Universidad de San Carlos de Guatemala
Escuela de Ciencias Físicas y Matemáticas
Departamento de Matemática

GEOMETRÍA DE UN ESPACIO DE HILBERT APLICADA A LAS SERIES DE TIEMPO

Joel Armando Juárez Martínez

Asesorado por Lic. William Roberto Gutierrez Herrera

Guatemala, marzo 2020

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

ESCUELA DE CIENCIAS FÍSICAS Y MATEMÁTICAS

GEOMETRÍA DE UN ESPACIO DE HILBERT APLICADA A LAS SERIES DE TIEMPO

TRABAJO DE GRADUACIÓN
PRESENTADO A LA JEFATURA DEL
DEPARTAMENTO DE MATEMÁTICA
POR

JOEL ARMANDO JUÁREZ MARTÍNEZ

ASESORADO POR LIC. WILLIAM ROBERTO GUTIERREZ HERRERA

AL CONFERÍRSELE EL TÍTULO DE

LICENCIADO EN MATEMÁTICA APLICADA

GUATEMALA, MARZO 2020

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

ESCUELA DE CIENCIAS FÍSICAS Y MATEMÁTICAS

CONSEJO DIRECTIVO

DIRECTOR	M.Sc. Jorge Marcelo Ixquiac Cabrera
SECRETARIO ACADÉMICO	M.Sc. Edgar Anibal Cifuentes Anleu

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

EXAMINADOR	Lic. William Roberto Gutierrez Herrera
EXAMINADOR	Lic. Rubén Darío Narciso Cruz
EXAMINADOR	Lic. Frank Jorge Frietzsche Barrios

Escuela de Ciencias Físicas y Matemáticas

Ref. D.DTG. 002-2020
Guatemala 09 de marzo de 2020

El Director de la Escuela de Ciencias Físicas y Matemáticas de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Coordinador de la Licenciatura en Matemática Aplicada, al trabajo de graduación Titulado: **"GEOMETRÍA DE UN ESPACIO DE HILBERT APLICADA A LAS SERIES DE TIEMPO"** presentado por el estudiante universitario **Joel Armando Juárez Martínez,** autoriza la impresión del mismo.

IMPRÍMASE.

"ID Y ENSEÑAD A TODOS"

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
ESCUELA DE CIENCIAS FÍSICAS Y MATEMÁTICAS
-DIRECTOR-
ECFM - USAC

M.Sc. Jorge Marcelo Ixquiac Cabrera
Director

201213274
JMIC/Paola

AGRADECIMIENTOS

A mis padres, por haberme brindado la oportunidad de estudiar y por sus consejas para llegar a ser una mejor persona.

A mis hermanos, por haberme apoyado y motivo a salir adelante.

A María José Maldonado, por compartir conmigo en todo momento y por sus grandes consejos.

A mis amigos, por compartir momentos de estudio y diversión en la carrera.

A mi asesor, Lic. Gutierrez por darme apoyo y paciencia en la elaboración de este trabajo.

A mis profesores, por impartir de la mejor manera todo su conocimiento.

A la USAC, por ser una gran casa de estudios.

DEDICATORIA

A mis padres Rafael Juárez y María Martínez, a mis hermanos Rafael, Daniel y Fabiola, a mi novia María José y a todo aquel que lea este trabajo.

ÍNDICE GENERAL

LISTA DE SÍMBOLOS

Símbolo	Significado
$\{E_n\}$	sucesión
$E_n\uparrow$	sucesión monótona creciente
$E_n\downarrow$	sucesión monótona decreciente
E^*	límite superior de la sucesión $\{E_n\}$
E_*	límite inferior de la sucesión $\{E_n\}$
$\subset$	subconjunto de
$\leq$	menor o igual a
$\cap$	intersección entre conjuntos
$\cup$	unión entre conjuntos
$\mathbf{R}$	anillo
$E \setminus F$	diferencia entre E y F
$E\Delta F$	diferencia simétrica entre E y F
$\in$	pertenece a
E^c	complemento de E
$\mathbf{E}$	clase de conjuntos
$\mathbf{R}(\mathbf{E})$	anillo generado por $\mathbf{E}$
$\mathbf{S}$	σ-anillo
$\mathbf{S}(\mathbf{E})$	σ-anillo generado por $\mathbf{E}$
μ	medida
μ^*	medida externa
$\mathbf{H}(\mathbf{E})$	σ-anillo hereditario generado por $\mathbf{E}$
$\varnothing$	conjunto vacío
χ_E	función característica de E
ínf	elemento ínfimo del conjunto
sup	elemento supremo del conjunto
$\sum_{n=1}^{\infty}$	sumatoria desde $n = 1$ hasta infinito
$\bar{\mu}$	medida extensión de μ

Símbolo	Significado
$(X, \mathbf{S})$	espacio medible
$(X, \mathbf{S}, \mu)$	espacio de medida
f^{-1}	imagen inversa de la función f
f^{-}	parte negativa de la función f
f^{+}	parte positiva de la función f
$\psi(f, g)$	distancia entre las funciones f y g
$\langle x, y \rangle$	producto interno entre x y y
$\\|\cdot\\|$	norma
$x \perp y$	x es ortogonal a y
H	espacio de Hilbert
$d(x, y)$	distancia entre x y y
$\bar{M}$	clausura del conjutno M
T	operador T
T^{*}	operador adjunto de Hilbert
$\mathscr{D}(T)$	dominio del operador T
$\mathscr{R}(T)$	rango del operador T
T^{-1}	operador inverso de T
P_M	operador proyección de H sobre M
Ω	espacio muestral
$\mathcal{F}$	σ-álgebra de eventos
$\Pr$	función de probabilidad
$(\Omega, \mathcal{F}, \Pr)$	espacio de probabilidad
$E(X)$	esperanza o media de la variable X
σ^2	varianza de una variable aleatoria
$Cov(X, Y)$	covarianza ente X y Y
ρ	coeficiente de correlación
$\gamma_X(\cdot, \cdot)$	función de autocovarianza de $\{X_t\}$

OBJETIVOS

General

Establecer los fundamentos matemáticos que sustentan las series de tiempo.

Específicos

1. Introducir los fundamentos matemáticos teóricos elementales de la teoría de la medida que sustentan la teoría de probabilidades.
2. Definir los elementos principales de los espacios de Hilbert, así como también las propiedades esenciales de los operadores y el teorema de proyección.
3. Establecer los fundamentos matemáticos teóricos elementales de la teoría de probabilidades, haciendo uso de la teoría de la medida.
4. Formalizar la teoría de las series de tiempo, mediante la teoría de probabilidades y los espacios de Hilbert, y definir los modelos principales para la solución de una serie de tiempo.

INTRODUCCIÓN

Las series de tiempo son utilizadas en diversas áreas tales como economía, finanzas, medicina, ingeniería, biología. Son de mucha utilidad en las investigaciones y desarrollo de nuevos modelos predictivos, a través de la recolección de datos, dependientes del tiempo, podemos generar un modelo capaz de cumplir con las necesidades de la investigación o proyecto. Dichos modelos se eligen dependiendo de la naturaleza de los datos recolectados, por lo que no todos los problemas utilizan el mismo modelo, esto genera que tengamos un mejor ajuste en las predicciones deseadas.

Para empezar el estudio de las series de tiempo es importante tener conocimiento de la teoría de probabilidades, la cual presenta conceptos fundamentales que se verán relacionados al momento de estar estudiando los datos. Sin embargo para dar a las series de tiempo de una estructura matemática formal es necesario estudiar otras áreas, como lo son: la teoría de la medida y los espacios de Hilbert. La teoría de la medida proporciona solidez a la teoría de probabilidades. Los espacios de Hilbert dan una estructura geométrica que se puede emplear en los modelos de las series de tiempo. El presente trabajo pretende dar los elementos matemáticos fundamentales a las series de tiempo.

En el primer capítulo se desarrolla el estudio de la teoría de la medida, se toma mucha más importancia a los temas que nos ayuden a formalizar la teoría de probabilidades, entre los cuales tenemos conceptos y propiedades de: anillos, σ-anillos, medidas y funciones medibles. En el segundo capítulo tenemos los conceptos de los espacios de Hilbert, siendo la parte más importante del capítulo la que presenta el teorema de proyección, ya que dicho teorema es utilizado principalmente en las series de tiempo, nos da condiciones necesarias para poder encontrar una solución del mejor ajuste a una serie de datos.

En el tercer capítulo tomamos los conceptos del capítulo 1 y los adecuamos de tal manera que podamos darle una estructura matemática a la teoría de probabilidades. Luego se presentan definiciones y teoremas que nos ayuden al estudio de las series de tiempo. En el cuarto capítulo se presentan los conceptos de series de tiempo, desarrollándolos con la ayuda de la teoría de probabilidades y los espacios de Hilbert. Además se desarrollan algunos modelos o métodos a utilizar para el estudio de una serie de tiempo, así como sus características y funciones.

1. TEORÍA DE LA MEDIDA

En este capítulo estudiaremos la noción de *medida* así como también sus principales resultados, los cuales nos ayuden a fundamentar de manera matemática la teoría de probabilidades. Cuando hablamos de *medida* nos referimos a una función que asigna un valor real a un intervalo, un área o un volumen, a los subconjuntos de un conjunto dado, de manera general, de un σ-anillo o un σ-álgebra. Para el desarrollo del capítulo tomaremos en cuenta su estructura a partir de σ-anillos, pero de igual forma se podría realizar a partir de σ-álgebras.

1.1. Anillos y σ-anillos

Antes de iniciar con el estudio de anillos y σ-anillos definiremos algunos conceptos fundamentales relacionados con sucesiones de conjuntos, los cuales utilizaremos en secciones posteriores. Además estudiaremos las clases de conjuntos conocidas como álgebras y σ-álgebras y la relación que tienen con anillos y σ-anillos.

Definición 1.1.1. Sea $\{E_n\}$ una sucesión de conjuntos en X, decimos que:

1. es **monótona creciente** si $E_1 \subseteq E_2 \subseteq E_3 \subseteq ...$, denotado como $E_n \uparrow$, o
2. es **monótona decreciente** si $E_1 \supseteq E_2 \supseteq E_3 \supseteq ...$, denotado como $E_n \downarrow$.

Y definimos el límite de una sucesión monótona como:

1. si $E_n\uparrow$, entonces $\lim\limits_{n\to\infty} E_n = \bigcup\limits_{n=1}^{\infty} E_n$, y
2. si $E_n\downarrow$, entonces $\lim\limits_{n\to\infty} E_n = \bigcap\limits_{n=1}^{\infty} E_n$.

Además si el conjunto E^* es el **límite superior** de la sucesión $\{E_n\}$ tal que:

$$E^* = \limsup_{n\to\infty} E_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} E_n,$$

decimos que el conjunto E_* es el **límite inferior** de la sucesión $\{E_n\}$ tal que:

$$E_* = \liminf_{n\to\infty} E_n = \bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} E_n.$$

Si se cumple que $E^* = E_*$ decimos que la sucesión $\{E_n\}$ tiene un conjunto límite y lo denotamos como: $\lim_{n\to\infty} E_n$.

Definición 1.1.2. Un **anillo** es una clase no vacía de conjuntos $\mathbf{R}$ tal que, para todo $E, F \in \mathbf{R}$, se cumple:

$$E \cup F \in \mathbf{R} \text{ y } E \setminus F \in \mathbf{R}.$$

Es decir, es cerrada bajo uniones y diferencias, y además implica que $E \cap F \in \mathbf{R}$.

Definición 1.1.3. Un **álgebra** es una clase no vacía de conjunto $\mathbf{R}$ tal que, para todo $E, F \in \mathbf{R}$, se cumple

$$E \cup F \in \mathbf{R} \quad \text{y} \quad E^c \in \mathbf{R}.$$

Es decir, es cerrada bajo uniones y complementos.

Dado que

$$E \setminus F = E \cap F^c = (E^c \cup F)^c,$$

se sigue que cada álgebra es un anillo. Un álgebra puede ser caracterizada como un anillo que contiene X, dado que $E^c = X \setminus E$.

Lema 1.1.1. *La intersección de cualquier colección de anillos o álgebras es un anillo o álgebra, respectivamente.*

Ejemplo 1.1.1. Sea X un conjunto no vacío, las siguientes clases de conjuntos son ejemplos de álgebras por lo que también lo son de anillos:

1. $C_1 = \{\emptyset, X\}$.

2. $C_2 = \{\text{todos los subconjuntos de } X\}$.

3. Sea X un conjunto infinito, contable o no contable, y sea C_3 la clase de subconjuntos finitos, o cuyo complemento sea finito, de X, es decir, $C_3 = \{E \subset X \mid E$ o E^c es finito$\}$.
   Mostraremos que C_3 es un álgebra.

a) Primero probaremos que para todo $E \in C_3$ se cumple que $E^c \in C_3$. Supongamos que $E \in C_3$, entonces E o E^c es finito. Si E es finito, entonces $(E^c)^c = E$ es finito, por lo que $E^c \in C_3$. Si E^c es finito, entonces $E^c \in C_3$.

b) Ahora probaremos que para todo $E, F \in C_3$ se cumple que $E \cup F \in C_3$. En efecto, supongamos que $E, F \in C_3$, entonces E o E^c es finito y F o F^c es finito:

1) Supongamos que E, F son ambos finitos, entonces $E \cup F$ es finito, por lo que $E \cup F \in C_3$.
2) Supongamos que E, F^c son finitos, entonces $(E \cup F)^c = E^c \cap F^c$ es finito dado que F^c es finito, por lo que $E \cup F \in C_3$.

Las otras 2 posibilidades se siguen de 2). Por lo tanto tenemos que C_3 es un álgebra.

Teorema 1.1.1. *Sea* $\mathbf{E}$ *cualquier clase de conjuntos, entonces existe un único anillo* $\mathbf{R}_0$ *y* $\mathbf{E} \subset \mathbf{R}_0$ *tal que, si* $\mathbf{R}$ *es cualquier otro anillo que contenga a* $\mathbf{E}$*, entonces* $\mathbf{R}_0 \subset \mathbf{R}$*. El anillo* $\mathbf{R}_0$ *es el anillo más pequeño que contiene a* $\mathbf{E}$ *y es llamado el* ***anillo generado*** *por* $\mathbf{E}$*, lo denotamos como* $\mathbf{R}(\mathbf{E})$*.*

Demostración. Dado que la clase de todos los subconjuntos de X es un anillo, es claro que siempre existe al menos un anillo que contiene a $\mathbf{E}$. Además sabemos por definición 1.1.1 que la intersección de cualquier colección de anillos es un anillo, por lo que la intersección de todos los anillos que contienen a $\mathbf{E}$ es el anillo $\mathbf{R}_0$ deseado. □

Teorema 1.1.2. *Sea* $\mathbf{E}$ *cualquier clase de conjuntos, entonces cualquier conjunto en* $\mathbf{R}(\mathbf{E})$ *puede ser cubierto por una unión finita de elementos en* $\mathbf{E}$*.*

Demostración. La clase de todos los conjuntos que puede ser cubierto por una unión finita de elementos de $\mathbf{E}$ es un anillo. Dado que este anillo contiene a $\mathbf{E}$, por teorema 1.1.1 también contiene a $\mathbf{R}(\mathbf{E})$. □

Teorema 1.1.3. *Sea* $\mathbf{E}$ *una clase contable de conjuntos, entonces* $\mathbf{R}(\mathbf{E})$ *es contable.*

Demostración. Consultar [9, p. 23]. □

Definición 1.1.4. Un **σ-anillo** es una clase no vacía de conjuntos $\mathbf{S}$, tal que:

1. para todo $E, F \in \mathbf{S}$, entonces $E \setminus F \in \mathbf{S}$.

2. Si $E_i \in \mathbf{S}$, $i = 1, 2, \ldots$, entonces $\displaystyle\bigcup_{i=1}^{\infty} E_i \in \mathbf{S}$.

Del teorema 1.1.1 si reemplazamos anillo con σ-anillo, podemos definir el σ-anillo, $\mathbf{S}(\mathbf{E})$ **generado** por cualquier clase de conjuntos $\mathbf{E}$ como el σ-anillo más pequeño que contiene a $\mathbf{E}$.

Teorema 1.1.4. *Sea $\mathbf{E}$ cualquier clase de conjuntos y sea E cualquier conjunto en $\mathbf{S} = \mathbf{S}(\mathbf{E})$, entonces existe una subclase contable de conjuntos $\mathbf{D}$ de $\mathbf{E}$ tal que $E \in \mathbf{S}(\mathbf{D})$.*

Demostración. La unión de todos los σ-subanillos que son generados por alguna subclase contable de $\mathbf{E}$ es un σ-anillo que contiene a $\mathbf{E}$ y contenido en $\mathbf{S}$; por lo que es idéntico a $\mathbf{S}$. $\square$

Para cada clase de conjuntos $\mathbf{E}$ de X y para cada conjunto fijo A de X, denotamos por $\mathbf{E} \cap A$, la clase de todos los conjuntos de la forma $E \cap A$ con $E \in \mathbf{E}$.

Teorema 1.1.5. *Sea $\mathbf{E}$ cualquier clase de conjuntos y sea A cualquier subconjunto de X, entonces:*

$$\mathbf{S}(\mathbf{E}) \cap A = \mathbf{S}(\mathbf{E} \cap A).$$

Demostración. Denotemos por $\mathbf{C}$ la clase de todos los conjuntos de la forma $B \cup (C - A)$, donde:

$$B \in \mathbf{S}(\mathbf{B} \cap A) \qquad \text{y} \qquad C \in \mathbf{S}(\mathbf{E});$$

es sencillo ver que $\mathbf{C}$ es un σ-anillo. Sea $E \in \mathbf{E}$, entonces la relación:

$$E = (E \cap A) \cup (E \setminus A),$$

junto con:

$$E \cap A \in \mathbf{E} \cap A \subset \mathbf{S}(\mathbf{E} \cap A),$$

muestra que $E \in \mathbf{C}$ y por lo tanto que $\mathbf{E} \subset \mathbf{C}$. Se sigue que $\mathbf{S}(\mathbf{E}) \subset \mathbf{C}$ y por lo tanto que:

$$\mathbf{S}(\mathbf{E}) \cap A \subset \mathbf{C} \cap A = \mathbf{S}(\mathbf{E} \cap A).$$

La otra parte de la igualdad se obtiene del hecho que $\mathbf{S}(\mathbf{E}) \cap A$ es un σ-anillo y:

$$\mathbf{E} \cap A \subset \mathbf{S}(\mathbf{E}) \cap A.$$

$\square$

Definición 1.1.5. Un **σ-álgebra** es una clase no vacía de conjuntos $\mathbf{S}$, tal que:

1. para todo $E \in \mathbf{S}$, entonces $E^c \in \mathbf{R}$.

2. Si $E_i \in \mathbf{S}$, $i = 1, 2, \ldots$, entonces $\displaystyle\bigcup_{i=1}^{\infty} E_i \in \mathbf{S}$.

Ejemplo 1.1.2. Las siguientes clases de conjuntos son ejemplos de σ-álgebras y por lo tanto de σ-anillos:

1. $C_1 = \{\emptyset, \mathbf{S}\}$.

2. $C_2 = \{\text{todos los subconjuntos de } \mathbf{S}\}$.

3. Sea $\mathbf{S}$ una clase no contable de conjuntos y sea $C_3 = \{$Todos los subconjuntos de $\mathbf{S}$ que son contables o cuyos complementos son contables$\}$. Mostraremos que C_3 es un σ-álgebra.

   a) Primero probaremos que para todo $E \in C_3$ se cumple que $E^c \in \mathbf{S}$. Supongamos que $E \in \mathbf{S}$, entonces E o E^c es contable. Si E es contable, entonces $(E^c)^c = E$ es contable, por lo que $E^c \in \mathbf{S}$. Si E^c es contable, entonces $E^c \in \mathbf{S}$.

   b) Ahora probaremos que si $E_i \in \mathbf{S}$, $i = 1, 2, \ldots$, entonces $\displaystyle\bigcup_{i=1}^{\infty} E_i \in \mathbf{S}$. En efecto, sea $E_i \in \mathbf{S}$, $i = 1, 2, \ldots$, entonces cualquier E_i es contable o E_i^c es contable. Si todos los E_i son contables, tenemos que unión contable de conjuntos contables es contable, por lo que se cumple lo que queríamos mostrar. Si no todos los E_i son contables, tenemos que:

   $$\left(\bigcup_{i=1}^{\infty} E_i\right)^c = \bigcap_{i=1}^{\infty} E_i^c$$

   el cual es contable, dado que es una intersección de conjuntos. Con lo que mostramos que C_3 es un σ-álgebra.

1.2. Medidas sobre anillos

En esta sección estudiaremos las propiedades fundamentales de medida sobre anillos y σ-anillos, así como también estudiaremos las medidas externas. Antes de iniciar con el estudio de esta sección cabe hacer la observación que todos los siguientes teoremas y definiciones funcionan también para álgebras y σ-álgebras, si consideramos que el anillo posee a todo el conjunto con el que estamos trabajando.

Definición 1.2.1. Sea X un conjunto y $\mathbf{R}$ un anillo sobre X. Una **medida** sobre $\mathbf{R}$ es una función $\mu\colon \mathbf{R} \longrightarrow [0,\infty]$, tal que:

1. $\mu(\emptyset) = 0$,
2. si $\{E_n\}$ es una sucesión disjunta de conjuntos de $\mathbf{R}$, entonces:

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

La propiedad 2) es llamada **contablemente aditiva**, esta propiedad implica la **aditividad finita**, es decir, sea $E_1, \dots, E_n$ conjuntos disjuntos en $\mathbf{R}$, entonces:

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

Definición 1.2.2. Sea μ una medida en un anillo $\mathbf{R}$, un conjunto $E \in \mathbf{R}$ tiene **medida finita** si $\mu(E) < \infty$. Un conjunto $E \in \mathbf{R}$ tiene **medida σ-finita** si existe una sucesión de conjuntos $\{E_n\}$ tal que:

$$E \subset \bigcup_{n=1}^{\infty} E_n \text{ y } \mu(E_n) < \infty,\ n = 1, 2, \dots$$

Si la medida para cada conjunto $E \in \mathbf{R}$ es finita (o σ-finita), decimos que μ es **una medida finita** (o **medida σ-finita**) en $\mathbf{R}$. Si $X \in \mathbf{R}$, es decir, $\mathbf{R}$ es un álgebra, y $\mu(X)$ es finita o σ-finita, decimos que μ es **totalmente finita** o **totalmente σ-finita** respectivamente. La medida μ es llamada **completa** si $E \in \mathbf{R}$, $F \subset E$ y $\mu(E) = 0$ implica que $F \in \mathbf{R}$.

Definición 1.2.3. Una función $\mu\colon \mathbf{E} \longrightarrow [0,\infty]$ es **monótona** si, para cualesquiera $E, F \in \mathbf{E}$ y $F \subset E$, se cumple que $\mu(F) \leq \mu(E)$. Decimos que μ es **sustractiva** si, para cualesquiera $E, F \in \mathbf{E}$, $E \subset F$, $F \setminus E \in \mathbf{E}$ y $|\mu(E)| < \infty$, se cumple que $\mu(F - E) = \mu(F) - \mu(E)$.

Teorema 1.2.1. *Sea μ una medida en un anillo $\mathbf{R}$, entonces μ es monótona y sustractiva.*

Demostración. Sea $E, F \in \mathbf{R}$ y $E \subset F$, entonces $F \setminus E \in \mathbf{R}$ y $\mu(F) = \mu(E) + \mu(F - E)$. La propiedad de ser monótona se sigue del hecho de ser no negativa, de donde $\mu(F) \leq \mu(E)$. La propiedad de ser sustractiva se obtiene del hecho que $\mu(E)$, si es finita, puede ser sustraída de ambos lados de la ecuación $\mu(F) = \mu(E) + \mu(F - E)$. $\square$

Teorema 1.2.2. *Sea μ una medida en un anillo $\mathbf{R}$, $E \in \mathbf{R}$ y sea $\{E_i\}$ una sucesión finita o infinita de conjuntos en $\mathbf{R}$, se cumplen las siguientes propiedades:*

1. *Si $E \subset \bigcup_i E_i$, entonces $\mu(E) \leq \sum_i \mu(E_i)$.*

2. *Si $\bigcup_i E_i \subset E$, donde los E_i son conjuntos disjuntos, entonces:*

$$\sum_i \mu(E_i) \leq \mu(E).$$

Demostración. 1. Sea $\{F_i\}$ cualquier sucesión de conjuntos en un anillo $\mathbf{R}$, entonces existe una sucesión disjunta $\{G_i\}$ de conjuntos en $\mathbf{R}$ tal que:

$$G_i \subset F_i \qquad \text{y} \qquad \bigcup_i G_i = \bigcup_i F_i.$$

Siendo $G_i = F_i - \bigcup \Big\{ F_j | 1 \leq j < i \Big\}$. Aplicando este resultado a la sucesión $\{E \cap E_i\}$, el resultado se sigue de la aditividad contable y la monotomía de μ.

