22

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES
ACATLAN

APLICACIONES DE LAS FUNCIONES CARACTERISTICAS

T E S I S

PARA OBTENER EL TITULO DE:

A C T U A R I O

P R E S E N T A :

O S W A L D O P A L M A C O C A

ASESOR: JOSE MARIA GONZALEZ BARRIOS MURGUIA

UNAM
CAMPUS ACATLÁN

ACATLAN, ESTADO DE MEXICO JUNIO DE 2000

Aplicaciones de las Funciones Características

Oswaldo Palma Coca

A mis padres:

Luis Palma Buendía

y

Eva Coca Medina

como un reconocimiento al
amor y cuidado que nos han
dado a mis hermanos y a mi

¡Gracias!

Concluir este trabajo me produce una doble satisfacción. En primer lugar porque constituye para mí el final de un sueño y el inicio de una preciada realidad. En segundo lugar porque me permite reconocer y expresar mi sincero agradecimiento a todas las personas que me brindaron su apoyo.

A mis padres por sus cuidados y consejos. En realidad gran parte de este logro les pertenece.

A todos mis hermanos pues su ejemplo ha sido definitivo en la dirección de mi vida.

El Dr. José María González Barrios Murguía, que aportó su tiempo, conocimientos y paciencia para dirigir este trabajo.

La Dra. Eliane Rodrígues que, con la accesibilidad y buena disposición que la caracterizan, revisó de manera minuciosa esta tesis y me hizo valiosos comentarios.

A mis sinodales M. en C. Eduardo Godoy, Dr. Sergio Chapa y Dr. Raúl Rueda que me regalaron parte de su tiempo al revisar y corregir esta tesis.

Fabiola V. Muedano por su comprensión, paciencia, entusiasmo,... en fin, por ser ella.

A mi amiga Yunuén Rodríguez, por sus consejos y comentarios, pero sobre todo por su amistad.

A todos ellos ¡*GRACIAS*!

Contenido

Introducción

La teoría de la probabilidad hace uso de diversas herramientas de análisis. Entre éstas herramientas se encuentra la transformada de Fourier. La transformada de Fourier de una función de distribución de probabilidad, se llama función característica.

La función característica tiene muchas propiedades que la hacen ser una herramienta analítica poderosa. El objetivo de este trabajo es mostrar las aplicaciones de las funciones características en la solución de problemas de la teoría de la probabilidad y de la estadística matemática.

Ya que la función característica ha sido ampliamente usada en muchos problemas desde ya hace tiempo, es prácticamente imposible incluir todas sus aplicaciones. Por ello esta tesis sólo incluye las aplicaciones que he considerado apropiadas y accesibles, considerado el nivel matemático adquirido a lo largo de los estudios de licenciatura.

En la actualidad existen algunas monografías sobre las aplicaciones de las funciones características. El presente trabajo pretende servir como una introducción elemental para el estudio posterior de textos más avanzados. Además se incluyen aplicaciones que aparecieron en fecha posterior a la publicación de tales monografías.

Como ocurre en todas las disciplinas, para poder comprender y hacer uso de las técnicas de la teoría de la probabilidad es necesario conocer sus fundamentos teóricos. Por tal razón el primer capítulo de esta tesis está dedicado a repasar las definiciones básicas de teoría de la probabilidad vistas dentro de la axiomática de Kolmogorov. El modelo de Kolmogorov hace uso de conceptos de la teoría de la medida por ello es que en el capítulo 1 también se pueden encontrar definiciones y resultados elementales de teoría de la medida.

Otra de las herramientas de las que se ha hecho amplio uso en probabilidad es la integral de Lebesgue. Dentro del contexto probabilístico, la integral de Lebesgue de una variable aleatoria, digamos X, respecto a su medida de probabilidad es conocida como el valor esperado, o la esperanza, de X. Ya que la función característica es un valor esperado, entonces se vuelve necesario tener un buen conocimiento de las propiedades del valor esperado. En el capítulo 2 se define y se prueban las propiedades más importantes del valor esperado.

En el capítulo 3 se define la función característica y se dan varios ejemplos del cálculo de funciones características para distribuciones conocidas. Este capítulo es fundamental para él estudio de las aplicaciones de la función característica pues en el también se incluyen teoremas importantes, como son los teoremas de inversión, de unicidad y de continuidad que hacen de la función característica una herramienta sumamente poderosa en las aplicaciones. También se incluyen algunas aplicaciones de la función característica como son: el cálculo de momentos, la determinación de la distribución de sumas de variables aleatorias independientes y la caracterización de propiedades como la simetría, la independencia y la convergencia débil.

En el capítulo 4, se ilustran las aplicaciones de la función característica a algunos problemas que surgen frecuentemente en probabilidad y estadística. En la teoría de la probabilidad, la función característica ha sido usada para estudiar el comportamiento límite de variable aleatorias. También ha demostrado ser de mucha utilidad cuando se estudian ciertas familias de distribuciones como son las distribuciones estables e infinitamente divisibles. En inferencia estadística a menudo es necesario conocer la distribución exacta de funciones de variables aleatorias, existen diferentes métodos para resolver este problema uno de ellos es usando la función característica. También se han propuesto algunos métodos de estimación y pruebas de bondad de ajuste basados en la función característica empírica.

Se incluyen tres apéndices. En el primer apéndice, apéndice A, se presentan las nociones básicas de convergencia estocástica, incluyendo leyes de los grandes números y teoremas del límite central. El apéndice B, proporciona un resumen de las propiedades básicas de los polinomios de Chebyshev, que son usados en la Sección 4.7. En el apéndice C se incluyen dos tablas con las distribuciones de probabilidad y sus funciones características, sólo se incluyen aquellas distribuciones que, además de ser muy usadas, tienen una expresión en términos de funciones elementales para su función característica.

Preliminares

A continuación describimos notación con la cual probablemente el lector no está familiarizado.

