Benemérita Universidad Autónoma de Puebla

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

UN ESPACIO DE VARIABLES ALEATORIAS NORMADO: BAJO UNA MEDIDA DE DISPERSIÓN

TESIS

QUE PARA OBTENER EL TÍTULO DE

LICENCIADO EN MATEMÁTICAS APLICADAS

PRESENTA

MARCO POLO GONZÁLEZ ZEPEDA

DIRECTOR DE TESIS

DR. FRANCISCO SOLANO TAJONAR SANABRIA

PUEBLA, PUE. 21 de noviembre de 2018

DEDICADO AL ÁVIDO LECTOR

Y es que no hay otro sino aquél que,

aunque sea porque el título le haya parecido atractivo,

lea el presente trabajo pues se ha interesado en el mismo.

Por esta razón, al lector, y a nadie más, va dedicado el presente trabajo;

con la esperanza de que le sirva como faro

e ilumine su desembarco en el análisis y la probabilidad.

Agradecimientos

Quiero agradecer, principalmente, a mis padres y hermanos pues han tenido que tolerar mi comportamiento y apatía a lo largo de la licenciatura, me han brindado su apoyo alegremente absorbiendo las dificultades que encontré en mi trayecto construyendo un gratificante y gustoso refugio; de esta manera, mis padres y hermanos son los pilares con los que se partió la construcción de este trabajo.

De la mano, agradezco a los más cercanos e íntimos amigos que tenemos mucho tiempo de conocernos: Manuel, Antonio, Anel, Abraham y Nava; personas cuáles sin su presencia este trabajo pudo haberse realizado mucho tiempo antes y concluir la carrera de mejor forma pero también hubiese tenido un trayecto menos satisfactorio, su compañía en buenos y malos momentos me hizó entender que son parte de mi familia.

He decidido no mencionar los nombres de los amigos que hice en la carrera pues son varios y tuvé temor de olvidar algunos ofendiendolos de esta manera; sin embargo, no puedo dejar pasar algunos nombres para un agradecimiento especial: para Armando que es el primer amigo que hice en la facultad cuyos consejos y apoyo fueron de gran utilidad, para Siddhartha que tenía la forma de hacerme reflexionar sobre varios temas, para Diana Aurora cuya amistad y recomendaciones sobre películas al igual que lectura y música ayudaron a eliminar el tedio de varias tardes, para Isaí Ivan y para Alberto que aunque nunca entendían de lo que les platicaba siempre me animaban a seguir estudiando y mejorando, para Erick y Hugo que juntos empezamos esta aventura.

Por último, quiero agradecer a mis maestros que tuve a lo largo de la carrera, sus conocimientos y apoyo son el medio móvil por el cual ahora puedo aventurarme a recorrer el mundo matemático; de igual forma, quiero hacer hincapié en los nombres de los maestros que más han influído en mí y el presente trabajo: la doctora Hortensia Reyes Cervantes que siempre me recibió con gran alegría y que

nunca me negó apoyo, al doctor Francisco Tajonar que me ha enseñado mucho y nunca me subestimo, al doctor Francisco Mendoza Torres pues no todos tienen la oportunidad de conocerlo y su técnica para trabajar en el análisis matemático me es de gran ayuda, al doctor Velasco Luna Fernando que tuvo participación en el presente trabajo aunque nunca me dió clase, a los doctores Iván Martínez y Carlos Andrade que siempre abrían un espacio para atenderme y resolver dudas sin importar cuanto trabajo tuvieran, al doctor Arrazola y los compañeros del laboratorio de Lógica Matemática que siempre tenían un consejo para cada problema, a la maestra Mónica Macías y al maestro Edgar Moyotl que me apoyaron en los sistemas computacionales, a la doctora Elsa Puente cuya actitud me motivaba a investigar más allá de lo visto en clase, por último al maestro Francisco Estrada y el doctor Julio Poisot que (aunque a veces a los alumnos no les parezca su forma de dar clase) gracias a ellos aprendí como aterrizar varios conceptos matemáticos a una forma práctica.

A todos ellos, les estaré eternamente agradecido.

Introducción

Una de las causas de interés para algunos investigadores que buscan comprender la teoría de probabilidad es que frecuentemente se enfrentan con problemas cuya solución es incierta y desearían tener al menos un poco de certeza. Es aquí, donde se empieza a implementar la teoría que sí da certeza como herramienta para enfrentar aquello que es incierto. Este proyecto de tesis es un claro ejemplo de ello, pues se emplea la teoría del análisis y álgebra para estudiar una parte de la teoría de probabilidad que suele confundir a varios estudiantes, la convergencia de variables aleatorias.

Durante los cursos de probabilidad y estadística se estudian diferentes criterios de convergencia para las variables aleatorias, los cuales pueden confundir al estudiante que no ha llevado un buen curso de análisis y no sabe de donde se originan estos criterios. En un curso de análisis se observa que en un espacio vectorial se puede definir diferentes normas y, por lo tanto, diferentes criterios de convergencia; estos criterios proporcionan información del comportamiento de los vectores, algunos criterios pueden dar información más efectiva que otros.

Es así como se pueden definir varias formas de convergencia en el espacio de variables aleatorias y, por esta razón, surge la pregunta sobre la existencia de algún espacio de variables aleatorias en el cual se pueda definir una forma de convergencia que pueda mejorar la información que nos proporciona el criterio de convergencia usual acerca de las variables aleatorias contenidas en éste.

De los estudios básicos de álgebra lineal y probabilidad se logra percibir que la covarianza es una forma bilineal para las variables aleatorias; aún más, se observa que cumple con casi todas la propiedades del producto interno. Por esta razón, el objetivo general de la investigación reside en encontrar un espacio de variables aleatorias que sea normado y completo, partiendo de otro espacio también normado, por un criterio de convergencia relacionado a una típica medida de dispersión, la covarianza. Para lograrlo, en esta tesis se estudia una parte de la teoría de análisis y una parte de la teoría de probabilidad, luego se utilizan para estudiar las variables aleatorias como un espacio métrico y revisar su forma de convergencia. De esta manera, la investigación que se realiza parece asemejarse a la forma que se construyen los espacios $\mathcal{L}_p$ de Lebesgue, sin embargo, no se trabaja sobre estos en sí.

Para tener una buen desarrollo del trabajo, en el primer capítulo se mencionan una serie de conceptos preliminares que sirven de preparación para lo que se desarrolla en el segundo capítulo, el cual es una revisión de la teoría de espacios de funciones integrables partiendo de los conceptos de espacios medibles para su mejor entendimiento, también se estudia de forma breve los espacios vectoriales y sus propiedades. El capítulo dos contiene toda la información necesaria para lograr el objetivo, por lo que se le pide al lector paciencia pues, de manera deductiva, del estudio de las funciones integrables como espacio vectorial surge el caso particular de las funciones definidas en un espacio con medida de probabilidad como estudio principal. En el tercer capítulo se revisan algunos conceptos básicos de probabilidad y medidas de dispersión, luego se desarrolla el proceso para encontrar el espacio con las características deseadas. A partir de esta información, será más fácil la construcción de un nuevo espacio que sirva como una propuesta alternativa para estudiar las variables aleatorias desde otro enfoque, donde la covarianza brinda información como criterio de convergencia..

Finalmente, se presentan las conclusiones obtenidas de este trabajo.

Índice general

Capítulo 1

Preliminares

Antes de comenzar a revisar la teoría de espacios de funciones medibles, se presenta una serie de conceptos preliminares de cálculo y probabilidad, no de manera detallada, para facilitar la comprensión de lo que se presenta más adelante.

1.1. Cálculo

Una herramienta básica para estudiar el Cálculo es la noción de límite, así que la primera tarea en este texto será presentar el concepto de una *sucesión*. Una manera intuitiva de entender una sucesión es a través de una lista ordenada e infinita de números reales en la cual existe una regla que permite generar un elemento de la lista en un determinado lugar; al igual que una función, un número real puede aparecer más de una vez en la lista pero un lugar en la lista no puede tener más de un valor. Se presenta la siguiente definición entendiendo que una *función* hace referencia a una regla de asociación que relaciona a cada elemento de un conjunto (*dominio*) con un único elemento de otro conjunto (*recorrido*).

Definición 1.1.1. Sea $x : \mathbb{N} \to \mathbb{R}$ una función, entonces se dice que x ***es una sucesión de números reales*** y se denota como $(x_n)_{n\in\mathbb{N}}$ o simplemente (x_n).

Si K es un subconjunto de $\mathbb{N}$ no finito y se define a x_{n_k} como la función x restringida a los enteros en K, se dice que $(x_{n_k})_{n_k\in K}$, o solo (x_{n_k}), es ***una subsucesión de*** (x_n).

Aún más, se pueden enunciar sucesiones de diferentes objetos dependiendo del recorrido, como lo es el conjunto de funciones reales definidas en algún conjunto no vacío. Esto es, si X es un conjunto no vacío y definimos al conjunto F como

$$F = \{h : X \to \mathbb{R} \mid h \text{ es una función}\},$$

entonces si $f : \mathbb{N} \to F$ es una función, se dice que (f_n) **es una sucesión de funciones reales**.

Recordando que dos funciones reales con dominio común pueden sumarse y multiplicarse, en particular sucede para las sucesiones reales. Esto es, si (x_n) y (y_n) son sucesiones reales entonces

$$(x_n + y_n)_{n\in\mathbb{N}} = (x_n)_{n\in\mathbb{N}} + (y_n)_{n\in\mathbb{N}} \quad \text{y} \quad (x_n \cdot y_n)_{n\in\mathbb{N}} = (x_n)_{n\in\mathbb{N}} \cdot (y_n)_{n\in\mathbb{N}}.$$

Las ideas de operaciones de sucesiones reales pueden extenderse a sucesiones de funciones reales, donde si f y g son funciones reales con dominio X se tiene que

$$(f + g)(x) = f(x) + g(x) \quad \text{y} \quad (fg)(x) = f(x)g(x)$$

para todo $x \in X$. Así, se genera la siguiente definición sobre sucesiones de funciones reales que son propiedades similares a las que tienen las sucesiones reales.

Definición 1.1.2. Sean (f_n) y (g_n) sucesiones de funciones reales. Entonces:

1. $(f_n) = (g_n)$ si y sólo si $f_n = g_n$ para todo $n \in \mathbb{N}$.

2. Si $c \in \mathbb{R}$ es una constante, entonces $(f_n) + (cg_n) = (f_n + cg_n)$.

3. $(f_n) \cdot (g_n) = (f_n \cdot g_n)$.

4. Se dice que (f_n) es una sucesión ***acotada superiormente de forma puntual*** si existe una función real g tal que para todo $n \in \mathbb{N}$ se tiene que $f_n(x) \leq g(x)$ para todo $x \in X$.

5. Se dice que (f_n) es una sucesión ***acotada puntualmente*** si existe una función real g tal que para todo $n \in \mathbb{N}$ se tiene que $|f_n(x)| \leq g(x)$ para todo $x \in X$.

6. Se dice que (f_n) es una sucesión ***acotada uniformemente*** si existe $M \in \mathbb{R}$ tal que para todo $n \in \mathbb{N}$ se tiene que $|f_n(x)| \leq M$.

7. Se dice que (f_n) es una sucesión *monótona creciente* (*decreciente*) si para $n < m$ se cumple que $f_n \leq f_m$ $(f_m \leq f_n)$ para todo $x \in X$.

En algunas sucesiones se puede observar que a partir de cierto entero los valores tienden a acercarse a un determinado elemento, esta es la noción del *límite*. Si (x_n) es una sucesión de números reales, se dice que $L \in \mathbb{R}$ es el **límite de** (x_n) si para todo $\epsilon > 0$ existe $N \in \mathbb{N}$ tal que si $n \geq N$ implica que $|x_n - L| < \epsilon$ y se dice que (x_n) **converge** L. De igual forma, se enuncia la definición del límite de una sucesión de funciones reales.

Definición 1.1.3. Sea (f_n) una sucesión de funciones reales. Se dice que f, una función real, es el ***límite de*** (f_n) si para todo $x \in X$ la sucesión $(f_n(x))$ converge a $f(x)$. En este caso, se dice que (f_n) ***converge puntualmente a*** f y se denota como $\lim f_n = f$.

Dado que una función real f_n puede ser la función constante x_n, las propiedades que se mencionen para una sucesión de funciones reales se cumplen también para las sucesiones reales.

A continuación, se presenta una serie de resultados importantes de sucesiones.

Proposición 1.1.1. *Sean (f_n) y (g_n) sucesiones de funciones reales convergentes. Entonces se cumple lo siguiente:*

1. *El límite es único.*

2. *La sucesión $(|f_n|)$ es convergente y $\lim |f_n| = |\lim f_n|$.*

3. *Si $\lim f_n = \lim g_n = L$ y (h_n) es una sucesión de funciones reales tal que para cada $n \in \mathbb{N}$, $f_n \leq h_n \leq g_n$ para todo $x \in X$, entonces $\lim h_n = L$.*

4. *Sea $c \in \mathbb{R}$, entonces $\lim(f_n + cg_n) = \lim f_n + c \lim g_n$.*

5. $\lim(f_n \cdot g_n) = (\lim f_n) \cdot (\lim g_n)$.

6. *Sea* $p \in \mathbb{N}$, *entonces* $\lim(f_n)^p = (\lim f_n)^p$. *Si para cada* $n \in \mathbb{N}$, $f_n \geq 0$ *para todo* $x \in X$, *entonces* $\lim(f_n)^{\frac{1}{p}} = (\lim f_n)^{\frac{1}{p}}$.

Como se observa, el límite de una sucesión de funciones es un poco diferente al concepto del límite de una función. Si f es una función real con dominio $\mathbb{R}$, se dice que $A \in \mathbb{R}$ es el ***límite de*** f ***en*** x_0 si dado $\epsilon > 0$ existe $\delta_\epsilon > 0$ tal que $|f(x) - A| < \epsilon$ siempre que $0 < |x - x_0| < \delta$ y se denota como $\lim_{x \to x_0} f(x) = A$.

Si $f(x_0)$ está definido, con $x_0 \in \mathbb{R}$, y si $\lim_{x \to x_0} f(x) = f(x_0)$, se dice que f ***es continua en*** x_0; si B es un subconjunto de $\mathbb{R}$ y si para cada $x \in B$ la función f es continua, se dice que f **es continua en** B.

Ahora, se le da sentido a la suma de una sucesión a través de la siguiente definición.

Definición 1.1.4. Sea (f_n) una sucesión de funciones reales convergente, entonces para cada $n \in \mathbb{N}$ se define la ***suma parcial de*** (f_n) como $S_n = f_1 + f_2 + ... + f_n$, o bien $S_n = \displaystyle\sum_{i=1}^{n} f_n$. Si la sucesión (S_n) converge se dice que (f_n) es ***sumable*** y se escribe como

$$\sum f_n = \lim S_n.$$

A este límite se le llama ***suma***.

Notar que el límite de (S_n) puede ser una función con valores que divergen a infinito, por lo que la función límite tiene como recorrido a $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. Si x_0 es tal que el límite de la sucesión $(f_n(x_0))$ toma el valor ∞ o $-\infty$ (en $\overline{\mathbb{R}}$), se dice que dicha sucesión **diverge a** ∞ **o** $-\infty$, según sea el caso.

Las series y sumas parciales ayudan a formalizar el concepto de *integral*. En este trabajo se define primero la integral de una función "escalonada" y luego la de una función cualquiera, de esta manera se concluirá esta sección.

Si $\varphi : [a, b] \to \mathbb{R}$ es una función con $[a, b] \subseteq \mathbb{R}$, se dice que es una **función escalonada** si existe una partición del intervalo $P = \{x_0, x_1, ..., x_n\}$ para la cual $\varphi(x) = a_k$ si $x \in (x_{k-1}, x_k)$ con a_k una constante real para $k = 1, 2, ..., n$.

Definición 1.1.5. La integral de una función escalonada φ sobre el intervalo

$[a, b]$, denotada como $\int_a^b \varphi dx$, se define como:

$$\int_a^b \varphi(x)dx = \sum_{k=1}^{n} a_k(x_k - x_{k-1}).$$

Esta definición exige que la función tenga como dominio un conjunto de números reales, esto es porque se requiere que la diferencia de los elementos del dominio exista y en un conjunto arbitrario no siempre está definida.

Si una función real f es acotada puntualmente en el intervalo $[a, b]$, entonces existen funciones reales g y h tales que $g(x) \leq f(x) \leq h(x)$ para todo $x \in [a, b]$. Estas funciones pueden ser escalonadas, basta considerar las funciones constantes con el valor más alto y más bajo de f.

Definición 1.1.6. Sea $f : [a, b] \to \mathbb{R}$ una función acotada puntualmente. Si g y h denotan funciones escalonadas con dominio $[a, b]$, se define los conjuntos $G = \{\int_a^b g(x)dx \mid g \leq f\}$ y $H = \{\int_a^b h(x)dx \mid f \leq h\}$, de esta manera f es acotada por funciones simples, $g \leq f \leq h$, se tiene que $\int_a^b g(x)dx \leq \int_a^b h(x)dx$ y por lo tanto

$$\int_a^b g(x)dx \leq \sup G \leq \text{ínf}\, H \leq \int_a^b h(x)dx.$$

Si $\sup G = \text{ínf}\, H$ se dice que f es ***integrable*** y se escribe como

$$\int_a^b f(x)dx = \sup G = \text{ínf}\, H.$$

1.2. Probabilidad

Los modelos matemáticos tratan de describir los fenómenos que se obtienen a través de la observación, estos se pueden clasificar en *modelos deterministas* y *modelos no deterministas*, también llamados *probabilísticos* o *estocásticos*. El modelo determinista estipula que el resultado de un experimento es el mismo siempre que las condiciones en las que se verifica dicho experimento se mantengan iguales. Por ejemplo: $s = -16t^2 + v_0t$, donde s es la distancia recorrida verticalmente de un objeto, t es el tiempo empleado y v_0 su velocidad inicial; lo que se quiere destacar de esta expresión es el hecho de que hay una relación definida entre t y v_0 que determinan unívocamente a s.

Sin embargo, hay veces en que una variable depende de factores que no se puede controlar o predecir su comportamiento, los juegos de azar son ejemplos típicos como lo es la “mano de 5 cartas” sacadas al azar de un mazo con 52 cartas; para estos casos es apropiado usar los modelos probabilísticos y a estos fenómenos se les conoce como *experimento aleatorio*. Un experimento aleatorio se caracteriza por lo siguiente: porque es posible repetir indefinidamente (al menos teóricamente) el experimento sin cambiar las condiciones escenciales, porque se puede describir el conjunto de todos los posibles resultados del experimento y porque va acompañado de una *regularidad estadística* (a medida que se repite el experimento los resultados individuales empiezan a ocurrir de forma regular) que permite construir el modelo.

Al conjunto de todos los posibles resultados del experimento aleatorio se le llama **espacio muestral** y usualmente se le denota como Ω. A un subconjunto A de Ω se le llama **evento**, este concepto surge en vista de que sólo algunos resultados podrían ser del interés de investigación; de esta forma, tanto el conjunto Ω como los resultados individuales son eventos. Recurriendo a la teoría de conjuntos, se dice que “ha ocurrido el evento A” si el resultado del experimento es algún $\omega \in A$, se dice que “no ocurre A” si el resultado es algún $\omega \in A^c = \Omega \setminus A$; de igual forma, si B es un evento entonces se dice que “ha ocurrido A y B” si el resultado es algún $\omega \in A \cap B$ y que “ocurre A o B” si es algún $\omega \in A \cup B$.

La noción intuitiva del modelo probabilístico que se ajusta al experimento surge de la regularidad estadística que permite relacionar a cada resultado individual ω con un número real en $[0, 1]$, de esta forma se vería reflejado el comportamiento del experimento a través de este modelo. Por ejemplo, si después de repetir un gran número de veces el experimento resulta que el resultado ω_k ocurre n_k por cada 100 repeticiones, entonces se relaciona a ω con $p_k = \dfrac{n_k}{100}$ y decimos que p_k *es la probabilidad de que ocurra* ω_k y suele denotarse como $P(\{\omega_k\})$. Bajo la misma idea, la *probabilidad de que ocurra el evento* A se define como

$$P(A) = P\left(\{\omega \in \Omega \mid \omega \in A\}\right) = \sum_{\omega \in A} P(\{\omega\}).$$

Dado que es un capítulo de conceptos preliminares, no se detallará más en el modelo probabilístico, solo destacar que está definido para cada posible evento del experimento por lo que es necesario decir que P es una función que relaciona a cada evento con un valor real entre cero y uno. Una σ-álgebra consiste en una colección de conjuntos de eventos $\mathscr{O}$ con la característica de que cada evento contenido en la colección implica que su complemento también pertenezaca a esta. Este concepto se detallará en el siguiente capítulo, sin embargo con esto se explica que si existe conocimiento de la probabilidad que un evento tiene de ocurrir entonces se conoce que tan probable es que no ocurra, por lo que si $A \in \mathscr{O}$ entonces $A^c \in \Omega$; aún más, se puede explicar que la unión de eventos ocurre. Dicho esto, se tiene que el límite superior e inferior de una sucesión de eventos también es un evento.