2. Si la sucesión $\{E_i\}$ es finita, entonces $\bigcup_i E_i \in \mathbf{R}$, y se sigue que:

$$\sum_i \mu(E_i) = \mu\left(\bigcup_i E_i \right) \leq \mu(E).$$

La validación de la desigualdad para una sucesión infinita de conjuntos es una consecuencia de su validación para cada subsucesión finita. $\square$

Teorema 1.2.3. *Sea μ una medida en un anillo $\mathbf{R}$ y sea $\{E_i\}$ una sucesión de conjuntos en $\mathbf{R}$ para la cual $\lim_n E_n \in \mathbf{R}$, se cumplen las siguientes propiedades:*

1. *Si $\{E_i\}$ es creciente, entonces $\mu(\lim_{n\to\infty} E_n) = \lim_{n\to\infty} \mu(E_n)$.*

2. *Si $\{E_i\}$ es decreciente y al menos un conjunto tiene medida finita, entonces $\mu(\lim_{n\to\infty} E_n) = \lim_{n\to\infty} \mu(E_n)$.*

Demostración. 1. Sea $E_0 = 0$, entonces:

$$\mu(\lim_{n\to\infty} E_n) = \mu\left(\bigcup_{i=1}^{\infty} E_i \right) = \mu\left(\bigcup_{i=1}^{\infty} (E_i - E_{i-1}) \right) =$$

$$= \sum_{i=1}^{\infty} \mu(E_i - E_{i-1}) = \lim_{n\to\infty} \sum_{i=1}^{n} \mu(E_i - E_{i-1}) =$$

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

2. Sea $\mu(E_m) < \infty$, entonces $\mu(E_n) \leq \mu(E_m)$ para $n \geq m$, y por lo tanto $\mu(\lim_{n\to\infty} E_n) < \infty$. Se sigue del teorema 1.2.1, el inciso anterior y el hecho que $\{E_m - E_n\}$ es una sucesión creciente que:

$$\mu(E_m) - \mu(\lim_{n\to\infty} E_n) = \mu(E_m - \lim_{n\to\infty} E_n) =$$

$$= \mu(\lim_{n\to\infty} (E_m - E_n)) = \lim_{n\to\infty} \mu(E_m - E_n) =$$

$$= \lim_{n\to\infty} (\mu(E_m) - \mu(E_n)) = \mu(E_m) - \lim_{n\to\infty} \mu(E_n).$$

Dado que $\mu(E_m) < \infty$, la demostración está completa. □

Definición 1.2.4. Una clase no vacía de conjuntos $\mathbf{E}$ es **hereditario** si para cualquier $E \in \mathbf{E}$ y cualquier $F \subset E$, se cumple que $F \in \mathbf{E}$. Denotamos por $\mathbf{H}(\mathbf{E})$ al σ-anillo hereditario **generado** por $\mathbf{E}$, es decir, el σ-anillo hereditario más pequeño que contiene a $\mathbf{E}$.

Definición 1.2.5. Una **medida externa** sobre $\mathbf{H}$ es una función $\mu^*\colon \mathbf{H} \longrightarrow [0, \infty]$, tal que:

1. $\mu^*(\emptyset) = 0$,

2. $\mu^*(A) \leq \mu^*(B)$ si $A \subset B$, $A, B \in \mathbf{H}$,

3. si $\{E_n\}$ es una sucesión de conjuntos de $\mathbf{H}$, entonces

$$\mu^*\left(\bigcup_{i=1}^{\infty} E_i\right) \leq \sum_{i=1}^{\infty} \mu^*(E_i).$$

La propiedad 3) es llamada **contablemente subaditiva**.

Teorema 1.2.4. *Sea μ una medida sobre un anillo $\mathbf{R}$ y si para cada $E \in \mathbf{H}(\mathbf{R})$,*

$$\mu^*(E) = \inf \left\{ \sum_{n=1}^{\infty} \mu(E_n) \mid E_n \in \mathbf{R}, E \subset \bigcup_{n=1}^{\infty} E_n \right\},$$

entonces μ^* es una extensión de μ a una medida externa sobre $\mathbf{H}(\mathbf{R})$. Si μ es σ-finito o totalmente σ-finita, entonces μ^* también lo es. La medida externa μ^* es llamada la medida externa ***inducida*** *por la medida μ.*

Demostración. Consultar [9, p. 42]. □

Definición 1.2.6. Sea μ^* una medida externa sobre un anillo hereditario $\mathbf{H}$. Un conjunto $E \in \mathbf{H}$ es **μ^*-medible** si para cada conjunto $A \in \mathbf{H}$,

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c).$$

Teorema 1.2.5. *Sea μ^* una medida externa sobre un σ-anillo hereditario $\mathbf{H}$ y sea $\bar{\mathbf{S}}$ la clase de todas los conjuntos μ^*-medibles, entonces se cumplen las siguientes propiedades:*

1. *$\bar{\mathbf{S}}$ es un anillo.*

2. *$\bar{\mathbf{S}}$ es un σ-anillo.*

3. *Sea $A \in \mathbf{H}$ y $\{E_n\}$ una sucesión de conjuntos disjuntos en $\bar{\mathbf{S}}$ con $\bigcup_{n=1}^{\infty} E_n = E$, entonces*
$$\mu^*(A \cap E) = \sum_{n=1}^{\infty} \mu^*(A \cap E_n).$$

4. *Cada conjunto de medida externa cero pertenece a $\bar{\mathbf{S}}$ y la función $\bar{\mu}$, definida para cada E en $\bar{\mathbf{S}}$ por $\bar{\mu}(E) = \mu^*(E)$, es una medida completa en $\bar{\mathbf{S}}$. La medida $\bar{\mu}$ es llamada la medida* ***inducida*** *por la medida externa μ^*.*

Demostración. Consultar [9, p. 45]. □

Teorema 1.2.6. *Sea μ una medida σ-finita en un anillo $\mathbf{R}$, entonces existe una única medida $\bar{\mu}$ en el σ-anillo $\mathbf{S}(\mathbf{R})$ tal que para cada $E \in \mathbf{R}$, $\bar{\mu}(E) = \mu(E)$; la medida $\bar{\mu}$ es σ-finita. La medida $\bar{\mu}$ es llamada la* ***extensión*** *de μ.*

Demostración. Consultar [9, p. 54]. □

Teorema 1.2.7. *Sea μ una medida sobre un σ-anillo $\mathbf{S}$, entonces la clase $\bar{\mathbf{S}}$ de conjuntos de la forma $E \Delta N$, donde $E \in \mathbf{S}$ y N es un subconjunto de un conjunto de medida cero en $\mathbf{S}$, es un σ-anillo, y la función $\bar{\mu}$ definida por $\bar{\mu}(E \Delta N) = \mu(E)$ es una medida completa en $\bar{\mathbf{S}}$. La medida $\bar{\mu}$ es llamada la* ***completación*** *de μ.*

Demostración. Sea $E \in \mathbf{S}$, $N \subset A \in \mathbf{S}$ y $\mu(A) = 0$, entonces las relaciones:

$$E \cup N = (E \setminus A)\Delta[A \cap (E \cup N)],$$

y

$$E\Delta N = (E \setminus A) \cup [A \cap (E\Delta N)],$$

muestran que la clase $\bar{\mathbf{S}}$ puede ser descrita como la clase de todos los conjuntos de la forma $E \cup N$, donde $E \in \mathbf{S}$ y N es un subconjunto de un conjunto de medida cero en $\mathbf{S}$. Dado que esto implica que la clase $\bar{\mathbf{S}}$, la cual es cerrada bajo la formación de diferencia simétrica, es cerrada también bajo la formación de contables uniones, por lo que tenemos que $\bar{\mathbf{S}}$ es un σ-anillo.

Sea $E_1\Delta N_1 = E_2\Delta N_2$, donde $E_i \in \mathbf{S}$ y N_i es un subconjunto de un conjunto de medida cero en $\mathbf{S}$, con $i = 1, 2$, entonces:

$$E_1\Delta E_2 = N_1\Delta N_2,$$

y por lo tanto $\mu(E_1\Delta E_2) = 0$. De donde se obtiene que $\mu(E_1) = \mu(E_2)$, y por lo tanto que $\bar{\mu}$ es definida sin ambigüidad por la relación:

$$\bar{\mu}(E\Delta N) = \bar{\mu}(E \cup N) = \mu(E).$$

Usando la representación de unión, en lugar de la diferencia simétrica, en $\bar{\mathbf{S}}$, es sencillo verificar que $\bar{\mu}$ es una medida; la completación de $\bar{\mu}$ es una consecuencia inmediata del hecho que $\bar{\mathbf{S}}$ contiene todos los subconjuntos de conjuntos de medida cero en $\mathbf{S}$. $\square$

Definición 1.2.7. Sea X la recta real, $\mathbf{P}$ la clase de todos los intervalos acotados semicerrados de la forma $[a, b)$, $a, b \in X$, sea $\mathbf{S}$ el σ-anillo generado por $\mathbf{P}$ y sea μ una función definida en $\mathbf{P}$ por $\mu([a, b)) = b - a$, decimos que el conjunto de los σ-anillos $\mathbf{S}$ son los **conjuntos de Borel** de la recta. Sea $\bar{\mu}$ en $\bar{\mathbf{S}}$ la completación de μ en $\mathbf{S}$, los conjuntos de $\bar{\mathbf{S}}$ son llamados los **conjuntos medibles de Lebesgue** de la recta; la medida $\bar{\mu}$ es la **medida de Lebesgue**. Dado que la recta X es la unión de conjuntos contables en $\mathbf{P}$, tenemos que $X \in \mathbf{S}$, es decir, que los σ-anillo $\mathbf{S}$ y $\bar{\mathbf{S}}$ son también σ-álgebras.

Teorema 1.2.8. *Cada conjunto contable es un conjunto de Borel con medida cero.*

Demostración. Para cualquier a, con $-\infty < a < \infty$, tenemos:

$$\{a\} = \{x \mid x = a\} = \bigcap_{n=1}^{\infty} \Big\{x \mid a \leq x < a + \frac{1}{n}\Big\},$$

y por lo tanto:

$$\mu(\{a\}) = \lim_{n\to\infty} \mu\Big(\Big[a, a + \frac{1}{n}\Big)\Big) = \lim_{n\to\infty} \frac{1}{n} = 0,$$

así que para cada conjunto de un elemento es un conjunto de Borel de medida cero. Dado que los conjuntos de Borel forman un σ-anillo y que μ es contable aditiva, se prueba el teorema. □

Definición 1.2.8. Un conjunto $G \subset \mathbb{R}$ es llamado **abierto** is para todo $x \in G$ existe un intervalo abierto que contiene a x y que está contenido en G. Sin perdida de generalidad, tales intervalos los consideraremos centrado en x.
De manera general sea $G \subset \mathbb{R}^m$, $m \geq 1$, es llamado **abierto** si para todo $x \in G$ existe un cubo abierto en $\mathbb{R}^m$ que contiene a x y que está contenido en G; por el termino **cubo abierto**, nos referimos al producto cartesiano de m intervalos abiertos de igual longitud. Además decimos que tales cubos están centrados en x.

Teorema 1.2.9. *La clase* $\mathbf{S}$ *de todos los conjuntos de Borel coincide con el* σ*-anillo generado por la clase de conjuntos abiertos* $\mathbf{U}$.

Demostración. Dado que para cualquier número real, el conjunto $\{a\}$ es un conjunto de Borel, se sigue de la relación $(a, b) = [a, b) - \{a\}$, que cada intervalo acotado abierto es un conjunto de Borel. Dado que cada conjunto abierto sobre la recta es una unión contable de intervalos abiertos acotados, se tiene que $\mathbf{U} \subset \mathbf{S}$ y consecuentemente que $\mathbf{S}(\mathbf{U}) \subset \mathbf{S}$. Para mostrar la contención inversa, observemos que, para cada número real a:

$$\{a\} = \bigcap_{n=1}^{\infty} \Big(a - \frac{1}{n}, a + \frac{1}{n}\Big),$$

por lo que $\{a\} \in \mathbf{S}(\mathbf{U})$. Se sigue de la relación $[a, b) = (a, b) \cup \{a\}$ que $\mathbf{P} \subset \mathbf{S}(\mathbf{U})$ y por lo tanto que:

$$\mathbf{S} = \mathbf{S}(\mathbf{P}) \subset \mathbf{S}(\mathbf{U}).$$

□

Teorema 1.2.10. *Sea* $\mathbf{U}$ *la clase de todos los conjuntos abiertos, entonces, para cada* $E \in X$*:*

$$\mu^*(E) = \inf \big\{\mu(U) \mid E \subset U, U \in \mathbf{U}\big\}.$$

Demostración. Dado que $\mu^*(E) = \text{ínf}\{\mu(F) \mid E \subset F \in \mathbf{S}\}$, se sigue del hecho $\mathbf{U} \subset \mathbf{S}$ que:

$$\mu^*(E) \leq \text{ínf}\{\mu(U) \mid E \subset U \in \mathbf{U}\}.$$

Por otro lado, sea $\epsilon > 0$, entonces se sigue de la definición de μ^* que existe una sucesión $\{[a_n, b_n)\}$ de conjuntos en $\mathbf{P}$ tal que:

$$E \subset \bigcup_{n=1}^{\infty}[a_n, b_n) \qquad \text{y} \qquad \sum_{n=1}^{\infty}(b_n - a_n) \leq \mu^*(E) + \frac{\epsilon}{2}.$$

Consecuentemente:

$$E \subset \bigcup_{n=1}^{\infty}\left(a_n - \frac{\epsilon}{2^{n+1}}, b_n\right) = U \in \mathbf{U},$$

y

$$\mu(U) \leq \sum_{n=1}^{\infty}(b_n - a_n) + \frac{\epsilon}{2} \leq \mu^*(E) + \epsilon.$$

El resultado deseado se sigue de la arbitrariedad de ϵ. $\square$

Definición 1.2.9. Sea C_0 el conjunto de todos los intervalos de $\mathbb{R}$, es decir:

$$C_0 = \Big\{(-\infty, x), (-\infty, x], (x.\infty), [x, \infty), (x, y), (x, y], [x, y), [x, y] | x, y \in \mathbb{R}, x < y\Big\}.$$

Por el teorema 1.1.1 y la definición 1.1.4 sabemos que existe un σ-anillo generado del conjunto C_0, además también tenemos una σ-álgebra generada de C_0, ya que $\mathbb{R}$ pertenece a nuestro conjunto C_0, denotamos está σ-álgebra como $\mathcal{B}$ y la llamaremos **σ-álgebra de Borel**, sobre la recta real, y al par $(\mathbb{R}, \mathcal{B})$ **la recta real de Borel**.

Teorema 1.2.11. *Cada una de las siguientes clases genera la σ-álgebra de Borel:*

$$C_1 = \Big\{(x, y] \mid x, y \in \mathbb{R}, x < y\Big\},$$

$$C_2 = \Big\{[x, y) \mid x, y \in \mathbb{R}, x < y\Big\},$$

$$C_3 = \Big\{[x, y] \mid x, y \in \mathbb{R}, x < y\Big\},$$

$$C_4 = \Big\{(x, y) \mid x, y \in \mathbb{R}, x < y\Big\},$$

$$C_5 = \Big\{(x, \infty) \mid x \in \mathbb{R}\Big\},$$

$$C_6 = \Big\{[x, \infty) | x \in \mathbb{R}\Big\},$$

$$C_7 = \Big\{(-\infty, x) \mid x \in \mathbb{R}\Big\},$$

$$C_8 = \{(-\infty, x] \mid x \in \mathbb{R}\}.$$

Además las clases C'_i, $i = 1, 2, ..., 8$, generan la σ-álgebra de Borel donde cada una de C'_i son definidas de la misma forma que C_i excepto que x, y son restringidas a números racionales.

Demostración. Sean C, C' dos clases de conjuntos, tales que $C \subseteq C_{\text{J}}$, entonces se tiene que la σ-álgebra generada por C está contenida en la σ-álgebra generada por C', $\sigma(C) \subseteq \sigma(C')$. Para mostrar el teorema, es suficiente mostrar que $\mathcal{B} \subseteq \sigma(C_j)$, $\mathcal{B} \subseteq \sigma(C'_j)$, con $j = 1, 2, \ldots, 8$, y para mostrarlo es suficiente mostrar que $C_0 \subseteq \sigma(C_j)$, $C_0 \subseteq \sigma(C'_j)$, con$j = 1, 2, \ldots, 8$. Como ejemplo, mostraremos que $C_0 \subseteq \sigma(C_7)$. Consideremos $x_n \downarrow x$. Entonces $(-\infty, x_n) \in \sigma(C_7)$ y por lo tanto $\cap_{n=1}^{\infty}(-\infty, x_n) \in \sigma(C_7)$, pero:

$$\bigcap_{n=1}^{\infty}(-\infty, x_n) = (-\infty, x].$$

Por lo que $(-\infty, x] \in \sigma(C_7)$ para todo $x \in \mathbb{R}$. Dado que:

$$(x, \infty) = (-\infty, x]^c, \qquad [x, \infty) = (-\infty, x)^c,$$

también se tiene que $(x, \infty), [x, \infty) \in \sigma(C_7)$. Luego:

$$(x, y) = (-\infty, y) - (-\infty, x] = (-\infty, y) \cap (x, \infty) \in \sigma(C_7),$$

$$(x, y] = (-\infty, y] \cap (x, \infty) \in \sigma(C_7), [x, y) = (-\infty, y) \cap [x, \infty) \in \sigma(C_7),$$

$$[x, y] = (-\infty, y] \cap [x, \infty) \in \sigma(C_7).$$

Por lo que $C_0 \subseteq \sigma(C_7)$. En el caso de C'_j, $j = 1, 2, \ldots, 8$, considere sucesiones monótonas de números racionales convergentes a un irracional x, y. $\square$

1.3. Funciones medibles e integración

En esta sección estudiaremos los conceptos de funciones medibles y sus propiedades, así como también la integraciones de éstas funciones. Una función medible es una función que conserva estructuras entre dos espacios medibles. Al igual que en el capítulo anterior definiremos la función medible haciendo uso de σ-anillos, pero de igual forma se podría trabajar con σ-álgebras.

Definición 1.3.1. Sea X un conjunto y $\mathbf{S}$ un σ-anillo de conjuntos en X con la propiedad que $\bigcup \mathbf{S} = X$, llamamos a la pareja $(X, \mathbf{S})$ un **espacio medible**. Decimos que los conjuntos $E \in X$ son **medibles** si pertenecen a $\mathbf{S}$. Un **espacio de medida** es un espacio medible $(X, \mathbf{S})$ con una medida μ en $\mathbf{S}$, denotado como $(X, \mathbf{S}, \mu)$.

Teorema 1.3.1. *Todo conjunto abierto en* $\mathbb{R}$ *es medible.*

Demostración. Sea $G \in \mathbb{R}$ un conjunto abierto, y para cada $x \in G$ considere un intervalo abierto centrado en x y contenido en G. Claramente, la unión sobre cada uno de los intervalos, de cada $x \in G$, es igual a G. Lo mismo ocurre si consideramos sólo los intervalos correspondientes a todos los racionales $x \in G$. Estos intervalos son contables y cada uno de ellos es medible, así mismo para la unión. $\square$

Teorema 1.3.2. *Todo conjunto abierto* $\mathbb{R}^m$ *es medible.*

Demostración. La demostración será análoga al teorema 1.3.1. Sea G un conjunto abierto en $\mathbb{R}^m$, y para cada $x \in G$ consideremos un cubo abierto centrado en x y contenido en G. La unión sobre cada uno de los cubos, de cada $x \in G$, es igual a G. Lo mismo ocurre si restringimos a las $x \in G$ cuyas coordenadas m son racionales. Entonces los cubos resultantes son contables y por lo tanto su unión es medible, ya que cada cubo lo es. $\square$

Definición 1.3.2. Una función $g\colon X \subseteq \mathbb{R} \to \mathbb{R}$ es continua en $x_0 \in X$ si para cada $\epsilon > 0$ existe un $\delta = \delta(\epsilon, x_0)$ tal que $|x - x_0| < \epsilon$ implica $|g(x) - g(x_0)| < \delta$. La función g es continua en X si es continua para todo $x \in X$.

Teorema 1.3.3. *Sea* g *una función continua* $g\colon \mathbb{R} \to \mathbb{R}$*. Entonces* g *es medible.*

Demostración. Por el teorema 1.2.11 es suficiente demostrar que $g^{-1}(G)$ son conjuntos medibles para todos los intervalos G en $\mathbb{R}$. Sea $B = g^{-1}(G)$, con lo que si $B = \emptyset$, el teorema es válido, así que sea $B \neq \emptyset$ y sea x_0 un punto arbitrario de B, así que $g(x_0) \in G$. La continuidad de g en x_0 implica que para cada $\epsilon > 0$ existe un $\delta = \delta(\epsilon, x_0) > 0$ tal que $|x - x_0| < \epsilon$ implica $|g(x) - g(x_0)| < \delta$. Equivalentemente, $x \in (x_0 - \epsilon, x_0 + \epsilon)$ implica $g(x) \in (g(x_0) - \delta, g(x_0) + \delta)$. Dado que $g(x_0) \in G$ y G es abierto, si elegimos un ϵ lo suficientemente pequeño, podemos hacer que δ sea tan pequeño tal que $(g(x_0) - \delta, g(x_0) + \delta)$ esté contenido en G. Pero $B = g^{-1}(G)$ es el conjunto de todos los $x \in \mathbb{R}$ para los cuales $g(x) \in G$. Como todos los $x \in (x_0 - \epsilon, x_0 + \epsilon)$ cumplen dicha propiedad, se tiene que $(x_0 - \epsilon, x_0 + \epsilon) \subset B$. Dado que x_0 es arbitrario en B, se tiene que B es abierto. Por el teorema 1.3.1 es medible. $\square$

Definición 1.3.3. Sea f una función de valores reales en un conjunto X y sea M cualquier subconjunto de la recta real. Decimos que $f^{-1}(M)$ es la **imagen inversa** de M, donde:

$$f^{-1}(M) := \{x \mid f(x) \in M\}.$$

Sea $(X, \mathbf{S})$ un espacio medible, si f es tal que para cada subconjutno de Borel M de la recta real el conjunto $N(f) \cap f^{-1}(M)$ es medible, donde $N(f) := \{x \mid f(x) \neq 0\}$, decimos que f es una **función medible**.