El símbolo ":=" quiere decir "*igual por definición*".

Si A y B son dos conjuntos cualesquiera el conjunto diferencia entre A y B se escribe como $A\backslash B$, es decir

$$A\backslash B := \{x : x \in A \text{ y } x \notin B\}.$$

El complemento del conjunto A se denota por A^c. El conjunto de todos los subconjuntos de A será denotado por 2^A.

El conjunto de los números reales será denotado por $\mathbb{R}$.

Para $a, b \in \mathbb{R}$, $a < b$, los intervalos abierto y cerrado con extremos a y b serán denotados por $]a, b[$, y $[a, b]$ respectivamente, es decir

$$]a, b[= \{x \in \mathbb{R} : a < x < b\}$$

y

$$[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}.$$

Análogamente,

$$[a, b[= \{x \in \mathbb{R} : a \leq x < b\}$$

y

$$]a, b] = \{x \in \mathbb{R} : a < x \leq b\}.$$

El conjunto de los reales extendidos, $[-\infty, \infty] := \{-\infty\} \cup \mathbb{R} \cup \{\infty\}$, será denotado por $\overline{\mathbb{R}}$. Para cualquier número real x, $-\infty + x := -\infty$ y $\infty + x := \infty$ mientras que $-\infty + \infty$ estará indefinido.

El conjunto de los números complejos es denotado por $\mathbb{C}$.

La función Gamma, denotada por $\Gamma(\cdot)$, se define como

$$\Gamma(z) := \int_0^\infty e^{-t} t^{z-1} dt.$$

Denotaremos por $B(\cdot,\cdot)$ la función Beta, es decir

$$B(z,\zeta) := \int_0^1 t^{z-1}(1-t)^{\zeta-1}\,dt.$$

Sean $u(x)$ y $v(x)$ dos funciones. La notación $u(x) = O(v(x))$ cuando $x \to L$, sirve para denotar que $\left|\frac{u(x)}{v(x)}\right|$ permanece acotado cuando $x \to L$. La notación $u(x) = o(v(x))$, $x \to L$, quiere decir que

$$\lim_{x \to L} \frac{u(x)}{v(x)} = 0.$$

Las letras X, Y, W y Z con o sin subíndices denotarán variables aleatorias.

La notación $X \sim F$ se lee como "*la variable aleatoria X tiene distribución F*".

Los símbolos ■ y ▲ denotan que se ha llegado al final de una demostración y de un ejemplo respectivamente.

Capítulo 1

Probabilidad y Variables Aleatorias

1.1 Introducción

Actualmente la teoría de la probabilidad es considerada como la rama de las matemáticas que se ocupa del análisis matemático de los eventos aleatorios.

Como ciencia, la teoría de la probabilidad nace a mediados del siglo XVII, cuando Pierre de Fermat (1601-1665) y Blaise Pascal (1623-1662), en su famosa correspondencia, establecen los fundamentos de la teoría elemental de la probabilidad. Sin embargo, aún a principios del siglo XX la teoría de la probabilidad todavía carecía de fundamentos teóricos rigurosos y era considerada parte de la física.

En el año de 1900, en París, se realizaron una serie de conferencias sobre matemáticas. En ellas se invitó al matemático David Hilbert para dar a conocer su opinión sobre la tarea de las matemáticas en el siglo que comenzaba. Hilbert, identificó 23 problemas[1] que consideraba tendrían que ser los temas centrales para los matemáticos de este siglo. En su sexto problema. Hilbert habló sobre la necesidad de axiomatizar aquellas ramas de la física que hacían uso de las matemáticas e hizo mención explícita de la mecánica y de la teoría de la probabilidad. A partir de este momento muchos matemáticos, entre ellos Emile Borel y Rademacher, hicieron diferentes propuestas. Desafortunadamente estas propuestas resultaban inadecuadas para tratar muchos problemas.

Fue en el año de 1931 que al fin, von Mises, dio un fundamento riguroso y firme de la teoría de la probabilidad, tomando frecuencias y leyes de los grandes números como ideas básicas. Sin embargo, este esquema fue completamente opacado en 1933 por la aparición de la monografía de A. N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung,* que daba

[1] Hoy en día, conocidos como los 23 problemas de Hilbert.

un fundamento de la teoría de la probabilidad basado en ideas de la teoría de la medida. Desde entonces casi todos los probabilistas han trabajado bajo el modelo de Kolmogorov y sólo unos cuantos han seguido la idea de von Mises.

El objetivo de este capítulo es presentar brevemente las definiciones e ideas básicas de la teoría de la probabilidad enmarcadas en la axiomática de Kolmogorov. En la sección 1.2 se estudian los conceptos básicos de la teoría de la medida y en la sección 1.3 se llevan estas definiciones al caso particular de un espacio de probabilidad. La sección 1.4 presenta las propiedades de las variables aleatorias y de las funciones de distribución. Cerramos el capítulo, en la sección 1.5, con un breve estudio sobre independencia.

1.2 Fundamentos de la Teoría de la Probabilidad

1.2.1 Clases de Conjuntos

En lo siguiente, Ω denota a un conjunto arbitrario no vacío al que llamaremos **espacio** y los elementos de Ω serán denotados genéricamente por ω.