Definición 1.2.1. Para una sucesión de eventos (A_n) se define el límite superior y el límite inferior como sigue:

- $\limsup(A_n) = \bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} (A_k)$.
- $\liminf(A_n) = \bigcup_{m=1}^{\infty} \bigcap_{k=m}^{\infty} (A_k)$.

Los límites inferior y superior son operaciones bien definidas; esto es, el resultado siempre existe y es único.

En resumen, para cada experimento aleatorio el modelo probabilístico asocia un espacio muestral Ω, la colección de los eventos $\mathscr{O}$ y una función real P; a la terna $(\Omega, \mathscr{O}, P)$ se le llama *espacio de probabilidad* y se explicará en el siguiente capítulo.

Para fines prácticos, es conveniente usar valores reales y aprovechar sus propiedades de campo para estudiar el fenómeno, por lo que se puede usar una función real X definida en Ω que además sea *medible*. Una vez dicho esto, es momento de cerrar este capítulo y comenzar a revisar la teoría de funciones medibles.

Capítulo 2

Funciones medibles

Nota 2.0.1. *Con la itención de estudiar el espacio de variables aleatorias es necesario comprender los conceptos de la funciones integrables, para esto se parte desde la idea de funciones y espacios medibles. Es así como se empieza este proyecto, mencionando conceptos y resultados para establecer una base teórica; algunas referencias importantes que forman esta base son:* "The Elements of Integración and Lebesgue Measure" *(ver [1]),* "Convergence of Probability and Measures" *(ver [2]) y* "Probability and Measure" *(ver [3]).*

Definición 2.0.1. Una familia de subconjuntos $\mathscr{X}$ de un conjunto X se dice que es una ***σ-álgebra*** (σ-campo) si cumple lo siguiente:

1. $X \in \mathscr{X}$.

2. Si $A \in \mathscr{X}$, entonces $A^c = (X \setminus A) \in \mathscr{X}$.

3. Si $\{A_n\}_{n\in\mathbb{N}}$ es una sucesión numerable de elementos en $\mathscr{X}$, entonces

$$\bigcup_{n=1}^{\infty} A_n \in \mathscr{X}.$$

Definición 2.0.2. La pareja ordenada de un conjunto no vacío X y una σ-álgebra $\mathscr{X}$ de X, $(X, \mathscr{X})$, se le llama ***espacio medible***. A cualquier elemento en $\mathscr{X}$ se le llama ***conjunto $\mathscr{X}$-medible***, o cuando no existe problema de confusión se le dice simplemente *conjunto medible*.

Observación 2.0.1. Recordando las leyes de *De Morgan*:

$$\left(\bigcup_{\lambda} A_{\lambda}\right)^{c} = \bigcap_{\lambda} A_{\lambda}^{c} \ \text{ y } \ \left(\bigcap_{\lambda} A_{\lambda}\right)^{c} = \bigcup_{\lambda} A_{\lambda}^{c},$$

con λ perteneciente a un conjunto indexado, posiblemente infinito numerable. Se tiene que la intersección de una sucesión numerable de conjuntos $\mathscr{X}$-medibles también es $\mathscr{X}$-medible.

Sea $A \subseteq X$ no vacío, la intersección de todas las σ-álgebras que contienen a A, denotada como $\sigma(A)$, es la σ-álgebra más pequeña que contiene a A y se le llama ***σ-álgebra generada por A***.

La ***σ-álgebra de Borel***, denotada como $\mathscr{B}$, es la σ-álgebra generada por todas las bolas abiertas (que en $\mathbb{R}$ son los intervalos abiertos (a, b)) y cualquier elemento en él es llamado ***conjunto Boreliano***, o simplemente *boreliano*. Si $E \in \mathsf{B}$ y sean $E_1 = E \cup \{-\infty\}$, $E_2 = E \cup \{\infty\}$ y $E_3 = E \cup \{-\infty, \infty\}$; la colección de todos los conjuntos $E,\ E_1,\ E_2,\ E_3$, para cada $E \in \mathscr{B}$, forman la ***σ-álgebra de Borel extendida*** y se denota como $\bar{\mathscr{B}}$.

La siguiente definición servirá para identificar el tipo de funciones que son convenientes para iniciar el estudio de nuestro interés.

Definición 2.0.3. Una función $f : X \to \mathbb{R}$ se dice ***$\mathscr{X}$-medible*** si

$$\forall \alpha \in \mathbb{R},\ \{x \in X \mid f(x) > \alpha\} \in \mathscr{X}. \tag{2.1}$$

Se puede demostrar que la condición (2.1) es equivalente con las siguientes condiciones:

(*i*) $\{x \in X \mid f(x) \geq \alpha\} \in \mathscr{X}$.

(*ii*) $\{x \in X \mid f(x) \leq \alpha\} \in \mathscr{X}$.

(*iii*) $\{x \in X \mid f(x) < \alpha\} \in \mathscr{X}$.

Ejemplo 2.0.1. A continuación, se presenta unos pequeños ejemplos de funciones medibles.

a) Cualquier función constante es $\mathscr{X}$-medible. Esto es, si $f(x) = c$ para todo $x \in X$ donde c es una constante real, entonces si $\alpha \geq c$ se tiene $\{x \in X \mid f(x) > \alpha\} = \emptyset \in \mathscr{X}$; si $\alpha < c$, se tiene que $\{x \in X \mid f(x) > \alpha\} = X \in \mathscr{X}$; por lo tanto, f es $\mathscr{X}$-medible.

b) La ***función característica*** es $\mathscr{X}$-medible. Esto es, si E es $\mathscr{X}$-medible, la función χ_E es tal que $\chi_E(x) = 1$ si $x \in E$ y $\chi_E(x) = 0$ si $x \notin E$, esta función es $\mathscr{X}$-medible; basta ver que $\{x \in X \mid \chi_E(x) > \alpha\}$ es X si $\alpha < 0$, que es vacío si $\alpha \geq 1$ y que es E si $0 \leq \alpha < 1$.

c) Si $f : \mathbb{R} \to \mathbb{R}$ es una función continua, como $\{x \in \mathbb{R} \mid f(x) > \alpha\}$ es un conjunto abierto y, por lo tanto, es la unión arbitraria de intervalos abiertos por lo que pertenece a $\mathscr{B}$, entonces f es Borel medible.

d) Si $f : \mathbb{R} \to \mathbb{R}$ es monótona, como $\{x \in \mathbb{R} \mid f(x) > \alpha\}$ consiste en una semirecta $\{x \in \mathbb{R} \mid x > a\}$, entonces f es Borel medible.

Lema 2.0.1. *Si f y g son funciones reales medibles y $c \in \mathbb{R}$, entonces las siguientes funciones son medibles:*

1. *La función cf, donde $(cf)(x) = cf(x)$ para todo $x \in X$.*

2. *La función f^2, donde $f^2(x) = (f(x))^2$ para todo $x \in X$.*

3. *La función $f + g$, donde $(f + g)(x) = f(x) + g(x)$ para todo $x \in X$.*

4. *La función fg, donde $(fg)(x) = f(x)g(x)$ para todo $x \in X$.*

5. *La función $|f|$, donde $|f|(x) = |f(x)|$ para todo $x \in X$.*

Demostración. Por hipótesis, f y g cumplen con la condición (2.1); es decir, para todo $\alpha \in \mathbb{R}$

$$\{x \in X \mid f(x) > \alpha\} \in \mathscr{X} \quad \text{y} \quad \{x \in X \mid g(x) > \alpha\} \in \mathscr{X}.$$

1. Sea $c > 0$, entonces

$$\{x \in X \mid cf(x) > \alpha\} = \left\{x \in X \mid f(x) > \frac{\alpha}{c}\right\}$$

es un conjunto $\mathscr{X}$-medible. De manera similar se muestra para $c < 0$ y el caso $c = 0$ es trivial. Por lo tanto, cf es medible.

2. Si $\alpha < 0$, se tiene que $\{x \in X \mid (f(x))^2 > \alpha\} = X$.

   Si $\alpha \geq 0$, entonces $(f(x))^2 > \alpha$ si y sólo si $f(x) > \sqrt{\alpha}$ ó $f(x) < -\sqrt{\alpha}$, así pues

$$\{x \in X \mid (f(x))^2 > \alpha\} = \{x \in X \mid f(x) > \sqrt{\alpha}\} \cup \{x \in X \mid f(x) < -\sqrt{\alpha}\}$$

   es un conjunto $\mathscr{X}$-medible dado que es la unión de conjuntos $\mathscr{X}$-medibles. Por lo tanto, f^2 es medible.

3. Se tiene que los conjuntos $\{x \in X \mid f(x) > n\}$ y $\{x \in X \mid g(x) > \alpha - n\}$ son $\mathscr{X}$-medibles para $n \in \mathbb{N}$; de aquí que

$$A_n = \{x \in X \mid f(x) > n\} \cap \{x \in X \mid g(x) > \alpha - n\} \in \mathscr{X}.$$

   Así pues,

$$\{x \in X \mid (f+g)(x) > \alpha\} = \{x \in X \mid f(x) + g(x) > n + \alpha - n\} = \bigcup_{n \in \mathbb{N}} A_n$$

   es un conjuto $\mathscr{X}$-medible pues es la unión de conjuntos $\mathscr{X}$-medibles. Por lo tanto, $f + g$ es función medible.

   Como se vió antes, si $c = -1$ entonces $-g$ es medible y, por lo tanto, $f - g$ es medible.

4. Ahora, se observa que

$$\frac{1}{4}\left((f+g)^2 - (f-g)^2\right) = \frac{1}{4}(f^2 + g^2 + 2fg - f^2 - g^2 + 2fg) = fg.$$

   Por lo que se ha mostrado, $(f+g)^2$ y $(f-g)^2$ son funciones medibles dado que las funciones $f + g$ y $f - g$ son medibles, de aquí que la resta de estas dos funciones es una función medible. Por lo tanto, fg es medible.

5. Si $\alpha < 0$, se tiene que $\{x \in X \mid |f(x)| > \alpha\} = X$.

   Si $\alpha \geq 0$, entonces $|f(x)| > \alpha$ si y sólo $f(x) > \alpha$ ó $f(x) < -\alpha$; así pues

$$\{x \in X \mid |f(x)| > \alpha\} = \{x \in X \mid f(x) > \alpha\} \cup \{x \in X \mid f(x) < -\alpha\} \in \mathscr{X}.$$

   Por lo tanto, $|f|$ es medible.

$\square$

Se define *la parte positiva de* f como $f^+(x) = \sup\{f(x), 0\}$ y *la parte negativa de* f como $f^-(x) = \sup\{-f(x), 0\}$, entonces se tiene que

$$f = f^+ - f^- \quad \text{y} \quad |f| = f^+ + f^-; \tag{2.2}$$

o de igual forma,

$$f^+ = \frac{1}{2}(|f| + f) \quad \text{y} \quad f^- = \frac{1}{2}(|f| - f). \tag{2.3}$$

De aquí se tiene que si f es medible, por tanto $|f|$ también lo es, implica que f^+ y f^- son medibles dado que son suma y resta de funciones medibles.

Por otro lado, si f^+ y f^- son medibles, entonces f es medible dado que es la resta de funciones medibles. Así pues, se ha probado el siguiente corolario.

Corolario 2.0.2. *Una función f es $\mathscr{X}$-medible si y sólo si f^+ y f^- son funciones $\mathscr{X}$-medibles.*

Definición 2.0.4. Una función real-extendida $f : X \to \bar{\mathbb{R}}$ es una función $\mathscr{X}$-medible si $\{x \in X \mid f(x) > \alpha\} \in \mathscr{X}$ para todo número real α. La colección de todas las funciones real-extendidas $\mathscr{X}$-medibles se denota por $M(X, \mathscr{X})$.

Observación 2.0.2. Si $f \in M(X, \mathscr{X})$, entonces los siguientes conjuntos pertenecen a $\mathscr{X}$:

(*i*) $\{x \in X \mid f(x) = \infty\} = \displaystyle\bigcap_{n=1}^{\infty} \{x \in X \mid f(x) > n\}$.

(*ii*) $\{x \in X \mid f(x) = -\infty\} = \left(\displaystyle\bigcup_{n=1}^{\infty} \{x \in X \mid f(x) > -n\} \right)^c$.

Lema 2.0.3. *Sea $A = \{x \in X \mid f(x) = \infty\}$ y $B = \{x \in X \mid f(x) = -\infty\}$. Una función real-extendida f es medible si y sólo si los conjuntos A y B pertenecen a $\mathscr{X}$ y la función*

$$f_1(x) = \begin{cases} f(x), & si \;\; x \notin A \cup B; \\ 0, & si \;\; x \in A \cup B. \end{cases}$$

es medible.

Demostración. (Ver referencia [1]). □

Por el lema anterior, se puede ver que si $f \in M(X;\mathscr{X})$ entonces los conjuntos A y B pertenecen a $\mathscr{X}$ y la función f_1 es medible, con A, B y f_1 descritos en dicho lema. Luego, por el Lema (2.0.1), las funciones $cf_1, f_1^2, |f_1|, f_1^+, f_1^-$ son medibles, con c una constante real; aún más, si g es una de estas funciones entonces los conjuntos $\{x \in \mathscr{X} \mid g(x) = \infty\}$ y $\{x \in \mathscr{X} \mid g(x) = -\infty\}$ se pueden describir como la intersección y/o la unión de los conjuntos A y B, por lo que también pertenecen a $\mathscr{X}$. De nuevo, por el lema anterior, se tiene que las funciones $cf, f^2, |f|, f^+, f^-$ pertenecen a $M(X;\mathscr{X})$.

Observación 2.0.3. La función $f+g$ no está definida en los conjuntos:

$$A_1 = \{x \in X \mid f(x) = \infty \ \text{ y } \ g(x) = -\infty\},$$

$$A_2 = \{x \in X \mid g(x) = \infty \ \text{ y } \ f(x) = -\infty\};$$

por lo que se define $f+g$ como

$$(f+g)(x) = \begin{cases} 0, & \text{si } \ x \in A_1 \cup A_2; \\ f(x) + g(x), & \text{de otra forma (d.o.f.)} \end{cases}$$

De esta manera, la función $f+g$ es medible y pertenece a $M(X,\mathscr{X})$.

Lema 2.0.4. *Sea $(f_n)_{n\in\mathbb{N}}$ una sucesión en $M(X,\mathscr{X})$, entonces las siguientes funciones también pertenecen a $M(X,\mathscr{X})$:*

1. *La función $f(x) = \inf_{n\in\mathbb{N}}(f_n(x))$.*

2. *La función $F(x) = \sup_{n\in\mathbb{N}}(f_n(x))$.*

3. *La función $f^*(x) = \liminf f_n(x)$.*

4. *La función $F^*(x) = \limsup f_n(x)$.*

Demostración. Como f_n es medible para cada $n \in \mathbb{N}$, entonces dado $\alpha \in \mathbb{R}$ el conjunto $A_n = \{x \in X \mid f_n(x) > \alpha\} \in \mathscr{X}$ para cada $n \in \mathbb{N}$; así, se tiene que el conjunto de todos los elementos $x \in X$ tales que $f_n(x)$ es mayor que α para algún $n \in \mathbb{N}$, el cual es $\bigcup_{n=1}^{\infty} A_n$, también pertenece a $\mathscr{X}$. Esto es,

$$\{x \in X \mid F(x) = \sup\left(f_n(x)\right) > \alpha\} = \bigcup_{n=1}^{\infty} A_n \in \mathscr{X}.$$

Por lo tanto, $F(x)$ es medible. De forma análoga, se muestra que $f(x) = \text{ínf}(f_n(x))$ es medible.

Por otro lado, se observa que

$$F^*(x) = \text{lím}\sup f_n(x) = \sup_{n\in\mathbb{N}}\left(\inf_{m\geq n} f_m(x)\right);$$

es decir, F^* es la función supremo de una sucesión de funciones medibles $(\text{ínf}(f_m))$ y, por lo tanto, $F^*(x) = \text{lím}\sup f_n(x)$ es medible. De nuevo, bajo argumento similar se muestra que $f^*(x) = \text{lím}\inf f_n(x)$ es una función medible. $\square$

Recuerde que si una sucesión $(f_n) \in M(X, \mathscr{X})$ converge a una función f, entonces

$$\lim_{n\to\infty} f_n = \liminf_n f_n = \limsup_n f_n;$$

así que la función f es una función medible. Este argumento muestra el siguiente resultado.

Corolario 2.0.5. *Si $(f_n)_{n\in\mathbb{N}}$ es una sucesión en $M(X, \mathscr{X})$ que converge a f en X, entonces $f \in M(X, \mathscr{X})$.*

Lema 2.0.6. *Si f es una función no negativa en $M(X, \mathscr{X})$, entonces existe una sucesión (φ_n) en $M(X, \mathscr{X})$ tal que:*

1. $0 \leq \varphi_n(x) \leq \varphi_{n+1}(x)$.
2. $f(x) = \lim_{n\to\infty} \varphi_n(x)$.
3. *Cada φ_n tiene una cantidad finita de valores reales.*

Demostración. Para cada $n \in \mathbb{N}$ se define el conjunto $E_{k,n}$, para $k = 0, ..., n2^n - 1$, como

$$E_{k,n} = \{x \in X \mid k2^{-n} \leq f(x) < (k+1)2^{-n}\},$$

y también

$$E_{n2^n,n} = \{x \in X \mid n \leq f(x)\}.$$

Como f es $\mathscr{X}$-medible estos conjuntos pertenecen a $\mathscr{X}$ para cada $n \in \mathbb{N}$; aún más, son conjuntos disjuntos y cuya unión es igual a X.

Para cada $n \in \mathbb{N}$ se define φ_n como la función tal que $\varphi_n(x) = k2^{-n}$ si $x \in E_{k,n}$. Como para cada $n \in \mathbb{N}$ la función φ_n es la suma de funciones características $\chi_{E_{k,n}}$ multiplicadas por una constante, entonces es la suma de funciones medibles y se tiene que φ_n es una función medible. Además, esta función tiene a lo más $2^n + 1$ valores diferentes, por lo que tiene una cantidad finita de valores reales, cumpliéndose así el inciso *(3)* del Lema.

Ahora bien, se ve que para cualquier $n \in \mathbb{N}$, la función φ_n es no negativa pues $0 \leq \varphi_n(x)$ para todo $x \in X$. Luego, como $X = \bigcup E_{k,n}$, para dado $x \in X$ se tiene que $x \in E_{k_n,n}$, donde k_n es el mayor entero k en $\{0, 1, ..., n2^n\}$ tal que $k2^{-n} \leq f(x)$; entonces se tiene que $\varphi_n(x) = k_n 2^{-n}$.

Por otro lado, $X = \bigcup E_{k,n+1}$, entonces $x \in E_{k_{n+1},n+1}$, donde k_{n+1} es el mayor entero k en $\{0, 1, ..., (n+1)2^{n+1}\}$, tal que $k2^{-(n+1)} \leq f(x)$; entonces se tiene que $\varphi_{n+1}(x) = k_{n+1}2^{-(n+1)}$. Por la definición de k_n, $2k_n$ es un entero tal que

$$2k_n 2^{-(n+1)} = k_n 2^{-n} \leq f(x);$$

así, $2k_n \leq k_{n+1}$ pue por la definición de k_{n+1} es el mayor entero que cumple esta condición para $n+1$ y se tiene que

$$\varphi_n(x) = k_n 2^{-n} = 2k_n 2^{-(n+1)} \leq k_{n+1} 2^{-(n+1)} = \varphi_{n+1}(x).$$

Por lo tanto, para cualquier $n \in \mathbb{N}$,

$$0 \leq \varphi_n(x) \leq \varphi_{n+1}(x)$$

para todo $x \in X$, mostrándose de esta forma el inciso *(1)* del Lema.

Por último, se muestra el inciso *(2)* del Lema. Si $x \in X$ es tal que $f(x) = \infty$, entonces $f(x) > n$ y $x \in E_{n2^n,n}$ para cada $n \in \mathbb{N}$; así pues,

$$\varphi_n(x) = n2^n \frac{1}{2^n} = n < f(x).$$

Aplicando el límite sobre n se tiene que

$$\lim_{n\to\infty} \varphi_n(x) = \lim_{n\to\infty} n = f(x).$$

Recordando la *propiedad arquimediana de los reales*, si $x \in X$ es tal que $f(x) < \infty$, entonces existe $n \in \mathbb{N}$ tal que $n2^n2^{-n} > f(x)$; es decir, existe n tal que $x \in E_{k_n,n}$ donde k_n es el mayor entero k en $\{0, 1, ..., n2^n - 1\}$ que cumple que $k2^{-n} \leq f(x)$. De esta forma, se tiene que

$$\frac{k_n}{2^n} \leq f(x) < \frac{k_n + 1}{2^n},$$

aplicando el límite sobre n se obtiene que

$$\lim_{n\to\infty} \frac{k_n}{2^n} \leq \lim_{n\to\infty} f(x) < \lim_{n\to\infty} \frac{k_n + 1}{2^n} = \lim_{n\to\infty} \frac{k_n}{2^n} + \lim_{n\to\infty} \frac{1}{2^n}$$

y esto sucede si y sólo si

$$\lim_{n\to\infty} \frac{k_n}{2^n} \leq f(x) \leq \lim_{n\to\infty} \frac{k_n}{2^n};$$

o bien,

$$\lim_{n\to\infty} \varphi_n(x) \leq f(x) \leq \lim_{n\to\infty} \varphi_n(x).$$

Por lo tanto, $f(x) = \lim_{n\to\infty} \varphi_n(x)$ para todo $x \in X$. $\square$

2.1. Medidas

Definición 2.1.1. Una ***medida*** es una función real extendida μ definida en una σ-álgebra $\mathscr{X}$ de un conjunto X tal que cumple lo siguiente:

1. $\mu(\emptyset) = 0$.