Teorema 1.3.4. *Una función real f de un espacio medible $(X, \mathbf{S})$ es medible si y sólo si, para cada número real c, el conjunto, $N(f) \cap \{x \mid f(x) < c\}$ es medible.*

Demostración. Sea M un intervalo semiabierto de la forma $M = (-\infty, c)$, entonces M es un conjunto de Borel y sea $f^{-1}(M) = \{x \mid f(x) < c\}$. Es claro que la primera condición es ciertamente necesaria ya que f es medible.
Supongamos que la siguiente condición se satisface. Sea $c_1, c_2 \in \mathbb{R}$ tales que $c_1 \leq c_2$, entonces:

$$\{x \mid f(x) < c_2\} - \{x \mid f(x < c_1)\} = \{x \mid c_1 \leq f(x) < c_2\}.$$

En otras palabras si M es cualquier intervalo semiabierto, entonces $N(f) \cap f^{-1}(M)$ es la diferencia de dos conjuntos medibles y por lo tanto es medible. Sea $\mathbf{E}$ la clase de todos los subconjuntos de M de la recta real para la cual $N(f) \cap f^{-1}(M)$ es medible. Dado que $\mathbf{E}$ es un σ-anillo, y dado que un σ-anillo contiene todos los intervalos semiabiertos, contiene también todos los conjuntos de Borel, con lo cual se completa la demostración del toreama. $\square$

Teorema 1.3.5. *Sean f y g funciones medibles de un espacio medible $(X, \mathbf{S})$, y sea c cualquier número real, entonces cada uno de los siguientes conjuntos tiene una intersección medible con cada conjunto medible:*

$$A = \{x \mid f(x) < g(x) + c\},$$
$$B = \{x \mid f(x) \leq g(x) + c\},$$
$$C = \{x \mid f(x) = g(x) + c\}.$$

Demostración. Sea M el conjunto de números racionales sobre la recta. Dado que:

$$A = \bigcup_{r \in M} [\{x \mid f(x) < r\} \cap \{x \mid r - c < g(x)\}],$$

se sigue que A tiene la propiedad deseada. La conclusión para B y C son consecuencias de las relaciones:

$$B = X \setminus \{x \mid g(x) < f(x) - c\} \qquad \text{y} \qquad C = B \setminus A,$$

respectivamente. $\square$

Teorema 1.3.6. *Sea ϕ una función medible de Borel de la recta real, tal que $\phi(0) = 0$, y sea f una función medible de un espacio medible $(X, \mathbf{S})$, entonces la función $\tilde{f}$, definida por $f := \phi(f(x))$, es una función medible sobre X.*

Demostración. Es conveniente hacer uso de la definición 1.3.3. Sea M cualquier conjunto de Borel sobre la recta, entonces:

$$N(\tilde{f}) \cap \tilde{f}^{-1}(M) = \{x | \phi(f(x)) \in M - \{0\}\} =$$

$$= \{x \mid f(x) \in \phi^{-}1(M - \{0\})\}.$$

Dado que $\phi(0) = 0$, tenemos que:

$$\phi^{-1}(M - \{0\}) = \phi^{-1}(M - \{0\}) - \{0\}.$$

Dado que ϕ es Borel medible, $\phi^{-1}(M - \{0\})$ es un conjunto de Borel y como el conjunto es medible:

$$N(\tilde{f}) \cap \tilde{f}^{-1}(M) = N(f) \cap f^{-1}(\phi^{-1}(M - \{0\})),$$

se sigue que f es medible. $\square$

Teorema 1.3.7. *Sean f y g funciones medibles de un espacio medible $(X, \mathbf{S})$, entonces $f + g$ y fg también son medibles.*

Demostración. Dado el comportamiento de $f + g$ y fg en los puntos x en los cuales al menos uno de los dos números, $f(x)$ y $g(x)$, es finito es fácil entenderlo, después de examinar un pequeño número de casos, dirigimos nuestra atención a funciones de valor finito. Podemos mencionar que si $f(x) = \pm\infty$ y $g(x) \pm \infty$, entonces $f(x) + g(x)$ no está definido. Dado que f, g son finitos y sea $c \in \mathbb{R}$, entonces:

$$\{x \mid f(x) + g(x) < c\} = \{x \mid f(x) < c - g(x)\},$$

la medibilidad de $f+g$ se sigue del teorema 1.3.5, con $-g$ en lugar de g. La medibilidad de fg es una consecuencia de la identidad:

$$fg = \frac{1}{4}[(f+g)^2 - (f-g)^2]. \qquad \square$$

Dado que f y g son finitas tenemos que:

$$f \cup g = \frac{1}{2}(f+g+|f-g|), \qquad f \cap g = \frac{1}{2}(f+g-|f-g|).$$

De los teoremas 1.3.6 y 1.3.7 tenemos que si f y g son medibles, entonces $f \cup g$ y $f \cap g$ también lo son. Si para cada función f como:

$$f^+(x) = \text{máx}(f(x), 0), \quad \text{y} \quad f^(x)- = \text{máx}(-f(x), 0),$$

entonces:

$$f = f^+ - f^-, \quad \text{y} \quad |f| = f^+ + f^-.$$

Las funciones f^+ y f^- son llamadas la **parte positiva** y la **parte negativa** de f respectivamente. La parte positiva y negativa de una función medible son ambas medibles; de igual manera, una función con partes positiva y negativa medibles es medible.

Teorema 1.3.8. *Sea $\{f_n\}$ una sucesión de funciones medibles de un espacio medible $(X, \mathbf{S})$, entonces cada una de las siguientes funciones son medibles:*

$$\begin{aligned} h(x) &:= \sup\{f_n(x) \mid n = 1, 2, \dots\}, \\ g(x) &:= \text{ínf}\{f_n(x) \mid n = 1, 2, \dots\}, \\ f^*(x) &:= \limsup_{n\to\infty} f_n(x), \\ f_*(x) &:= \liminf_{n\to\infty} f_n(x). \end{aligned}$$

Demostración. Es sencillo reducir el caso general al caso de funciones de valores finitos. La ecuación:

$$\{x \mid g(x) < c\} = \bigcup_{n=1}^{\infty} \{x \mid f_n(x) < c\},$$

implica la medibilidad de g. El resultado para h se sigue de la relación:

$$h(x) = \text{ínf}\{-f_n(x) \mid n = 1, 2, \dots\}.$$

La mediblidad de f^* y f_* es una consecuencia de la relación:

$$f^* = \inf_{n\geq 1} \sup_{m\geq n} f_m(x), \qquad f_*(x) = \sup_{n\geq 1} \inf_{m\geq n} f_m(x),$$

respectivamente. $\square$

Definición 1.3.4. Una función f, definida en un espacio medible $(X, \mathbf{S})$ es llamada **simple** si existe una clase finita de conjuntos disjuntos medibles $\{E_1, \ldots, E_n\}$ y un conjunto finito $\{\alpha_1, \ldots, \alpha_n\}$ de números reales, tales que:

$$f(x) = \begin{cases} \alpha_i, & \text{si} \quad x \in E_i, \quad i = 1, \ldots, n, \\ 0, & \text{si} \quad x \notin E_1 \cup \cdots \cup E_n. \end{cases}$$

Una función simples es siempre una función medible.

Teorema 1.3.9. *Toda función medible f es el límite de una sucesión $\{f_n\}$ de funciones simples; si f es no negativa, entonces cada una de las f_n puede tomarse no negativa y la sucesión $\{f_n\}$ puede ser asumida creciente.*

Demostración. Supongamos primero que $f \geq 0$. Para cada $n = 1, 2, \ldots,$ y para cada $x \in X$, tenemos:

$$f_n(x) = \begin{cases} \dfrac{i-1}{2^n}, & \text{si} \quad \dfrac{i-1}{2^n} \leq f(x) < \dfrac{i}{2^n}, \quad i = 1, \ldots, 2^n n, \\ \\ n, & \text{si} \qquad f(x) \geq n. \end{cases}$$

Claramente $f_n(x)$ es una función simple no negativa y la sucesión $\{f_n\}$ es creciente. Si $f(x) < \infty$, entonces tenemos que para algún n:

$$0 \leq f(x) - f_n(x) \leq \frac{1}{2^n},$$

si $f(x) = \infty$, entonces $f_n(x) = n$ para cada n. Esto demuestra la segunda mitad del teorema; la primera mitad se sigue, recordando que la diferencia de dos funciones simples es una función simple, aplicando el resultado anterior separadamente a f^+ y f^-. $\square$

Definición 1.3.5. Una función f es llamada es llamada **esencialmente acotada** si es acotada en casi todas partes, es decir, si existe una constante positiva finita c tal que $\{x \mid |f(x)| > c\}$ es un conjunto de medida cero.

Definición 1.3.6. Sea $(X, \mathbf{S})$ un espacio medible y sea $E \subset X$, la **función característica** χ_E de E está definida por:

$$\chi_E(x) = \begin{cases} 1, & \text{si} \quad x \in E, \\ 0, & \text{si} \quad x \notin E. \end{cases}$$

Definición 1.3.7. Decimos que una sucesión $\{f_n\}$ de funciones de valores reales es **fundamental en casi todas partes** si existe un conjunto E_0 de medida cero tal que, si $x \notin E_0$ y $\epsilon > 0$, entonces un entero $n_0 = n_0(x, \epsilon)$ puede ser encontrado con la propiedad que:

$$|f_n(x) - f_m(x)| < \epsilon, \quad \text{cuando} \quad n \geq n_0 \quad \text{y} \quad m \geq n_0.$$

Definición 1.3.8. Una sucesión $\{f_n\}$ de valores finitos casi en todas partes de funciones medibles, **converge en medida** a una función medible f si para cada $\epsilon > 0$, $\lim_n \mu(\{x \mid |f_(x) - f(x)| \geq \epsilon\}) = 0$. La sucesión $\{f_n\}$ es **fundamental en medida** si para cada $\epsilon > 0$,

$$\mu(\{x \mid |f_n(x) - f_m(x)| \geq \epsilon\}) \longrightarrow 0 \quad \text{cuando} \quad n, m \longrightarrow \infty.$$

Teorema 1.3.10. *Sea $\{f_n\}$ una sucesión de funciones que converge en medida a f, entonces $\{f_n\}$ es fundamental en medida. Si además $\{f_n\}$ converge en medida a g, entonces $f = g$ casi en todas partes.*

Demostración. La primera parte del teorema se sigue de la relación:

$$\{x \mid |f_n(x) - f_m(x)| \geq \epsilon\} \subset \left\{x \mid |f_n(x) - f(x)| \geq \frac{\epsilon}{2}\right\} \cup \cdots$$

$$\cdots \cup \left\{x \mid |f_m(x) - f(x)| \geq \frac{\epsilon}{2}\right\}.$$

Para demostrar la segunda parte, observemos que:

$$\{x \mid |f(x) - g(x)| \geq \epsilon\} \subset \left\{x \mid f_n(x) - f(x)| \geq \frac{\epsilon}{2}\right\} \cup \cdots$$

$$\cdots \cup \left\{x \mid |f_m(x) - g(x)| \geq \frac{\epsilon}{2}\right\}.$$

Dada una elección adecuada de n, la medida de ambos conjuntos de la derecha pueden ser arbitrariamente pequeña, tenemos que:

$$\mu(\{x \mid |f(x) - g(x)| \geq \epsilon\}) = 0,$$

para todo $\epsilon > 0$, esto implica que $f = g$ casi en todas partes. $\square$

Teorema 1.3.11. *Sea $\{f_n\}$ una sucesión de funciones medibles que es fundamental en medida, entonces existe una función medible f tal que $\{f_n\}$ converge en medida a f.*

Demostración. Consultar [9, p. 93] $\square$

Definición 1.3.9. Una función simple $f = \sum_{i=1}^{n} \alpha_i \chi_{E_i}$ de un espacio de medida $(X, \mathbf{S}, \mu)$ es integrable si $\mu(E_i) < \infty$ para cada subíndice i para el cual $\alpha_i \neq 0$. La integral de f está definida por:

$$\int f(x) d\mu(x) = \int f d\mu = \sum_{i=1}^{n} \alpha_i \mu(E_i).$$

El valor de la integral es independiente de la representación de f, es decir, si

$$f = \sum_{j=1}^{m} \beta_i \chi_{F_j}, \quad \text{entonces} \quad \int f d\mu = \sum_{j=1}^{m} \beta_j \mu(F_j).$$

Definición 1.3.10. Sea E un conjunto medible y f una función simple integrable, definimos la **integral** de f **sobre** E como:

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

La integral de la función característica de un conjunto medible E de medida finita, está dada por:

$$\int x_E d\mu = \int_E d\mu = \mu(E).$$

Teorema 1.3.12. *Sean f, g funciones simples integrables, entonces se cumplen las siguientes propiedades*

1. *Sean α, β números reales, entonces* $\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int f d\mu.$

2. *Si f es no negativa en casi todas partes, entonces* $\int f d\mu \geq 0.$

3. *Si $f \geq g$ en casi todas partes, entonces* $\int f d\mu \geq \int g d\mu.$

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

5. *$|\int f d\mu| \leq \int |f| d\mu$.*

6. *Sean α, β números reales y E un conjunto medible tal que, para $x \in E$, $\alpha \leq f(x) \leq \beta$, entonces $\alpha\mu(E) \leq \int_E f d\mu \leq \beta\mu(E)$.*

Demostración. 1. y 2. se siquen inmediatamente de la definición 1.3.10.

3. Aplicar 2. a $f - g$ en lugar de f.

4. Aplicar 3. a $|f| + |g|$ y $|f + g|$ en lugar de f y g respectivamente.

5. Aplicar 3. primero a $|f|$ y f y luego a $|f|$ y $-f$.

6. Podemos escribir la suposición principal de la forma $\alpha\chi_E \leq \chi_E f \leq \beta\chi_E$, el resultado deseado se sigue de 3. si $\mu(E) < \infty$; el caso en el cual $\mu(E) = \infty$ es sencillo conseguirlo de la aplicación directa de la definición 1.3.10. □

Definición 1.3.11. La **integral indefinida** de una función simple integrable f es la función ν definida para cada conjunto medible E como $\nu(E) = \int_E f d\mu$.

Teorema 1.3.13. *Sea f una función simple integrable no negativa casi en todas partes, entonces su integral indefinida es monótona.*

Demostración. Sean F, G conjuntos medibles tales que $E \subset F$, entonces $\chi_E f \leq \chi_F f$ casi en todas partes, y el resultado se sigue del teorema 1.3.12 inciso 3. □

Definición 1.3.12. Sean f, g funciones simples integrables, definimos la **distancia**, $\psi(f, g)$ entre ellas como:

$$\psi(f, g) = \int |f - g| d\mu.$$

La función d tiene las siguientes propiedades:

1. $\psi(f, f) = 0$,

2. $\psi(f, g) = \psi(g, f)$,

3. $\psi(f, g) \leq \psi(g, h) + \psi(h, f)$, donde h es una función integrable.

Sin embargo no es cierto que si $\psi(f, g) = 0$ entonces $f = g$.

Definición 1.3.13. Una sucesión $\{f_n\}$ de funciones simples integrables es **fundamental en media**, si:

$$\psi(f_n, f_m) \longrightarrow 0 \quad \text{cuando} \quad n, m \longrightarrow \infty.$$

Teorema 1.3.14. *Una sucesión de media fundamental $\{f_n\}$ de funciones simples integrables es fundamental en medida.*

Demostración. Si para cada $\epsilon > 0$ fijo:

$$E_{nm} = \{x \mid |f_n(x) - f_m(x)| \geq \epsilon\},$$

entonces:

$$\psi(f_n, f_m) = \int |f_n - f_m| d\mu \geq \int_{E_{nm}} |f_n - f_m| d\mu \geq \epsilon\mu(E_{nm}),$$

así que $\mu(E_{nm}) \longrightarrow 0$ cuando $n, m \longrightarrow \infty$. $\square$

Teorema 1.3.15. *Sea $\{f_n\}$ una sucesión de media fundamental de funciones simples integrables, y sea ν_n la integral indefinida de f_n, $n = 1, 2, \ldots$, entonces:*

$$\nu(E) = \lim_n \nu_n(E),$$

existe para cada conjunto medible E, y la función ν es finita y contablemente aditiva.

Demostración. Dado que:

$$|\nu_n(E) - \nu_m(E)| \leq \int |f_n - f_m| d\mu \longrightarrow 0,$$

cuando $n, m \longrightarrow \infty$, la existencia, la finalidad y la uniformidad del limite son claras, y se sigue de la aditividad finita de límites que ν es finita aditiva. Si $\{E_n\}$ es una sucesión disjunta de conjuntos medibles cuya unión es E, entonces tenemos que para cada par de enteros positivos n, k:

$$\left|\nu(E) - \sum_{i=1}^{k} \nu(E_i)\right| \leq |\nu(E) - \nu_n(E)| + \cdots$$

$$\cdots + \left|\nu_n(E) - \sum_{i=1}^{k} \nu_n(E_i)\right| + \left|\nu_n\Big(\bigcup_{i=1}^{k} E_i\Big) - \nu\Big(\bigcup_{i=1}^{k} E_i\Big)\right|.$$

El primer y tercer termino del lado derecho de esta desigualdad puede hacerse arbitrariamente pequeño eligiendo un n lo suficientemente grande, y para un n fijo, el termino central puede ser arbitrariamente pequeño escogiendo un k suficientemente grande. Esto demuestra que :

$$\nu(E) = \lim_{k\to\infty} \sum_{i=1}^{k} \nu(E_i) = \sum_{i=1}^{\infty} \nu(E_i). \qquad \square$$

Definición 1.3.14. Una función medible de valor finito casi en todas partes f de un espacio de medida $(X, \mathbf{S}, \mu)$ es **integrable** si existe una sucesión fundamental media $\{f_n\}$ de funciones simples integrables que convergen en medida a f, y está definida como:

$$\int f(x) d\mu(x) = \int f d\mu = \lim_{n\to\infty} \int f_n d\mu.$$

Cuyo valor de la integral es siempre finito. Las relaciones $f^+ = \frac{1}{2}(|f| + f)$ y $f^- = \frac{1}{2}(|f| - f)$ muestran que si f es integrable también f^+ y f^- son integrables. Definimos la **integral** de f **sobre** E como:

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

Definición 1.3.15. Decimos que una sucesión de funciones integrables $\{f_n\}$ **converge en la media** o **converge en media** a una función integrable f si:

$$\psi(f_n, f) = \int |f_n - f| d\mu \longrightarrow 0, \quad \text{cuando} \quad n \longrightarrow \infty.$$

Teorema 1.3.16. *Sea $\{f_n\}$ una sucesión de funciones integrables que converge en la media a f, entonces $\{f_n\}$ converge a f en medida.*

Demostración. Si, para cualquier $\epsilon > 0$ fijo, se cumple:

$$E_n = \{x \mid |f_n(x) - f(x)| \geq \epsilon\},$$

entonces:

$$\int |f_n - f| d\mu \geq \int_{E_n} |f_n - f| d\mu \geq \epsilon\mu(E_n),$$

tal que $\mu(E_n) \longrightarrow 0$ cuando $n \longrightarrow \infty$. □

Teorema 1.3.17. *Sea f una función integrable:*

1. *sea E un conjunto de medida cero, entonces:*

$$\int_E f d\mu = 0$$

2. *si f es positiva casi en todas partes sobre un conjunto medible E y si $\int_E f d\mu = 0$, entonces $\mu(E) = 0$.*

3. *si $\int_F f d\mu = 0$ para cada conjunto medible F, entonces $f = 0$ casi en todas partes.*

Demostración. 1. Dado que:

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

y dado que la ecuación característica de un conjunto de medida cero desaparece casi en todas partes se sigue el resultado deseado.

2. Escribimos:

$$F_0 = \{x \mid f(x) > 0\} \qquad \text{y} \qquad F_n = \left\{x \mid f(x) \geq \frac{1}{n}\right\},$$

$n = 1, 2, \ldots$, dado que la suposición de positividad implica que $E - F_0$ es un conjunto de medida cero, podemos simplemente demostrar que $E \cap F_0$ también es un conjunto de medida cero. Dado que:

$$\int_{E \cap F_n} f d\mu \geq \frac{1}{n}\mu(E \cap F_n) \geq 0,$$

y $F_0 = \bigcup_{n=1}^{\infty} F_n$, el resultado deseado se sigue de la relación:

$$\mu(E \cap F_0) \leq \sum_{n=1}^{\infty} \mu(E \cap F_n).$$

3. Si $E = \{x \mid f(x) > 0\}$, entonces por hipótesis $\int_E f d\mu = 0$, y por inciso inmediato anterior, E es un conjunto de medida cero. Aplicando el mismo resultado a $-f$ se muestra que $\{x \mid f(x) < 0\}$ es un conjunto de medida cero. □

Si f es una función medible tal que $f \geq 0$ casi en todas partes y si f no es integrable, entonces

$$\int f d\mu = \infty,$$

la clase de función más importantes de esta forma son las funciones f en las cuales al menos una de las dos funciones f^+ o f^- es integrable, entonces:

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

dado que al menos una de las dos integrables es infinita, el valor de $\int f d\mu$ es siempre $+\infty$, $-\infty$ o un número real.

Teorema 1.3.18. *Sea $\{f_n\}$ una sucesión de funciones simples fundamentales en media que convergen en medida a una función integrable f, entonces:*

$$\psi(f, f_n) = \int |f - f_n| d\mu \longrightarrow 0 \qquad \textit{cuando} \qquad n \longrightarrow \infty.$$

Por lo tanto, para cada función integrable f y para cada $\epsilon > 0$, corresponde una función simple integrable g tal que $\psi(f, g) < \epsilon$.

Demostración. Para cualquier entero fijo m, $\{|f_n - f_m\}$ es una sucesión fundamental en media de funciones simples integrables la cual converge en medida a $|f - f_m|$, y por lo tanto:

$$\int |f - f_m| d\mu = \lim_{n\to\infty} \int |f_n - f_m| d\mu.$$

El hecho que la sucesión $\{f_n\}$ fundamental en media implica el resultado deseado.

□

Teorema 1.3.19. *Sea $\{f_n\}$ una sucesión de funciones integrables fundamentales en media, entonces existe una función integrable f tal que $\psi(f_n, f) \longrightarrow 0$, y por consecuencia $\int f_n d\mu \longrightarrow \int f d\mu$, cuando $n \longrightarrow \infty$.*

Demostración. Por el teorema 1.3.18, para cada entero positivo n existe una función simple integrable g_n, tal que $\psi(f_n, g_n) < \frac{1}{n}$. Se sigue que $\{g_n\}$ es una sucesión fundamental en media de funciones simples integrables, sea f una función medible, por lo tanto integrable, tal que $\{g_n\}$ converge en medida a f. Dado que:

$$0 \leq \left| \int f_n d\mu - \int f d\mu \right| \leq \int |f_n - f| d\mu = \psi(f_n, f) \leq \cdots$$

$$\cdots \leq \psi(f_n, g_n) + \psi(g_n, f),$$

y el resultado deseado se sigue de 1.3.18. □

Teorema 1.3.20. *Sea f una función medible, g una función integrable y si $|f| \leq |g|$ casi en todas partes, entonces f es integrable.*

Demostración. Considerando las partes positiva y negativa de f muestra que es suficiente demostrar el teorema para las funciones no negativas de f. Si f es una función simple, el resultado es claro. En el caso general existe una sucesión creciente $\{f_n\}$ de funciones simples no negativas tal que $\lim_{n\to\infty} f_n(x) = f(x)$ para todo $x \in X$. Dado que $0 \leq f_n \leq |g|$, cada f_n es integrable y el resultado deseado se sigue de la convergencia acotada. $\square$

Teorema 1.3.21. *Sea $\{f_n\}$ una sucesión creciente de funciones medibles no negativas y si $\lim_n f_n(x) = f(x)$ casi en todas partes, entonces:*

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

Demostración. Si f es una función integrable, el resultado se sigue de la convergencia acotada y del teorema 1.3.20. La nueva característica del teorema es la aplicación del caso no necesariamente integrable. Tenemos que demostrar que si $\int f d\mu = \infty$, entonces $\lim_{n\to\infty} \int f_n d\mu = \infty$, o en otras palabras, que si $\lim_{n\to\infty} \int f_m d\mu < \infty$, entonces f es integrable. De la finitud del límite, podemos concluir que:

$$\lim_{m,n\to\infty} \left| \int f_m d\mu - \int f_n d\mu \right| = 0.$$

Dado que $f_m - f_n$ es de signo constante para cada valor fijo de m, n, tenemos que:

$$\left| \int f_m d\mu - \int f_n d\mu \right| = \int |f_m - f_n| d\mu,$$

así que la sucesión $\{f_n\}$ converge en media y por teorema 1.3.19, converge a una función integrable g. Dado que convergencia en media implica convergencia en medida, y por lo tanto casi en todas partes para alguna subsucesión, tenemos que $f = g$ casi en todas partes. $\square$

2. ESPACIOS DE HILBERT

2.1. Espacios con producto interno

Definición 2.1.1. Un **Espacio con producto interno** es un espacio vectorial X con un producto interno definido sobre X. Un **producto interno** sobre X es una aplicación: $\langle \cdot, \cdot \rangle \colon X \times X \longrightarrow \mathbb{C}$, esto es, para cada par de vectores x y y existe un escalar asociado, $\langle x, y \rangle$, llamado producto interno de x y y, tal que para todos los vectores x, y, z y escalares α se cumple que:

1. $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$,

2. $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$,

3. $\langle x, y \rangle = \overline{\langle y, x \rangle}$,

4. $\langle x, x \rangle \geq 0$,

5. $\langle x, y \rangle = 0 \Leftrightarrow x = 0$.