Definición 1.2.1 *Una clase $\mathcal{I}$ de subconjuntos de un espacio Ω se llama* ***semi-álgebra*** *en Ω si*

1. Ω, $\emptyset \in \mathcal{I}$.
2. *Si $A_1, A_2 \in \mathcal{I}$, entonces $A_1 \cap A_2 \in \mathcal{I}$.*
3. *Para cada $A \in \mathcal{I}$, existe una partición finita de A^c en $\mathcal{I}$.*

Ejemplo 1.2.1 *Sea $\Omega = \mathbb{R}$ y J la clase que contiene a $\emptyset$, Ω, y a todos los conjuntos de la forma*

$$\begin{array}{ll}]a,b] & -\infty < a < b < \infty \\ ]-\infty,b] & -\infty < b < \infty \\ ]a,\infty[& -\infty < a < \infty. \end{array}$$

Claramente J es cerrada bajo intersecciones finitas y además el complemento de cada elemento de la clase puede expresarse como la unión disjunta de un numero finito de elementos de J. Por lo tanto, J es un semi-álgebra. ▲

Ejemplo 1.2.2 *Sea $\Omega = \mathbb{R}$ y denotemos por $\mathcal{A}$ la clase de todos los subconjuntos de la forma $A \cap B$, donde A es un conjunto cerrado y B es un conjunto abierto de Ω. Entonces $\mathcal{A}$ es un semi-álgebra.* ▲

Definición 1.2.2 *Una clase $\mathcal{A}$ sobre Ω es llamada **álgebra** si*

1. *$\mathcal{A}$ es no vacía .*
2. *$A^c \in \mathcal{A}$ siempre que $A \in \mathcal{A}$.*
3. *$A_1 \cup A_2 \in \mathcal{A}$ siempre que $A_1, A_2 \in \mathcal{A}$.*

Sea $\mathcal{A}$ un álgebra, como $\mathcal{A}$ es no vacía, entonces existe algún subconjunto A de Ω, tal que $A \in \mathcal{A}$, así por 2 y 3 de la definición 1.2.2, se cumple que $A^c \in \mathcal{A}$, $A \cup A^c = \Omega \in \mathcal{A}$ y $\Omega^c = \emptyset \in \mathcal{A}$. Si además se proponen $A_1, A_2 \in \mathcal{A}$, entonces nuevamente, por 2 y 3 de la definición 1.2.2, A_1^c, $A_2^c \in \mathcal{A}$, $A_1^c \cup A_2^c \in \mathcal{A}$, pero, por las leyes de De Morgan $(A_1^c \cup A_2^c)^c = A_1 \cap A_2$. entonces $A_1 \cap A_2 \in \mathcal{A}$. Usando argumentos similares e inducción es fácil demostrar que $\mathcal{A}$ es cerrada bajo cualquier número finito de operaciones de conjuntos. Así, se tiene el

Teorema 1.2.1 *Si $\mathcal{A}$ es álgebra sobre un espacio Ω, entonces*

1. *Ω, $\emptyset \in \mathcal{A}$.*
2. *$\mathcal{A}$ es cerrada bajo un número finito de operaciones de conjuntos.*

Si para alguna clase $\mathcal{A}$ sobre Ω, que sea cerrada bajo uniones finitas, se tiene que $\Omega \in \mathcal{A}$ o $\emptyset \in \mathcal{A}$, entonces $\mathcal{A}$ es no vacía; recíprocamente, como ya se demostró, si $\mathcal{A}$ es no vacía, entonces $\Omega \in \mathcal{A}$ y $\emptyset \in \mathcal{A}$ de donde resulta claro que se puede definir un álgebra sustituyendo la condición de que $\mathcal{A}$ sea no vacía por la condición $\Omega \in \mathcal{A}$, o la condición $\emptyset \in \mathcal{A}$. Análogamente, se deduce que es indistinto pedir que $\mathcal{A}$ sea cerrada bajo uniones o intersecciones finitas. Lo anterior implica que si se quiere probar que una clase sobre Ω es álgebra, entonces probar que la clase es no vacía equivale a probar que Ω (o bien $\emptyset$) pertenece a la clase; de la misma forma, probar que la clase es cerrada bajo intersecciones es equivalente a probar que es cerrada bajo uniones finitas.

Ejemplo 1.2.3 *Para un conjunto Ω no vacío, la clase $\mathcal{A} = \{\emptyset, \Omega\}$ es un álgebra de conjuntos. Este es el álgebra más pequeño posible, con respecto a inclusión.* ▲

Ejemplo 1.2.4 *Para un conjunto Ω no vacío, su conjunto potencia, denotado por 2^Ω forma un álgebra. Este es el álgebra más grande posible.* ▲

Ejemplo 1.2.5 *Un conjunto $A \subset \Omega$ es llamado **cofinito** si su complemento es finito. Sea $\mathcal{A}$ la clase que contiene a todos los subconjuntos finitos y cofinitos de Ω. Entonces $\mathcal{A}$ es un álgebra.* ▲

Es claro que un álgebra siempre es un semi-álgebra; sin embargo, el ejemplo 1.2.2 muestra que el recíproco no es cierto.

Definición 1.2.3 *Sea Ω un espacio. Una clase $\mathcal{A}$ es llamada* ***$\sigma-$álgebra*** *si*

1. *$\mathcal{A}$ es no vacía.*
2. *Si $A \in \mathcal{A}$, entonces $A^c \in \mathcal{A}$.*
3. *Si para cualquier sucesión $\{A_n\}_{n\geq 1}$ en $\mathcal{A}$, se tiene $\bigcup_{n\geq 1} A_n \in \mathcal{A}$.*

Definición 1.2.4 *Si Ω es un espacio y $\mathcal{A}$ es un $\sigma-$álgebra en Ω, entonces la pareja $(\Omega, \mathcal{A})$ es llamada* ***espacio medible***.

Ejemplo 1.2.6 *Claramente las clases de los ejemplos 1.2.3 y 1.2.4, además de ser álgebras, son σ-álgebras.* ▲

Ejemplo 1.2.7 *La clase presentada en el ejemplo 1.2.5 es $\sigma-$álgebra si, y sólo si, Ω es finito.* ▲

Ejemplo 1.2.8 *Un conjunto $A \subset \Omega$, es llamado* ***conumerable*** *si su complemento es numerable. Si $\mathcal{A}$ es la clase que contiene a todos los subconjuntos numerables y conumerables de Ω. Entonces $\mathcal{A}$ es un $\sigma-$álgebra.* ▲

Al igual que en el caso de un álgebra, se sigue fácilmente que cada $\sigma-$álgebra es cerrado bajo un número contable de operaciones de conjuntos. Además, si en 3 de la definición 1.2.3, tomamos $A_n = \emptyset$, para $n \geq 3$, entonces $\bigcup_{j\geq 1} A_j = A_1 \cup A_2$, lo que demuestra que todo $\sigma-$álgebra es álgebra. Las propiedades anteriores quedan resumidas en el siguiente teorema.

