.d.* $y \in Y$, *la función* $x \longmapsto f(x, y)$ *es absolutamente integrable con respecto a* μ_X, *y en particular* $\displaystyle\int_X f(x, y) d\mu_X(x)$ *existe. Más aún, el mapeo definido* μ_Y-*c.d.* $y \longmapsto \displaystyle\int_X f(x, y) d\mu_X(x)$ *es absolutamente integrable con respecto a* μ_Y.

iii)

$$\int_{X\times Y} f(x,y)d\mu_X \bar{\times}\, d\mu_Y(x,y) = \int_X \left(\int_Y f(x,y)d\mu_Y(y)\right) d\mu_X(x)$$
$$= \int_Y \left(\int_X f(x,y)d\mu_X(x)\right) d\mu_Y(y).$$

Proposición A.0.31. *Considérese dos espacios métricos separables X_1 y X_2 y su producto $Y = X_1 \times X_2$. Sea Y equipado con la topología producto. Sea $\mathcal{B}_1$, $\mathcal{B}_2$ y $\mathcal{B}_Y$ las $\sigma-$álgebras Borelianas asociadas. Entonces $\mathcal{B}_Y = \mathcal{B}_1 \otimes \mathcal{B}_2$.*

Proposición A.0.32. *Sea $\{A_n\}_{n\in\mathbb{N}}$ una sucesión de conjuntos, y defínase*

$$A = \limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k.$$

a) Si $\sum_{n=1}^{\infty} P(A_n) < \infty$, entonces $P(A) = 0$.

b) Si $A_1, A_2, \ldots$ son independientes y $\sum_{n=1}^{\infty} P(A_n) = \infty$, entonces $P(A) = 1$.

Proposición A.0.33. *Sea $X_1, X_2, \ldots$ una sucesión de v.a's independientes e identicamente distribuidas tales que para cada natural n, $E(X_n) = \mu$ y $Var(X_n) = \sigma^2 < \infty$. Entonces*

$$\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}\sigma} \to \mathcal{N}(0,1)$$

en distribución, siendo $\mathcal{N}(0,1)$ la distribución normal con esperanza nula y varianza 1.

Proposición A.0.34. *Sea $\mathcal{B}_b$ un álgebra Booleana de subconjuntos de X, $\mathcal{M}$ la clase monótona generada por $\mathcal{B}_b$, y $\mathcal{B}$ la $\sigma-$álgebra generada por $\mathcal{B}_b$. Entonces $\mathcal{B} = \mathcal{M}$.*

Bibliografía

[1] BARTLE, ROBERT,*The Elements of Integration and Lebesgue Measure,* John Willey and Sons, págs. 31-32, 1995.

[2] BOURBAKY, NICOLAS, *Elements of Mathematics. General Topology,* Springer,pág. 21, 1966.

[3] BURRIS STANLEY, *A Course in Universal Algebra,* tercera edición, Springer-Verlag,págs. 152-155, 1981.

[4] CLAPP, MÓNICA,*Introducción al Análisis Real (Notas de Clase),* Instituto de Matemáticas, págs. 146-150, 2012.

[5] GRABINSKY GUILLERMO, *Teoría de la Medida,* primera reimpresión, La Prensa de Ciencias,págs. 7,31,37, 2011.

[6] LETAC, GÉRARD, Exercises and solutions manual for Integration and Probability, primera edición, Springer, págs. 80-83,85,86,89-90,122,125. 1995.

[7] MALLIAVIN, PAUL, *Graduation Texts in Matehematics: Integration and Probability,* primera edición, Springer, págs.20,23,34,98, 171-198. 1995.

[8] PRIETO DE CASTRO CARLOS, *Topología Básica,* tercera reimpresión, Fondo de Cultura Económica,págs. 51,201,210,244, 2012.

[9] RINCÓN LUIS, *Curso Intermedio de Probabilidad,* primera reimpresión, La Prensa de Ciencias, págs. 376-377, 2007.

[10] RINCÓN LUIS,*Introducción a los procesos estocásticos,*primera edición, La Prensa de Ciencias,págs. 200-201,273-292.

[11] Tao Terence,*An Introduction to Measure Theory (Graduate Studies in Mathematics),* primera edición, American Mathematic Society, págs. 203-206, 2011.

[12] Institut de France,
http://www.academie-sciences.fr/academie/membre/Malliavin_Paul.htm
Noviembre 2014.

[13] Garling D.J.H.,*A course in Analysis Mathematic Vol. III Complex Analysis, Measure and Integration,*primera edición,Cambridge University Press, págs. 826-827,899-901, 2014.

[14] Gasiñsky Leszek, Papageorgiou Nikolaos,*NonLinear Analisys Vol. 9,*primera edición,Chapman and Hall/CRC, págs. 174-176, 2005.