Definición 2.1.2. Una **norma** sobre un espacio vectorial complejo X es una aplicación: $\|\cdot\| \colon X \longrightarrow \mathbb{R}$, que cumple las siguientes propiedades:

1. $\|x\| \geq 0$,

2. $\|x\| = 0 \Leftrightarrow x = 0$,

3. $\|\alpha x\| = |\alpha| \|x\|$,

4. $\|x + y\| \leq \|x\| + \|y\|$ (desigualdad del triángulo).

Un producto interno sobre X define una norma sobre X dada por:

$$\|x\| = \sqrt{\langle x, x \rangle}.$$

Lema 2.1.1. *En un espacio con producto interno X, su producto interno y su norma correspondiente satisfacen la **desigualdad de Cauchy-Schwarz** y la **ley del paralelogramo**, es decir, para cada $x, y \in X$ se cumple:*

1. $|\langle x, y\rangle| \leq \|x\|\|y\|$ *(desigualdad de Cauchy-Schwarz),*

2. $\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$.

Demostración. Consultar [13, p. 137]. □

Lema 2.1.2. *La norma inducida por producto interno cumple todas las propiedades de una norma.*

Demostración.

1. Tenemos por la definición 2.1.2 inciso 4, $\langle x, x\rangle \geq 0$, por lo cual se cumple que $\|x\| \geq 0$.

2. De la definición 2.1.2 inciso 5, $\langle x, x\rangle = 0 \Leftrightarrow x = 0$, luego $\|x\| = \sqrt{\langle x, x\rangle} = 0$.

3. De la definición 2.1.2 inciso 2 y 4, tenemos que $\|\alpha x\| = \sqrt{\langle \alpha x, \alpha x\rangle} = \sqrt{\alpha^2 \langle x, x\rangle} = |\alpha|\|x\|$.

4. Usando la definición 2.1.2 y la desigualdad de Cauchy-Schawarz en el lema 2.1.1, tenemos:

$$\|x+y\|^2 = \langle x+y, x+y\rangle = \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle =$$

$$= \|x\|^2 + 2|\langle x, y\rangle| + \|y\|^2 \leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2,$$

de donde $\|x+y\| \leq \|x\| + \|y\|$. □

Definición 2.1.3. Un elemento x de un espacio con producto punto X es **ortogonal** a un elemento $y \in X$ si:

$$\langle x, y\rangle = 0.$$

También decimos que x y y son **ortogonales** y lo denotamos $x \perp y$. De manera similar, para subconjuntos $A, B \subset X$ escribimos $x \perp A$ si $x \perp a$ para todo $a \in A$ y $A \perp B$ si $a \perp b$ para todo $a \in A$ y para todo $b \in B$.

Definición 2.1.4. Una sucesión $\{x_n\}$ de elementos de un espacio con producto interno X es **convergente en norma** a un $x \in X$ si $\|x_n - x\| \longrightarrow 0$ cuando $n \longrightarrow \infty$.

Lema 2.1.3. *Sean* $\{x_n\}$ *y* $\{y_n\}$ *sucesiones de elementos de un espacio con producto interno* X*, tales que* $\|x_n - x\| \longrightarrow 0$ *y* $\|y_n - y\| \longrightarrow 0$*, entonces:*

1. $\|x_n\| \longrightarrow \|x\|$,

2. $\langle x_n, y_n \rangle \longrightarrow \langle x, y \rangle$.

Demostración. 1. De la desigualdad del triángulo tenemos que $||x|| \leq ||x-y|| + ||y||$ y $||y|| \leq ||y-x|| + ||x||$, juntando estos dos resultados tenemos:

$$||x-y|| \geq |\ ||x|| - ||y||\ |,$$

con lo cual queda demostrado.

2. Tenemos que:

$$|\langle x_n, y_n \rangle - \langle x, y \rangle| = |\langle x_n, y_n \rangle + \langle x_n - x, y \rangle| \leq |\langle x_n, y_n - y \rangle| + |\langle x_n - x, y \rangle| \leq$$

$$\leq ||x_n|| ||y_n - y|| + ||x_n - x|| ||y||,$$

la última linea se sigue de la desigualdad de Cauchy-Schwarz. Observemos que del inciso inmediato anterior $||x_n|| \to ||x||$ tenemos que:

$$|\langle x_n, y_n \rangle - \langle x, y \rangle| \longrightarrow 0 \qquad \text{cuando} \qquad n \longrightarrow \infty.$$

□

2.2. Espacios de Hilbert

Definición 2.2.1. Una sucesión $\{x_n\}_{n \in \mathbb{Z}^+}$ de elementos de un espacio con producto interno X es una **sucesión de Cauchy** si:

$$\|x_n - x_m\| \longrightarrow 0,$$

cuando $m, n \longrightarrow \infty$, es decir, para todo $\epsilon > 0$ existe un entero positivo $N(\epsilon)$ tal que:

$$\|x_n - x_m\| < \epsilon,$$

para todo $m, n > N(\epsilon)$.

Definición 2.2.2. Un **subespacio** de un espacio vectorial X es un conjunto no vacío Y de X tal que para todo y_1, y_2 y todos los escalares α, β tenemos que $\alpha y_1 + \beta y_2 \in Y$. Por lo que Y es en si mismo un espacio vectorial.

Definición 2.2.3. Un **espacio métrico** es un espacio vectorial X con una métrica definida sobre X. Una **métrica** es una aplicación: $d\colon X \times X \longrightarrow \mathbb{R}$, que cumple las siguientes propiedades:

1. $d \geq 0$,

2. $d(x,y) = 0$ si y sólo si $x = y$,

3. $d(x,y) = d(y,x)$,

4. $d(x,y) \leq d(x,z) + d(z,y)$.

Definición 2.2.4. Dado un punto $x_0 \in X$ y un número real $r > 0$, definimos los siguientes conjuntos:

1. **Bola abierta**: $B(x_0; r) = \{x \in X \mid d(x, x_0) < r\}$.

2. **Bola cerrada**: $\tilde{B}(x_0; r) = \{x \in X \mid d(x, x_0) \leq r\}$.

3. **Esfera**: $S(x_0; r) = \{x \in X \mid d(x, x_0) = r\}$.

Definición 2.2.5. Un subconjunto M de un espacio métrico X es **abierto** si contiene una bola para cada uno de sus puntos. Un subconjunto k de X es **cerrado** si su complemento, es X, es abierto.

Definición 2.2.6. Una bola abierta $B(x_0; \epsilon)$ de radio ϵ es llamada **ϵ-vecindad** de x_0. Por una **vecindad** de x_0 nos referimos a cualquier subconjunto de X que contiene una ϵ-vecindad de x_0.

Definición 2.2.7. Sea M un subconjunto de un espacio métrico X. Entonces un punto $x_0 \in X$, que puede o no ser un punto de M, es llamado **punto de acumulación** de M, o punto límite de M, si cada vecindad de x_0, contiene al menos un punto $y \in M$ distinto de x_0. El conjunto que consiste de todos los puntos de M y sus puntos de acumulación es llamado **clausura** de M y es denotado como $\overline{M}$.

Teorema 2.2.1. *Sea M un subconjunto no vacío de un espacio métrico (X, d) y su clausura $\overline{M}$. Entonces:*

1. *$x \in \bar{M}$ si y sólo si existe una sucesión $\{x_n\} \in M$ tal que $x_n \longrightarrow x$.*

2. *M es **cerrado** si y sólo si se cumple que si $x_n \in M$, $x_n \longrightarrow x$ implica que $x \in M$.*

Demostración.

1. Sea $x \in \overline{M}$. Si $x \in M$, una sucesión de ese tipo es $(x, x, \dots)$. Si $x \notin M$, es un punto de acumulación de M. Por lo que para cada $n = 1, 2, \dots$, la bola abierta $B(x; \frac{1}{n})$ contiene un $x_n \in M$, $x_n \longrightarrow x$ ya que $\frac{1}{n} \longrightarrow 0$ cuando $n \longrightarrow 0$.

2. M es cerrado si y sólo si $M = \overline{M}$, de tal forma que 2. se sigue del inciso inmediato anterior. □

Definición 2.2.8. Un subconjunto M de un espacio métrico $(X, d) = X$ es **denso** en X si:

$$\overline{M} = X.$$

Decimos que X es **separable** si tiene un subconjunto contable que es denso en X.

Definición 2.2.9. Un **espacio de Hilbert** H es un espacio con producto interno que es completo, es decir, es un espacio con producto interno en el cual toda sucesión de Cauchy converge en norma a un elemento $x \in H$.

Teorema 2.2.2. *Sea Y un subespacio de un espacio de Hilbert H. Entonces:*

1. *Y es completo si y sólo si Y es cerrado en H.*

2. *Si Y es de dimensión finita, entonces Y es completo.*

3. *Si H es separable, también Y. De manera general, cada subconjunto de un espacio separable con producto interno es separable.*

Demostración. Consultar [13, p. 140]. □

Definición 2.2.10. Un subespacio M de un espacio de Hilbert H, es un **subespacio cerrado** de H si M contiene todos sus puntos de acumulación.

Definición 2.2.11. El **complemento ortogonal** de un subconjunto M de un espacio de Hilbert H es definido como el conjunto $M^{\perp}$ de todos los elementos de H que son ortogonales a cada elemento de M. Así que: $x \in M^{\perp}$ si y sólo si $x \perp Y$ para todo $y \in M$.

Teorema 2.2.3. *Sea M cualquier subconjunto de un espacio de Hilbert H, entonces $M^{\perp}$ es un subespacio cerrado de H.*

Demostración. De la definición 2.1.1 inciso 5. es sencillo verificar que $0 \in M^{\perp}$. Además si $x, y \in M^{\perp}$ entonces todas sus combinaciones lineales pertenecen a $M^{\perp}$. Por lo tanto $M^{\perp}$ es un subespacio de H. Si $\|x_n - x\| \longrightarrow 0$, entonces por continuidad de producto interno, lema 2.1.3, $\langle x, y \rangle = 0$ para todo $y \in M$, por lo que $x \in M^{\perp}$ y $M^{\perp}$ es cerrado. □

2.3. Proyección en espacios de Hilbert

Definición 2.3.1. Un **operador** (o tranformación, aplicación) T de A en Y es obtenido al asociar cada elemento $x \in A$ un único elemento $y \in Y$, lo escribimos $y = Tx$ y lo llamamos la **imagen** de x respecto a T. El conjunto A es llamado el **dominio** de T y lo denotamos como $\mathscr{D}(T)$ y lo escribimos como:

$$T\colon \mathscr{D}(T) \longrightarrow Y,$$

$$x \longmapsto Tx.$$

El **rango** $\mathscr{R}(T)$ de T es el conjunto de todas las imágenes; así

$$\mathscr{R}(T) := \{y \in Y | y = Tx \text{ para algún } x \in \mathscr{D}(t)\}.$$

Definición 2.3.2. La **imagen inversa** de un $y_0 \in Y$ es el conjunto de todos los $x \in \mathscr{D}(T)$ tal que $Tx = y_0$. Similarmente la imagen inversa de un subconjunto $Z \subset Y$ es el conjunto de todos $x \in \mathscr{D}(T)$ tal que $Tx \in Z$.

Definición 2.3.3. Un operador T es **inyectivo** si para cada $x_1, x_2 \in \mathscr{D}(T)$,

$$x_1 \neq x_2 \text{ implica } Tx_1 \neq Tx_2.$$

Un operador T es **sobreyectivo** si $\mathscr{R}(T) = Y$. T es **biyectivo** si T es inyectivo y sobreyectivo. Entonces el **operador inverso** T^{-1} de $T : \mathscr{D}(T) \longrightarrow Y$ es el operador $T^{-1} : Y \longrightarrow \mathscr{D}(T)$ definido como $Tx_0 \longmapsto x_0$, esto es, T^{-1} asocia cada elemento $y_0 \in Y$ con un $x_0 \in \mathscr{D}(T)$ para el cual $Tx_0 = y_0$.

Definición 2.3.4. Sean X y Y espacio con norma y $T\colon \mathscr{D}(T) \longrightarrow Y$ un operador lineal, donde $\mathscr{D}(T) \subset X$. El operador T es **acotado** si existe un número real c tal que para todo $x \in \mathscr{D}(T)$,

$$\|Tx\| \leq c\|x\|.$$

Definición 2.3.5. Un operador lineal acotado $T\colon H \longrightarrow H$ sobre un espacio de Hilbert H es **auto-adjunto** si $T^* = T$, donde T^* es el operador **adjunto de Hilbert**, es decir, para todo $x, y \in H$:

$$\langle Tx, y\rangle = \langle x, T^*y\rangle.$$

Teorema 2.3.1 (Teorema de proyección). *Si M es un subespacio cerrado de un espacio de Hilbert H y sea $x \in H$, entonces:*

1. *existe un único elemento $\hat{x} \in M$ tal que:*

$$\|x - \hat{x}\| = \inf_{y \in M} \|x - y\|.$$

2. *$\hat{x} \in M$ y $\|x - \hat{x}\| = \inf_{y \in M} \|x - y\|$ si y sólo si $\hat{x} \in M$ y $(x - \hat{x}) \in M^{\perp}$.*

Demostración. Para mostrar 1 tenemos: sea $d = \inf_{y \in M} \|x - y\|^2$ entonces existe una sucesión $\{y_n\}$ de elementos de M tal que $\|y_n - x\|^2 \longrightarrow d$. Aplicando la ley del paralelogramo y usando el hecho de que $(y_m + y_n)/2 \in M$, tenemos que:

$$0 \leq \|y_m - y_n\|^2 = -4\|\frac{y_m + y_n}{2} - x\|^2 + 2(\|y_n - x\|^2 + \|y_m - x\|^2) \leq$$

$$\leq -4d + 2(\|y_n - x\|^2 + \|y_m - x\|^2) \longrightarrow 0, \qquad \text{cuando} \quad m, n \longrightarrow \infty.$$

Por el criterio de Cauchy, existe un $\hat{x} \in H$ tal que $||y_m - \hat{x}|| \longrightarrow 0$. Dado que M es cerrado sabemos que $\hat{x} \in M$, y por continuidad del producto interno:

$$||x - \hat{x}||^2 = \lim_{n \to \infty} ||x - y_n||^2 = d.$$

Para establecer la unicidad, supongamos que $\hat{y} \in M$ y que $||x - \hat{y}||^2 = ||x - \hat{x}||^2 = d$. Entonces, aplicando la ley del paralelogramo, tenemos:

$$0 \leq ||\hat{x} - \hat{y}||^2 = -4||\frac{\hat{x} + \hat{y}}{2} - x||^2 + 2(||\hat{x} - x||^2 + ||\hat{y} - x||^2) \leq -4d + 4d = 0.$$

Por lo que $\hat{y} = \hat{x}$.

Para mostrar 2: Si $\hat{x} \in M$ y $(x - \hat{x}) \in M^{\perp}$ entonces $\hat{x}$ es el único elemento de M definido en 1, dado que para cualquier $y \in M$:

$$||x - y||^2 = \|x - \hat{x} + \hat{x} - y, x - \hat{x} + \hat{x} - y\| =$$

$$= ||x - \hat{x}||^2 + ||\hat{x} - y||^2 \geq ||x - \hat{x}||^2,$$

la igualdad se cumple si y sólo si $y = \hat{x}$. Por otro lado si $\hat{x} \in M$ y $(x - \hat{x}) \notin M^{\perp}$, entonces $\hat{x}$ no es elemento de M más cercano a x, dado que:

$$\tilde{x} = \hat{x} + \frac{ay}{||y||^2},$$

es más cercano, donde y es cualquier elemento de M tal que $\|x - \hat{x}, y\| \neq 0$ y $a = \|x - \hat{x}, y\|$, tenemos que:

$$\begin{aligned}
||x - \tilde{x}||^2 &= \|x - \hat{x} + \hat{x} - \tilde{x}, x - \hat{x} + \hat{x} - \tilde{x}\| \\
&= ||x - \hat{x}||^2 + \frac{|a|^2}{||y||^2} + 2Re\|x - \hat{x}, \hat{x} - \tilde{x}\| \\
&= ||x - \hat{x}||^2 + \frac{|a|^2}{||y||^2} \\
&< ||x - \hat{x}||^2.
\end{aligned}$$

□

El elemento $\hat{x}$ es llamado **la proyección ortogonal** de x sobre M.

Teorema 2.3.2. *Un operador lineal acotado $P\colon H \longrightarrow H$ sobre un espacio de Hilbert H es una proyección si y sólo si P es auto-ajunto e idempotente, es decir, $P^2 = P$.*

Demostración. Para la primera parte de la demostración, supongamos que P es una proyección de H y denotemos $P(H)$ como Y. Entonces $P^2 = P$ porque para cada $x \in H$ y $Px = y \in Y$, tenemos que:

$$P^2 x = Py = y = Px.$$

Además, sea $x_1 = y_1 + z_1$ y $x_2 = y_2 + z_2$, donde $y_1, y_2 \in Y$ y $z_1, z_2 \in Y^\perp$. Entonces $\langle y_1, z_2\rangle = \langle y_2, z_1\rangle = 0$, ya que $Y \perp Y^\perp$, y lo auto-adjunto de P se sigue de:

$$\langle Px_1, x_2\rangle = \langle y_1, y_2 + z_2\rangle = \langle y_1, y_2\rangle = \langle y_1 + z_1, y_2\rangle = \langle x_1, Px_2\rangle.$$

Para la segunda parte de la demostración, supongamos que $P^2 = P = P^*$ y denotemos $P(H)$ como Y. Entonces para cada $x \in H$:

$$x = Px + (I - P)x.$$

La ortogonalidad de $Y \perp (I - P)(H)$ se sigue de:

$$\langle Px, (I - P)v\rangle = \langle x, P(I - P)v\rangle = \langle x, Pv - P^2 v\rangle = \langle x, 0\rangle = 0.$$

Y es el espacio nulo $N(I - P)$ de $(I - P)$, porque $Y \subset N(I - P)$ se sigue de:

$$(I - P)Px = Px - P^2 x = 0,$$

y $N(I-P) \subset Y$ se sigue si notamos que $(I-P)x = 0$ implica $x = Px$. Por lo que Y es cerrado. Finalmente $P|_Y$ es el operador identidad de Y dad que si $y = Px$, tenemos que $Py = P^2x = Px = y$. $\square$

Corolario 2.3.1. *Sea M un subespacio cerrado de un espacio de Hilbert H e I el operador identidad sobre H, entonces existe un único operador P_M de H sobre M tal que $I - P_M$ mapea H sobre $M^\perp$. P_M es llamado el **operador proyección** de H sobre M.*

Demostración. Por la demostración del teorema 2.3.1 tenemos que para cada $x \in H$ existe un único $\hat{x} \in M$ tal que $x - \hat{x} \in M^\perp$. El mapeo requerido es por lo tanto:

$$P_M x = \hat{x}, \qquad x \in H.$$ $\square$

Teorema 2.3.3. *Sea H un espacio de Hilbert y sea P_M el operador de proyección sobre un subespacio cerrado M. Entonces:*

1. $P_M(\alpha x + \beta y) = \alpha P_M x + \beta P_M y$, *para* $x, y \in H$, $\alpha, \beta \in \mathbb{C}$,
2. $\|x\|^2 = \|P_M x\|^2 + \|(I - P_M)x\|^2$,
3. *cada $x \in H$ tiene una única representación como suma de un elemento de M y un elemento de $M^\perp$, es decir, $x = P_M x + (I - P_M)x$,*
4. $P_M x_n \to P_M x$ *si* $\|x_n - x\| \to 0$,
5. $x \in M$ *si y sólo si* $P_M x = x$,
6. $x \in M^\perp$ *si y sólo si* $P_M x = 0$,
7. $M_1 \subseteq M_2$ *si y sólo si* $P_{M_1} P_{M_2} x = P_{M_1} x$ *para todo* $x \in H$.

Demostración.

1. $\alpha P_M x + \beta P_M y \in M$ dado que M es un subespacio lineal de H, además:

$$\alpha x + \beta y - (\alpha P_M x + \beta P_M y) = \alpha(x - P_M x) + \beta(y - P_M y) \in M^\perp,$$

dado que $M^\perp$ es un subespacio lineal de H. Estas dos propiedades identifican $\alpha P_M x + \beta P_M y$ como la proyección $P_M(\alpha x + \beta y)$.

2. Es una consecuencia de la ortogonalidad de $P_M x$ y $(I - P_M)x$.

3. Tenemos que una representación es $x = P_M x + (I - P_M)x$. Si $x = y + z$, $y \in M$, $z \in M^{\perp}$ es otra, entonces:

$$y - P_M x + z - (I - P_M)x = 0.$$

Tomando producto interno de ambos lados con $y - P_M x$ nos da $\|y - P_M x\|^2 = 0$, dado que $z - (I - P_M)x \in M^{\perp}$. Por lo tanto $y = P_M x$ y $z = (I - P_M)x$.

4. Del inciso 2, tenemos que $\|P_M(x_n - x)\|^2 \leq \|x_n - x\|^2 \longrightarrow 0$ si $\|x_n - x\|^2 \longrightarrow 0$.

5. $x \in M$ si y sólo si la única representación $x = y + z$, $y \in M$, $z \in M^{\perp}$, es tal que $y = x$ y $z = 0$, es decir, si y sólo si $P_M x = x$.

6. Usando el inciso 5, con $y = 0$ y $z = 0$.

7. $x = P_{M_2} + (I - P_{M_2})x$. Proyectando ambos lados sobre M_1 tenemos que:

$$P_{M_1} x = P_{M_1} P_{M_2} x + P_{M_1}(I - P_{M_2})x.$$

Por lo que $P_{M_1} x = P_{M_1} P_{M_2} x$ para todo $x \in H$, si y sólo si $P_M y = 0$ para todo $y \in M_2^{\perp}$, es decir, si y sólo si $M_1 \subset M_2$. $\square$

Definición 2.3.6. El **span cerrado**, $\overline{\text{span}}\{x_t, t \in T\}$, de cualquier subconjutno $\{x_t, t \in T\}$ de un espacio de Hilbert H es el subespacio cerrado más pequeño que contiene cada elemento $x_t, t \in T$.

Definición 2.3.7. Un conjunto $\{e_t, t \in T\}$ de elemento de un espacio con producto interno es **ortonormal** si para cada $s, t \in T$:

$$\langle e_s, e_t \rangle = \begin{cases} 1, & \text{si} \quad s = t \\ 0, & \text{si} \quad s \neq t \end{cases}$$

Teorema 2.3.4. *Sea $\{e_1, ..., e_k\}$ un subconjunto ortonormal de un espacio con producto interno H y $M = \overline{\text{span}}\{e_1, ..., e_k\}$, entonces:*

1. $P_M x = \sum_{i=1}^{k} \langle x, e_i \rangle e_i$ *para todo* $x \in H$,

2. $\|P_m x\|^2 = \sum_{i=1}^{k} |\langle x, e_i \rangle|^2$ *para todo* $x \in H$,

3. $\|x - \sum_{i=1}^{k} \langle x, e_i \rangle e_i\| \leq \|x - \sum_{i=1}^{k} c_i e_i\|$ *para todo* $x \in H$,

4. para todo $c_1, ..., c_k \in \mathbb{C}$, la desigualdad se mantiene si y sólo si $c_i = \langle x, e_i \rangle$, $i = 1, ..., k$. Los números $\langle x, e_i \rangle$ son llamados a veces los ***coeficientes de Fourier*** *de x relativos al conjunto $\{e_1, ..., e_k\}$.*

Demostración. 1. Para mostrarlo es suficiente mostrar que $P_M x$ satisface la ecuación de predicción, es decir:

$$\Big\langle \sum_{i=1}^{k} \langle x, e_i \rangle e_i, e_j \Big\rangle = \langle x, e_i \rangle, j = 1, \ldots, k.$$

Pero esto es una consecuencia inmediata de la definición 2.3.7.