Teorema 1.2.2 *Si $\mathcal{A}$ es un $\sigma-$álgebra sobre un espacio Ω entonces*

1. *$\mathcal{A}$ es cerrado bajo un número contable de operaciones de conjuntos.*
2. *$\mathcal{A}$ es álgebra.*

Del teorema anterior se concluye que si $\mathcal{A}$ es un $\sigma-$álgebra, entonces Ω, $\emptyset \in \mathcal{A}$. También en el caso de un $\sigma-$álgebra, el probar que la clase es no vacía equivale a probar que Ω (o bien $\emptyset$) pertenece a la clase, y probar que la clase es cerrada bajo intersecciones numerables equivale a probar que es cerrada bajo uniones numerables.

Sea Ω un espacio y $\mathcal{T}$ alguna familia de $\sigma-$álgebras en Ω. Para toda $\mathcal{A} \in \mathcal{T}$, $\Omega \in \mathcal{A}$ entonces $\Omega \in \bigcap\limits_{\mathcal{A}\in\mathcal{T}} \mathcal{A}$; además, si $A \in \bigcap\limits_{\mathcal{A}\in\mathcal{T}} \mathcal{A}$ entonces, para toda $\mathcal{A} \in \mathcal{T}$, $A \in \mathcal{A}$ y se tiene que $A^c \in \mathcal{A}$, de donde se sigue que $A^c \in \bigcap\limits_{\mathcal{A}\in\mathcal{T}} \mathcal{A}$; por último, si $A_1, A_2, ...$ es una sucesión en $\bigcap\limits_{\mathcal{A}\in\mathcal{T}} \mathcal{A}$, entonces, para cada $\mathcal{A} \in \mathcal{T}$, $A_1, A_2, ...$ es una sucesión en $\mathcal{A}$, de donde $\bigcup_{j\geq 1} A_j \in \mathcal{A}$, así $\bigcup_{j\geq 1} A_j \in \bigcap\limits_{\mathcal{A}\in\mathcal{T}} \mathcal{A}$ lo que demuestra el

Teorema 1.2.3 *Sea $\mathcal{T}$ una familia de $\sigma-$álgebras en Ω. Entonces $\bigcap\limits_{\mathcal{A}\in\mathcal{T}} \mathcal{A}$ es un $\sigma-$álgebra en Ω .*

Como la intersección arbitraria de $\sigma-$álgebras es un $\sigma-$álgebra, parece natural preguntarnos si, en general, la unión de $\sigma-$álgebras es un $\sigma-$álgebra. La respuesta es negativa. Un contraejemplo puede encontrarse en Stoyanov [1988, p. 6].

Definición 1.2.5 *Si C es una clase sobre Ω. El* ***mínimo $\sigma-$álgebra generado*** *por la clase C, denotado por $\sigma(C)$, es un $\sigma-$álgebra tal que*

1. *$C \subset \sigma(C)$.*
2. *Si $\mathcal{A}$ es $\sigma-$álgebra en Ω para la cual $C \subset \mathcal{A}$ entonces $\sigma(C) \subset \mathcal{A}$.*

Sea C una clase formada por subconjuntos de Ω y $\mathcal{T}_C := \{\mathcal{A} : C \subset \mathcal{A},\ \mathcal{A}$ es σ − álgebra en $\Omega\}$ y $\mathcal{F} := \bigcap\limits_{\mathcal{A}\in\mathcal{T}_C} \mathcal{A}$. Claramente $2^\Omega \in \mathcal{T}_C$, lo que implica que $\mathcal{T}_C$ es no vacía; además por el teorema 1.2.3, $\mathcal{F}$ es $\sigma-$álgebra y como para toda $\mathcal{A} \in \mathcal{T}_C$, $\bigcap\limits_{\mathcal{A}\in\mathcal{T}_C} \mathcal{A} \subset \mathcal{A}$ entonces $\mathcal{F}$ es minimal. Si además suponemos que $\mathcal{F}'$ es otra $\sigma-$álgebra que cumple con 1 y 2 de la definición 1.2.5, entonces $\mathcal{F}' \subset \mathcal{F}$ y $\mathcal{F} \subset \mathcal{F}'$ de donde $\mathcal{F}' = \mathcal{F}$ y se sigue que $\mathcal{F}$ es única. Así se tiene el siguiente resultado.

Teorema 1.2.4 *El mínimo $\sigma-$álgebra generado por la clase C en Ω siempre existe y es único.*

Definición 1.2.6 *Si Ω es el conjunto de los números reales $\mathbb{R}$, el $\sigma-$**álgebra de Borel** en $\mathbb{R}$, denotado por $\mathcal{B}(\mathbb{R})$, es el mínimo $\sigma-$álgebra generado por la clase $\mathcal{G}$ de los conjuntos abiertos de $\mathbb{R}$. Es decir $\mathcal{B}(\mathbb{R}) := \sigma(\mathcal{G})$.*

Teorema 1.2.5 *Sea $\mathcal{J} = \{]-\infty, x] : -\infty < x < \infty\}$. Entonces $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{J})$.*

Demostración. Como $\mathcal{B}(\mathbb{R})$ es $\sigma-$álgebra, entonces, para cualquier $x \in \mathbb{R}$, $]x, \infty[\in \mathcal{B}(\mathbb{R})$ de donde es claro que $\mathcal{J} \subset \mathcal{B}(\mathbb{R})$ y entonces, por la definición 1.2.5, se sigue que $\sigma(\mathcal{J}) \subset \mathcal{B}(\mathbb{R})$.