2. $\mu(E) \geq 0$, para todo $E \in \mathscr{X}$.

3. Es ***aditiva contable***; esto es, si (E_n) es una sucesión disjunta en $\mathscr{X}$ (si $n \neq m$ se tiene que $E_n \cap E_m = \emptyset$), entonces

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

Si una medida no toma el valor ∞ se dice que es ***finita***. Si X es la unión contable de conjuntos medibles y cada uno de estos tiene medida finita, entonces se dice que μ ***es σ-finita***.

Ejemplo 2.1.1. A continuación se presentan unos ejemplos de medida.

a) La ***medida unitaria concentrada en*** p***:*** sea $p \in X$ y el espacio medible $(X, \mathscr{X})$, entonces $\mu : \mathscr{X} \to \mathbb{R}$ es una medida si

$$\mu(E) = \begin{cases} 1, \text{ si } & p \in E; \\ 0, \text{ si } & p \notin E. \end{cases}$$

b) La ***medida contadora en*** $\mathbb{N}$***:*** Si μ es la función que da el valor cardinal (la cantidad de elementos que contiene el conjunto) para $E \subseteq \mathbb{N}$, es una medida σ-finita pero no finita.

c) La ***medida de Lebesgue (Borel):*** Sea $(\mathbb{R}, \mathscr{B})$ y $\lambda : \mathscr{B} \to \mathbb{R}$ donde $\lambda(a, b) = b - a$, entonces λ genera una medida que es σ-finita pero no finita.

Lema 2.1.1. *Sea $\mathscr{X}$ una σ-álgebra de un conjunto X y sea μ una medida definida en $\mathscr{X}$, entonces se cumplen las siguientes proposiciones:*

1. *Si $E \subseteq F \in \mathscr{X}$, se tiene que $\mu(E) \leq \mu(F)$; además, si $\mu(E) < \infty$, entonces $\mu(F \setminus E) = \mu(F) - \mu(E)$.*

2. *Si (E_n) es sucesión creciente en $\mathscr{X}$, se tiene que*

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \lim_{n \to \infty} \mu(E_n).$$

3. *Si (F_n) es sucesión decreciente en $\mathscr{X}$ y si $\mu(F_1) < \infty$, se tiene que*

$$\mu\left(\bigcap_{n=1}^{\infty} F_n\right) = \lim_{n \to \infty} \mu(F_n).$$

Demostración. Se observa que para probar los incisos *(2)* y *(3)* de este lema se puede apoyar del inciso *(1)*.

1. Como $E \subseteq F$ se observa que $F = E \cup (F \setminus E)$, además $E \cap (F \setminus E) = \emptyset$. Ya que μ es medida, entonces es aditiva contable, y así

$$\mu(F) = \mu(E) + \mu(F \setminus E).$$

Dado que $\mu(F \setminus E) \geq 0$, entonces $\mu(F) \geq \mu(E)$. Luego, si $\mu(E) < \infty$ permite sumar por su inverso aditivo en la igualdad anterior, por lo que

$$\mu(F) - \mu(E) = \mu(F \setminus E).$$

2. Sea (E_n) una sucesión creciente. Si para algún entero positivo m se tiene que $\mu(E_m) = \infty$, entonces para todo $n \geq m$ se tiene que $\mu(E_n) = \infty$ pues, por el inciso *(1)*, si $E_m \subseteq E_n$ implica que $\mu(E_m) \leq \mu(E_n)$; de igual forma sucede que $\mu\left(\bigcup E_n\right) = \infty$, por lo que

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

Si $\mu(E_n) < \infty$ para todo $n \in \mathbb{N}$, se crea una sucesión disjunta (A_n) donde $A_1 = E_1$ y $A_n = E_n \setminus E_{n-1}$ para todo $n > 1$. Se observa que $E_n = \bigcup_{j=1}^{n} A_j$, entonces

$$\bigcup_{n=1}^{\infty} E_n = \bigcup_{n=1}^{\infty} \bigcup_{j=1}^{n} A_j = \bigcup_{n=1}^{\infty} A_j.$$

Como μ es aditiva contable, se tiene que

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(A_n) = \lim \left(\sum_{j}^{n} \mu(A_j)\right);$$

por el inciso *(1)*, $\mu(A_n) = \mu(E_n) - \mu(E_{n-1})$, así que

$$\sum_{j=1}^{n} \mu(A_j) = \sum_{j=2}^{n} (\mu(E_j) - \mu(E_{j-1})) + \mu(E_1) = \mu(E_n).$$

Por lo tanto,

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \lim \left(\sum_{j}^{n} \mu(A_j)\right) = \lim \mu(E_n).$$

3. Sea (F_n) una sucesión decreciente con $\mu(F_1) < \infty$ (así $\mu(F_n) < \infty$ para todo n debido al inciso *(1)*). Se crea una sucesión creciente (E_n) donde para

cada $n \in \mathbb{N}$ se define $E_n = F_1 \setminus F_n$, por el inciso *(2)* de este lema se tiene que

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \lim \mu(E_n) = \lim(\mu(F_1) - \mu(F_n)) = \mu(F_1) - \lim \mu(F_n);$$

observar que la última igualdad es válida dado que para todo n, $\mu(F_n) < \infty$. Por otro lado, $\bigcup_{n=1}^{\infty} E_n = F_1 \setminus \bigcap_{n=1}^{\infty} F_n$, entonces

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \mu(F_1) - \mu\left(\bigcap_{n=1}^{\infty} F_n\right).$$

De las dos igualdades anteriores, se tiene

$$\mu(F_1) - \lim \mu(F_n) = \mu(F_1) - \mu\left(\bigcap_{n=1}^{\infty} F_n\right);$$

dado que $\mu(F_1) < \infty$, sumando por su inverso aditivo en la última igualdad y multiplicando por -1 se tiene que

$$\lim \mu(F_n) = \mu\left(\bigcap_{n=1}^{\infty} F_n\right).$$

□

Definición 2.1.2. Un ***espacio de medida*** (o con medida) es una terna $(X, \mathscr{X}, \mu)$ con un conjunto $X \neq \emptyset$, $\mathscr{X}$ una σ-álgebra de X y μ una medida definida en $\mathscr{X}$.

Nota 2.1.1. ***Se dice que una proposición se cumple μ-c.s. (lease "mu casi donde sea" o "mu casi donde quiera")*** *si existe $N \in \mathscr{X}$ tal que $\mu(N) = 0$ y dicha proposición se cumple en el complemento de N, $N^c = X \setminus N$.*

Por ejemplo: si $\mu(N) = 0$ y $f(x) = g(x)$ para todo $x \notin N$, entonces se dice que f y g son igulaes μ-c.s.; o bien, si $f(x) = \lim f_n(x)$ para todo $x \notin N$, entonces se dice que f es el límite de (f_n) μ-c.s.

2.1.1. Medida de probabilidad

Sea Ω el conjunto que denota todos los posibles resultados de algún experimento aleatorio. En probabilidad, a cualquier subconjunto de Ω se le llama ***evento*** y a un elemento $\omega \in \Omega$ se le llama ***punto muestral***.

De esta forma, $(\Omega, \mathscr{O})$ es un espacio medible donde $\mathscr{O}$ es la σ-álgebra de Ω que contiene a todos los eventos (todos los subconjuntos de Ω).

Definición 2.1.3. Una función real P definida en $\mathscr{O}$ se dice que es una ***medida de probabilidad*** si satisface lo siguiente:

1. Para todo $A \in \mathscr{O}$, se tiene que $0 \leq P(A) \leq 1$.

2. $P(\Omega) = 1$.

3. Es **aditiva contable**; es decir, si (A_n) es una sucesión de eventos disjunta dos a dos en $\mathscr{O}$, entonces
$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Observación 2.1.1. Hay que destacar que una medida de probablidad es finita. Además, $P(\emptyset) = 0$ pues
$$P(\emptyset) = P\left(\bigcup_{n=1}^{\infty} \emptyset\right) = \sum_{n=1}^{\infty} P(\emptyset),$$
y si $P(\emptyset) > 0$ la ecuación anterior tiende a infinito contradiciendo la propiedad 2.

Definición 2.1.4. Un ***espacio de medida de probabilidad*** (o sólo *espacio de probabilidad*) es una terna $(\Omega, \mathscr{O}, P)$ con Ω un conjunto no vacío llamado espacio muestral, $\mathscr{O}$ la σ-álgebra que contiene todos los eventos de Ω y P una medida de probabilidad definida en $\mathscr{O}$.

Se le llama ***soporte de*** P a cualquier elemento A de $\mathscr{O}$ tal que $P(A) = 1$.

Ejemplo 2.1.2. Sea $\mathscr{O}$ la σ-álgebra de todos los eventos de un espacio contable Ω y sea f una función no negativa real definida en Ω tal que $\sum_{\omega\in\Omega} f(\omega) = 1$. Entonces, si se define a P para $A \in \mathscr{O}$ como
$$P(A) = \sum_{\omega\in A} f(\omega),$$
entonces $P(A) \geq 0$ pues es la suma de elementos no negativos. Ya que $P(\Omega) = 1$, solo falta ver que P es aditiva contable y, así, se tiene que es una medida de probabilidad.

Supongamos que $A = \bigcup_{i=1}^{\infty} A_i$ donde A_i son disjuntos dos a dos y para cada A_i se denotan sus puntos como $\omega_{i1},\ \omega_{i2}, ...$; así,

$$P(A) = \sum_{i,j} f(\omega_{ij}) = \sum_i \sum_j f(\omega_{ij}) = \sum_i P(A_i).$$

En general, esta igualdad no es cierta pues no siempre se puede separar la suma de esta manera; en este caso, como cada A_i es subconjunto de Ω se tiene que $P(A_i) \leq P(\Omega) = 1$, por lo que fijando cada i o cada j las sumas son finitas y se permite esta igualdad. Por lo tanto, P es aditiva contable. Al espacio $(\Omega, \mathscr{O}, P)$ se le conoce como **espacio de probabilidad discreto**.

Una función real X definida en un espacio de probabilidad y que además sea $\mathscr{O}$-medible es llamada *variable aleatoria*. Esta definición se puede generalizar si la función parte de Ω a un espacio métrico S (más adelante se hablará sobre los espacios métricos).

Definición 2.1.5 (Elemento Aleatorio). Sea $(\Omega, \mathscr{O}, P)$ un espacio de probabilidad y sea X una función $\mathscr{O}$-medible definida sobre un espacio métrico S, entonces se dice que X es un ***elemento aleatorio de*** S. Si $S = \mathbb{R}^n$ se le llama **vector aleatorio** a X, si $S = \mathbb{R}^\infty$ se le llama **sucesión aleatoria** a X, si $S = C(a, b)$ o algún otro espacio de funciones se le llama **función aleatoria** a X; y si $S = \mathbb{R}$ se le llama a X simplemente como **variable aleatoria**; esto es, X es una *variable aleatoria* si

$$X^{-1}(H) = \{\omega \in \Omega \mid X(\omega) \in H\} \in \mathscr{O},$$

para todo conjunto boreliano $H \in \mathscr{B}$.

La intención de tener la definición de esta manera es que la preimágen de cada conjunto boreliano H, $\{\omega \in \Omega \mid X(\omega) \in H\}$, se le pueda asignar una medida de probabilidad; de esta manera, PX^{-1} define una medida de probabilidad para $(\mathbb{R}, \mathscr{B})$ o, en general, para cualquier espacio medible $(S, \mathscr{S})$ con S un espacio métrico.

Definición 2.1.6. Sea $(\Omega, \mathscr{O}, P)$ un espacio de probabilidad y sea X una variable aleatoria, se le llama ***distribución de*** X a la medida de probabilidad $P_X = PX^{-1}$

definida en el espacio medible $(\mathbb{R}, \mathscr{B})$ por

$$P_X(H) = P(X^{-1}(H)) = P[\omega \in \Omega \mid X(\omega) \in H],$$

para todo $H \in \mathscr{B}$.

También, se define a F_X como la ***función de distribución de*** X en $\mathbb{R}$ como

$$F_X(x) = P_X(-\infty, x] = P[\omega \in \Omega \mid X(\omega) \leq x].$$

Notar que P es la medida de probabilidad de $(\Omega, \mathscr{O}, P)$ y P_X está definida sobre la σ-álgebra de Borel de $\mathbb{R}$. La distribución P_X contiene información escencial sobre el comportamiento de la variable aleatoria, como se observará en la integral de una variable aleatoria.

2.2. La integral

Nota 2.2.1. *Se entenderá a $(X, \mathscr{X}, \mu)$ como el espacio de medida constituido por un conjunto no vacío X, $\mathscr{X}$ como una σ-álgebra de X y μ una medida definida en $\mathscr{X}$. Se denota a $M = M(X, \mathscr{X})$ como la colección de todas las funciones $\mathscr{X}$-medibles de X a $\bar{\mathbb{R}}$ y a $M^+ = M^+(X, \mathscr{X})$ como la colección de todas las funciones $\mathscr{X}$-medibles no negativas.*

Definición 2.2.1. Una función real f se dice que es ***simple*** si tiene valuación finita (tiene una cantidad finita de valores reales). Una función simple $\mathscr{X}$-medible φ se puede expresar como

$$\varphi = \sum_{j=1}^{n} a_j \chi_{E_j},$$

donde $a_j \in \mathbb{R}$ y χ_{E_j} la función característica de $E_j \in \mathscr{X}$. Esta es una ***representación estándar de*** φ bajo el hecho de que los a_j son distintos para cada j y que $X = \bigcup E_j$.

Definición 2.2.2. Si φ es una función simple en M^+, se define ***la integral de*** φ ***respecto a*** μ, denotada por $\int f d\mu$, como el valor en $\bar{\mathbb{R}}$ resultado de

$$\int \varphi d\mu = \sum_{j=1}^{n} a_j \mu(E_j),$$

donde los valores a_j y los conjuntos E_j definen una representación estándar, con $a_j \in \mathbb{R}^+$ para cada j.

Observación 2.2.1. $\displaystyle\int \bar{0} d\mu = 0 \in \mathbb{R}$ donde $\bar{0} \in M^+$.

Lema 2.2.1. *Sean $\varphi, \psi \in M^+$ funciones simples y sea $c \in \mathbb{R}^+$, entonces*

$$\int (\varphi + c\psi) d\mu = \int \varphi d\mu + c \int \psi d\mu.$$

Además, si

$$\lambda(E) = \int \varphi \chi_E d\mu,$$

para $E \in \mathscr{X}$, entonces λ es una medida en $\mathscr{X}$.

Demostración. Se escribe a las funciones en su representación estándar como

$$\varphi = \sum_{j=1}^{n} a_j \chi_{E_j} \quad \text{y} \quad \psi = \sum_{j=1}^{m} b_j \chi_{F_j},$$

donde

$$X = \bigcup_{j=1}^{n} E_j = \bigcup_{j=1}^{m} F_j.$$

Así pues, cada E_j se puede representar como $\displaystyle E_j = \bigcup_{i=1}^{m} E_j \cap F_i$, por lo que

$$\begin{aligned}
(\varphi + c\psi) &= \sum_{j=1}^{n} a_j \chi_{E_j} + c \sum_{i=1}^{m} b_i \chi_{F_i} = \sum_{j=1}^{n} \sum_{i=1}^{m} a_j \chi_{E_j \cap F_i} + \sum_{i=1}^{m} \sum_{j=1}^{n} c b_i \chi_{F_i \cap E_j} \\
&= \sum_{j=1}^{n} \sum_{i=1}^{m} a_j \chi_{E_j \cap F_i} + \sum_{j=1}^{n} \sum_{i=1}^{m} c b_i \chi_{E_j \cap F_i} = \sum_{j=1}^{n} \sum_{i=1}^{m} (a_j + c b_i) \chi_{E_j \cap F_i}.
\end{aligned}$$

Sea $G_h = \bigcup E_j \cap F_i$ para los índices i y j tales que $c_h = a_j + cb_i$, se tiene que G_h es una colección disjunta dos a dos para la cual $\mu(G_h) = \sum_{(h)} \mu(E_j \cap F_j)$. Hay que notar que se tiene que introducir c_h para tener una representación estándar de $\varphi + c\psi$ ya que no necesariamente los $a_j + cb_i$ son diferentes entre si. Luego, usando el hecho de que μ es aditiva contable, se tiene

$$\begin{aligned}
\int (\varphi + c\psi) d\mu &= \sum_{h=1}^{p} c_h \mu(G_h) = \sum_{h=1}^{p} \sum_{\{(i,j)|a_j+b_i=c_h\}} c_h \mu(E_j \cap F_i) \\
&= \sum_{j=1}^{n} \sum_{i=1}^{m} (a_j + cb_i) \mu(E_j \cap F_i) \\
&= \sum_{j=1}^{n} \sum_{i=1}^{m} a_j \mu(E_j \cap F_i) + \sum_{j=1}^{n} \sum_{i=1}^{m} cb_i \mu(E_j \cap F_i) \\
&= \sum_{j=1}^{n} a_j \sum_{i=1}^{m} \mu(E_j \cap F_i) + c \sum_{i=1}^{m} b_i \sum_{j=1}^{n} \mu(E_j \cap F_i) \\
&= \sum_{j=1}^{n} a_j \mu(E_j) + c \sum_{i=1}^{m} b_i \mu(F_i) \\
&= \int \varphi d\mu + c \int \psi d\mu.
\end{aligned}$$

Ahora, se observa que $\chi_E \chi_F = \chi_{E\cap F}$ para cualesquiera $E,\ F \in \mathscr{X}$, entonces

$$\varphi \chi_A = \sum_{j=1}^{n} a_j \chi_{E_j} \chi_A = \sum_{j=1}^{n} a_j \chi_{E_j \cap A}.$$

Así, como $\mu(\emptyset) = 0$, se tiene que

$$\lambda(\emptyset) = \int \varphi \chi_{\emptyset} d\mu = \sum_{j=1}^{n} a_j \mu(E_j \cap \emptyset) = \sum_{j=1}^{n} a_j \mu(\emptyset) = 0.$$

Dado que $\varphi \in M^+$, los a_j son no negativos y, como también la función μ es no negativa, se tiene que $a_j \mu(A) \geq 0$ para cualquier $A \in \mathscr{X}$, luego

$$\int \varphi \chi_A d\mu = \sum_{j=1}^{n} a_j \mu(A \cap E_j) \geq 0;$$

así que, $\lambda(A) \geq 0$ para cualquier $A \in \mathscr{X}$.

Finalmente, sea (A_n) una sucesión disjunta dos a dos en $\mathscr{X}$, entonces

$$\begin{aligned}
\lambda\left(\bigcup_{i=1}^{\infty} A_i\right) &= \int \varphi \chi_{\bigcup_{i=1}^{\infty} A_i} d\mu = \sum_{j=1}^{n} a_j \mu\left(E_j \cap \bigcup_{i=1}^{\infty} A_i\right) \\
&= \sum_{j=1}^{n} a_j \mu\left(\bigcup_{i=1}^{\infty} E_j \cap A_i\right) = \sum_{j=1}^{n} a_j \sum_{i=1}^{\infty} \mu\left(E_j \cap A_i\right) \\
&= \sum_{i=1}^{\infty} \sum_{j=1}^{n} a_j \mu\left(E_j \cap A_i\right) = \sum_{i=1}^{\infty} \int \varphi \chi_{A_i} = \sum_{i=1}^{\infty} \lambda(A_i).
\end{aligned}$$

Por lo tanto, λ es una medida definida en $\mathscr{X}$. $\square$

Observación 2.2.2. Se ha definido la integral en base a una medida arbitraria; es decir, sin importar si la medida es finita, medida no finita o medida de probabilidad, por lo que estos resultados prevalecen, en particular, en un espacio de probabilidad.