2. Usando propiedades del producto interno y asumiendo la ortonormalidad de $\{e_1, \ldots, e_k\}$ obtenemos el resultado deseado.

3. Del teorema 2.3.1 tenemos que $\|x - P_M x\| \leq \|x - y\|$ para todo $y \in M$, dicha desigualdad es la deseada. La igualdad existe si y sólo si:

$$\sum_{i=1}^{k} c_i e_i = P_M x = \sum_{i=1}^{k} \langle x, e_i \rangle e_i.$$

4. Tomando producto interno en ambos con e_j, en ambos lados de la igualdad anterior, y asumiendo la ortonormalidad, tenemos que la igualdad anterior es equivalente a $c_j = \langle x, e_j \rangle$, $j = 1, \ldots, k$. □

Ejemplo 2.3.1. A continuación presentamos ejemplos de los coeficientes de Fourier y de la proyección de un vector.

1. Sea V el espacio de funciones continuas sobre $[-\pi, \pi]$. Sea f la función dada por $f(x) = \sin kx$, donde k es un entero mayor a cero. Entonces:

$$\|f\| = \langle f, f \rangle^{1/2} = \left(\int_{-\pi}^{pi} \sin^2 kx dx \right)^{1/2} = \sqrt{\pi}.$$

En el presente ejemplo de un espacio vectorial de funciones, la componente de g sobre f es llamada **coeficiente de Fourier** de g **respecto de** f. Sea g cualquier función continua sobre $[-\pi, \pi]$, entonces el coeficiente de Fourier de g con respecto a f es:

$$\frac{\langle f, g \rangle}{\langle f, f \rangle} = \frac{1}{\pi} \int_{-\pi}^{pi} g(x) \sin kx dx.$$

2. En el siguiente diagrama tenemos un ejemplo de la proyección de un vector sobre un plano en $\mathbb{R}^3$.

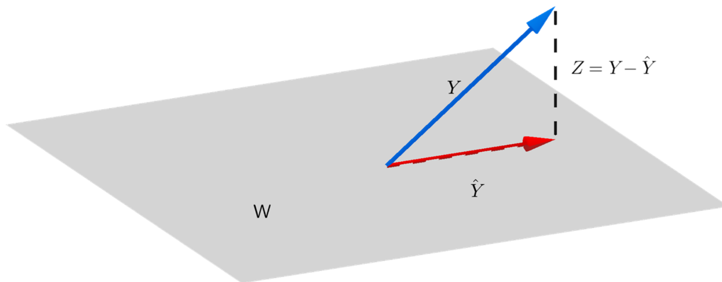

Figura 2.1. Proyección ortogonal en $\mathbb{R}^3$ de Y sobre el plano W.
Fuente: elaboración propia con el programa Geogebra.

El vector proyección de la gráfica anterior, $\hat{Y}$, viene dado por:

$$\hat{Y} = Y - \hat{Y}_N = Y - \frac{\langle Y, N\rangle}{\|N\|^2}N,$$

donde el vector N es el **vector normal al plano** W.

3. TEORÍA DE PROBABILIDADES

Los conceptos de *probabilidad* se empezaron a desarrollar desde antes del siglo XVI, los cuales surgen para fines de juegos de azar, sin darle una estructura matemática que sustentara todos los resultados. No fue hasta que se desarrolló la teoría de la medida en el siglo XX cuando se le dió una estructura matemática para formalizar todos sus resultados. Antes de iniciar el estudio de la teoría de probabilidades haremos una pequeña comparación entre conceptos de teoría de la medida, ver capítulo 1, y conceptos de teoría de probabilidades y posteriormente definiremos cada uno de esos conceptos y sus propiedades fundamentales.

3.1. Teoría de la medida y probabilidad

En la siguiente tabla podemos observar algunos terminos de la teoría de la medida y como estos se conocen en la teoría de probabilidades, así como también bajo cuales condiciones se presentan.

Tabla 1: Conceptos de teoría de la medida aplicables en teoría de probabilidad

Teoría de la medida	**Teoría de probabilidades**
Espacio medible, $(X, \mathbf{S}, \mu)(\mu(X) = 1)$	Espacio de probabilidad, $(\Omega, \mathcal{F}, \mathrm{Pr})$
Conjuntos medibles	Eventos
$(\sigma$-)álgebra	$(\sigma$-)campo
Función medible de valor real f	Variable aleatoria X
Integral de f, $\int f d\mu$	Esperanza o media de X, $\mathrm{E}(X)$
Convergencia en medida	Convergencia en probabilidad
Casi en todas partes	Casi seguramente

3.2. Conceptos fundamentales

Empezaremos el estudio de esta sección con los axiomas de probabilidad de Kolmogorov, los cuales nos definen lo que conocemos como espacio de probabilidad.

Posteriormente daremos las principales propiedades de medidas de probabilidad para finalizar la sección con la definición de independencia de eventos en un espacio de probabilidad.

Definición 3.2.1. Un **espacio de probabilidad** es un espacio de medida $(\Omega, \mathcal{F}, \Pr)$, donde: Ω es un conjunto no vacío, $\mathcal{F}$ es un σ-campo sobre Ω, cuyos elementos son llamados **eventos**, Pr es una **medida o función de probabilidad** sobre $\mathcal{F}$, es decir, cumple las siguientes propiedades, para cualquier evento $A \in \mathcal{F}$, $\Pr(A)$ es llamada la **probabilidad** de A:

1. Pr es no negativa, es decir, $\Pr(A) \geq 0$, para todo evento A.
2. Pr es normado, es decir, $\Pr(\Omega) = 1$.
3. Pr es contablemente aditiva, para toda colección disjunta de eventos $\{A_i\}$, se cumple que $\Pr(\sum_i A_i) = \sum_i \Pr(A_i)$.

Las anteriores propiedades son conocidas como los **axiomas de probabilidad de Kolmogorov**.

Definición 3.2.2. Dado que $\mathcal{F}$ es una σ-campo y de la definición de Pr se siguen las siguientes propiedades:

1. $\Pr(\emptyset) = 0$, cualquier otro evento distinto del vacio cuya probabilidad sea igual a cero es llamado **evento nulo**.
2. Pr es finitamente aditivo, es decir, que para cada evento A_j, $j = 1, 2, \ldots, n$ tal que $A_i \cap A_j$, $i \neq j$:
$$\Pr\left(\sum_{j=1}^{n} A_j\right) = \sum_{j=1}^{n} \Pr(A_j).$$
3. El **complemento** del evento A, $A^c := \Omega \backslash A$ está en $\mathcal{F}$, y su probabilidad viene dada por:
$$\Pr(A^c) = 1 - \Pr(A).$$
4. Pr no es decreciente, es decir, si $A_1 \subseteq A_2$ entonces $\Pr(A_1) \leq \Pr(A_2)$.
5. $0 \leq \Pr(A) \leq 1$ para cualquier evento A.
6. Para cualesquiera eventos A_1, A_2, se cumple:
$$\Pr(A_1 \cup A_2) = \Pr(A_1) + \Pr(A_2) - \Pr(A_1 \cap A_2).$$

7. Pr es subaditiva, es decir:

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

Teorema 3.2.1. *Para cualquier número de eventos finitos tenemos que:*

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \Pr(A_i) - \sum_{1\leq i_1<i_2\leq n} (\Pr(A_{i_1} \cap A_{i_2}))+$$

$$+ \sum_{1\leq i_1<i_2<i_3\leq n} (\Pr(A_{i_1} \cap A_{i_2} \cap A_{i_3})) - \ldots + (-1)^{n+1} \Pr(A_1 \cap A_2 \cap \ldots \cap A_n).$$

Demostración. Para demostrar el teorema usaremos inducción sobre n. Para $n = 1$ el resultado es trivial, de la definición 3.2.2 inciso 6 tenemos el caso para $n = 2$. Ahora asumamos que el resultado es válido para $n = k$, y demostremos el caso para $n = k + 1$. Tenemos que:

$$\Pr\left(\bigcup_{j=1}^{k+1} A_j\right) = \Pr\left(\left(\bigcup_{j=1}^{k} A_j\right) \cup A_{k+1}\right) =$$

$$= \Pr\left(\bigcup_{j=1}^{k} A_j\right) + \Pr(A_{k+1}) - \Pr\left(\left(\bigcup_{j=1}^{k} A_j\right) \cap A_{k+1}\right) =$$

$$= \left[\sum_{j=1}^{k} \Pr(A_j) - \sum_{1\leq j_1<j_2\leq k} \Pr(A_{j_1} \cap A_{j_2}) + \sum_{1\leq j_1<j_2<j_3\leq k} \Pr(A_{j_1} \cap A_{j_2} \cap A_{j_3}) - \cdots +\right.$$

$$\left. +(-1)^{k+1} \Pr(A_1 \cap A_2 \cap \cdots \cap A_k)\right] + \Pr(A_{k+1}) - \Pr\left(\bigcup_{j=1}^{k} (A_j \cap A_{k+1})\right) =$$

$$= \sum_{j=1}^{k+1} \Pr(A_j) - \sum_{1\leq j_1<j_2\leq k} \Pr(A_{j_1} \cap A_{j_2}) + \sum_{1\leq j_1<j_2<j_3\leq k} \Pr(A_{j_1} \cap A_{j_2} \cap A_{j_3}) - \cdots +$$

$$+ (-1)^{k+1} \Pr(A_1 \cap A_2 \cap \cdots \cap A_k) - \Pr\left(\bigcup_{j=1}^{k} (A_j \cap A_{k+1})\right). \tag{3.1}$$

Pero:

$$\Pr\left(\bigcup_{j=1}^{k} (A_j \cap A_{k+1})\right) = \sum_{j=1}^{k} (A_j \cap A_{k+1}) - \sum_{1\leq j_1<j_2\leq k} \Pr(A_{j_1} \cap A_{j_2} \cap A_{k+1})+$$

$$+\sum_{1\le j_1<j_2<j_3\le k} \Pr(A_{j_1}\cap A_{j_2}\cap A_{j_3}\cap A_{k+1})-$$
$$-\cdots+(-1)^k \sum_{1\le j_1<j_2<\ldots<j_{k-1}\le k} \Pr(A_{j_1}\cap\cdots\cap A_{j_{k-1}}\cap A_{k+1})+$$
$$+(-1)^{k+1}\Pr(A_1\cap\cdots\cap A_k\cap A_{k+1}).$$

Reemplazando lo anterior en 3.1, tenemos que:

$$\Pr\left(\bigcup_{j=1}^{k+1} A_j\right)=\sum_{j=1}^{k+1}\Pr(A_j)-\left[\sum_{1\le j_1<j_2\le k}\Pr(A_{j_1}\cap A_{j_2})+\sum_{j=1}^{k}\Pr(A_j\cap A_{k+1})\right]+$$
$$+\left[\sum_{1\le j_1<j_2<j_3\le k}\Pr(A_{j_1}\cap A_{j_2}\cap A_{j_3})+\sum_{1\le j_1<j_2\le k}\Pr(A_{j_1}\cap A_{j_2}\cap A_{k+1})\right]-$$
$$-\cdots+(-1)^{k+1}\left[\Pr(A_1\cap\cdots\cap A_k)+\sum_{1\le j_1<j_2<\ldots<j_{k-1}\le k}\Pr(A_{j_1}\cap\cdots\cap A_{j_{k-1}}\cap A_{k+1})\right]+$$
$$+(-1)^{k+2}\Pr(A_1\cap\cdots\cap A_k\cap A_{k+1})=\sum_{j=1}^{k}\Pr(A_j)-\sum_{1\le j_1<j_2\le k+1}\Pr(A_{j_1}\cap A_{j_2})+$$
$$+\sum_{1\le j_1<j_2<j_3\le k+1}\Pr(A_{j_1}\cap A_{j_2}\cap A_{j_3})-\cdots+(-1)^{k+2}\Pr(A_1\cap\cdots\cap A_{k+1}).\qquad\square$$

Teorema 3.2.2. *Sea $\{A_i\}$ una sucesión de eventos tales que, cuando $n\to\infty$, $A_n\uparrow$ o $A_n\downarrow$, entonces:*

$$\Pr\left(\lim_{n\to\infty} A_n\right)=\lim_{n\to\infty}\Pr(A_n).$$

Demostración. Para realizar la demostración primero asumamos que $A_n\uparrow$. Entonces:

$$\lim_{n\to\infty} A_n=\bigcup_{j=1}^{\infty} A_j.$$

Como $A_n\uparrow$, tenemos que:

$$\bigcup_{j=1}^{\infty} A_j=A_1=(A_1^c\cap A_2)+(A_1^c\cap A_2^c\cap A_3)+\cdots=A_1+(A_2-A_1)+(A_3-A_2)+\ldots.$$

Por lo que:

$$\Pr(\lim_{n\to\infty} A_n)=\Pr\left(\bigcup_{j=1}^{\infty} A_j\right)=\Pr(A_1)+\Pr(A_2-A_1)+\Pr(A_3-A_2)+\cdots+$$

$$+\Pr(A_n - A_{n-1}) + \cdots = \lim_{n\to\infty} \Big[\Pr(A_1) + \Pr(A_2 - A_1) + \ldots \Pr(A_n - A_{n-1}) \Big] =$$

$$= \lim_{n\to\infty} \Big[\Pr(A_1) + \Pr(A_2) - \Pr(A_1) + \Pr(A_3) - \Pr(A_2) + \cdots + \Pr(A_n) - \Pr(A_{n-1}) \Big] =$$

$$= \lim_{n\to\infty} \Pr(A_n).$$

Así que:

$$\Pr\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \Pr(A_n).$$

Ahora sea $A_n \downarrow$. Entonces $A_n^c \uparrow$, así que:

$$\lim_{n\to\infty} A_n^c = \bigcup_{j=1}^{\infty} A_j^c.$$

Por lo que:

$$\Pr\left(\lim_{n\to\infty} A_n^c\right) = \Pr\left(\bigcup_{j=1} A_j^c\right) = \lim_{n\to\infty} \Pr(A_n^c),$$

o de manera equivalente:

$$\Pr\left[\left(\bigcap_{j=1}^{\infty} A_j\right)^c\right] = \lim_{n\to\infty} \Big[1 - \Pr(A_n)\Big], \qquad \text{o} \qquad 1 - \Pr\left(\bigcap_{j=1}^{\infty} A_j\right) = 1 - \lim_{n\to\infty} \Pr(A_n).$$

Así que:

$$\lim_{n\to\infty} \Pr(A_n) = \Pr\left(\bigcap_{j=1}^{\infty} A_j\right) = \Pr\left(\lim_{n\to\infty} A_n\right). \qquad \square$$

Definición 3.2.3. Sea A un evento tal que $\Pr(A) > 0$, entonces la **probabilidad condicional**, dado A, es la función (de conjunto) denotado por $\Pr(\cdot|A)$ y se define para cada evento B tal que:

$$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)}.$$

Teorema 3.2.3. *Sea $\{A_i\}$ una sucesión de eventos tales que:*

$$\Pr\left(\bigcap_{i=1}^{n-1} A_i\right) > 0,$$

entonces:

$$\Pr\left(\bigcap_{i=1}^{n-1} A_i\right) = \Pr\big(A_n | A_1 \cap A_2 \cap \cdots \cap A_{n-1}\big) \times$$

$$\times \Pr\left(A_{n-1}|A_1 \cap A_2 \cap \cdots \cap A_{n-2}\right) \times \cdots \times \Pr(A_2|A_1) \times \Pr(A_1).$$

Demostración. Consultar [17, p. 23]. $\square$

Definición 3.2.4. Sea $\{A_i\}$ una sucesión de eventos de un espacio muestral Ω, tales que $A_i \cap A_j = \emptyset$, para $i \neq j$, y $\sum_i A_i = \Omega$, entonces decimos que la sucesión $\{A_i\}$ es una **partición** de Ω. La partición es *finita* o infinita de acuerdo si los eventos A_i son finitos o infinitos.

Teorema 3.2.4. *Sea* $\{A_i\}$,$i = 1, 2, \ldots$, *una partición de* Ω *con* $\Pr(A_i) > 0$, *entonces para todo* $B \in \mathcal{A}$, *tenemos que:*

$$\Pr(B) = \sum_i \Pr(B|A_i)P(A_i).$$

Demostración. Como $\{A_i\}$,$i = 1, 2, \ldots$, es una partición de Ω, tenemos que $A_i \cap A_j = \emptyset$, con $i \neq j$ y $\sum_j A_j = \Omega$. Para cualquier evento B tenemos que:

$$B = \sum_j (B \cap A_j).$$

Por lo tanto:

$$\Pr(B) = \sum_j (B \cap A_j) = \sum_j \Pr(B|A_j)\Pr(A_j).$$ $\square$

Consideremos ahora dos experimentos $\mathcal{E}_1$ y $\mathcal{E}_2$ con respectivos espacios vectoriales $(\Omega_1, \mathcal{A}_1, \Pr_1)$ y $(\Omega_2, \mathcal{A}_2, \Pr_2)$, entonces el experimento compuesto $(\mathcal{E}_1, \mathcal{E}_2) = \mathcal{E}_1 \times \mathcal{E}_2$ tiene espacio muestral $\Omega = \Omega_1 \times \Omega_2$. Para obtener el σ-campo apropiado $\mathcal{A}$ de eventos en Ω, primero definimos:

$$C = \{A_1 \times A_2 \mid A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\},$$

donde:

$$A_1 \times A_2 = \{(s_1, s_2) \mid s_1 \in A_1, s_2 \in A_2\}.$$

Entonces $\mathcal{A}$ es el σ-campo generado por C. Ahora definamos en C la función de conjuntos denotada como $\Pr(A_1 \times A_2) = \Pr(A_1)\Pr(A_2)$, la cual define una única medida de probabilidad sobre $\mathcal{A}$. Esta medida de probabilidad es usualmente denotada por $\Pr_1 \times \Pr_2$ y es llamada **medida de probabilidad producto** y su espacio de probabilidad $(\Omega, \mathcal{A}, \Pr)$ es llamada **espacio de probabilidad producto**. Los eventos de $\mathcal{E}_1$ son de la forma $B_1 = A_1 \times \Omega_1, A_1 \in \mathcal{A}_1$ y los eventos de $\mathcal{E}_2$ son de

la forma $B_2 = A_2 \times \Omega_2, A_2 \in \mathcal{A}_2$. Los experimentos $\mathcal{E}_1$ y $\mathcal{E}_2$ son **independientes** si $\Pr(B_1 \cap B_2) = \Pr(B_1)\Pr(B_2)$ para todos los eventos B_1 y B_2.
Para n experimentos $\mathcal{E}_i$, con $i = 1, 2, \ldots, n$, con correspondientes espacios de probabilidad $(\Omega_i, \mathcal{A}_i, \Pr_i)$, el experimento compuesto $(\mathcal{E}_1, \ldots, \mathcal{E}_n) = \mathcal{E}_1 \times \cdots \times \mathcal{E}_n$ tiene espacio de probabilidad $(\Omega, \mathcal{A}, \Pr)$, donde:

$$\Omega = \Omega_1 \times \cdots \times \Omega_n = \{(s_1, \ldots, s_2) \mid s_i \in \Omega_i, i = 1, 2, \ldots, n\},$$

$\mathcal{A}$ es el σ-campo generado por la clase de conjuntos C, donde:

$$C = \{A_1 \times \cdots \times A_n \mid A_i \in \mathcal{A}_i, i = 1, 2, \ldots, n\},$$

y Pr es la única medida de probabilidad definida en $\mathcal{A}$ a través de la relación:

$$\Pr(A_1 \times \cdots \times A_n) = \Pr(A_1) \times \cdots \times \Pr(A_n),$$

con $A_i \in \mathcal{A}_i, i = 1, 2, \ldots, n$. La medida de probabilidad Pr es usualmente denotada por $\Pr_1 \times \ldots \times \Pr_n$ y es llamada **medida de probabilidad producto** y su espacio de probabilidad $(\Omega, \mathcal{A}, \Pr)$ es llamado **espacio de probabilidad producto**. Los experimentos $\mathcal{E}_i$, $i = 1, 2, \ldots, n$, son independientes si $\Pr(B_1 \cap \cdots \cap B_n) = \Pr(B_1) \times \cdots \times \Pr(B_n)$, donde los B_i están definidos por:

$$B_i = \Omega_1 \times \cdots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \cdots \times \Omega_n,$$

con $i = 1, 2, \ldots, n$. La definición de independencia de eventos se sigue de los siguiente: sean $\mathcal{A}_1$, $\mathcal{A}_2$ dos σ-campos de $\mathcal{A}$. Decimos que $\mathcal{A}_1$, $\mathcal{A}_2$ son independientes si $\Pr(A_1 \cap A_2) = \Pr(A_1) \times \Pr(A_2)$ para cualquier $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$. De manera más general, los σ-campos $\mathcal{A}_i$, $i = 1, 2, \ldots, n$, σ-campos de $\mathcal{A}$, son independientes si:

$$\Pr\left(\bigcap_{1=1}^{n} A_i\right) = \prod_{i=1}^{n} \Pr(A_i),$$

para cualquier $A_i \in \mathcal{A}_i$, $i = 1, 2, \ldots, n$.

3.3. Variables Aleatorias

En esta sección estudiaremos las principales propiedades de las variables aleatorias. En la tabla de la sección **3.1** observamos que podemos considerar a las variables aleatorias como funciones medibles de valores reales, las cuales fueron definidas en el capítulo 1, con ayuda de eso fundamentaremos las variables aleatorias dentro de un espacio de probabilidad, dicho de otra forma podemos considerar a las variables aleatorias como un caso especial de las funciones medibles de valor real.

Sea $(\Omega, \mathcal{A}, \Pr)$ un espacio de probabilidad, $\mathcal{T}$ un espacio y X una función definida sobre Ω hacia $\mathcal{T}$, es decir, $X\colon \Omega \to \mathcal{T}$. Para $T \in \mathcal{T}$, definimos la imagen inversa de T, bajo X, denotado por $X^{-1}(T)$, como:

$$X^{-1}(T) = \{s \in \Omega \mid X(s) \in \mathcal{T}\}.$$

Este conjunto también es denotado como $[X \in T]$ o $(X \in T)$. Con esta definición, y el hecho de que X es una función, tenemos de forma inmediata que las siguientes propiedades se cumplen:

$$X^{-1}\Big(\bigcup_i T_i\Big) = \bigcup_i X^{-1}(T_i). \tag{3.2}$$

$$\text{Si} \qquad T_1 \cap T_2 = \emptyset, \text{entonces} X^{-1}(T_1) \cap X^{-1}(T_2) = \emptyset. \tag{3.3}$$

De estas dos propiedades, tenemos:

$$X^{-1}\left(\sum_i T_i\right) = \sum_i X^{-1}(T_i), \tag{3.4}$$

$$X^{-1}\left(\bigcap_i T_i\right) = \bigcap_i X^{-1}(T_i), \tag{3.5}$$

$$X^{-1}(T^c) = \left[X^{-1}(T)\right]^c, \tag{3.6}$$

$$X^{-1}(\mathcal{T}) = \Omega, \tag{3.7}$$

$$X^{-1}(\emptyset) = \emptyset. \tag{3.8}$$

Sea $\mathcal{D}$ un σ-campo de subconjuntos de $\mathcal{T}$ y definamos la clase $X^{-1}(\mathcal{D})$ de subconjuntos de Ω como:

$$X^{-1}(\mathcal{D}) = \{A \subseteq \Omega \mid A = X^{-1}(T) \text{ para algún } T \in \mathcal{D}\}.$$

Con las propiedades 3.2, 3.6 y 3.7 hemos demostrado el siguiente teorema, el cual nos indica la razón por la cual necesitamos medida en nuestra definición de variables aleatorias, ya que garantiza que la función de distribución de probabilidad de un vector aleatorio está bien definido.