Ahora, sea $\mathcal{G}$ la clase formada por todos los conjuntos abiertos de $\mathbb{R}$ y supongamos que $A \in \mathcal{G}$, entonces, como A es abierto, para cada $x \in A$ existe $\delta > 0$ tal que

$$x \in]x-\delta, x+\delta[\,;$$

además existen números racionales p y q tales que

$$]p, q[\subset]x-\delta, x+\delta[\qquad \text{y} \qquad x \in]p, q[\,.$$

Como el conjunto de los números racionales es numerable, entonces el conjunto de los intervalos $]p, q]$ es numerable y por lo tanto, también lo es el subconjunto de los subintervalos $]p_k, q_k]$, $k = 1, 2, ...$, que están contenidos en A.

Ahora, como $]-\infty, q_k]$ y $]-\infty, p_k]$ pertenecen a $\sigma(\mathcal{J})$, para $k = 1, 2, \ldots$, también

$$]p_k, q_k] =]-\infty, q_k] \setminus]-\infty, p_k]$$

pertenece a $\sigma(\mathcal{J})$. Finalmente notamos que A puede ser escrito como

$$\bigcup_{k=1}^{\infty}]p_k, q_k] ;$$

entonces, para cada conjunto abierto A, $A \in \sigma(\mathcal{J})$, y, como $\emptyset$ es abierto, $\mathcal{G} \subset \sigma(\mathcal{J})$ de donde, por la definición 1.2.5, $\mathcal{B}(\mathbb{R}) \subset \sigma(\mathcal{J})$ y por lo tanto, $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{J})$. ■

Consideremos dos clases $\mathcal{C}$ y $\mathcal{C}'$ en un espacio Ω tales que $\mathcal{C} \subset \mathcal{C}'$; ya que, $\mathcal{C}' \subset \sigma(\mathcal{C}')$, entonces, por transitividad, $\mathcal{C} \subset \sigma(\mathcal{C}')$ y se sigue que $\sigma(\mathcal{C}) \subset \sigma(\mathcal{C}')$. Si tuviéramos la condición adicional de que $\mathcal{C}' \subset \sigma(\mathcal{C})$, entonces también $\sigma(\mathcal{C}')$ estaría contenida en $\sigma(\mathcal{C})$ y por lo tanto $\sigma(\mathcal{C}) = \sigma(\mathcal{C}')$. Así, podemos establecer el

Teorema 1.2.6 *Sean $\mathcal{C}$ y $\mathcal{C}'$ clases en un espacio Ω.*

1. *Si $\mathcal{C} \subset \mathcal{C}'$ entonces $\sigma(\mathcal{C}) \subset \sigma(\mathcal{C}')$.*

2. Si $\mathcal{C} \subset \mathcal{C}' \subset \sigma(\mathcal{C})$ *entonces* $\sigma(\mathcal{C}) = \sigma(\mathcal{C}')$.

Ejemplo 1.2.9 *Sea J la clase definida en el ejemplo 1.2.1 y $\mathcal{J}$ la clase considerada en el teorema 1.2.5. Claramente $\mathcal{J} \subset J$ y además $J \subset \sigma(\mathcal{J}) = \mathcal{B}(\mathbb{R})$, entonces, por el teorema 1.2.6, tenemos que $\sigma(J) = \mathcal{B}(\mathbb{R})$.* ▲

1.2.2 Transformaciones Medibles

Definición 1.2.7 *Sean Ω_1 y Ω_2 espacios y $f : \Omega_1 \to \Omega_2$. Para cada subconjunto A en Ω_2 se define la* ***imagen inversa*** *de A como*

$$f^{-1}(A) := \{\omega \in \Omega_1 : f(\omega) \in A\}.$$

Análogamente, si $\mathcal{C}$ es una clase de subconjuntos de Ω_2, la ***imagen inversa de la clase*** *$\mathcal{C}$ se define como*

$$f^{-1}(\mathcal{C}) := \left\{f^{-1}(A) : A \in \mathcal{C}\right\}.$$

La definición 1.2.7 implica que $\omega \in f^{-1}(A)$ si, y sólo si, $f(\omega) \in A$. De este hecho y de la definición 1.2.7 se siguen las siguientes propiedades.

Teorema 1.2.7 *Si A y B son subconjuntos de Ω_2 y $f : \Omega_1 \to \Omega_2$, entonces*

1. $f^{-1}(\Omega_2) = \Omega_1$.
2. $f^{-1}(A \backslash B) = f^{-1}(A) \backslash f^{-1}(B)$.
3. $f^{-1}(A^c) = \left(f^{-1}(A)\right)^c$.
4. *Si para cada $t \in T$ se tiene que $A_t \subset \Omega_2$ entonces*

$$f^{-1}\left(\bigcup_{t\in T} A_t\right) = \bigcup_{t\in T} f^{-1}(A_t)$$
$$f^{-1}\left(\bigcap_{t\in T} A_t\right) = \bigcap_{t\in T} f^{-1}(A_t).$$

5. *Si $A \subset B$ entonces $f^{-1}(A) \subset f^{-1}(B)$.*
6. *Si $\mathcal{C}$ y $\mathcal{D}$ son dos clases no vacías en Ω_2 tal que $\mathcal{C} \subset \mathcal{D}$, entonces $f^{-1}(\mathcal{C}) \subset f^{-1}(\mathcal{D})$.*

Otras propiedades importantes de la imagen inversa se dan en el siguiente teorema.