Definición 2.2.3. Sea $f \in M^+$, se define ***la integral de** f **respecto a** μ* como el valor en $\bar{\mathbb{R}}$ resultado de

$$\int f d\mu = \sup_{\varphi \in M^+ \text{ función simple}} \left(\int \varphi d\mu \right),$$

donde $0 \leq \varphi(x) \leq f(x)$ para todo $x \in X$. Si $E \in \mathscr{X}$, entonces $f\chi_E \in M^+$ y se define la ***integral de** f **sobre** E **respecto a** μ* como

$$\int_E f d\mu = \int f \chi_E d\mu.$$

Lema 2.2.2. *Sea $f, g \in M^+$ tal que $f \leq g$, entonces*

$$\int f d\mu \leq \int g d\mu;$$

además, si $E \subseteq F \in \mathscr{X}$, se tiene que

$$\int_E f d\mu \leq \int_F f d\mu.$$

Demostración. Si φ es una función simple tal que $0 \leq \varphi \leq f$, entonces $0 \leq \varphi \leq g$ por lo que

$$\sup_{0 \leq \varphi \leq f} \int \varphi d\mu \leq \sup_{0 \leq \varphi \leq g} \int \varphi d\mu;$$

o bien,

$$\int f d\mu \leq \int g d\mu.$$

Luego, dado que $E \subseteq F$ se observa que $\chi_E \leq \chi_F$ y, como $f \in M^+$, se tiene que $f\chi_E \leq f\chi_F$, por lo que

$$\int_E f d\mu = \int f \chi_E d\mu \leq \int f \chi_F d\mu = \int_F f d\mu.$$

□

Teorema 2.2.3 (convergencia monótona). *Si* (f_n) *es sucesión monótona creciente en* M^+ *que converge a* f*, entonces*

$$\int f d\mu = \text{lím} \int f_n d\mu.$$

Demostración. Anteriormente, se obtuvó el resultado que f es medible ya que es el límite de una sucesión de funciones medibles (Corolario (2.0.5)). Por el lema anterior, como para todo $n \in \mathbb{N}$ se tiene que $f_n \leq f_{n+1} \leq f$, entonces

$$\int f_n d\mu \leq \int f_{n+1} d\mu \leq \int f d\mu.$$

Por lo que se obtiene una nueva sucesión creciente cuya cota superior es $\int f d\mu$ y donde cada elemento es el valor de una integral; así,

$$\lim_{n\to\infty} \int f_n d\mu \leq \int f d\mu.$$

Ahora se mostrará la desigualdad inversa. Para ello, se toma a una función simple medible φ tal que $0 \leq \varphi \leq f$ y sea $A_n = \{x \in X \mid f_n(x) \geq \varphi(x)\}$. De esta manera, A_n es medible y (A_n) es una sucesión creciente que converge a X; esto es, $X = \bigcup_{n=1}^{\infty} A_n$. Por el lema (2.2.1), λ es una medida en $\mathscr{X}$ si

$$\lambda(A_n) = \int_{A_n} \varphi d\mu$$

y, de nuevo, por el lema anterior se observa que

$$\int_{A_n} \varphi d\mu \leq \int_{A_n} f_n d\mu \leq \int_{\bigcup A_n} f_n d\mu,$$

por lo que al aplicar el límite sobre n se tiene

$$\lim_{n\to\infty} \int_{A_n} \varphi d\mu \leq \lim_{n\to\infty} \int_{A_n} f_n d\mu \leq \lim_{n\to\infty} \int_X f_n d\mu.$$

Por el lema (2.1.1),

$$\int_{\bigcup A_n} \varphi d\mu = \lambda\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \lambda(A_n) = \lim_{n\to\infty} \int_{A_n} \varphi d\mu,$$

sutituyendo en la desigualdad anterior se obtiene

$$\int_X \varphi d\mu \leq \int_X f_n d\mu \leq \lim_{n\to\infty} \int f_n d\mu;$$

o bien,

$$\int_X \varphi d\mu \leq \lim_{n\to\infty} \int f_n d\mu.$$

Como se habla de una función simple arbitraria φ tal que $\varphi \leq f$, por la definición de la integral de f respecto a μ y de la última desigualdad se tiene que

$$\int f d\mu = \sup_{\varphi} \int \varphi d\mu \leq \lim_{n\to\infty} \int f_n d\mu.$$

De esta manera, se muestra la desigualdad deseada. Por lo tanto,

$$\lim \int f_n d\mu = \int f d\mu.$$

□

En este último teorema, no se restringió a la integral a tener valor finito pues la sucesión $\int f_n d\mu$ es una sucesión de valores reales extendidos que tiene límite en $\bar{\mathbb{R}}$ pero podría no tenerlo en $\mathbb{R}$.

Por otro lado, si $f \in M^+(X, \mathscr{X})$ siempre existe una sucesión creciente de funciones medibles de valuación finita (funciones simples) menores o iguales a f y que converge a f; esto por el Lema (2.0.6). Así pues, añadiendo el teorema de convergencia monótona, **la integral de una función $f \in M^+$ es el límite de una sucesión creciente de integrales de funciones simples, el cual siempre existe**.

Por ejemplo, si $c > 0$ y $f, g \in M^+$, entonces $(cf + g) \in M^+$. Si (φ_n) es una sucesión creciente de funciones simples en M^+ que converge a f, entonces $(c\varphi_n)$ es una sucesión creciente de funciones simples en M^+ que converge a cf. Luego, existe una sucesión creciente de funciones simples $(\psi_n) \in M^+$ que converge a g, entonces $(c\varphi_n + \psi_n) \in M^+$ y es una sucesión creciente de funciones simples que converge a $cf + g$. Por el Lema (2.2.1), la integral de la suma de funciones simples es la suma de las integrales de estas funciones y además $\int c\varphi d\mu = c \int \varphi d\mu$; de aquí que,

$$\begin{aligned} \int (cf + g) d\mu &= \lim \int (c\varphi_n + \psi_n) d\mu = \lim \left(c \int \varphi_n d\mu + \int \psi_n d\mu \right) \\ &= c \lim \int \varphi_n d\mu + \lim \int \psi_n d\mu = c \int f d\mu + \int g d\mu. \end{aligned}$$

Con esto, se ha probado el siguiente resultado.

Corolario 2.2.4. *Si $f, g \in M^+$, $c \geq 0$, entonces $(cf + g) \in M^+$ y*

$$\int (cf+g)d\mu = c\int f d\mu + \int g d\mu.$$

Si λ es una función definida en $\mathscr{X}$ como

$$\lambda(E) = \int_E f d\mu,$$

para $E \in \mathscr{X}$, entonces λ es una medida.

Lema 2.2.5 (de Fatou). *Si (f_n) es una sucesión en M^+, entonces*

$$\int \text{lím}\inf(f_n)d\mu \leq \text{lím}\inf\left(\int f_n d\mu\right).$$

Demostración. Para probar el Lema de Fatou conviene apoyarse del Teorema (2.2.3), por lo que se construye una subsucesión (g_m) de (f_n) tal que $g_m \leq f_n$ siempre que $m \leq n$, para esto se define a g_m como $g_m = \text{ínf}\{f_m, f_{m+1}, f_{m+2}, ...\}$. Así, para todo $n \geq m$ se tiene que $\int g_m d\mu \leq \int f_n d\mu$, entonces

$$\int g_m d\mu \leq \text{lím}\inf \int f_n d\mu.$$

Por otro lado, se observa que (g_m) es una sucesión monótona creciente que converge a $\text{lím}\inf(f_n)$, por lo que el Teorema (2.2.3) implica que

$$\text{lím}\left(\int g_m d\mu\right) = \int \text{lím}(g_m)d\mu = \int \text{lím}\inf(f_n)d\mu.$$

En resumen,

$$\int \text{lím}\inf(f_n)d\mu = \text{lím}\left(\int g_m d\mu\right) \leq \text{lím}\inf \int f_n d\mu.$$

□

Proposición 2.2.6. *Sea $f \in M^+$. Entonces $f(x) = 0$ μ-c.s. en X si y sólo si*

$$\int f d\mu = 0.$$

Demostración. Se muestra primero la condición necesaria. Sea E tal que

$$E = \{x \in X \mid f(x) > 0\},$$

como $f(x) = 0$ casi donde quiera entonces $\mu(E) = 0$; entonces todo subconjunto E_j de E también tiene medida cero. Como $f(x) = 0$ en $X \setminus E$ se puede expresar a f como el producto $0 \cdot f$, integrando f y suponiendo que φ denota una función simple tal que $0 \le \varphi \le f$ se tiene que

$$\begin{aligned}
\int f d\mu = \int_E f d\mu + \int_{X\setminus E} f d\mu &= \int f \chi_E d\mu + \int 0 \cdot f \chi_{X\setminus E} \\
&= \sup\left(\int \varphi d\mu\right) + 0 \int f \chi_{X\setminus E} d\mu = \sup\left(\sum_{j=1}^{n} a_j \mu(E_j)\right) + 0 \\
&= \sup\left(\sum_{j=1}^{n} a_j * 0\right) = \sup\left(\sum_{j=1}^{n} 0\right) = 0.
\end{aligned}$$

Ahora, suponiendo que la integral de f respecto a μ es cero y sea E_n el conjunto tal que

$$E_n = \{x \in X \mid f(x) > n^{-1}\}.$$

De esta manera $f \ge n^{-1}\chi_{E_n}$, por lo que al integrar se tiene

$$0 \le \frac{1}{n}\mu(E_n) = \int n^{-1}\chi_{E_n} d\mu \le \int f d\mu = 0;$$

así que $\mu(E_n) = 0$ para cualquier $n \in \mathbb{N}$. Luego, el conjunto

$$\{x \in X \mid f(x) > 0\} = \bigcup_{n=1}^{\infty} E_n$$

también tiene medida cero. Por lo tanto, $f(x) = 0$ μ-c.s. $\square$

Observación 2.2.3. Si f y g son funciones medibles no negativas y suponiendo que $f \le g$ casi donde quiera, entonces el Lema (2.2.2) igual se cumple. Aún más, si (f_n) es una sucesión monótona creciente en M^+ que converge a f casi donde quiera, el Teorema (2.2.3) también se cumple.

2.3. Funciones integrables

Definición 2.3.1. La colección de ***las funciones integrables (o sumables)***, denotada por $L = L(X, \mathscr{X}, \mu)$, consiste en las funciones reales $\mathscr{X}$-medibles definidas en X tales que su parte positiva y negativa, f^+ y f^-, tienen integral finita respecto a μ.

La ***integral de*** f ***respecto a*** μ, con $f \in L$, se define como

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu.$$

Si $E \in \mathscr{X}$, entonces

$$\int_E f d\mu = \int_E f^+ d\mu - \int_E f^- d\mu.$$

En otras palabras, $f \in L(X, \mathscr{X}, \mu)$ si y sólo si f^+ y f^- son medibles y tienen integral finita respecto a μ. Aún más, $|f|$ tiene integral finita pues

$$\int |f| d\mu = \int (f^+ + f^-) d\mu = \int f^+ d\mu + \int f^- d\mu < \infty;$$

así que $|f| \in L(X, \mathscr{X}, \mu)$ si y sólo si $f \in L(X, \mathscr{X}, \mu)$.

De esta manera, algunos resultados adicionales se apoyarán de estos hechos, ya que se ha definido la integral sólo para funciones medibles no negativas; así que la integral de una función medible cualquiera y sus resultados se realiza a través de f^+ y f^- que son funciones positivas evitando así complicaciones.

Lema 2.3.1. *La integral indefinida en* L *es aditiva contable. Esto es, si* $f \in L$ *y* (E_n) *es sucesión disjunta en* $\mathscr{X}$ *cuya unión es* E*, entonces*

$$\int_E f d\mu = \sum_{n=1}^{\infty} \int_{E_n} f d\mu.$$

Demostración. Como λ_1 y λ_2 son medidas si para todo conjunto $A \in \mathscr{X}$

$$\lambda_1(A) = \int_A f^+ d\mu$$

y

$$\lambda_2(A) = \int_A f^- d\mu,$$

ya que $f^+, f^- \in M^+$, entonces

$$\lambda_1(E) = \lambda_1\left(\bigcup E_n\right) = \sum_{n=1}^{\infty} \lambda(E_n) = \sum_{n=1}^{\infty} \int_{E_n} f^+ d\mu,$$

de igual forma se tiene que

$$\lambda_2(E) = \sum_{n=1}^{\infty} \int_{E_n} f^- d\mu;$$

Por definición,

$$\int_E f d\mu = \int_E f^+ d\mu - \int_E f^- d\mu = \lambda_1(E) - \lambda_2(E);$$

o bien,

$$\begin{aligned}\int_E f d\mu &= \sum_{n=1}^{\infty} \int_{E_n} f^+ d\mu - \sum_{n=1}^{\infty} \int_{E_n} f^- d\mu \\ &= \sum_{n=1}^{\infty} \left(\int_{E_n} f^+ d\mu - \int_{E_n} f^- d\mu \right) \\ &= \sum_{n=1}^{\infty} \int_{E_n} f d\mu.\end{aligned}$$

□

Teorema 2.3.2 (Propiedad de integrabilidad absoluta). *Sea f una función integrable, entonces*

$$\left| \int f d\mu \right| \leq \int |f| d\mu.$$

Demostración. Como f^+ y f^- son funciones no negativas, se tiene que

$$f^+ - f^- \leq f^+ + f^-;$$

por el Lema (2.2.4) se cumple que

$$\int (f^+ - f^-) d\mu \leq \int (f^+ + f^-) d\mu;$$

aún más,

$$\left| \int (f^+ - f^-) d\mu \right| \leq \left| \int (f^+ + f^-) d\mu \right| = \int (f^+ + f^-) d\mu.$$

Se observa que $|f|^+ = (f^+ + f^-)$ y $|f|^- = 0$, por lo que

$$\int |f| d\mu = \int (f^+ + f^-) d\mu - \int 0 d\mu = \int (f^+ + f^-) d\mu.$$

Por otro lado,

$$\left| \int f d\mu \right| = \left| \int f^+ d\mu - \int f^- d\mu \right| = \left| \int (f^+ - f^-) d\mu \right|,$$

donde la última igualdad se debe al Corolario (2.2.4). Finalmente, sustituyendo estas integrales en la desigualdad anterior se tiene que

$$\left| \int f d\mu \right| \leq \int |f| d\mu.$$

□

Corolario 2.3.3. *Si f es una función medible, g una función integrable y $|f| \leq |g|$* (o bien, si $|f| \leq |g|$ casi donde quiera)*. Entonces f es integrable y*

$$\int |f| d\mu \leq \int |g| d\mu.$$

Demostración. Si $N \in \mathscr{X}$ es tal que $\mu(N) = 0$, entonces

$$\int |g| d\mu = \int_N |g| d\mu + \int_{X \setminus N} |g| d\mu = \int_{X \setminus N} |g| d\mu < \infty,$$

por el Lema (2.2.2)

$$\int |f| d\mu = \int_{X \setminus N} |f| d\mu \leq \int_{X \setminus N} |g| d\mu = \int |g| d\mu < \infty.$$

□

Teorema 2.3.4 (Operador lineal). *Si $f, g \in L$ y $\alpha \in \mathbb{R}$, entonces $(\alpha f + g) \in L$ y*

$$\int (\alpha f + g) d\mu = \alpha \int f d\mu + \int g d\mu.$$

Demostración. Como $f, g \in L(X, \mathscr{X}, \mu)$ entonces $|f|, |g| \in L(X; \mathscr{X}, \mu)$; es decir, sus integrales son finitas. Luego

$$\int |\alpha f + g| d\mu \leq \int (|\alpha f| + |g|) d\mu = |\alpha| \int |f| d\mu + \int |g| d\mu < \infty,$$

por lo que $(\alpha f + g)$ es función integrable y pertenece a $L(X; \mathscr{X}, \mu)$.

Si $\alpha = 0$,

$$\int (\alpha f + g) d\mu = \int \bar{0} d\mu + \int g d\mu = \int g d\mu.$$

Ahora, si $\alpha \neq 0$, entonces

$$(\alpha f + g) = \begin{cases} (\alpha f^+ - \alpha f^-) + (g^+ - g^-), & \text{si } \alpha > 0, \\ (-\alpha f^+ + \alpha f^-) + (g^+ - g^-), & \text{si } \alpha < 0. \end{cases}$$

Se tomará el caso en que $\alpha > 0$, el otro caso se maneja de forma similar. Ahora, se reescribe la función como

$$(\alpha f + g) = (\alpha f^+ + g^+) - (\alpha f^- + g^-),$$

donde $(\alpha f^{+}+g^{+})$ y $(\alpha f^{-}+g^{-})$ son funciones positivas. Con estas expresiones, se puede aplicar la integral respecto a μ y se obtiene

$$\int(\alpha f+g)d\mu=\int(\alpha f^{+}+g^{+})-\int(\alpha f^{-}+g^{-}).$$

Luego, por el Corolario (2.2.4), se tiene

$$\begin{aligned}\int(\alpha f+g)d\mu&=\alpha\int f^{+}d\mu-\alpha\int f^{-}d\mu+\int g^{+}d\mu-\int g^{-}d\mu\\&=\alpha\int fd\mu+\int gd\mu.\end{aligned}$$

□

Observación 2.3.1. Si $f\in L$, $g\in M$ tal que $f(x)=g(x)$ μ-a.e en X, entonces $g\in L$ y

$$\int fd\mu=\int gd\mu.$$

Para la siguiente definición, se retoma los conceptos de la Definición (2.1.5) (de "Elemento aleatorio").

Definición 2.3.2. Sea X una variable aleatoria en $M(\Omega,\mathscr{O},P)$, con $(\Omega,\mathscr{O},P)$ un espacio de probabilidad. Se dice que X es una **variable aleatoria simple** si la función X es simple.

Sin embargo, el interés es que la función también sea integrable y al resultado de esta integral lo llamarémos de una manera específica, por lo que se retoma la Definición (2.1.6).

Definición 2.3.3 (Esperanza de una variable aleatoria). Sea $(\Omega,\mathscr{O},P)$ un espacio de probabilidad y sea X una variable aleatoria, se le llama ***valor esperado*** o ***esperanza de*** X ***(o de*** $X(\omega)$***)*** a la integral de la función X respecto a P (o de P_X), y se denota por $E(X)$. Es decir,

$$E(X)=\int_{\Omega}XdP=\int_{\mathbb{R}}X(\omega)dP_X=E\left(X(\omega)\right).$$

Debido a esta última definición, el espacio de funciones reales integrables de un espacio de probabilidad, $L(\Omega,\mathscr{O},P)$, es *el espacio de variables aleatorias con esperanza finita.*

2.4. Espacios vectoriales y nociones de convergencia

Un espacio abstracto se define mediante un conjunto de axiomas que enuncian las propiedades básicas de los elementos pertenecientes al espacio. Los elementos pueden ser puntos, vectores, funciones, variables o funciones estocásticas, entre otros. Los espacios vectoriales estan asociados a un campo y están conformados por dos operaciones, la suma y producto por un escalar. La suma relaciona dos elementos del espacio, que llamamos ***vectores***, por lo que hablar de suma puede causar confusión con la suma de elementos del campo, que llamamos ***escalares***, por lo que a veces es mejor tener una notación específica para esta suma. De igual forma, se debe evitar confusión con el producto de un vector por un escalar y el producto definido en el campo, por lo que también se puede denotar de forma específica este producto. Ninguna notación diferente es necesaria siempre que exista claridad de que operación se este realizando. Las referencias que principalmente se consideraron para la elaboración de esta sección son: *"Apuntes de Álgebra Lineal"* (ver [4]), *"Topics In Algebra"* (ver [6]) y *"The Elements of Integration and Lebesgue Measure"* (ver [1]). Así pues, se reconoce un espacio vectorial de la siguiente manera.

Definición 2.4.1. Sea V un conjunto no vacío y sea $\mathbb{F}$ un campo, se dice que V ***es un espacio vectorial sobre el campo*** $\mathbb{F}$, y lo denotamos como $V_{(\mathbb{F})}$, si están definidas dos operaciones de asociación: ***suma*** (o *adición*) y ***producto por un escalar***; tales que para $\vec{u}, \vec{v}, \vec{w} \in V_{\mathbb{F}}$ y $\alpha, \beta \in \mathbb{F}$ se cumple que:

1. La suma de $\vec{u}$ y $\vec{v}$ asigna a un solo elmento $(\vec{u}+\vec{v}) \in V_{(\mathbb{F})}$.
2. El producto de α por $\vec{u}$ asigna a un solo elemento $\alpha\vec{u} \in V_{(\mathbb{F})}$.
3. Existe un elemento $\vec{0} \in V_{(\mathbb{F})}$ tal que $\vec{0}+\vec{u} = \vec{u}$.
4. Para cada vector $\vec{u}$ existe un elemento $-\vec{u} \in V_{(\mathbb{F})}$ tal que $-\vec{u}+\vec{u} = \vec{0}$.
5. La suma es asociativa, $\vec{u}+(\vec{v}+\vec{w}) = (\vec{u}+\vec{v})+\vec{w}$.

6. La suma es abeliana, $\vec{u} + \vec{v} = \vec{v} + \vec{u}$.

7. El producto distribuye escalar sobre la suma de vectores, $\alpha(\vec{u}+\vec{v}) = \alpha\vec{u}+\alpha\vec{v}$.

8. El producto distribuye vector sobre la suma de escalares, $(\alpha+\beta)\vec{u} = \alpha\vec{u}+\beta\vec{u}$.

9. El producto por escalar asocia a los valores escalares, $\alpha(\beta\vec{u}) = (\alpha\beta)\vec{u}$.

10. Si $1 \in \mathbb{F}$ es la unidad del campo, entonces $1\vec{u} = \vec{u}$.

Ejemplo 2.4.1. Se aborda el espacio de funciones integrables de un espacio con medida $L(X, \mathscr{X}, \mu)$ como ejemplo.