Teorema 3.3.1. *La clase $X^{-1}(\mathcal{D})$ es un σ-campo de subconjuntos de Ω.*

Definición 3.3.1. Si $X^{-1}(\mathcal{D}) \subseteq \mathcal{A}$ entonces decimos que X es $(\mathcal{A}, \mathcal{D})$-medible, o simplemente medible. Si $(\mathcal{T}, \mathcal{D}) = (\mathbb{R}, \mathcal{B})$ y X es $(\mathcal{A}, \mathcal{B})$-medible, decimos que X es una **variable aleatoria**. De forma más general, si $(\mathcal{T}, \mathcal{D}) = (\mathbb{R}^k, \mathcal{B}^k)$, donde $\mathbb{R}^k = \mathbb{R} \times \mathbb{R} \times ... \times \mathbb{R}$, k copias de $\mathbb{R}$, y X es $(\mathcal{A}, \mathcal{B}^k)$-medible, decimos que X es un vector aleatorio de **k**-dimensión. Una variable aleatorio es un vector aleatorio unidimensional.

Teorema 3.3.2. *Sea C^* una clase de subconjuntos de $\mathcal{T}$ de la forma $C^* = \{T \subseteq \mathcal{T} \mid X^{-1}(T) = A$ para algún $A \in \mathcal{A}\}$. Entonces C^* es un σ-campo.*

Demostración. El teorema se demustra con las propiedades 3.2 a 3.8. □

Corolario 3.3.1. *Sea $\mathcal{D} = \sigma(C)$, donde C es una clase de subconjuntos de $\mathcal{T}$. Entonces X es $(\mathcal{A}, \mathcal{D})$-medible si y sólo si $X^{-1}(C) \subseteq \mathcal{A}$. En particular, X es una variable aleatoria si y sólo si $X^{-1}(C_o)$, $X^{-1}(C_i)$ o $X^{-1}(C_i') \subseteq \mathcal{A}$, $i = 1, 2, \ldots, 8$, y similarmente para cualquier caso de vectores aleatorios k-dimensionales.*

Demostración. El σ-campo C^* del teorema 3.3.2 tiene la propiedad de $C \subseteq C^*$. Entonces $\sigma(C) = \mathcal{D} \subseteq C^*$ y por lo tanto $X^{-1}(\mathcal{D}) \subseteq X^{-1}(C^*)$. Pero $X^{-1}(C^*) \subseteq \mathcal{A}$. Así que $X^{-1}(\mathcal{D}) \subseteq \mathcal{A}$. El converso es una consecuencia directa de la definición de $(\mathcal{A}, \mathcal{D})$ es medible. □

Con la definición de un vector aleatorio $X : (\Omega, \mathcal{A}, \Pr) \to (\mathbb{R}, \mathcal{B}^k)$, definamos la función de conjuntos P_x como:

$$P_x(B) = \Pr\left[X^{-1}(B)\right] = \Pr(X \in B) = \Pr\left(\left\{s \in \Omega | X(s) \in B\right\}\right).$$

Por el corolario 3.3.1, los conjuntos $X^{-1}(B) \in \Omega$ son en realidad eventos dado que asumimos que X es un vector aleatorio. Por lo que P_x está bien definido al igual que $\Pr[X^- - 1(B)]$. Con lo que mostramos que P_x es una medida de probabilidad sobre $\mathcal{B}^k$. De hecho, $P_x(B) \geq 0$, $B \in \mathcal{B}^k$, ya que Pr es una medida de probabilidad. Luego, $P_x(\mathbb{R}^k) = \Pr\left[X^{-1}(\mathbb{R}^k)\right] = \Pr(\Omega) = 1$, y por último:

$$P_x\left(\sum_{i=1}^{\infty} B_i\right) = \Pr\left[X^{-1}\left(\sum_{i=1}^{\infty} B_i\right)\right] = \Pr\left[\sum_{i=1}^{\infty} X^{-1}(B_i)\right] =$$

$$= \sum_{i=1}^{\infty} \Pr\left[X^{-1}(B_i)\right] = \sum_{i=1}^{\infty} P_x(B_i).$$

La medida de probabilidad P_x es conocida como **función de distribución de probabilidad**, o distribución, de X.

Definición 3.3.2. Sea $P_x(B)$ una función de distribución de probabilidad y consideremos que B es un intervalo en $\mathbb{R}^k$, es decir, $B = \{y \in \mathbb{R}^k, y \leq x\}$, en el sentido de que si $x = (x_1, \ldots, x_k)'$ y $y = (y_1, \ldots, y_k)$, entonces $y_i \leq x_i$, $i = 1, \ldots, k$. Para esta elección de B, denotamos $P_x(B)$ como $F_x(X)$ y es llamada la **función de distribución acumulada** de X.

Teorema 3.3.3. *La distribución F de una variable aleatoria X satisface las siguientes propiedades:*

1. $0 \leq F(x) \leq 1$, $x \in \mathbb{R}$.

2. F *es no decreciente.*

3. F *es continua a la derecha.*

4. $F(x) \to 0$ *cuando* $x \to -\infty$ *y* $F(x) \to 1$ *cuando* $x \to +\infty$.

Demostración. En la presente demostración colocamos a Q para la distribución de P_X de X, por motivo de simplicidad. Tenemos que:

1. El resultado es obvio.

2. Esto significa que $x_1 < x_2$ implica que $F(x_1) < F(x_2)$, ciertamente:

$$x_1 < x_2 \qquad \text{implica} \qquad (-\infty, x_1] \subset (-\infty, x_2],$$

y por lo tanto:

$$Q(-\infty, x_1] \leq Q(-\infty, x_2]; \qquad \text{equivalentemente} \qquad F(x_1) \leq F(x_2).$$

3. Esto significa que, si $x_n \downarrow x$, entonces $F(x_n) \downarrow F(x)$. ciertamente:

$$x_n \downarrow x \qquad \text{implica} \qquad (-\infty, x_n] \downarrow (-\infty, x],$$

y por lo tanto:

$$Q(-\infty, x_n] \downarrow Q(-\infty, x],$$

por teorema 3.2.2; equivalentemente $F(x_n) \downarrow F(x)$.

4. Sea $x \longrightarrow -\infty$ y asumimos que $x_n \downarrow -\infty$. Entonces:

$$(-\infty, x_n] \downarrow \emptyset \qquad \text{así que} \qquad Q(-\infty, x_n] \downarrow Q(\emptyset) = 0,$$

por teorema 3.2.2. Equivalentemente, $F(x_n) \longrightarrow 0$. De manera similar, si $x_n \longrightarrow +\infty$. Asumiendo que $x_n \uparrow \infty$. Entonces:

$$(-\infty, x_n] \uparrow \mathbb{R} \qquad \text{y por lo tanto} \qquad Q(-\infty, x_n] \uparrow Q(\mathbb{R}) = 1;$$

equivalentemente $F(x_n) \longrightarrow 1$. □

Generalizando la definición 1.3.2 y el teorema 1.3.3, a espacios euclidianos de dimensión mayor tenemos:

Definición 3.3.3. Una función $g\colon S \subseteq \mathbb{R}^k \to \mathbb{R}^m$, $k, m \geq 1$, es continua en $x_0 \in \mathbb{R}^k$ si para cada $\epsilon > 0$ existe un $\delta = \delta(\epsilon, x_0) > 0$ tal que $\|x - x_0\| < \epsilon$ implica $\|g(x) - g(x_0)\| < \delta$. La función g es continua en S si es continua para todo $x \in S$.

Teorema 3.3.4. *Sea $g\colon \mathbb{R}^k \to \mathbb{R}^m$ continua, entonces g es medible.*

Demostración. La demostración es similar a la demostración del teorema 1.3.3. Es suficiente mostrar que $g^{-1}(G)$ son conjuntos medibles para todos los cubos abiertos $G \in \mathbb{R}^m$. Sea $B = g^{-1}(G)$. Si $B = \emptyset$, el teorema es válido, así que supongamos que $B \neq \emptyset$ y sea x_0 un punto arbitrario de B. Continuidad de g en x_0 implica que para cada $\epsilon > 0$ existe un $\delta = \delta(\epsilon, x_0) > 0$ tal que $\|x - x_0\| < \epsilon$ implica $\|g(x) - g(x_0)\| < \delta$; equivalentemente, $x \in S(x_0, \epsilon)$ implica $g(x) \in S(g(x_0), \delta)$, donde $S(c, r)$ es una esfera abierta con centro en c y radio r. Dado que $g(x_0) \in G$ y G es abierto, podemos elegir un ϵ lo suficientemente pequeño de tal forma que el δ correspondiente sea lo suficientemente pequeño para implicar que $g(x) \in S(g(x_0), \delta)$. Así que, para tal elección de ϵ y δ, $x \in S(x_0, \epsilon)$ implica que $g(x) \in S(g(x_0), \delta)$. Dado que $B = g^{-1}(G)$ es el conjunto de todos los $x \in \mathbb{R}^k$ para los cuales $g(x) \in G$ y $x \in S(x_0, \epsilon)$ implica que $g(x) \in S(g(x_0), \delta)$, se sigue que $S(x_0, \epsilon) \subset B$. En este punto, observemos que es claro que exista un cubo que contenga a x_0 y completamente contenido en $S(x_0, \epsilon)$; llamemoslo $C(x_0, \epsilon)$. Entonces $C(x_0, \epsilon) \subset B$, y por lo tanto B es abierto y por teorema 1.3.1 es medible. □

Teorema 3.3.5. *Sea $X\colon (\Omega, \mathcal{A}) \to (\mathbb{R}^k, \mathcal{B}^k)$ un vector aleatorio y sea $g\colon (\mathbb{R}^k, \mathcal{B}^k) \to (\mathbb{R}^m, \mathcal{B}^m)$ medible. Entonces $g(X)\colon (\Omega, \mathcal{A}) \to (\mathbb{R}^m, \mathcal{B}^m)$ es un vector aleatorio, es decir, funciones medibles de vectores aletarios son vectores aleatorio.*

Demostración. Para demostrar que $[g(X)(B)]^{-1}(B) \in \mathcal{A}$ si $B \in \mathcal{B}^m$, tenemos que:

$$\big[g(X)\big]^{-1}(B) = X^{-1}\big[g^{-1}(B)\big] = X^{-1}(B_1) \qquad \text{donde} \qquad B_1 = g^{-1}(B) \in \mathcal{B}^k,$$

por la medibilidad de g. Además, $X^{-1}(B_1) \in \mathcal{A}$ dado que X es medible. $\square$

Corolario 3.3.2. *Sea X como en el teorema anterior y g continua. Entonces $g(X)$ es un vector aleatorio, es decir, funciones continuas de vectores aleatorios son vectores aleatorios.*

Demostración. La continuidad de g implica su mediblidad por teorema 1.3.3 y aplicando el teorema 3.3.5 se demuestra el resultado. $\square$

Definición 3.3.4. Para $i = 1, \dots, k$, la *i-ésima* función proyección g_i es definida como $g_i \colon \mathbb{R}^k \to \mathbb{R}$ y $g_i(x) = g_i(x_1, \dots, x_k) = x_i$.

Teorema 3.3.6. *Las funciones coordenadas g_i, $i = 1, \dots, k$, definidas como antes, son continuas.*

Demostración. Para un punto arbitrario $x_0 \in \mathbb{R}^k$, consideremos $x \in \mathbb{R}^k$ tal que $\|x - x_0\| < \epsilon$ para algún $\epsilon > 0$. Esto es equivalente a $\|x - x_0\|^2 < \epsilon^2$ o $\sum_{j=1}^{k}(x_j - x_{0_j})^2 < \epsilon^2$ lo cual implica que $(x_j - x_{0_j})^2 < \epsilon^2$ para $j = 1, \dots, k$ o $|x_j - x_{0_j}| < \epsilon$, $j = 1, \dots, k$. Esta última expresión es equivalente a $|g_j(x_j) - g_j(x_{0_j})| < \epsilon$, $j = 1, \dots, k$. Por lo que la definición de continuidad de g_j se cumple para $\delta = \epsilon$. $\square$

Teorema 3.3.7. *Sea $X = (X_1, \dots, X_k)' \colon (\Omega, \mathcal{A}) \to (\mathbb{R}^k, \mathcal{B}^k)$. Entonces X es un vector aleatorio si y sólo si X_i, $i = 1, \dots, k$ son variables aleatorias.*

Demostración. Supongamos que X es un vector aleatorio y sean g_j, $j = 1, \dots, k$ las funciones coordenadas definidas sobe $\mathbb{R}+$. Entonces las g_j son continuas por teorema 3.3.6 y por lo tanto medibles por teorema 3.3.4. Entonces para cada $j = 1, \dots, k$, $g_j(X) = g_j(X_1, \dots, X_k) = X_j$ es medible y por lo tanto una variable aleatoria.

Luego asumamos que X_j, $j = 1, \dots, k$ son variables aleatorias. Para demostrar que X es un vector aleatorio, es suficiente mostrar que $X^{-1}(B) \in \mathcal{A}$ para cada $B = (-\infty, x_1] \times \cdots \times (-\infty, x_k]$, $x_1, \dots, x_k \in \mathbb{R}$. En efecto:

$$X^{-1}(B) = (X \in B) = (X_j \in (-\infty, x_j], j = 1, \dots, k) = \bigcap_{j=1}^{k} X_j^{-1}((-\infty, x_j]) \in \mathcal{A}. \ \square$$

3.4. Momentos de Variables Aleatorias

Definición 3.4.1. Sea $X = (X_1, \dots, X_k)'$ un vector aleatorio con función de distribución de probabilidad f y consideremos la función medible $g\colon \mathbb{R}^k \to \mathbb{R}$, es decir que $g(X) = g(X_1, \dots, X_k)$ es una variable aleatoria, tenemos que:

1. Para $n = 1, 2, \dots$, el **n-ésimo momento** de $g(X)$, denotado como $\mathrm{E}[g(X)]^n$, es:
$$\mathrm{E}[g(X)]^n = \sum_x \big[g(X)\big]^n f(x), x = (x_1, \dots, x_k)'.$$
Para $n = 1$ tenemos:
$$\mathrm{E}[g(X)] = \sum_x g(x) f(x),$$
a la expresión anterior la llamamos **esperanza matemática** o **valor medio** de $g(X)$. Otra notación para $\mathrm{E}[g(X)]$ que es usada comúnmente es $\mu_{g(X)}$, $\mu[g(X)]$ o simplemente μ.

2. Para $r > 0$, el **r-ésimo momento absoluto** de $g(X)$ es denotado por $\mathrm{E}|g(X)|^r$ y es definido por:
$$\mathrm{E}|g(X)|^r = \sum_x |g(x)|^r f(x), x = (x_1, \dots, x_k)'.$$

3. Para una constante arbitraria c, el n-ésimo momento y el r-ésimo momento absoluto de $g(X)$ sobre c son denotados como $E[g(X) - c]^n$, $E|g(X) - c|^r$, respectivamente, son definidos como:
$$\mathrm{E}[g(X) - c]^n = \sum_x [g(x) - c]^n f(x), x = (x_1, \dots, x_k)',$$
y
$$\mathrm{E}|g(x) - c|^r = \sum_x |g(x) - c|^r f(x), x = (x_1, \dots, x_k)'.$$
Para $c = \mathrm{E}[g(X)]$, los momentos son llamados **momentos centrales**. El segundo momento central de $g(X)$, esto es:
$$\mathrm{E}\{g(X) - \mathrm{E}[g(X)]\}^2 = \sum_x [g(x) - \mathrm{E}g(X)]^2 f(x), x = (x_1, \dots, x_k)',$$
es llamado la **varianza** de $g(X)$, también es denotado como $\sigma^2[g(X)]$, $\sigma^2_{g(x)}$ o

simplemente σ^2. La cantidad $+\sqrt{\sigma^2[g(X)]} = \sigma[g(X)]$ es llamada la **desviación estándar** de $g(X)$, también es denotada como $\sigma_{g(X)}$ o simplemente σ.

Caso especial:

Sea $g(X_1, \ldots, X_k) = X_1^{n_1} \cdots X_k^{n_k}$, $n_i \geq 0$ son enteros. Entonces $E(X_1^{n_1} \cdots X_k^{n_k})$ es llamada el $(n_1, \ldots, n_k)$**-momento conjunto** de $X_1, \ldots, X_k$. En particular, para $n = \cdots = n_{i-1} = n_{i+1} = \cdots = n_k = 0$, $n_i = n$, tenemos:

$$\mathrm{E}(X_i^n) = \sum_x x_i^n f(x) = \sum_{(x_1, \ldots, x_k)} x_i^n f(x_1, \ldots, x_k),$$

la cual es el n-ésimo momento de la variable aleatoria X_i. Por lo tanto el n-ésimo momento de una variable aleatoria X con función de distribución de probabilidad f es:

$$\mathrm{E}(X^n) = \sum_x x^n f(x).$$

Para $n = 1$ tenemos:

$$\mathrm{E}(X) = \sum_x x f(x),$$

que es la **esperanza matemática** o **valor medio** de X, también denotada como μ_X, $\mu(X)$ o μ.

Para $g(X_1, \ldots, X_k) = X_1^{n_1} \cdots X_k^{n_k}$, $n = \cdots = n_{i-1} = n_{i+1} = \cdots = n_k = 0$, $n_i = n$, y $c = E(X_i)$ tenemos:

$$\mathrm{E}(X_i - \mathrm{E}X_i)^n = \sum_x (x_i - \mathrm{E}X_i)^n f(x), x = (x_1, \ldots, x_k)',$$

es llamado el **n-ésimo momento central** de la variable aleatoria X_i, o el **n-ésimo momento** de X_i sobre su media. Por lo que el n-ésimo momento central de una variable aleatoria X con función de distribución de probabilidad f y media μ es:

$$\mathrm{E}(X - \mathrm{E}X)^n = \mathrm{E}(X - \mu)^n = \sum_x (x - \mathrm{E}X)^n f(x) = \sum_x (x - \mu)^n f(x).$$

Cuando $n = 2$ el segundo momento central de X se conoce como **varianza**, denotada como σ_X^2, $\sigma^2(X)$ o σ, y su raíz positiva es llamada **desviación estándar** de X.

Definición 3.4.2. La esperanza matemática de una variable aleatoria, $\mathrm{E}[g(X)]$, cumple las siguientes propiedades:

1. $\mathrm{E}(c) = c$, donde c es una constante.

2. $\mathrm{E}[cg(X)] = c\mathrm{E}[g(X)]$, y en particular $\mathrm{E}(cX) = c\mathrm{E}(X)$ si X es una variable aleatoria.

3. $\mathrm{E}[g(X+d)] = \mathrm{E}[g(X)]+d$, donde d es una constante. En particular $\mathrm{E}(X+d) = \mathrm{E}(X) + d$ si X es una variable aleatoria.

4. $\mathrm{E}\left[\sum_{i=1}^{n} c_i g_i(X)\right] = \sum_{i=1}^{n} c_i \mathrm{E}[g_i(X)]$. En particular $\mathrm{E}\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i \mathrm{E}(X_i)$.

5. Si $X \geq 0$, entonces $\mathrm{E}(X) \geq 0$.

6. Si $X \geq Y$, entonces $\mathrm{E}(X) \geq \mathrm{E}(Y)$, donde X, Y son variables aleatorias, con esperanzas finitas.

7. $|\mathrm{E}[g(X)]| \leq \mathrm{E}|g(X)|$.

8. Si $\mathrm{E}|X|^r < \infty$ para algún $r > 0$, donde X es una variable aleatoria, entonces $\mathrm{E}|X|^{r'} < \infty$ para todo $0 < r' < r$.

9. Si $\mathrm{E}(X^n)$ existe, es decir, $E(X^n) < \infty$, para algún $n = 2, 3, \ldots,$ entonces $\mathrm{E}(X^{n'})$ existe, para todo $n' = 1, 2, \ldots,$ con $n' < n$.

Definición 3.4.3. La varianza de una variable aleatoria, $\sigma^2[g(X)]$, cumple las siguientes propiedades:

1. $\sigma^2(c) = 0$, donde c es una constante.

2. $\sigma^2[cg(X)] = c^2\sigma^2[g(X)]$, en particular, $\sigma^2(cX) = c^2\sigma^2(X)$, si X es una variable aleatoria.

3. $\sigma^2[g(X)+d] = \sigma^2[g(X)]$, donde d es una constante. En particular $\sigma^2(X+d) = \sigma^2(X)$, si X es una variable aleatoria.

4. $\sigma^2[g(X)] = E[g(X)]^2 - [Eg(X)]^2$. En particular $\sigma^2(X) = E(X^2) - (E(X))^2$, si X es una variable aleatoria.

5. $\sigma^2(X) = E[X(X-1)] + EX - (EX)^2$, si X es una variable aleatoria.

Definición 3.4.4. Sea X una variable aleatoria y $g(X) = |X - \mu|^r$, $\mu = E(X)$, $r > 0$, entonces:

$$\Pr\left[|X-\mu| \geq c\right] = \Pr\left[|X-\mu|^r \geq c^r\right] \leq \frac{\mathrm{E}|X-\mu|^r}{c^r},$$

que es conocida como **desigualdad de Markov**. Si substituimos r por 2, obtenemos:

$$\Pr\left[|X-\mu| \geq c\right] = \Pr\left[|X-\mu|^2 \geq c^2\right] \leq \frac{\mathrm{E}|X-\mu|^2}{c^2} = \frac{\sigma^2(X)}{c^2} = \frac{\sigma^2}{c^2},$$

que es conocida como **desigualdad de Tchebichev**. En particular, si $c = k\sigma$, entonces:

$$\Pr\left[|X-\mu| \geq k\sigma\right] \leq \frac{1}{k^2}.$$

Teorema 3.4.1. *Sea X, Y variables aleatorias tales que:*

$$\mathrm{E}(X) = \mathrm{E}(Y) = 0, \quad \sigma^2(X) = \sigma^2(Y) = 1,$$

entonces:

$$\mathrm{E}^2(XY) \leq 1, \ \textit{equivalentemente,} \ -1 \leq \mathrm{E}(XY) \leq 1,$$

y

$$\mathrm{E}(XY) = 1, \ \textit{si y sólo si,} \ \Pr(Y = X) = 1,$$

$$\mathrm{E}(XY) = -1, \ \textit{si y sólo si,} \ \Pr(Y = -X) = 1.$$

Demostración. Tenemos que:

$$0 \leq \mathrm{E}(X-Y)^2 = \mathrm{E}(X^2 - 2XY + Y^2) = \mathrm{E}X^2 - 2\mathrm{E}(XY) + \mathrm{E}Y^2 = 2 - 2\mathrm{E}(XY).$$

Y:

$$0 \leq \mathrm{E}(X+Y)^2 = \mathrm{E}(X^2 + 2XY + Y^2) = \mathrm{E}X^2 + 2\mathrm{E}(XY) + \mathrm{E}Y^2 = 2 + 2\mathrm{E}(XY).$$

Por lo tanto $\mathrm{E}(XY) \leq 1$ y $-1 \leq \mathrm{E}(XY)$, así que $-1 \leq \mathrm{E}(XY) \leq 1$. Ahora sea $\Pr(Y = X) = 1$. Entonces $\mathrm{E}(XY) = \mathrm{E}Y^2 = 1$ y si $\Pr(Y = -X) = 1$, entonces $\mathrm{E}(XY) = -\mathrm{E}Y^2 = -1$. Conversamente, sea $\mathrm{E}(XY) = 1$. Entonces:

$$\sigma^2(X-Y) = \mathrm{E}(X-Y)^2 - [\mathrm{E}(X-Y)]^2 = \mathrm{E}(X-Y)^2 =$$

$$= \mathrm{E}X^2 - 2\mathrm{E}(XY) + \mathrm{E}Y^2 = 1 - 2 + 1 = 0,$$

así que $\Pr(X = Y) = 1$. Finalmente, sea $\mathrm{E}(XY) = -1$. Entonces $\sigma^2(X+Y) = 2 + 2\mathrm{E}(XY) = 2 - 2 = 0$, así que $\Pr(X = -Y) = 1$. □

Teorema 3.4.2. *Sean X, Y variables aleatorias con medias μ_1, μ_2 y varianzas, positivas, σ_1^2, σ_2^2 respectivamente. Entonces se cumple la siguiente igualdad, conocida* *como desigualdad de Cauchy-Schwarz:*

$$\mathrm{E}^2\big[(X-\mu_1)(Y-\mu_2)\big] \leq \sigma_1^2\sigma_2^2,$$

equivalentemente:

$$-\sigma_1\sigma_2 \leq \mathrm{E}\big[(X-\mu_1)(Y-\mu_2)\big] \leq \sigma_1\sigma_2.$$

Además se cumple que:

$$\mathrm{E}\big[(X-\mu_1)(Y-\mu_2)\big] = \sigma_1\sigma_2,$$

si y sólo si:

$$\Pr\left[Y = \mu_2 + \frac{\sigma_2}{\sigma_1}(X-\mu_1)\right] = 1,$$

y:

$$\mathrm{E}\big[(X-\mu_1)(Y-\mu_2)\big] = -\sigma_1\sigma_2,$$

si y sólo si:

$$\Pr\left[Y = \mu_2 - \frac{\sigma_2}{\sigma_1}(X-\mu_1)\right] = 1.$$

Demostración. Coloquemos:

$$X_1 = \frac{X-\mu_1}{\sigma_1}. \qquad Y_1 = \frac{Y-\mu_2}{\sigma_2}.$$

Entonces X_1, Y_1 son como el teorema 3.4.1 y por lo tanto:

$$\mathrm{E}^2(X_1Y_1) \leq 1,$$

si y sólo si:

$$-1 \leq \mathrm{E}(X_1Y_1) \leq 1,$$

por lo que se convierte en:

$$\frac{\mathrm{E}^2[(X-\mu_1)(Y-\mu_2)]}{\sigma_1^2\sigma_2^2} \leq 1,$$

si y sólo si:

$$-\sigma_1\sigma_2 \leq \mathrm{E}[(X-\mu_1(Y-\mu_2)] \leq \sigma_1\sigma_2.$$

La segunda mitad del teorema se sigue de manera similar. $\square$

Definición 3.4.5. Sean X, Y variables aleatorias con medias μ_1, μ_2 la **covarianza** de X y Y es:

$$\mathrm{Cov}(X, Y) = \mathrm{E}[(X - \mu_1)(Y - \mu_2)].$$

Sean σ_1, σ_2 las desviaciones estándar de X, Y, y asumamos que son positivas, entonces la covarianza de $\frac{X-\mu_1}{\sigma_1}$, $\frac{Y-\mu_2}{\sigma_2}$ es llamado el **coeficiente de correlación** de X, Y y es denotado por $\rho(X, Y), \rho_{X,Y}, \rho_{1,2}$ o simplemente ρ. Tenemos que:

$$\rho = \mathrm{E}\left[\left(\frac{X - \mu_1}{\sigma_1}\right)\left(\frac{Y - \mu_2}{\sigma_2}\right)\right] = \frac{\mathrm{E}\big[(X - \mu_1)(Y - \mu_2)\big]}{\sigma_1\sigma_2} = \frac{\mathrm{Cov}(X, Y)}{\sigma_1\sigma_2} =$$

$$= \frac{\mathrm{E}(XY) - \mu_1\mu_2}{\sigma_1\sigma_2}.$$

De la desigualdad de Cauchy-Schwarz, tenemos que $\rho^2 \leq 1$, y $\rho = 1$ si y sólo si:

$$Y = \mu_2 + \frac{\sigma_2}{\sigma_1}(X - \mu_1),$$

con probabilidad 1, y $\rho = -1$ si y sólo si:

$$Y = \mu_2 - \frac{\sigma_2}{\sigma_1}(X - \mu_1),$$

con probabilidad 1. Con lo que ρ es una medida de dependencia lineal entre X y Y. Si $\rho = \pm 1$ significa que X, Y están linealmente relacionados y decimos que están completamente correlacionados, si $\rho = 0$ decimos que X, Y no están correlacionados.