Teorema 1.2.8 *Sean Ω_1 y Ω_2 espacios y $f : \Omega_1 \to \Omega_2$.*

1. *Si $\mathcal{F}_2$ es $\sigma-$álgebra en Ω_2, entonces $\mathcal{F}_1 = f^{-1}(\mathcal{F}_2)$ es $\sigma-$álgebra en Ω_1.*
2. *Si $\mathcal{F}_1$ es $\sigma-$álgebra en Ω_1*
$$\mathcal{F}_2 := \left\{A : A \subset \Omega_2 \quad y \quad f^{-1}(A) \in \mathcal{F}_1\right\}$$
es $\sigma-$álgebra en Ω_2.
3. *Si C es una clase de conjuntos en Ω_2*
$$\sigma\left(f^{-1}(C)\right) = f^{-1}(\sigma(C)).$$

Demostración.

1. Como $\Omega_2 \in \mathcal{F}_2$, entonces, por 1 del teorema 1.2.7, $f^{-1}(\Omega_2) = \Omega_1 \in \mathcal{F}_1$. Si A y A^c pertenecen a $\mathcal{F}_2$, entonces $f^{-1}(A)$ y $f^{-1}(A^c) = \left(f^{-1}(A)\right)^c$ pertenecen a $\mathcal{F}_1$. Además, si $A_1, A_2, ... \in \mathcal{F}_2$, se tiene que $\bigcup_{j\geq 1} A_j \in \mathcal{F}_2$, entonces $f^{-1}(A_1), f^{-1}(A_2), ... \in \mathcal{F}_1$ y
$$f^{-1}\left(\bigcup_{j\geq 1} A_j\right) = \bigcup_{j\geq 1} f^{-1}(A_j) \in \mathcal{F}_1.$$
Por lo tanto, $\mathcal{F}_1 = f^{-1}(\mathcal{F}_2)$ es $\sigma-$álgebra en Ω_1.
2. Claramente $\Omega_2 \in \mathcal{F}_2$. Si $A \in \mathcal{F}_2$, entonces $A^c \subset \Omega_2$ y, como $\mathcal{F}_1$ es $\sigma-$álgebra, $f^{-1}(A^c) \in \mathcal{F}_1$, de donde $A^c \in \mathcal{F}_2$. Ahora, si $A_1, A_2, ... \in \mathcal{F}_2$ obviamente $\bigcup_{j\geq 1} A_j \subset \Omega_2$; además, como $f^{-1}(A_1), f^{-1}(A_2), ... \in \mathcal{F}_1$ y $\mathcal{F}_1$ es $\sigma-$álgebra, entonces $f^{-1}\left(\bigcup_{j\geq 1} A_j\right) \in \mathcal{F}_1$, así $\bigcup_{j\geq 1} A_j \in \mathcal{F}_2$. Por lo tanto, $\mathcal{F}_2$ es $\sigma-$álgebra en Ω_2.
3. Por la afirmación 1 de este teorema, $f^{-1}(\sigma(C))$ es $\sigma-$álgebra en Ω_1, entonces, como $C \subset \sigma(C)$,
$$f^{-1}(C) \subset f^{-1}(\sigma(C))$$
y como $\sigma\left(f^{-1}(C)\right)$ es mínima se cumple que
$$\sigma\left(f^{-1}(C)\right) \subset f^{-1}(\sigma(C)). \tag{1.1}$$

Ahora, sea

$$\mathcal{H} := \left\{ A \subset \Omega_2 : f^{-1}(A) \subset \sigma\left(f^{-1}(\mathcal{C})\right) \right\}.$$

Como para cada elemento $B \in \mathcal{C}$ se tiene que $B \subset \Omega_2$ y $f^{-1}(B) \in f^{-1}(\mathcal{C})$, entonces $\mathcal{C} \subset \mathcal{H}$. Además, la afirmación 2 implica que $\mathcal{H}$ es $\sigma-$álgebra, y ya que $\sigma(\mathcal{C})$ es mínima,

$$\sigma(\mathcal{C}) \subset \sigma(\mathcal{H}),$$

de donde

$$f^{-1}(\sigma(\mathcal{C})) \subset f^{-1}(\sigma(\mathcal{H}));$$

pero, por definición de $\mathcal{H}$, cada elemento de $\mathcal{H}$ está en $\sigma\left(f^{-1}(\mathcal{C})\right)$, entonces

$$f^{-1}(\mathcal{H}) \subset \sigma\left(f^{-1}(\mathcal{C})\right)$$

y por transitividad

$$f^{-1}(\sigma(\mathcal{C})) \subset \sigma\left(f^{-1}(\mathcal{C})\right). \tag{1.2}$$

Así, de las ecuaciones (1.1) y (1.2), se concluye que

$$f^{-1}(\sigma(\mathcal{C})) = \sigma\left(f^{-1}(\mathcal{C})\right).$$

■

Definición 1.2.8 *Sean $(\Omega_1, \mathcal{A}_1)$ y $(\Omega_2, \mathcal{A}_2)$ espacios medibles y $f : \Omega_1 \to \Omega_2$. Se dice que f es una* ***transformación medible*** *de $(\Omega_1, \mathcal{A}_1)$ a $(\Omega_2, \mathcal{A}_2)$ si $f^{-1}(\mathcal{A}_2) \subset \mathcal{A}_1$.*

Definición 1.2.9 *Una transformación medible f de $(\Omega_1, \mathcal{A}_1)$ a $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ es llamada* ***función real-medible****, o simplemente* ***medible****, sobre $(\Omega_1, \mathcal{A}_1)$.*

Es posible demostrar que la definición 1.2.9 es equivalente a decir que f es medible si para cada $x \in \mathbb{R}$ se tiene que

$$f^{-1}(]-\infty, x]) \in \mathcal{A}_1; \tag{1.3}$$

o equivalentemente, f es medible, si para cada $x \in \mathbb{R}$ la condición (1.3) se cambia por alguna de las siguientes

$$f^{-1}(]-\infty, x[) \in \mathcal{A}_1,$$

$$f^{-1}(]x, +\infty[) \in \mathcal{A}_1,$$

$$f^{-1}([x, +\infty[) \in \mathcal{A}_1,$$

ya que cada uno de los intervalos, $]-\infty, x[$, $]x, +\infty[$, $[x, +\infty[$, está en $\mathcal{B}(\mathbb{R})$.