Primero, se observa que los vectores son funciones y los escalares son números reales aunque el dominio de las funciones pueda no ser un conjunto relacionado con los reales. Las operaciones definidas en $L(X, \mathscr{X}, \mu)$ son las operaciones comunes para cualquier función; estas son, si $f, g \in L(X, \mathscr{X}, \mu)$ y α es un escalar:

(i) La suma de vectores $(f + g)(x) = f(x) + g(x)$ para todo $x \in X$.

(ii) Producto por un escalar $(\alpha f)(x) = \alpha f(x)$ para todo $x \in X$.

Por el Lema (2.3.4) la suma de funciones integrables y el producto escalar por una función también son funciones integrables, por lo que se cumplen las propiedades *(1)* y *(2)* de espacio vectorial. Luego, para probar las demás propiedades se supondrá que $\alpha, \beta \in \mathbb{R}$ y $f, g, h \in L(X, \mathscr{X} \mu)$.

El elemento $\vec{0}$ es la función $\bar{0}(x) = 0 \in \mathbb{R}$ para todo $x \in X$, pues

$$(f + \bar{0})(x) = f(x) + 0 = f(x),$$

por lo que $f + \bar{0} = f$. El elemento $-f = -1(f)$ es tal que $-f + f = \vec{0}$, pues

$$(-f + f)(x) = -f(x) + f(x) = 0$$

para todo $x \in X$. También se tiene que para $x \in X$,

$$\begin{aligned}(f + (g + h))(x) &= f(x) + (g + h)(x) = f(x) + g(x) + h(x) \\ &= (f + g)(x) + h(x) = ((f + g) + h)(x),\end{aligned}$$
$$(f + g)(x) = f(x) + g(x) = g(x) + f(x) = (g + f)(x);$$

lo cuál implica que $(f+g)+h = f+(g+h)$ y $f+g = g+f$. Así, se han mostrado las propiedades *(3)*, *(4)*, *(5)* y *(6)* de espacio vectorial.

Se observa que las propiedades de campo que tienen los reales son quienes permiten la comprobación de las propiedades del espacio vectorial. De la misma forma, es la propiedad abeliana y distributiva de los reales que verifican las propiedades *(7)*, *(8)* y *(9)* de espacio vectorial.

Por último, notar que $f(x) = 1 \cdot f(x)$ para todo $x \in X$ por lo que $f = 1f$, mostrando así la propiedad *(10)*. Por lo tanto, $L(X, \mathscr{X}, \mu)$ es un espacio vectorial sobre $\mathbb{R}$.

Observación 2.4.1. Bajo una combinación de las propiedades de espacio vectorial se puede ver que el elemento $\vec{0}$ y $-\vec{u}$ para $\vec{u}$ en $V_{(\mathbb{F})}$ son elementos únicos en el espacio. También que $\alpha\vec{0} = \vec{0}$ y que $0\vec{u} = \vec{0}$ para cualquier vector $\vec{u}$ y escalar α.

Como estas propiedades se cumplen para cualquier vector del espacio y para cualquier escalar del campo, en particular se deben cumplir en cualquier subconjunto del espacio. Si el producto escalar es cerrado en este subconjunto, digamos S, entonces contiene al elemento $\vec{0}$ y, dado $\vec{u} \in S$, contiene también al elemento $-\vec{u}$, pues $0\vec{u} = \vec{0}$ y $-1\vec{u} = -\vec{u} \in S$.

Así que, si en un subconjuto sucede que las operaciones son cerradas, se tiene que este subconjunto también es un espacio vectorial; es decir, si $S \subseteq V$ y para todo $\vec{u}, \vec{v} \in S$ y $\alpha \in \mathbb{F}$ sucede que $(\vec{u}+\vec{v}), (\alpha\vec{u}) \in S$ entonces $S_{(\mathbb{F})}$ es también un espacio vectorial y lo llamamos ***subespacio de*** $V_{(\mathbb{F})}$.

Ejemplo 2.4.2. Se denota como $L_p = L_p(X, \mathscr{X}, \mu)$ a la colección de ***las funciones integrables de orden p respecto a*** μ ***definidas en*** $(X, \mathscr{X})$, con p un entero mayor o igual que 1; estas son las funciones $f \in M$ tales que $\int |f|^p d\mu < \infty$. Se tiene que para cada $p \in \mathbb{N}$, L_p es un espacio vectorial

Si μ es una medida finita, es fácil ver que L_p es un subconjunto de L_r, donde $1 \le r \le p$, pues si $f \in L_p$ entonces $|f|^r \le 1 + |f|^p$ se cumple para $r = 1, ..., p$ y al integrar respecto a μ se tiene

$$\int |f|^r d\mu \le \int (1+|f|^p) d\mu = \int d\mu + \int |f|^p d\mu = \mu(X) + \int |f|^p d\mu < \infty.$$

Tampoco son vacíos pues cada función constante es integrable de orden p (si μ es finita), para verificar que es un subespacio de L_r hay que revisar que las operaciones son cerradas en L_p. Sea $f, g \in L_p(X, \mathscr{X}, \mu)$, entonces $\int |f|^p d\mu < \infty$ y $\int |g|^p d\mu < \infty$; luego, dado $\alpha \in \mathbb{R}$ se tiene que

$$\int |\alpha f|^2 d\mu = \int |\alpha|^2 |f|^2 d\mu = |\alpha|^2 \int |f|^2 d\mu < \infty.$$

Por otro lado, se observa que $|f+g|^p \leq (2 \sup\{|f|, |g|\})^p \leq 2^p \left(|f|^p + |g|^p\right)$, y en virtud del Corolario (2.3.3) se tiene

$$\int |f+g|^p d\mu \leq 2^p \int (|f|^p + |g|^p) d\mu = 2^p \int |f|^p d\mu + 2^p \int |g|^p d\mu < \infty;$$

por lo que las funciones αf y $(f+g)$ pertencen a $L_p(X, \mathscr{X}, \mu)$ y, así, ambas operaciones son cerradas en este subconjunto. Por lo tanto, $L_p(X, \mathscr{X}, \mu)$ **es un subespacio de** $L_r(X, \mathscr{X}, \mu)$ para $r = 1, ..., p$.

Aún más, si $(\Omega, \mathscr{O}, P)$ es un espacio de probabilidad, entonces es de medida finita por lo que el **espacio de variables aleatorias con esperanza finita de orden p**, $L_p(\Omega, \mathscr{O}, P)$, es un subespacio de $L(\Omega, \mathscr{O}, P)$. De esta forma, se tiene que $X \in L_p(\Omega, \mathscr{O}, P)$ si y sólo si $E(|X|^p) < \infty$.

Nota 2.4.1. *Se ha usado L_p para denotar el espacio de las funciones integrables de orden p, no se deben confundir con los espacios normados de Lebesgue que usualmente se denotan como $\mathscr{L}_p$ o $\mathcal{L}_p$ (donde la "L" es escrita con un estilo curveado).*

Ya que se ha introducido los espacios de funciones integrables de orden p, L_p, se muestran algunos resultados que servirán de apoyo para estudiar la convergencia.

Proposición 2.4.1. *Sea $f \in L_p$ y sea $g \in L_q$, donde $p > 1$ y $\frac{1}{p} + \frac{1}{q} = 1$, entonces $fg \in L_1$.*

Demostración. Si $f \in L_p$ y $g \in L_q$, entonces $\int |f|^p d\mu < \infty$ y $\int |g|^q d\mu < \infty$ y por hipótesis $\frac{1}{q} = 1 - \frac{1}{p}$. Se mostrará que $\int |fg| d\mu < \infty$.

Sea φ una función real definida como $\varphi(t) = \alpha t - t^\alpha$ para $0 \leq t$, donde $\alpha \in (0,1)$; derivando φ, se tiene que $\varphi'(t) = \alpha - \alpha t^{\alpha-1}$. Luego, se muestra que φ alcanza su valor mínimo en $t = 1$, primero se observa que para $t > 0$ se tiene:

- $\varphi'(t) > 0 \iff \alpha - \alpha t^{\alpha-1} > 0 \iff 1 > t^{\alpha-1} > 0 \iff t^{1-\alpha} > 1 \iff t > 1$.
- $\varphi'(t) < 0 \iff \alpha - \alpha t^{\alpha-1} < 0 \iff 1 < t^{\alpha-1} \iff 0 < t^{1-\alpha} < 1 \iff t < 1$.
- $\varphi'(1) = \alpha - \alpha 1^{\alpha-1} = \alpha - \alpha = 0$.

Comparando con $t = 0$, que es el otro punto crítico aunque la derivada de la función en 0 no está definida, se tiene que

$$\varphi(0) = 0 > \alpha - 1 = \varphi(1);$$

por lo tanto, $\varphi(1) \leq \varphi(t)$ para $t \geq 0$. Es decir, dado $\alpha \in (0,1)$ y para todo $t \geq 0$ se tiene que

$$\alpha - 1 \leq \alpha t - t^{\alpha}.$$

Si $\alpha = \frac{1}{p}$ y escribiendo a t como el cociente de dos reales positivos A^p y B^q, $t = \dfrac{A^p}{B^q} > 0$, se tiene que

$$-\frac{1}{q} = \frac{1}{p} - 1 \leq \frac{1}{p}\frac{A^p}{B^q} - \left(\frac{A^p}{B^q}\right)^{\frac{1}{p}} = \frac{A^p}{p}\frac{1}{B^q} - \frac{A^{p^{1/p}}}{B^{q^{1/p}}}.$$

Ahora, multiplicando la desigualdad por B^q, se tiene

$$-\frac{B^q}{q} \leq \frac{A^p}{p} - \frac{AB^q}{B^{q/p}}.$$

Por otro lado,

$$\frac{B^q}{B^{q/p}} = B^{q-q/p} = B^{q(1-1/p)} = B^{q^{1/q}} = B,$$

por lo que la desigualdad se puede ver como

$$AB \leq \frac{A^p}{p} + \frac{B^q}{q}. \tag{2.4}$$

Como t es el cociente de dos reales positivos, se puede ver que si $A = |f(x)|$ y $B = |g(x)|$ entonces

$$|f(x)g(x)| \leq \frac{|f(x)|^p}{p} + \frac{|g(x)|^q}{q}$$

para todo $x \in X$. Por lo tanto,

$$\int |fg| d\mu \leq \frac{1}{p}\int |f|^p d\mu + \frac{1}{q}\int |g|^q d\mu < \infty.$$

Por lo que $fg \in L$. $\square$

Aún más, retomando la expresión (2.4) de la demostración anterior y suponiendo que f y g son diferentes de la función cero casi donde quiera (así las integrales de $|f|^p$ y $|g|^q$ son mayores que cero), se tiene que si

$$A = \frac{|f(x)|}{\left(\int |f|^p d\mu\right)^{1/p}} \quad \text{y} \quad B = \frac{|g(x)|}{\left(\int |g|^q d\mu\right)^{1/q}}$$

entonces

$$\frac{|f(x)g(x)|}{\left(\int |f|^p d\mu\right)^{1/p}\left(\int |g|^q d\mu\right)^{1/q}} \leq \frac{|f(x)|^p}{p\left(\int |f|^p d\mu\right)} + \frac{|g(x)|^q}{q\left(\int |g|^q d\mu\right)};$$

para todo $x \in X$. Luego, por el Teorema (2.3.4), al integrar respecto a μ se obtiene

$$\frac{\int |fg| d\mu}{\left(\int |f|^p d\mu\right)^{1/p}\left(\int |g|^q d\mu\right)^{1/q}} \leq \frac{\int |f|^p d\mu}{p\left(\int |f|^p d\mu\right)} + \frac{\int |g|^q d\mu}{q\left(\int |g|^q d\mu\right)} = \frac{1}{p} + \frac{1}{q} = 1;$$

pues por hipótesis $\frac{1}{p}+\frac{1}{q} = 1$. Ahora, multiplicando la desigualdad por los cocientes de A y B se obtiene

$$\int |fg| d\mu \leq \left(\int |f|^p d\mu\right)^{\frac{1}{p}} \left(\int |g|^q d\mu\right)^{\frac{1}{q}}. \tag{2.5}$$

A la expresión (2.5) se le conoce como ***la desigualdad de Hölder*** y será de utilidad más adelante. Notar que si las funciones son cero o iguales a la función cero casi donde quiera la desigualdad se sigue cumpliendo.

Observación 2.4.2. La proposición anterior implica que el producto de una función en L_p con otra función en L_q es integrable si $p > 1$ y $p + q = pq$. Estos índices se llaman ***índices conjugados*** siendo $p = 2$ el único conjugado *propio* y, por lo tanto, el **producto de dos funciones en L_2 es integrable**.

Corolario 2.4.2. *Sean $f, g \in L_2$, entonces fg es integrable ($fg \in L$).*

Demostración. Este es un caso particular de la Proposición (2.4.1), basta ver que $2 > 1$ y que $\frac{1}{2}+\frac{1}{2} = 1$, por lo que si dos funciones f y g pertenecen a L_2, entonces fg pertenece a L. □

En consecuencia de este corolario, si $(\Omega, \mathscr{O}, P)$ es un espacio de probabilidad, entonces si dos variables aleatorias tienen esperanza finita de orden 2 implica que su producto tiene esperanza finita. Esto es, si $E(|X|^2) < \infty$ y $E(|Y|^2) < \infty$ entonces $X, Y \in L_2(\Omega, \mathscr{O}, P)$, por lo que $XY \in L(\Omega, \mathscr{O}, P)$ y $E(|XY|) < \infty$.

La idea de *magnitud* (o tamaño) de un vector se introduce formalmente en un espacio vectorial a través del concepto de *norma*.

Definición 2.4.2 (Norma). Sea $V_{(\mathbb{R})}$ un espacio lineal (vectorial), a una función N se le dice ***norma sobre*** $V_{(\mathbb{R})}$ si es una función real no negativa tal que cumple lo siguiente:

1. $N(\vec{v}) = 0$ si y sólo si $\vec{v} = \vec{0}$.

2. $N(\alpha\vec{v}) = |\alpha| N(\vec{v})$ para todo $\vec{v} \in V_{(\mathbb{R})}$ y cualquier $\alpha \in \mathbb{R}$.

3. $N(\vec{u} + \vec{v}) \leq N(\vec{u}) + N(\vec{v})$ para todo $\vec{u}, \vec{v} \in V_{(\mathbb{R})}$.

Al espacio (V, N) se le llama **espacio vectorial normado**.

Si la función N no satisface la propiedad (1) pero si las demás propiedades se dice que es una ***seminorma*** o ***pseudonorma***.

Ejemplo 2.4.3. A continuación, se presenta tres ejemplos de normas.

a) Sea $V = \mathbb{R}^n$ y $\vec{u} = (u_1, ..., u_n) \in V$, se pueden definir estas tres normas:

(*i*) $N_1(u_1, ..., u_n) = \sum_{i=1}^{n} |u_i|$.

(*ii*) $N_p(u_1, ..., u_n) = \sum_{i=1}^{n} |u_i|^p$ con $p \geq 1$.

(*iii*) $N_\infty(u_1, ..., u_n) = \sup\{|u_1|, ..., |u_n|\}$.

b) Sea $B(X)$ la colección de funciones reales acotadas en X, es un espacio vectorial normado bajo la función

$$N(f) = \sup\{|f(x)| \mid x \in X\}.$$

c) Sea $(X, \mathscr{X}, \mu)$ un espacio de medida, si $f \in L$ entonces N_μ es una seminorma para el espacio $L(X, \mathscr{X}, \mu)$ si

$$N_\mu(f) = \int |f| d\mu.$$

La siguiente definición presenta el concepto de producto interno que surge como una generalización a un espacio vectorial cualquiera del producto escalar en $\mathbb{R}^n$. Por esta razón, se contempla el campo complejo $\mathbb{C}$ o un subcampo de este.

Definición 2.4.3 (Producto interno). Sea $V_{(\mathbb{C})}$ un espacio vectorial, se define el ***producto interno*** en $V_{(\mathbb{C})}$ como una función que asigna a cada pareja ordenada de vectores un valor escalar del campo, $\langle \cdot \mid \cdot \rangle : V_{(\mathbb{C})} \times V_{(\mathbb{C})} \to \mathbb{C}$, y que, para $\vec{u}, \vec{v}, \vec{w} \in V_{(\mathbb{C})}$ y $\alpha \in \mathbb{C}$, satisface lo siguiente:

1. $\langle \vec{u} \mid \vec{v} \rangle = \overline{\langle \vec{v} \mid \vec{u} \rangle}$.

2. $\langle \vec{u} \mid \vec{v} + \vec{w} \rangle = \langle \vec{u} \mid \vec{v} \rangle + \langle \vec{u} \mid \vec{w} \rangle$.

3. $\langle \alpha\vec{u} \mid \vec{v} \rangle = \alpha \langle \vec{u} \mid \vec{v} \rangle$.

4. $\langle \vec{u} \mid \vec{u} \rangle > 0$ si $\vec{u} \neq 0$ y $\langle \vec{u} \mid \vec{u} \rangle = 0$ si y sólo si $\vec{u} = 0$.

Nota 2.4.2. *Cuando el espacio se restringe al campo $\mathbb{R}$, la propiedad* (1) *se resume a una abelianidad: $\langle \vec{u} \mid \vec{v} \rangle = \langle \vec{v} \mid \vec{u} \rangle$. En lo que sigue, se limitará el espacio al campo $\mathbb{R}$ aunque los resultados obtenidos puedan extenderse al campo $\mathbb{C}$.*

Cuando un espacio vectorial cuenta con producto interno, se puede inducir un espacio normado donde la norma es inferida por el producto interno. Para verificar esto, primero se requiere revisar el siguiente resultado.

Teorema 2.4.3 (Desigualdad de Cauchy-Schwarz). *Sea $V_{(\mathbb{R})}$ un espacio vectorial con producto interno definido, entonces para todo $\vec{u}, \vec{v} \in V_{(\mathbb{R})}$ se tiene que*

$$|\langle \vec{u} \mid \vec{v} \rangle|^2 \leq \langle \vec{u} \mid \vec{u} \rangle \langle \vec{v} \mid \vec{v} \rangle ;$$

o bien,

$$-\langle \vec{u} \mid \vec{u} \rangle^{\frac{1}{2}} \langle \vec{v} \mid \vec{v} \rangle^{\frac{1}{2}} \leq \langle \vec{u} \mid \vec{v} \rangle \leq \langle \vec{u} \mid \vec{u} \rangle^{\frac{1}{2}} \langle \vec{v} \mid \vec{v} \rangle^{\frac{1}{2}} .$$

Demostración. Si alguno de los vectores es el vector cero se obtendría la igualdad. Por otro lado, si $\vec{u} = \alpha\vec{v}$ con $\alpha \in \mathbb{R}$ se tiene que

$$\begin{aligned} |\langle \vec{u} \mid \vec{v} \rangle|^2 &= |\langle \alpha\vec{v} \mid \vec{v} \rangle|^2 = |\alpha \langle \vec{v} \mid \vec{v} \rangle|^2 = |\alpha||\alpha| \, |\langle \vec{v} \mid \vec{v} \rangle| \, |\langle \vec{v} \mid \vec{v} \rangle| \\ &= |\langle \alpha\vec{v} \mid \alpha\vec{v} \rangle| \, |\langle \vec{v} \mid \vec{v} \rangle| = |\langle \vec{u} \mid \vec{u} \rangle| \, |\langle \vec{v} \mid \vec{v} \rangle| \\ &= \langle \vec{u} \mid \vec{u} \rangle \langle \vec{v} \mid \vec{v} \rangle . \end{aligned}$$

En este caso, también se cumple la igualdad. Ahora, si ninguno de los vectores es cero y $\vec{u} \neq -\alpha\vec{v}$ para cualquier $\alpha \in \mathbb{R}$, entonces $\vec{u} + \alpha\vec{v} \neq 0$ y por la propiedad *(4)* de la Definición (2.4.3) se tiene

$$\langle \vec{u} + \alpha\vec{v} \mid \vec{u} + \alpha\vec{v} \rangle > 0.$$

Luego, desarrollando el producto interno bajo uso de sus propiedades, se obtiene

$$\begin{aligned} \langle \vec{u} + \alpha\vec{v} \mid \vec{u} + \alpha\vec{v} \rangle &= \langle \vec{u} + \alpha\vec{v} \mid \vec{u} \rangle + \langle \vec{u} + \alpha\vec{v} \mid \alpha\vec{v} \rangle \\ &= \langle \vec{u} \mid \vec{u} \rangle + \langle \alpha\vec{v} \mid \vec{u} \rangle + \langle \vec{u} \mid \alpha\vec{v} \rangle + \langle \alpha\vec{v} \mid \alpha\vec{v} \rangle \\ &= \langle \vec{u} \mid \vec{u} \rangle + 2\alpha \langle \vec{u} \mid \vec{v} \rangle + \alpha^2 \langle \vec{v} \mid \vec{v} \rangle ; \end{aligned}$$

en resumen, se tiene

$$\langle \vec{u} \mid \vec{u} \rangle + 2\alpha \langle \vec{u} \mid \vec{v} \rangle + \alpha^2 \langle \vec{v} \mid \vec{v} \rangle > 0.$$