4. SERIES DE TIEMPO

En este capítulo estudiaremos los temas principales relacionados con las series de tiempo. Entendemos una serie de tiempo como un conjunto de datos registrados de manera ordenada respecto al tiempo, el cual puede ser discreto o continuo. La idea fundamental para determinar una serie de tiempo esta en la estrategia para construir modelos que se adecuen a nuestros datos. Podemos resumir los pasos de la estrategia en los siguiente:

1. Identificación del modelo.
2. Estimación de los parámetros implícitos en el modelo.
3. Verificación de supuestos.
4. Uso del modelo.

4.1. Espacios de Hilbert L^2

Definición 4.1.1. Sea $(X, \mathbf{S}, \mu)$ un espacio de medida, f una función medible en X y $0 < p < \infty$, definimos la norma de f como:

$$\|f\|_p = \left[\int |f|^p d\mu \right]^{1/p},$$

y definimos:

$$L^p(X, \mathbf{S}, \mu) = \{f \colon X \to \mathbb{C} \mid f \text{ es medible y } \|f\|_p < \infty\},$$

dichos espacios son conocidos como **espacios** $L^p(X, \mathbf{S}, \mu)$ o simplemente L^p.

Los espacios L^p son espacios de Banach, ver [5], sin embargo el espacio L^2 es el único espacio que es de Hilbert, para mostrarlo haremos uso de la ley del paralelogramo 2.1.1. Sean A, B dos conjuntos disjuntos de X. Para cualquier conjunto

A, sea:

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

Calculando la norma de 1_A tenemos que:

$$\|1_A\|_p = \left(\int 1_A(x)^p d\mu \right)^{1/p} = \mu(A)^{1/p},$$

donde $\mu(A)$ es la medida del conjunto A. Dado que A, B son disjuntos, tenemos:

$$1_A + 1_B = \mid 1_A - 1_B \mid = 1_{A \cup B},$$

y

$$\mu(A \cup B) = \mu(A) + \mu(B).$$

Usando la ley del paralelogramo obtenemos:

$$2(\mu(A))^{2/p} + 2(\mu(B))^{2/p} = 2(\mu(A) + \mu(B))^{2/p},$$

de donde la igualdad se cumple cuando $p = 2$ y falla para cualquier $p \neq 2$, por lo que el espacio L^2 es un espacio de Hilbert.

Definición 4.1.2. Consideremos un espacio de probabilidad $(\Omega, \mathcal{F}, \Pr)$ y la colección C de todas las variables aleatorias X definidas sobre Ω y que satisfacen la condición:

$$\mathrm{E}X^2 = \int_\Omega X(w)^2 \Pr(dw) < \infty.$$

Dicho espacio es conocido como **espacio** $L^2(\Omega, \mathcal{F}, \Pr)$.

Con la notación usual de multiplicación por un número escalar y adición de variables aleatorias, tenemos que C es un espacio vectorial, dado que:

$$\mathrm{E}(aX)^2 = a^2\mathrm{E}X^2 < \infty, \text{ para todo } a \in \mathbb{R}, X \in C,$$

y de la desigualdad $(X + Y)^2 \leq 2X^2 + 2Y^2$,

$$\mathrm{E}(X + Y)^2 \leq 2\mathrm{E}X^2 + 2\mathrm{E}Y^2 < \infty, \text{ para todo } \; X, Y \in C.$$

Las demás propiedades de un espacio vectorial se verifican de manera sencilla. En particular C tiene un elemento cero, las variables aleatorias que son identicamente cero sobre Ω. Para cualesquiera dos elementos $X, Y \in C$ tenemos que:

$$\mathrm{E}(XY) = \langle X, Y \rangle.$$

El espacio L^2 es la colección de todas estas clases con el producto interno definido anteriormente.

Definición 4.1.3. Convergencia en norma de X_n a X en un espacio L^2 es llamada **convergencia de media cuadrada** y es representada como $X_n \to X$, es decir:

$$\|X_n - X\|^2 = \mathrm{E} \mid X_n - X \mid^2 \to 0 \text{ cuando } n \to \infty.$$

En este capítulo consideraremos que las series de tiempo X_n, son elementos del espacio $L^2(\Omega, \mathcal{F}, \Pr)$.

4.2. Series de tiempo estacionarias

Iniciaremos el estudio de esta sección definiendo algunos conceptos estadísticos que nos permitirán describir y analizar las series de tiempo, haciendo uso de dos enfoques. El descriptivo, que se ocupa de de resumir y describir en forma concisa, mediante gráficas o medidas descriptivas, la información con que se cuente. Y el enfoque inferencial, que se ocupa en utilizar los datos muestrales para realizar inferencias, que sean válidas para toda la población que se obtuvo de la muestra. El siguiente paso es la selección de un modelo matemático adecuado para los datos.

Definición 4.2.1. Un **proceso estocástico** es una familia de variables aleatorias $\{X_t, t \in T\}$ definida en un espacio de probabilidad $(\Omega, \mathcal{F}, \Pr)$. El conjunto de índices, o parámetros, T es un conjunto de puntos de tiempo, tales como $\{0, \pm 1, \pm 2, \dots\}$, $\{1, 2, 3, \dots\}$, $[0, \infty)$ o $(-\infty, \infty)$.

Definición 4.2.2. Diremos que las funciones $\{X_t(w), w \in \Omega\}$ sobre T son las **realizaciones** o **caminos de muestra** del proceso $\{X_t, t \in T\}$.

Definición 4.2.3. Una serie de tiempo $\{x_t, t \in T_0\}$ es la realización de una familia de variables aleatorias $\{X_t, t \in T_0\}$. Frecuentemente usamos el termino serie de tiempo para referirnos a los datos y al proceso en el cual hay una realización.

Ejemplo 4.2.1. A continuación se presentan algunos ejemplos de modelos con media cero.

1. Uno de los modelos más sencillos de series de tiempo es uno en el cual no existen componentes temporales y de tendencia, en las cuales las observaciones son **variables aleatorias independientes e idénticamente distribuidas**, **iid**, con media cero. Nos referimos a tal sucesión de variables aleatorias $X_1, X_2, \dots,$ como ruido iid. Por definición podemos representarlo, para cualquier entero positivo n y números reales $x_1, \dots, x_n$,

$$\Pr[X_1 \leq x_1, \dots, X_n \leq x_n] = \Pr[X_1 \leq x_1] \dots \Pr[X_n \leq x_n] = F(x_1) \dots F(x_n),$$

   donde $F(\cdot)$ es la función de distribución acumulada de cada una de las varialbes iid $X_1, X_2, \dots, X_n$.

2. Continuando con ruido iid, consideremos la sucesión de variables aleatorias iid $\{X_t, t = 1, 2, \dots\}$ con:

$$\Pr[X_t = 1] = p, \qquad \Pr[X_t = -1] = 1 - p,$$

   donde $p = \frac{1}{2}$. La serie de tiempo obtenida al lanzar una moneda repetidas veces e identificando $+1$ para las caras y -1 para los escudos. Este tipo de procesos se conoce como un **proceso binario**.

3. Una **caminata aleatoria** $\{S_t, t = 0, 1, 2, \dots\}$, empezando en cero, es obtenida al sumar acumulativamente variables iid. Así que una caminata aleatoria con media cero es obtenida al definir $S_0 = 0$ y:

$$S_t = X_1 + X_2 + \cdots + X_t, \qquad \text{para } t = 1, 2, \dots,$$

   donde $\{X_t\}$ es ruido iid.

Definición 4.2.4. Sea $\mathcal{T}$ el conjunto de todos los vectores $\{t = (t_1, \dots, t_n) \in T^n \mid t_1 < t_2 < \dots < t_n, n = 1, 2, \dots\}$. Entonces las **funciones de distribución**, de dimensión finita, de $\{X_t, t \in T\}$ son las funciones $\{F_t(\cdot), t \in \mathcal{T}\}$ definidas para $t = (t_1, \dots, t_n)$ por:

$$F_t(x) = \Pr(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n), \qquad x = (x_1, \dots, x_n)' \in \mathbb{R}^n.$$

Teorema 4.2.1. *Las funciones de distribución de probabilidad* $\{F_t(\cdot), t \in \mathcal{T}\}$ *son la función de distribución de algún proceso estocástico si y sólo si para cualquier*

$n \in \{1, 2, \dots\}, t = (t_1, \dots, t_n) \in \mathcal{T}$ y $1 \leq i \leq n$,

$$\lim_{x_i \to \infty} F_t(x) = F_{t(i)}(x(i)),$$

donde $t(i)$ y $x(i)$ son los $(n-1)$-vectores componentes obtenidos al remover el i-ésimo componente de t y x respectivamente.

Definición 4.2.5. Sea $\{X_t, t \in T\}$ un proceso tal que $\sigma^2(X_t) < \infty$ para cada $t \in T$, entonces la **función de autocovarianza**, $\gamma_X(\cdot, \cdot)$, de $\{X_t\}$ está definida por:

$$\gamma_X(r, s) = \mathrm{Cov}(X_r, X_s) = \mathrm{E}[(X_r - \mathrm{E}X_r)(X_s - \mathrm{E}X_s)],$$

para $r, s \in T$.

Definición 4.2.6. La serie de tiempo $\{X_t, t \in \mathbb{Z}\}$, con conjunto de índices $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ es **estacionaria** si:

1. $\mathrm{E}|X|^2 < \infty$ para todo $t \in \mathbb{Z}$,

2. $\mathrm{E}X_t = m$ para todo $t \in \mathbb{Z}$,

3. $\gamma_X(r, s) = \gamma_X(r + t, s + t)$ para todo $r, s, t \in \mathbb{Z}$.

Si $\{X_t, t \in \mathbb{Z}\}$ es estacionaria, entonces $\gamma_X(r, s) = \gamma_X(r - s, 0)$ para todo $r, s \in \mathbb{Z}$.

Definición 4.2.7. Un proceso es **ergódico** si los parámetros estadísticos calculados en un conjunto de realizaciones son iguales que los parámetros estadísticos calculados en una única realización.

Podemos redefinir la función de autocovarianza de un proceso estacionario como una función de una variable:

$$\gamma_X(h) = \gamma_X(h, 0) = \mathrm{Cov}(X_{t+h}, X_t) \qquad \text{para todo } t, h \in \mathbb{Z}.$$

La función $\gamma_X(\cdot)$ se referirá como una función de autocovarianza de $\{X_t\}$ y $\gamma_X(h)$ como su valor en el **retraso** h. La función de **autocorrelación** de $\{X_t\}$ es definida análogamente como la función cuyo valor de retraso h es:

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \mathrm{Corr}(X_{t+h}, X_t) \qquad \text{para todo } t, h \in \mathbb{Z}.$$

La función de autocorrelación sirve para especificar, junto con su media y varianza, un proceso estocástico estacionario; por otro lado, desde un punto de vista práctico de series de tiempo, debe recordarse que se cuenta sólo con una realización finita del proceso, a partir de la cual se hace necesario obtener una estimación de ρ_X, para obtenerla se supondrá que el proceso estacionario posee ciertas propiedades ergódicas, que permiten la equivalencia entre valores esperados y promedios muestrales obtenidos de una realización suficientemente larga del proceso.

Definición 4.2.8. Decimos que la serie de tiempo $\{X_t \in \mathbb{Z}\}$ es **estrictamente estacionaria** si la distribución conjunta de $(X_{t_1}, \ldots, X_{t_k})'$ y $(X_{t_{1+h}}, \ldots, X_{t_{k+h}})'$ son iguales, para todo entero positivo k y para todo $t_1, \ldots, t_k, h \in \mathbb{Z}$. Estrictamente estacionaria significa intuitivamente que la gráfica sobre dos intervalos de tiempo de igual longitud de la realización de la serie de tiempo deben de mostrar similares características estadísticas.

Ejemplo 4.2.2. En los siguientes ejemplos usaremos la **propiedad de linealidad de la covarianza**, es decir, si $\mathrm{E}X^2 < \infty, \mathrm{E}Y^2 < \infty, \mathrm{E}Z^2 < \infty$, y $a, b, c \in \mathbb{R}$, entonces:

$$\mathrm{Cov}(aX + bY + c, Z) = a\mathrm{Cov}(X, Z) + b\mathrm{Cov}(Y, Z).$$

1. Si $\{X_t\}$ es ruido iid y $E(X_t^2) = \sigma^2 < \infty$, entonces sabemos que $E(X_t) = 0$ para todo t. Asumiendo la independencia tenemos que su función de covarianza es:

$$\gamma_X(t+h, t) = \begin{cases} \sigma^2, & \text{si} \quad h = 0, \\ \\ 0, & \text{si} \quad h \neq 0, \end{cases}$$

   la cual no depende de t. Por lo tanto el ruido iid con segundo momento finito es estacionario. Usamos la siguiente notación:

$$\{X_t\} \sim \mathrm{IID}(0, \sigma^2),$$

   para indicar que las variables aleatorias X_t son variables aleatorias independientes e idénticamente distribuidas, cada una con media 0 y varianza σ^2.

2. Sea $\{X_t\}$ una sucesión de variables aleatorias no correlacionadas, cada una con media 0 y varianza σ^2, entonces claramente $\{X_t\}$ es estacionaria con la misma función de covarianza del ejemplo anterior. Tal sucesión se conoce como **ruido blanco**, con media 0 y varianza σ^2. Usamos la notación:

$$\{X_t\} \sim \mathrm{WN}(0, \sigma^2).$$

Claramente, cada sucesión $\mathrm{IID}(0, \sigma^2)$ es $\mathrm{WN}(0, \sigma^2)$ pero no al contrario.

3. Sea $\{S_t\}$ una caminata aleatoria y $\{X_t\}$ un ruido iid, entonces tenemos que $\mathrm{E}S_t = 0$ y $\mathrm{E}(S_t^2) = t\sigma^2$ para todo t y $h \geq 0$,

$$\gamma_S(t+h, t) = \mathrm{Cov}(S_{t+h}, S_t) =$$

$$= \mathrm{Cov}(S_t + X_{t+1} + \cdots + X_{t+h}, S_t) = \mathrm{Cov}(S_t, S_t) = t\sigma^2.$$

Dado que $\gamma_S(t+h, t)$ depende de t, la serie $\{S_t\}$ no es estacionaria.

4. **Media móvil de primer orden o proceso MA.** Considere la sucesión definida por:

$$X_t = Z_t + \theta Z_{t-1}, \qquad t = 0, \pm 1, \ldots,$$

donde $\{Z_t\}$ $\sim$WN$(0, \sigma^2)$ y θ es una constante de valor real. De la ecuación inmediata anterior podemos observar que $\mathrm{E}X_t = 0$, $\mathrm{E}X_t^2 = \sigma^2(1+\theta^2) < \infty$, y:

$$\gamma_X(t+h, t) = \begin{cases} \sigma^2(1+\theta^2), & \text{si} \quad h = 0, \\ \sigma^2\theta, & \text{si} \quad h = \pm 1, \\ 0, & \text{si} \quad |h| > 1. \end{cases}$$

Por lo que $\{X_t\}$ es estacionaria. La función de autocorrelación de $\{X_t\}$ es:

$$\rho_X(h) = \begin{cases} 1, & \text{si} \quad h = 0, \\ \dfrac{\theta}{1+\theta^2} & \text{si} \quad h \pm 1, \\ 0, & \text{si} \quad |h| > 1. \end{cases}$$

5. **Autoregresión de primer orden o proceso AR.** Asumamos que $\{X_t\}$ es una serie estacionaria que satisface la ecuación:

$$X_t = \phi X_{t-1} + Z_t, \qquad t = 0, \pm 1, \ldots,$$

donde $\{Z_t\}$ $\sim$WN$(0, \sigma^2)$, $|\phi| < 1$, y Z_t no es correlacionada con X_s para cada $s < t$. Tomando la esperenza matemática en ambos lados de la ecuación inmediata anterior y asumiendo que $\mathrm{E}Z_t = 0$, podemos ver que:

$$\mathrm{E}X_t = 0.$$

Para encontrar la función de autocorrelación de $\{X_t\}$ multiplicamos de ambos lados de $X_t = \phi X_{t-1} + Z_t$ por $X_{t-h}, (h > 0)$, y luego tomando la esperanza matemática para obtener:

$$\gamma_X(h) = \mathrm{Cov}(X_t, X_{t-h}) = \mathrm{Cov}(\phi X_{t-1}, X_{t-h}) + \mathrm{Cov}(Z_t, X_{h-t}) =$$

$$= \phi\gamma_X(h-1) + 0 = \cdots = \phi^h \gamma_X(0).$$

Observemos que $\gamma(h) = \gamma(-h)$, encontramos que:

$$\rho_X(g) = \frac{\gamma_X(h)}{\gamma_X(0)} = \phi^{|h|}, \qquad h = 0, \pm 1, \ldots.$$

Se sigue de la linealidad de la función de covarianza en cada uno de sus argumentos y el hecho que Z_t no es correlacionada con X_{t-1} que:

$$\gamma_X(0) = \mathrm{Cov}(X_t, X_t) = \mathrm{Cov}(\phi X_{t-1} + Z_t, \phi X_{t-1} + Z_t) = \phi^2\gamma_X(0) + \sigma^2,$$

y por lo tanto:

$$\gamma_X(0) = \frac{\sigma^2}{1-\phi^2}.$$

4.2.1. Relación entre una serie de tiempo estacionaria y una estrictamente estacionaria

Sea $\{X_t\}$ una serie de tiempo estrictamente estacionaria, se sigue que si tomamos $k = 1$ en la definición 4.2.8, es decir X_t tiene la misma distribución para cada $t \in \mathbb{Z}$. Si $\mathrm{E}|X_t|^2 < \infty$ implica en particular que $\mathrm{E}X_t$ y $\mathrm{Var}(X_t)$ son ambas constantes. Más aún, tomando $k = 2$ en la definición4.2.8, tenemos que X_{t+h} y X_t tienen la misma distribución conjunta y por lo tanto tienen la misma covarianza para todo $h \in \mathbb{Z}$. Por lo que un proceso estrictamente estacionario con segundo momento finito es estacionario. El converso no es valido.

Por ejemplo, sea $\{X_t\}$ una sucesión de variables aleatorias independientes tal que X_t está distribuido exponencialmente con media 1 cuando t es impar y distribuido normalmente con media 1 y varianza 1 cuando t es par, entonces $\{X_t\}$ es estacionaria con $\gamma_X(0) = 1$ y $\gamma_X(h) = 0$ para $h \neq 0$. Sin embargo dado que X_1 y X_2

tienen diferentes distribuciones, $\{X_t\}$ no puede ser estrictamente estacionaria.

4.3. Estimación y eliminación de los componentes temporales y de tendencia

Lo primero que se debe de realizar en el análisis de cualquier serie de tiempo es graficar los datos. Si existen discontinuidades aparentes en la serie, tales como repentinos cambios de nivel, es aconsejable analizar primero la serie separándola en segmentos homogeneos. Si existen observaciones exteriores, deben ser estudiadas cuidadosamente si hay alguna justificación para descartarlos (por ejemplo, si una observación ha sido colocada de algún otro proceso por equivocación).

Inspeccionar una gráfica también puede sugerir la posibilidad de representar los datos como una realización de un proceso (el modelo de **descomposición clásico**),

$$X_t = m_t + s_t + Y_t, \tag{4.1}$$

donde m_t es una función de cambio lento conocida como una **componente de tendencia**, s_t es una función con período conocido d referido como una **componente temporal**, y Y_t es una **componente de ruido aleatorio** la cual es estacionaria en el sentido de la definición 4.2.6. Si las fluctuaciones de ruido y temporalidad parecen incrementar con el nivel del proceso, entonces una transformación preliminar de los datos es frecuentemente usada para hacer los datos transformados compatible con el modelo 4.1.

Nuestro objetivo es estimar y extraer los componentes determinísticos m_t y s_t esperando que el componente residual o de ruido Y_t resultará ser un proceso aleatorio estacionario. Entonces podemos usar la teoría de tal proceso para encontrar un modelo probabilístico estacionario para el proceso $\{Y_t\}$, de esta forma analizar sus propiedades y usarlo en conjunto con m_t y s_t para propósitos de predicción y control de $\{X_t\}$.

4.3.1. Eliminación de una tendencia en ausencia de temporalidad

En la ausencia de una componente de temporalidad el modelo 4.1 se convierte en:

$$X_t = m_t + Y_t, \qquad t = 1, \ldots, n, \tag{4.2}$$

donde, sin perdida de generalidad, podemos asumir que $\mathrm{E}Y_t = 0$.

1. **Primer método (estimación de mínimos cuadrados de m_t):** en este procedimiento intentaremos ajustar una familia paramétrica de funciones, por ejemplo:
$$m_t = a_0 + a_1 t + a_2 t^2, \tag{4.3}$$
a los datos, escogiendo los parámetros, en el caso de 4.3 a_0, a_1 y a_2, para minimizar $\Sigma_t (x_t - m_t)^2$.