Definición 1.2.10 *Una transformación medible del espacio* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *en sí mismo es llamada función* ***Borel-medible.***

Teorema 1.2.9 *Sea* $(\Omega, \mathcal{A})$ *un espacio medible. La función* $f : \Omega \to \mathbb{R}$ *es medible si y sólo si*

$$f^{-1}(\mathcal{B}(\mathbb{R})) \subset \mathcal{A} \tag{1.4}$$

Demostración. Para cada $x \in \mathbb{R}$ se tiene que $]-\infty, x] \in \mathcal{B}(\mathbb{R})$, entonces, si se cumple (1.4), $f^{-1}(]-\infty, x]) \in \mathcal{A}$ y por lo tanto f es medible. Por otro lado, si f es medible entonces

$$f^{-1}(\mathcal{J}) \subset \mathcal{A}$$

y

$$\sigma\left(f^{-1}(\mathcal{J})\right) \subset \mathcal{A}$$

y, como $\sigma\left(f^{-1}(\mathcal{J})\right) = f^{-1}(\sigma(\mathcal{J}))$ y $\sigma(\mathcal{J}) = \mathcal{B}(\mathbb{R})$, se sigue que

$$f^{-1}(\mathcal{B}(\mathbb{R})) \subset \mathcal{A}.$$

■

En el siguiente teorema se dan algunos resultados básicos relacionados con funciones medibles; la demostración es trivial a partir de la definición de medibilidad.

Teorema 1.2.10 *Si* $(\Omega, \mathcal{A})$ *es un espacio medible y* $f : \Omega \to \mathbb{R}$ *una función medible. Entonces*

1. *Toda función constante es medible.*
2. *Si* c *es una constante,* cf *es medible.*
3. $|f|$ *es medible.*
4. f^2 *es medible.*
5. *Si* $f \neq 0$ *sobre* Ω *,* $\dfrac{1}{f}$ *es medible.*

Supongamos que $(\Omega, \mathcal{A})$ es un espacio medible y que $f_1 : \Omega \to \mathbb{R}$ y $f_2 : \Omega \to \mathbb{R}$ son funciones medibles. Entonces, para $x \in \mathbb{R}$, tenemos que

$$\begin{aligned} (f_1 + f_2)^{-1}(]-\infty, x[) &= \{\omega \in \Omega : f_1(\omega) + f_2(\omega) < x\} \\ &= \{\omega \in \Omega : f_1(\omega) < x - f_2(\omega)\}, \end{aligned}$$

entonces existe un número racional q tal que

$$f_1(\omega) < q < x - f_2(\omega);$$

ya que los racionales son numerables, se tiene que

$$\begin{aligned}(f_1 + f_2)^{-1}(]-\infty, x[) &= \bigcup_q \{\omega \in \Omega : f_1(\omega) < q < x - f_2(\omega)\} \\ &= \bigcup_q [\{\omega \in \Omega : f_1(\omega) < q\} \cap \{\omega \in \Omega : q < x - f_2(\omega)\}].\end{aligned}$$

Finalmente notamos que $\{\omega \in \Omega : f_1(\omega) < q\}$ y $\{\omega \in \Omega : q < x - f_2(\omega)\}$ están en $\mathcal{A}$, de donde

$$(f_1 + f_2)^{-1}(]-\infty, x[) \in \mathcal{A}.$$

Por lo tanto, la suma de funciones medibles es una función medible.

Ya que $f_1 + f_2$ es medible si f_1 y f_2 son funciones medibles, entonces, por el teorema 1.2.10, se sigue que $(f_1 + f_2)^2$ y $(f_1 - f_2)^2$ son medibles, por lo tanto,

$$f_1 f_2 = \frac{1}{4}\left[(f_1 + f_2)^2 - (f_1 - f_2)^2\right]$$

es una función medible.

Lo anterior demuestra el

Teorema 1.2.11 *Si $(\Omega, \mathcal{A})$ es un espacio medible, $f_1 : \Omega \to \mathbb{R}$ y $f_2 : \Omega \to \mathbb{R}$ son funciones medibles. Entonces*

1. *$f_1 + f_2$ es medible*
2. *$f_1 f_2$ es medible*

Supongamos ahora que $f : \Omega \to \mathbb{R}$ es una función medible en $(\Omega, \mathcal{A})$ y $g : \mathbb{R} \to \mathbb{R}$ es una función Borel-medible, entonces, para todo $\omega \in \Omega$,

$$[g \circ f](\omega) = g(f(\omega));$$

además, es fácil ver que para cualquier conjunto $G \in \mathcal{B}(\mathbb{R})$

$$[g \circ f]^{-1}(G) = \left[f^{-1} \circ g^{-1}\right](G).$$

Entonces, por la medibilidad de g, se tiene que $g^{-1}(G) \in \mathcal{B}(\mathbb{R})$ y se sigue, de la medibilidad de f , que

$$f^{-1}\left(g^{-1}(G)\right) \in \mathcal{A}.$$

Así, queda demostrado el

Teorema 1.2.12 *Si $f : \Omega \to \mathbb{R}$ es una función medible en $(\Omega, \mathcal{A})$ y $g : \mathbb{R} \to \mathbb{R}$ es una función Borel-medible, entonces $g \circ f$ es medible.*

Consideremos dos funciones medibles $f_1 : \Omega \to \mathbb{R}$ y $f_2 : \Omega \to \mathbb{R}$ en un espacio medible $(\Omega, \mathcal{A})$. Observamos que

$$\begin{aligned} \max{}^{-1}(f_1, f_2)(]-\infty, x]) &= \{\omega \in \Omega : [\max(f_1, f_2)](\omega) \leq x\} \\ &= \{\omega \in \Omega : f_1(\omega) \leq x\} \cap \{\omega \in \Omega : f_2(\omega) \leq x\}, \end{aligned}$$

entonces es claro que $\max(f_1, f_2)$ es medible.