Como esto sucede para cualquier escalar, en particular sucede para $\alpha = -\dfrac{\langle \vec{u} \mid \vec{v} \rangle}{\langle \vec{v} \mid \vec{v} \rangle}$, que sustituyendo en la desigualdad anterior se obtiene

$$\langle \vec{u} \mid \vec{u} \rangle - \frac{\langle \vec{u} \mid \vec{v} \rangle^2}{\langle \vec{v} \mid \vec{v} \rangle} > 0;$$

o bien

$$\langle \vec{u} \mid \vec{u} \rangle \langle \vec{v} \mid \vec{v} \rangle > \langle \vec{u} \mid \vec{v} \rangle^2 .$$

De esta manera, la demostración queda completa. □

Ahora, se define la norma a través del producto interno de la siguiente manera:

$$||\vec{v}|| = \langle \vec{v} \mid \vec{v} \rangle^{\frac{1}{2}} . \tag{2.6}$$

Se verifica que efectivamente es una norma. Por la propiedad *(4)* de la Definición de producto interno (2.4.3), el producto es mayor o igual que 0, por lo que $\langle \vec{v} \mid \vec{v} \rangle^{\frac{1}{2}} \geq 0$; es decir, la función definida es nonegativa. Luego,

$$||\vec{v}|| = \langle \vec{v} \mid \vec{v} \rangle^{\frac{1}{2}} = 0 \iff \langle \vec{v} \mid \vec{v} \rangle = 0 \iff \vec{v} = 0.$$

Así que la propiedad *(1)* de la norma (Definición (2.4.2)) se cumple. Después

$$||\alpha\vec{v}|| = \langle \alpha\vec{v} \mid \alpha\vec{v} \rangle^{\frac{1}{2}} = \left(\alpha^2\right)^{\frac{1}{2}} \langle \vec{v} \mid \vec{v} \rangle^{\frac{1}{2}} = |\alpha| \langle \vec{v} \mid \vec{v} \rangle^{\frac{1}{2}} = |\alpha| \; ||v||,$$

por lo que la propiedad *(2)* también se cumple. Por último, se muestra la tercera propiedad de la norma. Se observa que

$$\begin{aligned}||\vec{u}+\vec{v}||^2 &= \langle \vec{u}+\vec{v} \mid \vec{u}+\vec{v}\rangle = \langle \vec{u}+\vec{v} \mid \vec{u}\rangle + \langle \vec{u}+\vec{v} \mid \vec{v}\rangle \\ &= \langle \vec{u} \mid \vec{u}\rangle + \langle \vec{u} \mid \vec{v}\rangle + \langle \vec{v} \mid \vec{u}\rangle + \langle +\vec{v} \mid \vec{v}\rangle \\ &= \langle \vec{u} \mid \vec{u}\rangle + 2\langle \vec{u} \mid \vec{v}\rangle + \langle \vec{v} \mid \vec{v}\rangle ,\end{aligned}$$

aplicando el Teorema (2.4.3), se obtiene

$$\begin{aligned}\langle \vec{u} \mid \vec{u}\rangle + 2\langle \vec{u} \mid \vec{v}\rangle + \langle \vec{v} \mid \vec{v}\rangle &\leq \langle \vec{u} \mid \vec{u}\rangle + 2\langle \vec{u} \mid \vec{u}\rangle^{\frac{1}{2}}\langle \vec{v} \mid \vec{v}\rangle^{\frac{1}{2}} + \langle \vec{v} \mid \vec{v}\rangle \\ &= \left(\langle \vec{u} \mid \vec{u}\rangle^{\frac{1}{2}} + \langle \vec{v} \mid \vec{v}\rangle^{\frac{1}{2}}\right)^2 .\end{aligned}$$

Así pues,

$$||\vec{u}+\vec{v}||^2 \leq \left(\langle \vec{u} \mid \vec{u}\rangle^{\frac{1}{2}} + \langle \vec{v} \mid \vec{v}\rangle^{\frac{1}{2}}\right)^2$$

si y sólo si

$$||\vec{u}+\vec{v}|| \leq \left(\langle \vec{u} \mid \vec{u}\rangle^{\frac{1}{2}} + \langle \vec{v} \mid \vec{v}\rangle^{\frac{1}{2}}\right) = ||\vec{u}|| + ||\vec{v}||.$$

Por lo tanto, la función definida en la ecuación (2.6) es una norma sobre un espacio vectorial con producto interno; por lo que se ha probado el siguiente corolario.

Corolario 2.4.4. *Sea $V_{(\mathbb{R})}$ un espacio vectorial con producto interno, entonces la función $||\vec{v}|| = \langle \vec{v} \mid \vec{v}\rangle^{\frac{1}{2}}$ define una norma en este espacio.*

Empleando el concepto de norma, se puede introducir la idea de métrica (distancia) entre vectores.

Definición 2.4.4. Sea $V_{(\mathbb{R})}$ un espacio vectorial, se le llama ***función distancia*** o **métrica** a la función real $d : V_{(\mathbb{R})} \times V_{(\mathbb{R})} \to [0, \infty)$ tal que si $\vec{u}, \vec{v}, \vec{w} \in V_{(\mathbb{R})}$ se cumple lo siguiente:

1. $d(\vec{u}, \vec{v}) \geq 0$.
2. $d(\vec{u}, \vec{v}) = 0$ si y sólo si $\vec{u} = \vec{v}$.
3. $d(\vec{u}, \vec{v}) = d(\vec{v}, \vec{u})$.
4. $d(\vec{u}, \vec{v}) \leq d(\vec{u}, \vec{w}) + d(\vec{w}, \vec{v})$.

A los espacios con esta característica se les llama ***espacios métricos*** y suele denotarse por (V, d).

Observación 2.4.3. Si $V_{(\mathbb{R})}$ es un espacio vectorial normado, la distancia en este puede definirse de manera inducida por la norma como

$$d(\vec{u}, \vec{v}) = ||\vec{v} - \vec{u}||. \tag{2.7}$$

La distancia definida de esta manera será utilizada cuando se hable de convergencia.

En los espacios con producto interno real, tambíen es posible definir el concepto de *ángulo* entre dos vectores de la siguiente manera.

Definición 2.4.5. Sea $V_{(\mathbb{R})}$ un espacio vectorial con producto interno real y sean $\vec{u}$ y $\vec{v}$ dos vectores no nulos (diferentes de $\vec{0}$). Se llama ángulo entre $\vec{u}$ y $\vec{v}$ al número real θ tal que $0 \leq \theta \leq \pi$ y que

$$\cos\theta = \frac{\langle \vec{u} \mid \vec{v} \rangle}{||\vec{u}||\ ||\vec{v}||}.$$

Este ángulo está bien definido pues, a partir de la desigualdad de Cauchy-Schwarz (2.4.3), si $\vec{u}$ y $\vec{v}$ son no nulos, entonces

$$-1 \leq \frac{\langle \vec{u} \mid \vec{v} \rangle}{||\vec{u}||\ ||\vec{v}||} \leq 1,$$

por lo que existe un **único** número real θ en el intervalo $[0, \pi]$ para el cual el coseno es igual a dicho cociente.

Nota 2.4.3. *Los conceptos de norma, distancia y ángulo que se han definido en esta sección constituyen una generalización a espacios vectoriales de sus ideas geómetricas que estas representan.*

Lema 2.4.5 (Espacio cociente). *Sea $V_{(\mathbb{F})}$ un espacio vectorial sobre el campo $\mathbb{F}$ y sea $W_{(\mathbb{F})}$ un subespacio de $V_{(\mathbb{F})}$. Entonces $^{V}/_{W}$ es un espacio vectorial sobre el campo $\mathbb{F}$, donde las operaciones para $(\vec{u}+W), (\vec{v}+W) \in {}^{V}/_{W}$ y $\alpha \in \mathbb{F}$ se definen a continuación.*

1. ***Suma:*** $(\vec{u} + W) + (\vec{v} + W) = (\vec{u} + \vec{v}) + W$.

2. ***Producto por un escalar:*** $\alpha(\vec{u}+W)=\alpha\vec{u}+W$.

El espacio ${}^{V}/_{W}$ *es llamado el* **espacio cociente de** V **sobre** W.

Demostración. (Ver referencia [6]). □

Observación 2.4.4. El espacio cociente es interpretado como el espacio de **las clases de vectores equivalentes respecto a** W; esto es, si $\vec{u}\in V_{(\mathbb{F})}$, $\vec{u}+W=W$ si existe $\vec{w}\in W_{(\mathbb{F})}$ para el cual $\vec{u}-\vec{0}=\vec{w}$ y así $\vec{u}\in W_{(\mathbb{F})}$. Por otro lado, si $\vec{v}\in V_{(\mathbb{F})}$ se tiene que

$$\vec{u}+W=\vec{v}+W \iff \vec{u}-\vec{v}+W=W \iff (\vec{u}-\vec{v})=\vec{w}\in W_{(\mathbb{F})}.$$

Los elementos en ${}^{V}/_{W}$ suelen denotarse de forma distinta para enfatizar que pertenecen a un espacio de clases que es distinto del espacio de origen $V_{(\mathbb{F})}$, como ejemplo se tiene:

$$\vec{u}+W, \quad [\vec{u}] \quad \text{o} \quad \overline{u}.$$

Ejemplo 2.4.4. Se ha mencionado que N_μ es una seminorma para el espacio $L(X,\mathscr{X},\mu)$ si

$$N_\mu(f)=\int |f|d\mu,$$

para $f\in L$. En efecto, pues:

- Para todo $f\in L$, se tiene

$$N_\mu(f)=\int |f|d\mu\geq 0.$$

- Si $\alpha\in\mathbb{R}$,

$$N_\mu(\alpha f)=\int|\alpha f|d\mu=\int|\alpha||f|d\mu=|\alpha|\int|f|d\mu=|\alpha|N_\mu(f).$$

- Si $g\in L$,

$$\int|f+g|d\mu\leq\int(|f|+|g|)\,d\mu=\int|f|d\mu+\int|g|d\mu,$$

  por lo que $N_\mu(f+g)\leq N_\mu(f)+N_\mu(g)$.

Así, N_μ es una seminorma. De hecho, por la Proposición (2.2.6), $N_\mu(f) = 0$ si y sólo si f es equivalente a la función cero μ-casi donde quiera. Aún más, $f = g$ μ-c.s. si y sólo si $f - g = 0$ μ-c.s; es decir, $f - g$ es una función equivalente a la función cero μ-c.s. Estos argumentos insinúan que si consideramos las clases de funciones μ-equivalentes se puede obtener un espacio vectorial normado. Para esto, se considera el conjunto K como

$$K = \{f \in L \mid f = 0, \quad \mu - \text{casi donde quiera}\},$$

se mostrará que K es subespacio de L. Primero, es evidente que $\vec{0} \in K$. Luego, sean f y g funciones en K, entonces $f(x) = g(x) = 0$ para todo $x \in X \setminus M$ donde $M \in \mathscr{X}$ tiene medida cero; entonces, $(f + g)(x) = f(x) + g(x) = 0$ para todo $x \in X \setminus M$, por lo que $(f + g) \in K$. Por último, si $\alpha \in \mathbb{R}$ se tiene que $(\alpha f)(x) = \alpha f(x) = \alpha \cdot 0 = 0$ para todo $x \in X \setminus M$, por lo que $\alpha f \in K$.

Por lo tanto, K es un subespacio de L y así, por el Lema (2.4.5), ${}^{L}/_{K}$ es ***el espacio vectorial de las clases de funciones integrables μ-equivalentes*** **(o iguales μ-casi donde quiera)**; además, como $[\vec{0}]$ es el elemento neutro y por la Proposición (2.2.6), se tiene que N_μ define una norma en este ${}^{L}/_{K}$.

Nota 2.4.4 (**Sobre la notación de las clases de funciones μ-equivalentes**). *Cada elemento del espacio ${}^{L}/_{K}$ es un conjunto y la integral de un conjunto $[f]$ no está definida; sin embargo, dado que para cada función $g \in [f]$ el valor de la integral es igual a $\int f d\mu$, bastará con definir la* ***integral de una clase*** *como*

$$\int [f] d\mu = \int f d\mu.$$

Se ha visto que en varios resultados basta que la hipótesis se cumpla μ casi donde quiera. Así, para fines prácticos, en lo que sigue se considera al espacio $L(X, \mathscr{X}, \mu)$ como el espacio de las clases de funciones integrables μ-equivalentes sin olvidar que L denotará ${}^{L}/_{K}$ y que un elemento f denotará la clase de las funciones μ-equivalentes a f. En general, los espacios L_p que ahora denotarán el ***espacio de las funciones integrables de orden p μ-equivalentes*** *con $p \geq 1$ y donde la función que describe la norma en forma general es*

$$||f||_p = \left(\int |f|^p d\mu \right)^{\frac{1}{p}}. \tag{2.8}$$

Observación 2.4.5. Se puede expresar la **desigualdad de Hölder** (2.5) en términos de la norma. Esto es, si $f \in L_p$ y $g \in L_q$, con p y q índices conjugados, entonces

$$||fg||_1 \leq ||f||_p ||g||_q. \tag{2.9}$$

Si $(\Omega, \mathscr{O}, P)$ es un espacio de probabilidad y $f \in L_p$, entonces $|f|^r \in L_{\frac{p}{r}}$ para $r = 1, ..., p$ y usando esta desigualdad se obtiene que

$$||f||_r \leq ||f||_p \cdot (P(\Omega))^{\frac{p-r}{pr}} = ||f||_p.$$

Antes de seguir, es adecuado mostrar que la función definida en (2.8) es una norma para L_p. Es evidente que es una función no negativa y que es cero si y sólo si $f = [\vec{0}]$, luego

$$||\alpha f||_p = \left(\int |\alpha f|^p d\mu \right)^{\frac{1}{p}} = \left(|\alpha|^p \int |f|^p d\mu \right)^{\frac{1}{p}} = |\alpha| ||f||_p,$$

cumpliendo así la segunda propiedad de la Definición (2.4.2). Lo que no es evidente es que cumpla la tercera propiedad de la misma definición, para mostrar esto se introduce la ***desigualdad de Minkowski***, la cual dice que si $f, g \in L_p$ $(p \geq 1)$ entonces

$$||f + g||_p \leq ||f||_p + ||g||_p. \tag{2.10}$$

De esta forma, la función $|| \cdot ||_p$ es una norma para L_p.

Para mostrar que esta desigualdad se cumple, primero se observa que

$$|f + g|^p = |f + g||f + g|^{p-1} \leq |f||f + g|^{p-1} + |g||f + g|^{p-1},$$

como $|f + g|$ es integrable de orden p, la función $|f + g|^{p-1}$ es integrable de orden $\frac{p}{p-1}$ (que es mayor que 1); es decir, $|f + g|^{p-1} \in L_{\frac{p}{p-1}}$. Se destaca que p y $\frac{p}{p-1}$ son índices conjugados pues

$$\frac{1}{p} + \frac{1}{p/_{p-1}} = \frac{1}{p} + \frac{p-1}{p} = \frac{1 + p - 1}{p} = \frac{p}{p} = 1,$$

por lo que se puede aplicar la desigualdad de Hölder (2.9) a $|f|$ y $|f + g|^{p-1}$

obteniendo

$$\begin{aligned}||f(f+g)^{p-1}|| &= \int |f||f+g|^{p-1}d\mu \\ &\leq \left(\int |f|^p d\mu\right)^{\frac{1}{p}} \left(\int (|f+g|^{p-1})^{p/_{p-1}} d\mu\right)^{\frac{1}{p/_{p-1}}} \\ &= ||f||_p \left(||f+g||_p\right)^{p-1}.\end{aligned}$$

De forma análoga, se aplica la desigualdad de Hölder (2.9) a $|g|$ y $|f+g|^{p-1}$ obteniendo

$$||g(f+g)^{p-1}|| \leq ||g||_p \left(||f+g||_p\right)^{p-1}.$$

Así que, integrando $|f+g|^p$ respecto a μ se tiene

$$\int |f+g|^p d\mu \leq \int |f(f+g)^{p-1}| d\mu + \int |g(f+g)^{p-1}| d\mu;$$

o dicho de otra forma,

$$\left(||f+g||_p\right)^p \leq |f||_p \left(||f+g||_p\right)^{p-1} + ||g||_p \left(||f+g||_p\right)^{p-1}.$$

Suponiendo que $||f+g||_p \neq 0$, se divide la desigualdad por $\left(||f+g||_p\right)^{p-1}$ y se obtiene

$$||f+g||_p \leq |f||_p + ||g||_p;$$

mostrando de esta forma que la desigualdad de Minkowski (2.10) se cumple. Se nota que si $||f+g||_p = 0$, la expresión se cumple trivialmente.

Ahora, se involucra el concepto de *convergencia*, el cual existen varios criterios (o modos) caracterizados por la función distancia, por lo cual se hablará de sucesiones de vectores en espacios métricos.

Definición 2.4.6. Sea $V_{(\mathbb{F})}$ un espacio vectorial con función distancia definida y sea $(\vec{v_n})$ una sucesión de vectores en este espacio. La sucesión se dice ser una ***sucesión de Cauchy*** si para todo $\epsilon > 0$ existe $N_\epsilon \in \mathbb{N}$ tal que si $m, n \geq N_\epsilon$ implica que $|\vec{v_m} - \vec{v_n}| < \epsilon$.

La sucesión es ***convergente a*** $\vec{v}$ ***en*** $V_{(\mathbb{F})}$ si para todo $\epsilon > 0$ existe $N_\epsilon \in \mathbb{N}$ tal que si $n \geq N_\epsilon$ implica que $|\vec{v} - \vec{v_n}| < \epsilon$. Este hecho se escribe como

$$\lim \vec{v_n} = \vec{v}.$$

Definición 2.4.7. Un espacio vectorial con función distancia definida se dice ***completo*** si toda sucesión de Cauchy converge a un elemento en el espacio.

Observación 2.4.6. Como se mencionó anteriormente, existen varias formas de convergencia por lo que en la Definición (2.4.6) se usó la notación del valor absoluto real, $|\cdot|$, aunque son vectores en el espacio; esto fue para indicar que se trata de la distancia entre los elementos.

Se construyen bases teóricas que dependen de la forma en que este definida la función distancia, tal es el caso en el que la distancia es inducida por la norma como se hizo en (2.7). Para este caso, se puede reformular la condición de una sucesión de Cauchy como sigue: una sucesión $(\vec{v_n})$ *es de Cauchy* si para todo $\epsilon > 0$ existe $N_\epsilon \in \mathbb{N}$ tal que si $m, n \geq N_\epsilon$ implica que

$$d(\vec{v_m}, \vec{v_n}) = ||\vec{v_n} - \vec{v_m}|| < \epsilon;$$

donde la distancia es inducida por la norma. De igual forma, se puede reformular la condición de una sucesión que converge a un elemento.

En algunos espacios, existen sucesiones que pueden converger a un elemento que no está en el espacio, pero cuando si pertenece a él se encuentran resultados importantes como la siguiente proposición.

Proposición 2.4.6. *Una sucesión convergente en un espacio métrico es una sucesión de Cauchy.*

Demostración. Sea $V_{(\mathbb{F})}$ un espacio vectorial con métrica d y sea $(\vec{v_n})$ una sucesión de vectores que converge a $\vec{v} \in V_{(\mathbb{F})}$; es decir, $\lim \vec{v_n} = \vec{v}$. Sea $\epsilon > 0$, entonces existe $N_\epsilon \in \mathbb{N}$ tal que si $n \geq N_\epsilon$ implica que

$$d(\vec{v_n}, \vec{v}) < \frac{\epsilon}{2};$$

por otro lado, si $m \geq N_\epsilon$ se tiene que

$$d(\vec{v_m}, \vec{v}) < \frac{\epsilon}{2}.$$

Luego, por el punto (iv) de la Definición (2.4.4), se observa que para dado $\epsilon > 0$ existe $N_\epsilon \in \mathbb{N}$ tal que si $n, m \geq N_\epsilon$ implica que

$$d(\vec{v_n}, \vec{v_m}) \leq d(\vec{v_n}, \vec{v}) + d(\vec{v}, \vec{v_m}) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Por lo tanto, una sucesión convergente en un espacio métrico es una sucesión de Cauchy. □

Definición 2.4.8 (Espacio de Banach). Un espacio vectorial normado y completo bajo la distancia definida por su norma es llamado ***espacio de Banach***.