2. **Segundo método (suavizado mediante media móvil):** Sea q un entero no negativo y consideremos la media móvil de dos caras:
$$W_t = (2q+1)^{-1} \sum_{j=-q}^{q} X_{t+j}, \tag{4.4}$$
del proceso $\{X_t\}$ definido en 4.2. Entonces para $q+1 \leq t \leq n-q$,
$$W_t = (2q+1)^{-1} \sum_{j=-q}^{q} m_{t+j} + (2q+1)^{-1} \sum_{j=-q}^{q} Y_{t+j} \simeq m_t, \tag{4.5}$$
asumiendo que m_t es aproximadamente lineal sobre un intervalo $[t-q, t+q]$ y que la media de los errores sobre el intervalo es cercano a cero. Así que la media móvil nos proporciona las estimaciones:
$$\hat{m}_t = (2q+1)^{-1} \sum_{j=-q}^{q} X_{t+j}, \qquad q+1 \leq t \leq n-q. \tag{4.6}$$
Dado que X_t no se observa para $t \leq 0$ o $t > n$ no podemos usar 4.6 para $t \leq q$ o $t > n-q$. Sin embargo podemos usar la media móvil de una cara tal que:
$$\hat{m}_t = \sum_{j=0}^{n-1} \alpha (1-\alpha)^j X_{t+j}, \qquad t = 1, \ldots, q, \tag{4.7}$$
y
$$\hat{m}_t = \sum_{j=0}^{t-1} \alpha (1-\alpha)^j X_{t-j} \qquad t = n-q+1, \ldots, n. \tag{4.8}$$

3. **Tercer método (diferenciación para generar datos estacionarios):** en lugar de intentar eliminar el ruido suavizándolo como en el método 2, ahora intentaremos eliminar el elemento de tendencia mediante diferenciación.

Definimos el **operador diferencia** ∇ como:

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t, \tag{4.9}$$

donde B es el **operador de retraso**,

$$BX_t = X_{t-1}. \tag{4.10}$$

Potencias del operador B y ∇ están definidas de manera similar, es decir, $B^j(X_t) = X_{t-j}$ y $\nabla^j(X_t) = \nabla(\nabla^{j-1}(X_t))$, $j \geq 1$ con $\nabla^0(X_t) = X_t$. Polinomios en B y ∇ son operados de la misma manera que funciones polinomiales de variable real. Por ejemplo:

$$\nabla^2 X_t = \nabla(\nabla X_t) = (1-B)(1-B)X_t = (1-2B+B^2)X_t = X_t - 2X_{t-1} + X_{t-2}.$$

Si el operador ∇ es usado en una función de tendencia lineal $m_t = at + b$, obtenemos la función constante $\nabla m_t = a$. De la misma manera, cualquier polinomio de tendencia de grado k puede ser reducido a una constante mediante la aplicación del operador ∇^k. Comenzando con el modelo $X_t = m_t + Y_t$ donde $m_t \sum_{j=0}^{k} a_j t^j$ y Y_t es estacionario con media cero, obtenemos:

$$\nabla^k X_t = k! a_k + \nabla^k Y_t,$$

un proceso estacionario con media a_k. Éstas consideraciones sugieren la posibilidad, dada una sucesión $\{x_t\}$ de datos, de aplicar un operador ∇ repetidamente hasta encontrar una sucesión $\{\nabla^k x_t\}$, la cual puede ser aceptablemente modelada como una realización de un proceso estacionario. Es frecuentemente encontrado en la práctica que el orden de k de diferenciación es bastante pequeño, uno o dos. (Esto depende del hecho que muchas funciones pueden ser bien aproximadas, sobre un intervalo de dimensión finita, por un polinomio de razonablemente grado menor.)

4.3.2. Eliminación de tendencia y temporalidad

Los métodos descritos para remover la tendencia pueden ser adaptados de manera natural para eliminar la tendencia y temporalidad del modelo general:

$$X_t = m_t + s_t + Y_t, \tag{4.11}$$

donde $\mathrm{E}Y_t = 0$, $s_{t+d} = s_d$ y $\sum_{j=1}^{d} s_j = 0$.

4.4. La función de autocovarianza de un proceso estacionario

Teorema 4.4.1. *Sea $\gamma(\cdot)$ la función de autocovarianza de un proceso estacionario $\{X_t, t \in \mathbb{Z}\}$, entonces:*

$$\gamma(0) \geq 0,$$

$$|\gamma(h)| \leq \gamma(0) \qquad \textit{para todo}\ h \in \mathbb{Z},$$

$$\gamma(h) = \gamma(-h) \qquad \textit{para todo}\ h \in \mathbb{Z}.$$

Demostración. La propiedad de $\gamma(0) \geq 0$ es obvia debido a que $\mathrm{Var}(X_t) \geq 0$. Para demostrar la segunda propiedad hacemos uso de la desigualdad de Cauchy-Schwarz:

$$|\mathrm{Cov}(X_{t+h}, X_t)| \leq (\mathrm{Var}(X_{t+h}))^{\frac{1}{2}}(\mathrm{Var}(X_t)^{\frac{1}{2}}),$$

y la tercera propiedad se demuestra con lo siguiente:

$$\gamma(-h) = \mathrm{Cov}(X_{t-h}, X_t) = \mathrm{Cov}(X_t, X_{t+h}) = \gamma(h). \qquad \square$$

Las funciones de autocovarianza también tienen la propiedad de definidas no negativas.

Definición 4.4.1. Una función de valor real sobre los enteros, $\kappa\colon \mathbb{Z} \to \mathbb{R}$ es **definida no negativa** si y sólo si:

$$\sum_{i,j=1}^{n} a_i \kappa(t_i - t_j) a_j \geq 0,$$

para todos los enteros positivos n y todos los vectores $a = (a_1, \ldots, a_n)' \in \mathbb{R}^n$ y $t = (t_1, \ldots, t_n)' \in \mathbb{Z}^n$.

Teorema 4.4.2. *Una función par de valor real definida en el conjunto $\mathbb{Z}$ de todos los enteros es definida no negativa si y sólo si es la función de autocovarianza de una serie de tiempo estacionaria.*

Demostración. Para demostrar que la función de autocovarianza de cualquier serie de tiempo estacionaria es definida no negativa, simplemente observemos que si $a = (a_1, \ldots, a_n)' \in \mathbb{R}^n$, $t = (t_1, \ldots, t_n) \in \mathbb{Z}$, y $Z_t = (X_{t_1} - \mathrm{E}X_{t_1}, \ldots, X_{t_n} - \mathrm{E}X_{t_n})'$, entonces:

$$0 \leq \mathrm{Var}(a' Z_t) = a' E Z_t Z_t' a =$$

$$= a'\Gamma_n a = \sum_{i,j=1}^{n} a_i \gamma(t_i - t_j) a_j,$$

donde $\Gamma_n = [\gamma(t_i - t_j)]_{i,j=1}^n$ es la matriz de covarianzas de $(X_{t_1}, \ldots, X_{t_n})'$. Para establecer el converso, sea $\kappa\colon \mathbb{Z} \longrightarrow \mathbb{R}$ una función definida no negativa.

Necesitamos demostrar que existe un proceso estacionario con $\kappa(\cdot)$ como su función de autocovarianza y para esto usaremos el teorema 4.2.1. Para cada entero positivo n y para cada $t = (t_1, \ldots, t_n) \in \mathbb{Z}^n$ tal que $t_1 < t_2 < \cdots < t_n$, sea F_t la función de distribución sobre $\mathbb{R}^n$ con función característica:

$$\phi(u) = \mathrm{e}^{\dfrac{-u'Ku}{2}},$$

donde $u = (u_1, \ldots, u_n)' \in \mathbb{R}^n$ y $K = [\kappa(t_i - t_j)]_{i,j=1}^n$. Dado que κ es definida no negativa la matriz K es también definida no negativa y consecuentemente ϕ_t es la función característica de una distribución normal $n-$variada con media cero y matriz de covarianza K.

Usando la notación del teorema 4.2.1:

$$\phi_{t(i)}(u(i)) = \lim_{u_t \to 0} \phi_t(u), \text{ para cada } \quad t \in \mathcal{T},$$

es decir, las funciones de distribución F_t son consistentes, y por el teorema 4.2.1 existe una serie de tiempo $\{X_t\}$ con funciones de distribución F_t y funciones características ϕ_t, $t \in \mathcal{T}$. En particular la distribución conjunta de X_i y X_j es normal bivariada con media 0 y matriz de covarianza:

$$\begin{bmatrix} \kappa(0) & \kappa(i-j) \\ \kappa(i-j) & \kappa(0) \end{bmatrix},$$

lo que muestra que $\mathrm{Cov}(X_i, X_j) = \kappa(i-j)$ como se requiere. $\square$

Definición 4.4.2. La **función de autocovarianza muestral** de $\{x_1, \ldots, x_n\}$ es definida por:

$$\hat{\gamma} = \frac{1}{n} \sum_{j=1}^{n-k} (x_{j+h} - \bar{x})(x_j - \bar{x}), \qquad 0 \le h < n,$$

y $\hat{\gamma}(h) = \hat{\gamma}(-h)$, $-n < h \le 0$, donde $\bar{x}$ es la media muestral, $\bar{x} = \dfrac{1}{n} \displaystyle\sum_{j=1}^{n} x_j$.

4.5. Modelo lineal general

En esta sección mostraremos algunas de las propiedades principales del modelo general lineal a partir del teorema de proyección y sus aplicaciones. Considere el modelo lineal general:

$$Y = X\Theta + Z,$$

donde $Y = (Y_1, \dots, Y_n)'$ es el vector de observaciones, X es una matriz conocida de $n \times m$ de rango $n < m$, $\Theta = (\Theta_1, \dots, \Theta_m)'$ es un vector de dimensión m de valores paramétricos, y $Z = (Z_1, \dots, Z_n)'$ es el vector de variables de ruido. Asumimos que $Z \sim N(0, \sigma^2 I_n)$, donde I_n es la matriz identidad de dimensión n. El estimador de mínimos cuadrados viene dado por:

$$\hat{\Theta} = (X'X)^{-1}X'Y.$$

1. $Y \sim N(X\Theta, \sigma^2 I_n)$. Calculando el valor medio de Y tenemos que:

$$\begin{aligned} \mathrm{E}Y &= \mathrm{E}(X\Theta + Z) \\ &= \mathrm{E}(X\Theta) + EZ \\ &= \mathrm{E}(X\Theta) = X\Theta. \end{aligned}$$

   Para la varianza:

$$\begin{aligned} V(Y) &= V(X\Theta + Z) \\ &= \mathrm{E}[(X\Theta + Z - \mathrm{E}(X\Theta + Z))^2] \\ &= \mathrm{E}[(X\Theta + Z - X\Theta)^2] \\ &= \mathrm{E}[Z^2] = \mathrm{E}[(Z - \mathrm{E}Z)^2] \\ &= V(Z) = \sigma^2 I_n. \end{aligned}$$

2. $\hat{\Theta} \sim N(\Theta, \sigma^2 (X'X)^{-1})$.

   Calculando el valor medio de $\hat{\Theta}$ tenemos que:

$$\begin{aligned} \mathrm{E}\hat{\Theta} &= \mathrm{E}[(X'X)^{-1}X'Y] \\ &= \mathrm{E}[(X'X)^{-1}X'(X\Theta + Z)] \\ &= \mathrm{E}[(X'X)^{-1}X'X\Theta] + \mathrm{E}[(X'X)^{-1}X'Z] \\ &= \mathrm{E}\Theta + (X'X)^{-1}X'\mathrm{E}Z \\ &= \Theta. \end{aligned}$$

Para la varianza:

$$\begin{aligned}V(\hat{\Theta}) &= V[(X'X)^{-1}X'Y]\\ &= [(X'X)^{-1}X']V(Y)[(X'X)^{-1}X']'\\ &= [(X'X)^{-1}X']\sigma^2 I_n[(X'X)^{-1}X']'\\ &= \sigma^2[(X'X)^{-1}X'][X(X'X)^{-1}] = \sigma^2(X'X)^{-1}.\end{aligned}$$

3. Mostraremos que la matriz proyección $P_M = X(X'X)^{-1}X'$ es definida no negativa y tiene m valores propios todos iguales a 1. Para mostrar que una matriz es definida no negativa se debe de cumplir que todos sus valores propios λ sean iguales o mayores a 0. De la definición de un valor propio tenemos que, para todo $V \neq 0$, donde V es el vector propio asociado a cada λ, se cumple que $P_M V = \lambda V$.

   Aplicamos de nuevo el vector proyección en ambos lados de la igualdad y usando los teoremas 2.3.2 y 2.3.3, tenemos:

$$P_M(P_M V) = P_M(\lambda V),$$

$$P_M^2 V = \lambda P_M V,$$

$$P_M V = \lambda(\lambda V),$$

$$\lambda V = \lambda^2 V,$$

   como $V \neq 0$, se cumple que $\lambda = \lambda^2$ y por lo tanto:

$$\lambda^2 - \lambda = 0,$$

$$\lambda(\lambda - 1) = 0,$$

   de donde $\lambda = 0$ o $\lambda = 1$, y esto se cumple para todo λ, por lo tanto la matriz P_M es definida no negativa. Para mostrar que todos sus valores propios son iguales a 1 haremos uso de la traza de una matriz:

$$\begin{aligned}\mathrm{tr}(P_M) &= \mathrm{tr}(X(X'X)^{-1}X)\\ &= \mathrm{tr}(X'X(X'X)^{-1})\\ &= \mathrm{tr}(I_m),\end{aligned}$$

   donde I_m es una matriz identidad de dimensión m. Además tenemos que para calcular la traza de una matriz podemos hacer la sumatoria de todos sus valores propios, por lo que:

$$\begin{aligned}\sum_i \lambda_i &= \operatorname{tr}(P_M)\\ &= \operatorname{tr}(I_m)\\ &= m,\end{aligned}$$

pero todos los valores propios λ_i son iguales a 1 o 0, por lo que deben haber exactamente m valores propios iguales a 1. Similarmente mostraremos que que $I_n - P_M$ es definida no negativa con $(n-m)$ valores propios iguales a 1. Haciendo uso de la definición de valores propios tenemos que, para todo $V \neq 0$:

$$(I_n - P_M)V = \lambda V,$$

usando de nuevo los teoremas 2.3.2 y 2.3.3 tenemos:

$$(I_n - P_M)(I_n - P_M)V = (I_n - P_M)\lambda V,$$

$$\begin{aligned}(I_n - P_M)^2 V &= \lambda(I_n - P_M)V,\\ (I_n - P_M)V &= \lambda(\lambda V),\\ \lambda V &= \lambda^2 V,\end{aligned}$$

como $V \neq 0$ se cumple que $\lambda = \lambda^2$, por lo que λ es igual a 1 o 0, para todo λ, por lo tanto la matriz $(I_n - P_M)$ es definida no negativa. Usando la traza de una matriz tenemos que:

$$\begin{aligned}\operatorname{tr}(I_n - P_M) &= \operatorname{tr}(I_n) - \operatorname{tr}(P_m)\\ &= n - m,\end{aligned}$$

usando la sumatoria de sus valores propios para calcular la traza, tenemos que:

$$\begin{aligned}\sum_i \lambda_i &= \operatorname{tr}(I_n - P_M)\\ &= n - m,\end{aligned}$$

pero todos los valores propios λ_i son iguales a 1 o 0, por lo que deben haber exactamente $n - m$ valores propios iguales a 1.

4. Mostrar que los dos vectores de variables aleatorias, $P_M(Y - X\Theta)$ y $(I_n - P_M)Y$ son independientes. Para mostar que $P_M(Y - X\Theta)$ y $(I_n - P_M)Y$ son independientes, tenemos que mostrar que:

$$\langle P_M(Y - X\Theta), (I_n - P_M)Y \rangle = 0.$$

Aplicamos la matriz proyección al producto punto de las variables aleatorias y usando el teorema 2.3.2, obtenemos:

$$\begin{aligned}\langle P_M(P_M(Y - X\Theta)), P_M((I_n - P_M)Y) \rangle &= \langle P_M^2(Y - X\Theta), (P_M - P_M^2)Y \rangle \\ &= \langle P_M(Y - X\Theta), (P_M - P_M)Y \rangle \\ &= \langle P_M(Y - X\Theta), (0)Y \rangle \\ &= 0,\end{aligned}$$

por lo tanto los dos vectores de variables aleatorias son independientes.

Ahora mostraremos que $\sigma^{-2}\|P_M(Y - X\Theta)\|^2$ y $\sigma^{-2}\|(I_n - P_M)Y\|^2$ son variables aleatorias independientes de chi-cuadrado, χ^2, con m y $(n - m)$ grados de libertad respectivamente, donde $\|Y\|$ denota la norma **euclidiana** de Y, es decir:

$$\|Y\| = \left(\sum_{i=1}^{n} Y_i^2 \right)^{1/2}.$$

Para hacer la demostración haremos uso del siguiente teorema.

Teorema 4.5.1. *Sean $(Z_1, Z_2, \ldots, Z_n)$ una sucesión de variables aleatorias independientes, $n \in \mathbb{N}$. Entonces la suma de los cuadrados:*

$$V = \sum_{i=1}^{n} Z_i^2,$$

tiene la distribución chi-cuadrada con n grados de libertad.

Demostración. Consultar [15, p. 432]. □

La variable aleatoria $\sigma^{-2}\|P_M(Y - X\Theta)\|^2$ la podemos expresar como:

$$\frac{1}{\sigma^2}\|P_M(Y - X\Theta)\|^2 = \frac{1}{\sigma^2}\|P_M Z\| = \frac{1}{\sigma^2} \sum_{i=1}^{m} P_M Z_i^2,$$

como sabemos los Z_i siguen una distribución normal, usando el teorema 4.5.1 tenemos que la variable aleatoria $\sigma^{-2}\|P_M(Y - X\Theta)\|^2$ sigue una distribución χ^2 con m grados de libertad.

La variable aleatoria $\sigma^{-2}\|(I_n - P_M)Y\|^2$ la podemos expresar, utilizando el teorema 2.3.3 como:

$$\frac{1}{\sigma^2}\|Y\|^2 = \frac{1}{\sigma^2}\|P_M Y\|^2 + \frac{1}{\sigma^2}\|(I_n - P_M)Y\|^2,$$

despejando tenemos que:

$$\begin{aligned}\frac{1}{\sigma^2}\|(I_n - P_M)Y\|^2 &= \frac{1}{\sigma^2}\|Y\|^2 - \frac{1}{\sigma^2}\|P_M Y\|^2 \\ &= \frac{1}{\sigma^2}\sum_{i=1}^{n} Y_i^2 - \frac{1}{\sigma^2}\sum_{i=1}^{m} P_M Y_i^2,\end{aligned}$$

y usando el teorema 4.5.1 tenemos que $\sigma^{-2}\|(I_n - P_M)Y\|^2$ sigue una distribución de χ^2 con grados de libertad $(n - m)$.

5. Por último mostraremos que:

$$\frac{(n-m)\|P_M(Y - X\Theta)\|^2}{m\|Y - P_M Y\|^2},$$

tiene una distribución F con m y $(n - m)$ grados de libertad. Al igual que en el inciso anterior haremos uso de otro teorema.

Teorema 4.5.2. *Sean U_1, U_2 variables aleatorias con distribuciones de χ^2 y grados n_1 y n_2 respectivamente. Entonces:*

$$F = \frac{U_1/n_1}{U_2/n_2},$$

tiene una distribución F con n_1 y n_2 grados de libertad.

Demostración. Consultar [3, p. 598]. □

Vemos que la variable aleatoria definida anteriormente cumple con las condiciones del teorema 4.5.2 por lo que tiene una distribución F con grados de libertad m y $(n - m)$.

CONCLUSIONES

1. Se estudiaron las definiciones y teoremas de los anillos, σ-anillos, como estos se relacionan con las σ-álgebras. También se estudió las propiedades principales de las medidas y de las funciones medibles.

2. Se definieron las propiedades principales de los espacios de Hilbert, de los operadores y se demostró el teorema de proyección, el cual garantiza la existencia de un único elemento $\hat{x} \in M$ que es la mejor aproximación para algún $y \notin M$.

3. Se axiomatizó la teoría de probabilidades a partir de los conceptos de la teoría de la medida, adaptándolos de manera apropiada para definir los conceptos y teoremas. Se enunciaron los conceptos de variables aleatorias así como también sus propiedades, para el estudio de las series de tiempo.

4. Se describió el espacio que contiene a las series de tiempo, el espacio $L^2(\Omega, \mathcal{F}, \mathrm{Pr})$, que además es un espacio de Hilbert. Se describieron los conceptos de las series de tiempo y los modelos principales para su solución. Se presentó una demostración del modelo general lineal, base de toda la teoría de regresión lineal y estimación de series de tiempo.

RECOMENDACIONES

1. Comparar la parte geométrica de las series de tiempo con sus métodos de resolución, para establecer similitudes y ver la importancia de conocer su interpretación geométrica.

2. Utilizar el capítulo 1 como material introductorio para el curso de teoría de la medida y el capítulo 3 para el curso de estadística.

3. Tomar este trabajo como material formal de apoyo para estudiantes de nivel de licenciatura en el estudio de teoría de probabilidades y series de tiempo.

BIBLIOGRAFÍA

[1] Bowerman, B. L, O'Connell R. T. y Koehler B, (2007). *Pronósticos, series de tiempo y regresión: un enfoque aplicado*. México: 4"ta ed. Cengage Learning. (Orig. 2005).

[2] Brockwell, P. J. y Davis, R. A, (1987). *Time series: theory and methods.* Estados Unidos: 1.ª ed. Springer-Verlag New York Inc.

[3] DeGroot, M. H. y Schervish, M. J, (2012). *Probability and Statistics.* Estados Unidos: 4"ta. ed. Pearson Education, Inc.

[4] Enders, W. (2015). *Applied econometric time series.* Estados Unidos: 4.ª ed. John Wiley & Sons, Inc.

[5] Folland, G. B, (1999). *Real analysis: modern techniques and their applications.* Estados Unidos: 2.ª ed. John Wiley & Sons, Inc.

[6] Guajarati, D. y Porter, D. (2010). *Econometría.* México: 5.ª ed. McGraw-Hill, Interamericana editores S.A. (Orig. 2009).

[7] Guerrero Guzmán, V. (2003). *Análisis estadístico de series de tiempo económicas.* México: 2.ª ed. International Thomson Editores, S.A.

[8] Haaser, N. y Sullivan, J. (1992). *Análisis real.* México: 2.ª ed. Editorial Trillas. (Orig. 1964).

[9] Halmos, P. (1973). *Teoría intuitiva de los conjuntos.* México: 8.ª ed. Compañía Editorial Continental, S.A. (Orig. 1960).

[10] Hausdorff, F. (1962). *Set theory.* Estados Unidos: 2.ª ed. Chelsea Publishing Company.

[11] Hewitt, E. y Stromberg, K. (1965). *Real and abstract analysis.* Estados Unidos: Springer-Verlag.

[12] Kolmogorov, A. y Fomin, S. (1975). *Elementos de la teoría de funciones y del análisis funcional.* URSS: Editorial MIR, Moscú. (Orig. 1975).

[13] Kreyszig, E. (1978). *Introductory functional analysis with applications.* Estados Unidos: 1.ª ed. John Wiley & Sons. Inc.

[14] Meza Carvajalino, C. A, (2012). *Econometría de Series de tiempo: elementos y fundamentos.* España: Editorial Académica Española.

[15] Miller, S. J, (2017). *The Probability Lifesaver: All the Tools You Need to Understand Chance.* Estados Unidos: Princeton University Press.

[16] Pérez Ramírez, F.O, (2007). *Introducción a las series de tiempo. Métodos paramétricos.* Colombia: Universidad de Medellín.

[17] Roussas, G. (1997). *A course in mathematical stadistics.* Estados Unidos: 2.ª ed. Academic Press.

[18] Wackerly, Mendenhall y Scheaffer. (2010). *Estadística matemática.* México: 7.ª ed. Cengage Leaning Editores, S.A. (Orig. 2008).