Ejemplo 2.4.5. En esta ocasión, se mostrará que el espacio L_p es un espacio de Banach para $p \geq 1$. Ya se ha visto que L_p es un espacio vectorial normado bajo

$$||f||_p = \left(\int |f|^p d\mu\right)^{\frac{1}{p}},$$

por lo que falta ver que cada sucesión de Cauchy en L_p es una sucesión que converge a un elemento en L_p. Para verificar esta parte, supóngase que (f_n) es una sucesión de Cauchy en L_p y sea $\epsilon > 0$, entonces existe $N_\epsilon \in \mathbb{N}$ tal que si $n, m \geq N_\epsilon$ implica que

$$d(f_m, f_n) = \left(\int |f_n - f_m|^p d\mu\right)^{\frac{1}{p}} = ||f_n - f_m|| < \epsilon.$$

Se construye una subsucesión (h_n) de (f_n) de la siguiente manera: para cada $k \in \mathbb{N}$ existe un entero n_k tal que

$$||f_{n_k+1} - f_{n_k}|| < 2^{-k}$$

y se toma $h_k = f_{n_k}$, esto es posible bajo el argumento de que (f_n) es una sucesión de Cauchy. De esta forma, se tiene que para cada $k \in \mathbb{N}$

$$||h_{k+1} - h_k|| < 2^{-k}.$$

Ahora, se considera las funciones g_n definidas como

$$g_n(x) = |h_1(x)| + \sum_{k=1}^{n} |h_{k+1}(x) - h_k(x)|,$$

por lo que para cada entero n, $g_n \in M^+$. Luego, se halla el límite inferior de la sucesión (g_n), esto es

$$\begin{aligned} \text{lím inf}(g_n(x)) &= \sup_m \left(\inf_{n\geq m}\left(|h_1(x)| + \sum_{k=1}^{n} |h_{k+1}(x) - h_k(x)|\right)\right) \\ &= \sup_m \left(|h_1(x)| + \sum_{k=1}^{m} |h_{k+1}(x) - h_k(x)|\right) \\ &= |h_1(x)| + \sum_{k=1}^{\infty} |h_{k+1}(x) - h_k(x)|. \end{aligned}$$

Se denota a este límite como $g \in M^+$ y, elevando a la potencia p, se tiene que

$$(g(x))^p = \liminf (g_n(x))^p = \left(|h_1(x)| + \sum_{k=1}^{\infty} |h_{k+1}(x) - h_k(x)|\right)^p.$$

Ahora, por el Lema de Fatou (2.2.5), se obtiene

$$\int \liminf (g_n)^p d\mu \leq \liminf \left(\int (g_n)^p d\mu\right);$$

es decir,

$$\int |g|^p d\mu \leq \liminf \left(\int \left(|h_1| + \sum_{k=1}^{m} |h_{k+1} - h_k|\right)^p d\mu\right).$$

Se eleva esta última desigualdad a la potencia $\frac{1}{p}$ y se obtiene

$$\begin{aligned}\left(\int |g|^p d\mu\right)^{\frac{1}{p}} &\leq \liminf \left(\int \left(|h_1| + \sum_{k=1}^{n} |h_{k+1} - h_k|\right)^p d\mu\right)^{\frac{1}{p}} \\ &= \liminf \left(\left|\left||h_1| + \sum_{k=1}^{n} |h_{k+1} - h_k|\right|\right|\right) \\ &\leq \liminf \left(||h_1|| + \sum_{k=1}^{n} ||h_{k+1} - h_k||\right),\end{aligned}$$

esta última desigualdad es debido a que la propiedad *(3)* de la Definición (2.4.2).

Luego, como para cada entero k se cumple que $||h_{k+1} - h_k|| \leq 2^{-k}$ entonces

$$\sum_{k=1}^{\infty} ||h_{k+1} - h_k|| \leq \sum_{k=1}^{\infty} \frac{1}{2^k},$$

donde la suma del lado derecho converge a 1 (*suma geométrica*). Así, se tiene que

$$\liminf \left(||h_1|| + \sum_{k=1}^{n} ||h_{k+1} - h_k||\right) = ||h_1|| + \sum_{k=1}^{\infty} ||h_{k+1} - h_k|| \leq ||h_1|| + 1.$$

Finalmente, se concluye que

$$\left(\int |g|^p d\mu\right)^{\frac{1}{p}} \leq ||h_1|| + 1 < \infty,$$

por lo que $g \in L_p$.

Sea $M = \{x \in X \mid g(x) < \infty\}$, así que $\mu(X \setminus M) = 0$. Luego, se define a f como

$$f(x) = \begin{cases} h_1(x) + \sum_{k=1}^{\infty} (h_{k+1}(x) - h_k(x)), & \text{si } \ x \in M; \\ 0, & \text{si } \ x \notin M. \end{cases}$$

Es evidente que $|f| \leq g$ por lo que

$$\int |f|^p d\mu \leq \int (g)^p d\mu < \infty,$$

entonces $f \in L_p$. Por otro lado, se observa que f es el límite de (h_k) pues para cada $k \geq 2$ se puede expresar a h_k como

$$h_k = h_1 + \sum_{j=1}^{k-1} (h_{j+1} - h_j);$$

de aquí que $|h_k - f|^p$ converge a cero casi donde quiera. Aplicando de nuevo el Lema de Fatou (2.2.5), al integrar se obtiene

$$\int 0 d\mu \leq \liminf \left(\int |h_k - f|^p d\mu \right)$$

y también

$$\int 0 d\mu \geq -\liminf \left(\int |h_k - f|^p d\mu \right) = \limsup \left(\int |h_k - f|^p d\mu \right).$$

Es decir,

$$0 = \lim \left(\int |h_k - f|^p d\mu \right)$$

y elevando a la potencia $\frac{1}{p}$ se tiene que $0 = \lim ||f - h_k|| < \epsilon$, por lo que la sucesión h_k converge a f con respecto a la norma en L_p.

Luego, por hipótesis, para $m, n_k \geq N_\epsilon$ se tiene que

$$d(f_m, f_{n_k}) = d(f_m, h_k) = \left(\int |f_m - h_k|^p d\mu \right)^{\frac{1}{p}} < \epsilon;$$

se observa que, fijando el entero m, la sucesión $|f_m - h_k|$ converge a $|f_m - f|$, por lo que al aplicar una vez más el Lema de Fatou (2.2.5) sobre la variable k se obtiene

$$\left(\int |f_m - f|^p d\mu \right)^{\frac{1}{p}} \leq \liminf_k \left(\int |f_m - h_k|^p d\mu \right)^{\frac{1}{p}} < \epsilon.$$

En resumen, dado $\epsilon > 0$ existe N_ϵ tal que si $m \geq N_\epsilon$ implica que

$$d(f_m, f) = \left(\int |f_m - f|^p d\mu \right)^{\frac{1}{p}} < \epsilon,$$

con f en L_p definida anteriormente. De esta manera, se mostró que una sucesión de Cauchy es una sucesión convergente en L_p; por lo tanto, L_p ***es un espacio de Banach***.

Desde luego, existe interés de estudio cuando la norma del espacio está definida por el producto escalar.

Definición 2.4.9 (Espacio de Hilbert). Se llama ***espacio de Hilbert*** a un conjunto H si es un *espacio vectorial con producto interno* sobre $\mathbb{R}$ o $\mathbb{C}$, que además es *completo* por la norma definida por su producto interno.

Ejemplo 2.4.6. (*i*) **Espacio Euclideano n-dimensional:** Sea H_n el espacio constituido por las n-tuplas sobre $\mathbb{C}$ (o $\mathbb{R}$), es un espacio de Hilbert bajo el producto interno definido como

$$(x, y) = \sum_{i=1}^{n} x_i \bar{y}_i.$$

(*ii*) **Espacio de funcionales de cuadrado-sumables L_2:** consiste en las funciones tales que $\int_a^b |f(x)^2| dF(x) < \infty$, donde $F(x)$ es no decreciente y el producto interno es

$$(f, g) = \int_a^b f(x) g(\bar{x}) dF(x).$$

(*iii*) **Espacios estocásticos de Hilbert:** es el espacio de variables aleatorias de un espacio de probabilidad con segundo momento finito, es decir,

$$E[|x^2|] = \int_\Omega |f(\omega)|^2 m d\omega < \infty,$$

en donde la variable aleatoria tiene medida $m(d\omega)$ y $x = f(\cdot)$. En este espacio, dos variables aleatorias se dicen *iguales* si $E[|x - y|^2] = 0$. El producto interno se define como

$$(x, y) = E[x\bar{y}].$$

Se puede consultar detalles de este espacio en la referencia [10].

Capítulo 3

Construcción de un nuevo espacio a través de una medida de dispersión

A lo largo del capítulo, se tratará a Ω como el conjunto no vacío de los posibles resultados de un experimento aleatorio (*espacio muestral*), a $\mathscr{O}$ como la σ-álgebra que contiene a todos los subconjuntos de Ω y a P como una medida de probabilidad definida en $\mathscr{O}$; por lo que $(\Omega, \mathscr{O}, P)$ denotará un espacio de probabilidad. Se entenderá que las funciones $\mathscr{O}$-medibles pertenecen a $M(\Omega, \mathscr{O}, P)$, por lo que se limitará a escribir que la función pertenece a M, o a M^+ si la función es no negativa. De igual forma, se abreviará como L_p (con p un entero positivo) al espacio $L_p(\Omega, \mathscr{O}, P)$ que, por lo visto en el capítulo anterior, es el espacio de las clases de variables aleatorias P-equivalentes con esperanza finita de orden p.

El objetivo de este capítulo es construir un espacio de variables aleatorias que tenga las características específicas para ser estudiado a través de una medida de dispersión, por lo que se recurre a la teoría de probabilidad.

3.1. Conceptos básicos de la teoría de probabilidad

Se retoma las definiciones (2.1.5), (2.1.6), (2.3.2) y (2.3.3) escritas en el capítulo anterior. Con la intención de agilizar la escritura y lectura, cuando se escriba

dP_X se entenderá que la integral se realiza sobre $\mathbb{R}$ respecto a la medida de probabilidad P_X en $(\mathbb{R}, \mathscr{B})$, denotada por $E(X(\omega))$, y en caso de que se escriba dP se entenderá que la integral se realiza sobre Ω respecto a la medida de probabilidad P en $(\Omega, \mathscr{O})$, denotada por $E(X)$; recordando que ambas integrales tienen el mismo valor.

Como se puede observar, una variable aleatoria es integrable si y sólo si $E[|X|] < \infty$, no es diferente a la noción de esperanza en cualquier libro de probabilidad. De esta manera, $X \in L_p$ si y sólo si $E[|X|^p] < \infty$; además, es evidente que si $X \in L_p$ entonces $X + c \in L_p$ con c una constante real. De hecho, como la definición está en base al concepto de la integral, la esperanza cumple con las propiedades de operador lineal (Teorema (2.3.4)).

Proposición 3.1.1. *Si $C \in M$ es la función constante tal que $C(\omega) = c$ para $\omega \in \Omega$ donde c es un número real, entonces tiene esperanza definida y es igual al valor de la constante; esto es,*

$$E(C) = c.$$

Demostración. Sea c número real y sea $C(\omega) = c$ para todo $\omega \in \Omega$. Dado que P es medida de probabilidad, $P(\Omega) = 1$, entonces

$$E(C) = \int CdP = \int cdP = \sum cP(\Omega) = c \cdot 1 = c.$$

□

En la teoría de probabilidad, se tienen también los conceptos de varianza y covarianza, que son *medidas de dispersión* para las variables aleatorias y se definen de la siguiente forma.

Definición 3.1.1. Sean X y Y dos variables aleatorias en L_2, se define:

1. ***Covarianza de*** X ***y*** Y como $Cov(X, Y) = E\left[(X - E[X])(Y - E[Y])\right]$.

2. ***Varianza de*** X como $Var[X] = E\left[(X - E[X])^2\right]$.

3. ***Desviación estándar de*** X como $DE[X] = \sqrt{Var[X]}$.

Hay que aclarar que, en términos de las medidas de probabilidad en $(\mathbb{R}, \mathscr{B})$, la esperanza del producto de dos variables aleatorias X y Y es igual a la integral sobre $\mathbb{R}$ respecto a la distribución P_{XY}; esto es,

$$E[(XY)(\omega)] = E[X(\omega)Y(\omega)] = \int_{\mathbb{R}} X(\omega)Y(\omega)dP_{XY} = \int_{\Omega} XY dP = E[XY].$$

Se observa que la covarianza y la varianza exigen que la integral de un producto de funciones sea finita, esto se argumenta por la Propocisión (2.4.1). Es decir, estas medidas de dispersión están definidas en $L_p \times L_q$ donde p y q son índices conjugados. Nuestro caso de interés es L_2, por lo que en particular se pide en este trabajo que las variables aleatorias pertenezcan a L_2 y así asegurar que el producto de elementos en el espacio sea integrable; de esta forma, se tiene

$$\begin{aligned} Cov(\cdot,\cdot) :& L_2 \times L_2 \to \mathbb{R} \\ (X,Y) \to & \int (X - E(X))(Y - E(Y))dP. \end{aligned}$$

De forma similar se puede expresar la varianza como una función de $L_2 \times L_2$ a $\mathbb{R}$. Estos conceptos se definen a través de la esperanza, así que usando propiedades de la esperanza se puede obtener una forma equivalente para estos que resulta más fácil de manipular en situaciones prácticas.

$$\begin{aligned} E\left[(X - E[X])(Y - E[Y])\right] &= E[XY - E[Y]X - E[X]Y + E[X]E[Y] \\ &= E[XY] - E\left[E[X]Y\right] - E\left[E[Y]X\right] + E\left[E[X]E[Y]\right] \\ &= E[XY] - E[X]E\left[Y\right] - E[Y]E\left[X\right] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y]. \end{aligned}$$

Por lo que

$$Cov(X,Y) = E[XY] - E[X]E[Y].$$

Cabe mencionar que $Var[X] = Cov(X,X)$ por lo que

$$Var[X] = E[X^2] - E^2[X].$$

Aún más, se puede mostrar que la varianza de una variable aleatoria es 0 si y sólo si es una variable aleatoria constante, la cual es el valor esperado de esta. Una variable aleatoria con estas características se le llama ***variable aleatoria degenerada***. Para mostrar esto, se hace uso del siguiente resultado.

Teorema 3.1.2 (Desigualdad de Chebyshev). *Sea X una variable aleatoria con esperanza finita y sea $c \in \mathbb{R}$. Entonces, si $E(X-c)^2$ es finito y dado $\epsilon > 0$ se tiene que*

$$P\left[\omega \in \Omega \mid |X(\omega) - c| \geq \epsilon\right] \leq \frac{1}{\epsilon^2} E\left(X - c\right)^2;$$

o de forma equivalente,

$$P\left[\omega \in \Omega \mid |X(\omega) - c| < \epsilon\right] \geq 1 - \frac{1}{\epsilon^2} E\left(X - c\right)^2.$$

Demostración. (Ver referencia [8]). □

En particular, si $\mu = E(X)$, de la desigualdad de chebyshev se encuentra que

$$P\left[\omega \in \Omega \mid |X(\omega) - \mu| \geq \epsilon\right] \leq \frac{1}{\epsilon^2} E\left(X - \mu\right)^2 = \frac{Var(X)}{\epsilon^2}. \tag{3.1}$$

Por lo que si $Var(X) = 0$, se tiene que para cualquier $\epsilon > 0$

$$P\left[\omega \in \Omega \mid |X(\omega) - \mu| \geq \epsilon\right] \leq \frac{Var(X)}{\epsilon^2} = \frac{0}{\epsilon^2} = 0;$$

o equivalentemente,

$$P\left[\omega \in \Omega \mid |X(\omega) - \mu| < \epsilon\right] \geq 1 - \frac{Var(X)}{\epsilon^2} = 1.$$

Por lo que $X(\omega) = \mu$ para todo $\omega \in \Omega$ respecto a la medida P y es una variable aleatoria degenerada. Ahora bien, si X es variable aleatoria constante c,

$$Var(X) = E(c)^2 - E^2(c) = c^2 - (c)^2 = 0;$$

por lo que su varianza es cero y c es el valor esperado de X. De esta forma, se ha probado el siguiente resultado.

Teorema 3.1.3. *Sea X una variable aleatoria, entonces X es una variable aleatoria degenerada con $X(\omega) = E(X)$ si y sólo si $Var(X) = 0$.*

3.2. Construcción de un espacio normado a través de una medida de dispersión

Ahora, se construye un nuevo espacio a partir del espacio L_2 y usando la covarianza como producto interno, por lo que se recurrirá al Lema (2.4.5).

Proposición 3.2.1. *El espacio de variables aleatorias degeneradas (constantes) es un subespacio vectorial de L_p. A este espacio se le denotará como $L_D(\Omega, \mathscr{O}, P)$ o simplemente L_D.*

Demostración. Sea $A \in L_D$ y sea $p \in \mathbb{N}$, entonces $A^p(\omega) = a^p$ para todo $\omega \in \Omega$, con $a \in \mathbb{R}$, por lo que A^p es una función constante. Por la Proposición (3.1.1), se tiene que

$$\int |A^p| dP = |a^p| < \infty;$$

por lo que L_D es un subconjunto de L_p para cualquier entero positivo p.

Sea $B \in L_D$, entonces $B(\omega) = b$ para todo $\omega \in \Omega$. Sea $\alpha \in \mathbb{R}$, entonces αB es constante pues

$$(\alpha B)(\omega) = \alpha B(\omega) = \alpha b$$

para todo $\omega \in \Omega$. Luego $A(\omega) = a$ para todo $\omega \in \Omega$, entonces

$$(A + \alpha B)(\omega) = A(\omega) + \alpha B(\omega) = a + \alpha b$$

para todo $\omega \in \Omega$; por lo que $(A + \alpha B) \in L_D$. Es decir, la suma y producto escalar son operaciones cerradas en este conjunto.

Luego, es claro que no es vacío pues $\vec{0} \in L_D$ donde $\vec{0}(\omega) = 0$ para todo $\omega \in \Omega$; por lo tanto, L_D es un subespacio vectorial de L_p para cualquier $p \in \mathbb{N}$. □

Existen propiedades de interés en las medidas de dispersión y, siendo la varianza definida a través de la covarianza, se concentrará en la función covarianza bajo el conocimiento de que las propiedades se heredarán a la varianza.

Proposición 3.2.2. *Sean X, Y y Z variables aleatorias en L_2 y $\alpha \in \mathbb{R}$. Entonces:*

1. $Cov(X, Y) = Cov(Y, X)$.

2. $Cov(X, Y + Z) = Cov(X, Y) + Cov(X, Z)$.

3. $Cov(\alpha X, Y) = \alpha Cov(X, Y)$.

4. $Cov(X, C) = 0$ *con* $C \in L_D$.

5. $Cov(X, X) > 0$ *si* $X \notin L_D$ *y* $Cov(X, X) = 0$ *si y sólo si* $X \in L_D$.

Demostración. Se mostrará cada uno de los postulados apoyándose de la segunda forma presentada de la covarianza.

1. Usando la propiedad abeliana de las funciones reales:

$$Cov(X,Y) = E[XY] - E[X]E[Y] = E[YX] - E[Y]E[X] = Cov(Y,X).$$

2. Usando la propiedad distributiva de las funciones reales y la linealidad de la esperanza (integral):

$$\begin{aligned} Cov(X,Y+Z) &= E[X(Y+Z)] - E[X]E[Y+Z] \\ &= E[XY+XZ] - E[X]\left(E[Y]+E[Z]\right) \\ &= E[XY] + E[XZ] - E[X]E[Y] - E[X]E[Z] \\ &= E[XY] - E[X]E[Y] + E[XZ] - E[X]E[Z] \\ &= Cov(X,Y) + Cov(X,Z). \end{aligned}$$

3. Usando de nuevo la linealidad de la esperanza:

$$\begin{aligned} Cov(\alpha X,Y) &= E[\alpha XY] - E[\alpha X]E[Y] = \alpha E[XY] - \alpha E[X]E[Y] \\ &= \alpha\left(E[XY] - E[X]E[Y]\right) = \alpha Cov(X,Y). \end{aligned}$$

4. $Cov(X,C) = E[XC] - E[X]E[C] = cE[X] - cE[X] = 0.$

5. Se ha visto que $Cov(X,X) = 0$ si y sólo si X es degenerada. Luego, la covarianza es una función positiva para $X \notin L_D$,

$$Cov(X,X) = Var[X] = \int (X - E[X])^2 dP > 0.$$

□

Comparando con la definición de producto interno (2.4.3), se nota que la covarianza es casi un producto interno en L_2, la complicación es la última propiedad en que el producto interno es cero si y sólo si la variable aleatoria es degenerada.

Así, la covarianza es propuesta para ser producto interno en un espacio donde las variables aleatorias degeneradas sean el elemento neutro.

Definición 3.2.1. Se nombrará el ***espacio de clases de variables aleatorias integrables de orden 2 P-equivalentes salvo constante*** al espacio vectorial cociente ${}^{L_2}/_{L_D}$ y se denotará como D_2. Un elemento en D_2 se denota como $\overline{X}$ y representa la clase de todas las variables aleatorias en L_2 iguales P-c.s. a la variable aleatoria X salvo constante, esto es

$$\overline{X} = \{Y \in L_2 \mid Y = X + c, \text{ con } c \text{ una constante}\}.$$

Esta definición se logra gracias al Lema (2.4.5), lo que implica que los elementos en D_2 son iguales a excepción de una constante. Esto es, si $\overline{X}, \overline{Y} \in D_2$, entonces $\overline{X} = \overline{Y}$ (o bien $X + L_D = Y + L_D$) si y sólo si $X - Y \in L_D$. Así que, $X - Y = C$ para alguna variable aleatoria degenerada C. Luego,

$$\begin{aligned}
\overline{X+Y} &= \{f \in L_2 \mid f = X + Y + (a+b), \text{ con } a \text{ y } b \text{ constantes}\} \\
&= \{f_1 + f_2 \in L_2 \mid f_1 = X + a \ \text{ y } \ f_2 = Y + b\} \\
&= \{f_1 \in L_2 \mid f_1 = X + a\} + \{f_2 \in L_2 \mid f_2 = Y + b\} \\
&= \overline{X} + \overline{Y}.
\end{aligned}$$

De igual forma

$$\begin{aligned}
\overline{\alpha X} &= \{f \in L_2 \mid f = \alpha X + c, \text{ con } c \text{ una constante}\} \\
&= \left\{f \in L_2 \mid f = \alpha \left(X + \frac{c}{\alpha}\right), \text{ con } c \text{ una constante}\right\} \\
&= \{\alpha g \in L_2 \mid g = X + c' \text{ con } c' \text{ una constante}\} \\
&= \alpha\{g \in L_2 \mid g = X + c' \text{ con } c' \text{ una constante}\} \\
&= \alpha\overline{X}.
\end{aligned}$$

A continuación, se presentan algunas observaciones inmediatas que surgen de este espacio.

1. Si dos clases $\overline{X}$ y $\overline{Y}$ son iguales implica que para toda función X y Y en estas clases, su diferencia $X - Y$ pertenece a L_D; es decir, $X - Y = c$ con c alguna constante real.

2. Para cada clase en D_2 existe una única función tal que su integral es cero. En efecto, pues si $X \in \overline{X}$ es tal que $E(X) = \alpha \in \mathbb{R}$, entonces $(X - \alpha) \in \overline{X}$

es tal que $E(X-\alpha)=0$. Luego, si $E(X_1)=E(X_2)$ para $X_1, X_2 \in \overline{X}$ entonces $X_1 - X_2 = c$ y se tiene que

$$c = E(X_1 - X_2) = E(X_1) - E(X_2) = 0,$$

por lo que $c=0$ y así $X_1 = X_2$.

3. El elemento cero en el espacio D_2, que se denotará como $\overline{0}$, es la clase de variables aleatorias degeneradas (L_D).

4. Como se mencionó anteriormente, $Cov(X,X)=0$ si y sólo si X es una variable aleatoria degenerada, por lo que la función covarianza es un producto interno en D_2.

5. Para cualesquiera clases $\overline{X}$ y $\overline{Y}$, la covarianza de sus elementos son iguales. En efecto, pues si consideramos como "representantes" de clase a X_0 y a Y_0 respectivamente, entonces para $X_0 + c = X \in \overline{X}$ y $Y_0 + d = Y \in \overline{Y}$, con c y d constantes reales, se tiene que

$$\begin{aligned} Cov(X,Y) &= Cov(X_0 + c, Y_0 + d) = Cov(X_0, Y_0 + d) + Cov(c, Y_0 + d) \\ &= Cov(X_0, Y_0) + Cov(X_0, d) + 0 = Cov(X_0, Y_0) + 0 \\ &= Cov(X_0, Y_0). \end{aligned}$$

Como una clase tienen varios elementos, la clase no tiene notación única pues si $X, Y \in \overline{X}$ entonces $\overline{X} = \overline{Y}$, por lo que es conveniente tener una forma única de denotar la clase cuando sea posible para evitar confusiones.

Existen varias formas de definir un **representante de clase** para poder describir cualquier elemento en términos del representante, como puede ser aquella función cuya integral es cero. Por esta razón, es adecuado hacer la primera definición en el espacio D_2.

Definición 3.2.2. Sea una clase $\overline{X} \in D_2$, se define el ***representante de clase de*** $\overline{X}$ como la variable aleatoria $X_0 \in \overline{X}$ tal que

$$\int X_0 d\mu = 0.$$

Así, la forma adecuada de escribir la clase será $\overline{X_0}$.

Nota 3.2.1. *En adelante, para agilizar la lectura y escritura, cuando se hable de una clase $\overline{X} \in D_2$ se entenderá que el* ***representante de la clase*** *es $X \in L_2$ y que su integral es cero.*

Como para cualesquiera elementos de dos clases en D_2 la covarianza es el mismo valor, se da la siguiente definición.

Definición 3.2.3. Sean $\overline{X}, \overline{Y} \in D_2$ (sus representantes son X y Y), se define la ***covarianza de las clases*** $\overline{X}$ ***y*** $\overline{Y}$ como

$$Cov(\overline{X}, \overline{Y}) = Cov(X, Y).$$

De igual forma, se define la ***varianza de la clase*** $\overline{X}$ como

$$Var(\overline{X}) = Var(X).$$

Se observa que $Cov(\overline{X}, \overline{X}) = 0$ si y sólo si $\overline{X} = L_D$. De esta manera, la covarianza no es producto interno en L_2 pero si lo es para D_2 y, apoyándose de la ecuación (2.6), se aprovechará para inducir una norma y, por consecuencia, una distancia entre clases. Es decir, dado $\overline{X} \in D_2$ se expresa la **norma de** $\overline{X}$ como

$$||\overline{X}|| = Cov(\overline{X}, \overline{X})^{\frac{1}{2}} = Cov(X, X)^{\frac{1}{2}};$$

o bien,

$$||\overline{X}|| = Var(\overline{X})^{\frac{1}{2}} = Var(X)^{\frac{1}{2}} = DE(X).$$

Luego, si $\overline{X}, \overline{Y} \in D_2$ se expresa la **distancia entre** $\overline{Y}$ **y** $\overline{X}$ como

$$d(\overline{Y}, \overline{X}) = Cov(\overline{X} - \overline{Y}, \overline{X} - \overline{Y})^{\frac{1}{2}} = Cov(X - Y, X - Y)^{\frac{1}{2}};$$

o bien,

$$d(\overline{Y}, \overline{X}) = Var(\overline{X} - \overline{Y})^{\frac{1}{2}} = Var(X - Y)^{\frac{1}{2}} = DE(X - Y).$$

De esta manera, se enuncia la siguiente proposición.

Proposición 3.2.3. *El espacio de las clases de variables aleatorias con esperanza finita de orden 2 P-equivalentes salvo constantes, D_2, es un espacio vectorial normado por*

$$||\overline{X}|| = Cov(\overline{X}, \overline{X})^{\frac{1}{2}},$$

donde $Cov(\cdot, \cdot)$ es el producto interno definido.

En consecuencia, si una sucesión $\overline{X_n}$ en D_2 converge a una clase $\overline{X}$, entonces para un entero n suficientemente grande se tiene que $\overline{X_n} - \overline{X}$ converge a la clase L_D. Dicho en términos del espacio L_2, para un entero n suficientemente grande las diferencias de los elementos de $\overline{X_n}$ y los elementos de $\overline{X}$ se asemejan a una variable aleatoria degenerada.

Una vez dicho esto, se menciona los primeros resultados importantes encontrados sobre este nuevo espacio.

Proposición 3.2.4. *Sea una sucesión de variables aleatorias (X_n) en L_2. Si (X_n) es una sucesión de Cauchy, entonces la sucesión $\left(\overline{Y}_n\right)$ en D_2 es una suceción de Cauchy, donde $\overline{Y_n}$ es la clase de variables aleatorias en L_2 $P-$equivalentes de X_n salvo constante.*

Demostración. Sea $\epsilon > 0$, como (X_n) es sucesión de Cauchy, entonces existe $N_\epsilon \in \mathbb{N}$ tal que para todo $n, m \geq N_\epsilon$ se tiene que

$$\left(\int (X_n - X_m)^2 \, dP\right)^{\frac{1}{2}} < \epsilon;$$

o bien,

$$\int (X_n - X_m)^2 \, dP < \epsilon^2. \tag{3.2}$$

Por otro lado, la varianza es una función no negativa pues $(X - E[X])^2 \geq 0$, por lo que

$$0 \leq \int (X - E[X])^2 dP = Var(X);$$

o mejor dicho,

$$0 \leq Var(X) = E[X^2] - E^2(X) = \int X^2 dP - \left(\int X dP\right)^2.$$

De esta forma, se tiene que

$$0 \leq \int (X_n - X_m - E(X_n - X_m))^2 \, dP = Var(X_n - X_m); \tag{3.3}$$

o bien,

$$0 \leq \int (X_n - X_m)^2 dP - \left(\int (X_n - X_m) dP\right)^2 = Var(X_n - X_m). \tag{3.4}$$

Luego, de las expresiones (3.2) y (3.4) se obtiene

$$0 \leq \int (X_n - X_m)^2 dP - \left(\int (X_n - X_m) dP \right)^2 < \epsilon^2 - \left(\int (X_n - X_m) dP \right)^2 \leq \epsilon^2; \tag{3.5}$$

es decir, de forma resumida, se obtiene

$$0 \leq Var(X_n - X_m) < \epsilon^2. \tag{3.6}$$

Ahora, considerando la clase de variables aleatorias P-equivalentes salvo constante para cada X_n se obtiene la sucesión $(\overline{Y}_n)$ en D_2, donde cada Y_n es el representante de clase para cada X_n. Así, se tiene que

$$Var(\overline{Y_n} - \overline{Y_m}) = Var(Y_n - Y_m) = Var(X_n - X_m).$$

Sustituyendo en (3.6) y como $Var(X) = Cov(X, X)$, así $Var(\overline{X}) = Cov(\overline{X}, \overline{X})$, se tiene que

$$0 \leq Var(\overline{Y_n} - \overline{Y_m}) = Cov(\overline{Y_n} - \overline{Y_m}, \overline{Y_n} - \overline{Y_m}) < \epsilon^2. \tag{3.7}$$

Finalmente, elevando a la potencia $\frac{1}{2}$ la expresión (3.7), se concluye que si $\epsilon > 0$ entonces existe $N_\epsilon \in \mathbb{N}$ tal que para todo $n, m \geq N_\epsilon$ se cumple que

$$0 \leq ||\overline{Y_n} - \overline{Y_m}|| = Cov(\overline{Y_n} - \overline{Y_m}, \overline{Y_n} - \overline{Y_m})^{\frac{1}{2}} < \epsilon. \tag{3.8}$$

Por lo tanto, $(\overline{Y_n})$ es una sucesión de Cauchy en D_2. □

Proposición 3.2.5. *Sea una sucesión de variables aleatorias (X_n) en L_2. Si (X_n) converge a un elemento X de L_2, entonces la sucesión $(\overline{Y}_n)$ en D_2 converge a $\overline{Y}$ en D_2, donde $\overline{Y_n}$ es la clase de variables aleatorias P-equivalentes de X_n salvo constante y $\overline{Y}$ es la clase de variables aleatorias P-equivalentes de X salvo constante.*

Demostración. La demostración es similar a la demostración de la Proposición (3.2.4), por lo que se omitirán algunos detalles.

Sea $\epsilon > 0$, como (X_n) converge a X, entonces existe $N_\epsilon \in \mathbb{N}$ tal que para todo $n \geq N_\epsilon$ se tiene que

$$\left(\int (X_n - X)^2 \, dP \right)^{\frac{1}{2}} < \epsilon;$$

o bien,

$$\int (X_n - X)^2 \, dP < \epsilon^2. \tag{3.9}$$

Por otro lado, como X y X_n son integrables de orden 2, entonces $X_n - X$ también es integrable de orden 2, por lo que su varianza existe y así

$$0 \leq \int (X_n - X - E(X_n - X))^2 \, dP = Var(X_n - X);$$

o bien,

$$0 \leq \int (X_n - X)^2 dP - \left(\int (X_n - X) dP \right)^2 = Var(X_n - X). \tag{3.10}$$

Luego, de la expresión (3.9) y (3.10) se obtiene

$$0 \leq Var(X_n - X) = \int (X_n - X)^2 dP - \left(\int (X_n - X) dP \right)^2 < \epsilon^2. \tag{3.11}$$

Ahora, considerando la clase de variables aleatorias P-equivalentes salvo constante para cada X_n se obtiene la sucesión $(\overline{Y}_n)$ en D_2, donde cada Y_n es el representante de clase para cada X_n; de igual forma $\overline{Y}$ es la clase de X en D_2 y Y es el representante. De esta manera, se tiene que

$$Var(\overline{Y_n} - \overline{Y}) = Var(Y_n - Y) = Var(X_n - X).$$

Dicho lo anterior y de la expresión (3.11), se tiene que

$$0 \leq Var(\overline{Y_n} - \overline{Y}) = Cov(\overline{Y_n} - \overline{Y}, \overline{Y_n} - \overline{Y}) < \epsilon^2. \tag{3.12}$$

Finalmente, elevando a la potencia $\frac{1}{2}$ la expresión (3.12), se concluye que si $\epsilon > 0$ entonces existe $N_\epsilon \in \mathbb{N}$ tal que para todo $n \geq N_\epsilon$ se cumple que

$$0 \leq ||\overline{Y_n} - \overline{Y}|| = Cov(\overline{Y_n} - \overline{Y}, \overline{Y_n} - \overline{Y})^{\frac{1}{2}} < \epsilon. \tag{3.13}$$

Por lo tanto, $(\overline{Y_n})$ converge a $\overline{Y}$ en D_2. □

En el capítulo anterior, se mostró que el espacio L_p es un *espacio de Banach*; en particular, sucede que L_2 es un espacio normado completo. Esto implica que, en base a la Proposición (3.2.4) y la Proposición (3.2.5), **si X_n es una sucesión de Cauchy en L_2 entonces converge a algún elemento X en L_2, por lo**

tanto $\overline{Y_n}$ es una suceción de Cauchy en D_2 que converge a $\overline{Y}$ en D_2, donde $\overline{Y_n}$ y $\overline{Y}$ son las clases de X_n y X en D_2 respectivamente.

Esto **no** significa, de ninguna manera, que el espacio D_2 sea completo. Lo que se probado hasta el momento es una muestra de como los resultados que se conocen de L_2 tienen influencia en el nuevo espacio D_2 y se pueden generar resultados similares, por lo que se puede aplicar la teoría obtenida acerca de L_2 y generar conocimientos sobre D_2. Sin embargo, existe el interés sobre la *completes* del espacio D_2, por lo que existe la alternativa de estudiar directamente el espacio y a partir de ahí verificar su influencia sobre L_2.

Proposición 3.2.6. *Sea $(\overline{X_n})$ una sucesión en D_2. Si $(\overline{X_n})$ es una sucesión de Cauchy, entonces la sucesión de sus representantes (X_n) es una sucesión de Cauchy en L_2.*

Demostración. Sea $\epsilon > 0$, como $(\overline{X_n})$ es una sucesión de Cauchy, entonces existe $N_\epsilon \in \mathbb{N}$ tal que para todo $n, m \geq N_\epsilon$ se tiene que

$$||\overline{X_n} - \overline{X_m}|| = Cov(\overline{X_n} - \overline{X_m}, \overline{X_n} - \overline{X_m})^{\frac{1}{2}} < \epsilon;$$

dicho de otra forma

$$Var(\overline{X_n} - \overline{X_m}) < \epsilon^2. \tag{3.14}$$

Luego, como $Var(\overline{X_n} - \overline{X_m}) = Var(X_n - X_m)$ se tiene que

$$0 \leq Var(X_n - X_m) = \int (X_n - X_m - E(X_n - X_m))^2 dP < \epsilon^2. \tag{3.15}$$

Por la Defnición (3.2.2), el representante es tal que su valor esperado es cero por lo que

$$E(X_n - X_m) = E(X_n) - E(X_m) = 0 - 0 = 0;$$

por lo que la expresión (3.15) se resume a

$$0 \leq \int (X_n - X_m)^2 dP < \epsilon^2. \tag{3.16}$$

Finalmente, elevando a la potencia $\frac{1}{2}$ la expresión (3.16), se concluye que si $\epsilon > 0$ entonces existe $N_\epsilon \in \mathbb{N}$ tal que para todo $n, m \geq N_\epsilon$ se cumple que

$$0 \leq ||X_n - X_m||_{L_2} = \left(\int (X_n - X_m)^2 dP \right)^{\frac{1}{2}} < \epsilon. \tag{3.17}$$

Por lo tanto, la sucesión de los representantes de clase (X_n) en L_2 es una sucesión de Cauchy. □

Esta proposición muestra que toda sucesión de Cauchy $(\overline{X_n})$ en D_2 implica que la sucesión de los representantes (X_n) es una sucesión de Cauchy en L_2 y, como el espacio L_2 es completo, esta sucesión converge a un elemento X de L_2. Por la Proposición (3.2.5), como la sucesión de los representantes en L_2 converge a un elemento en L_2, la sucesión de las clases de estos representantes de clase, $(\overline{X_n})$, converge a la clase $\overline{X}$ en D_2.

Esto implica que D_2 ***es un espacio completo*** pues toda sucesión de Cauchy en D_2 converge a un elemento del mismo espacio. Aún más, como la norma en D_2 está definida a través de la covarianza que es un producto interno, se tiene que D_2 ***es un espacio métrico normado por su producto interno***; esto es, D_2 ***es un espacio de Hilbert***. De esta manera, se ha probado el siguiente resultado el cual servirá para culminar el presente capítulo.

Teorema 3.2.7. *El espacio de clases de variables aleatorias integrables de orden 2 P-equivalentes salvo constantes, D_2, es un espacio de Hilbert.*

Conclusión

La covarianza resultó ser un buen candidato para crear un espacio normado pues, además de que D_2 resulto ser normado y completo, vista como función es menor o igual que la norma definida en L_2, por lo que podría resultar que es un criterio más potente de convergencia en el sentido de que si una sucesión de variables aleatorias converge en base la covarianza implica que converge bajo la norma de L_2; por esta razón, se seguirá explorando el alcanze de la función covarianza vista como criterio de convergencia.

Usualmente, se estudia el espacio de variables aleatorias como funciones medibles y con la norma definida se obtiene información de la distribución que tiene y así, como consecuencia, generar información de sus medidas de dispersión. Con el nuevo espacio obtenido, se puede estudiar las variables aleatorias relacionadas a un espacio muestral como las clases de variables aleatorias equivalentes salvo constantes a través de las medidas de dispersión y, como consecuencia, generar información de las variables aleatorias como función medible. Es decir, como la varianza es menor o igual que la norma definida en L_2, todos los conocimientos que se desarrollen en D_2 pueden tener implicaciones en L_2 y así es posible estudiar las variables aleatorias desde otro enfoque distinto al usual en donde los conocimientos desarrollados en L_2 tienen implicaciones en las medidas de dispersión.

El Teorema (3.2.5) muestra que dada una sucesión de Cauchy en D_2 implica que la sucesión de representantes es de Cauchy en L_2; sin embargo, implica la existencia de una gran cantidad de sucesiones de Cauchy en L_2, basta fijar una constante real y construir la sucesión de variables aleatorias cuya esperanza es igual a esta constante y que sean pertenecientes a las clases de la sucesión original en D_2. Esto es una muestra de como los resultados en D_2 pueden extenderse y desarrollar conocimientos en L_2.

Aún más, el presente trabajo muestra poca información del nuevo espacio por lo que queda abierto a seguir desarrollando conocimientos a través de la teoría conocida, incluyendo la teoría que no fue mencionada aquí, y ver sus implicaciones sobre L_2. De esta forma, se concluye el proyecto de tesis proponiendo *el espacio de las clases de variables aleatorias con esperanza finita de orden 2 P-equivalentes salvo constante* como un enfoque alternativo de estudio de la teoría de probabilidad.

Bibliografía

[1] Bartle, R. G. (1966). *"The Elements of Integration and Lebesgue Measure"*. (ed. Wiley Classics Library). New York: John Wiley & Sons, Inc.

[2] Billingsley, P. (1999). *"Convergence of Probability and Measures"*. (2^a ed.). New York: John Wiley & Sons, Inc.

[3] Billingsley, P. (1995). *"Probability and Measure"*. (3^a ed.). New York: John Wiley & Sons, Inc.

[4] González, E. & Speziale, L. (1985). *"Apuntes de Álgebra Lineal"*. México: División de Ciencias Básicas UNAM.

[5] Haaser, N. B. & Sullivan, J. A. (1971). *"Real Analysis"*. New York: Van Nostrand Reinhold Company.

[6] Herstein, I. N. (1975). *"Topics In Algebra"*. (2^a ed.). New York: John Wiley & Sons, Inc.

[7] Lang, S. (1970). *"Introduction to Linear Algebra"*. (2^a ed.). New York: Springer.

[8] Meyer, P. (1970). *"Introductory Probability and Statistical Applications"*. (2^a ed.). Massachusetts: Addison-Wesley Publishing Company.

[9] Strang, G. (1988). *"Linear Algebra and Its Applications"*. (3^a ed.). Connecticut: Thomson Learning, Inc.

[10] Urbelz Ibarrola, F. J. (1981). *"Aplicaciones del Espacio de Hilbert a la Estadística"*. Revista Estadística Española, Núm. 91, 23-52.