También se tiene

$$\min{}^{-1}(f_1, f_2)(]-\infty, x]) = \{\omega \in \Omega : f_1(\omega) \leq x\} \cup \{\omega \in \Omega : f_2(\omega) \leq x\}$$

y se sigue que $\min(f_1, f_2)$ es medible. Así queda establecido el

Teorema 1.2.13 *Si $f_1 : \Omega \to \mathbb{R}$ y $f_2 : \Omega \to \mathbb{R}$ son funciones medibles en un espacio medible $(\Omega, \mathcal{A})$. Entonces $\max(f_1, f_2)$ y $\min(f_1, f_2)$ son medibles.*

Corolario 1.2.14 *Si $(\Omega, \mathcal{A})$ es un espacio medible y $f : \Omega \to \mathbb{R}$ es medible entonces*

$$f^{+} := \max(0, f) \qquad y \qquad f^{-} := \min(0, -f)$$

son medibles.

Si $\{f_n\}_{n\geq 1}$ es una sucesión de funciones medibles en un espacio medible $(\Omega, \mathcal{A})$, entonces

$$\begin{aligned} [\sup f_n]^{-1}(]-\infty, x]) &= \left\{\omega \in \Omega : \sup_n f_n(\omega) \leq x\right\} \\ &= \bigcap_{n=1}^{\infty} \{\omega \in \Omega : f_n(\omega) \leq x\} \end{aligned}$$

y como para cada n se tiene que $\{\omega \in \Omega : f_n(\omega) \leq x\} \in \mathcal{A}$, se sigue que $\sup f_n$ es medible. Así mismo,

$$\begin{aligned} [\inf f_n]^{-1}(]-\infty, x[) &= \{\omega \in \Omega : \inf f_n(\omega) < x\} \\ &= \bigcup_{n=1}^{\infty} \{\omega \in \Omega : f_n(\omega) < x\}, \end{aligned}$$

de donde $\inf_n f_n$ es medible. Esto se resume en el

Teorema 1.2.15 *Si $\{f_n\}_{n\geq 1}$ es una sucesión de funciones medibles en un espacio medible $(\Omega, \mathcal{A})$, entonces*

1. $\sup f_n$ *es medible.*
2. $\inf f_n$ *es medible.*

Corolario 1.2.16 *Si $\{f_n\}_{n\geq 1}$ es una sucesión de funciones medibles y acotadas en un espacio medible $(\Omega, \mathcal{A})$, entonces*

1. $\limsup f_n$ *es medible.*
2. $\liminf f_n$ *es medible.*
3. *Si $\lim_{n\to\infty} f_n(\omega) = f(\omega)$ para toda $\omega \in \Omega$ entonces f es medible.*

Definición 1.2.11 *Sea $(\Omega, \mathcal{A})$ un espacio medible y X, Y funciones reales en Ω. La función compleja $Z = X + iY$ sobre Ω se llama* ***$\mathcal{A}$-medible*** *o simplemente* ***medible*** *siempre y cuando ambas, X y Y, sean medibles.*

Sea Ω un conjunto y sea $(\Psi, \mathcal{G})$ un espacio medible. Supongamos que $f : \Omega \to \Psi$ es una función y $\mathcal{A}$ un $\sigma-$álgebra en Ω tal que f sea una transformación medible. Entonces. por definición, $f^{-1}(\mathcal{G}) \subset \mathcal{A}$. Además, por 1 del teorema 1.2.8, sabemos que $f^{-1}(\mathcal{G})$ es un $\sigma-$álgebra en Ω; por lo tanto, $f^{-1}(\mathcal{G})$ es el $\sigma-$álgebra más pequeño en Ω que hace que f sea medible.

Definición 1.2.12 *Sea $(\Psi, \mathcal{G})$ un espacio medible y $f : \Omega \to \Psi$ una transformación medible. El* ***$\sigma-$álgebra generado por la función*** *f se define como*

$$\sigma(f) := f^{-1}(\mathcal{G}) = \left\{f^{-1}(G) : G \in \mathcal{G}\right\}.$$

De la misma forma si suponemos que $(\Psi_j, \mathcal{G}_j)$ es espacio medible para $j \in \{1, 2, ..., n\}$ y que $f_j : \Omega \to \Psi_j$ son funciones para $j \in \{1, 2, ..., n\}$; entonces el **$\sigma-$álgebra generado por las funciones** $f_1, ..., f_n$ es

$$\sigma(f_1, ..., f_n) := \sigma\left\{f_j^{-1}(G_j) : j \in \{1, 2, ..., n\}, G_j \in \mathcal{G}_j\right\}$$

Más generalmente, sea $(\Psi_j, \mathcal{G}_j)$ espacio medible y sea $f_j : \Omega \to \Psi_j$ una función para toda j en algún conjunto J. Si $\mathcal{M} = \{f_j : j \in J\}$, entonces

$$\sigma(\mathcal{M}) = \sigma(f_j : j \in J) := \sigma\left\{f_j^{-1}(G_j) : j \in J, G_j \in \mathcal{G}_j\right\}$$

es el $\sigma-$álgebra más pequeño en Ω que hace que todas las funciones f_j sean medibles para $j \in J$.

Definición 1.2.13 *Sea $(\Omega_j, \mathcal{A}_j)$ espacio medible para $j \in \{1, 2, ..., n\}$ y consideremos el* ***conjunto producto***

$$\mathop{\times}_{j=1}^{n} \Omega_j = \Omega_1 \times \cdots \times \Omega_n := \{(\omega_1, ..., \omega_n) : \omega_j \in \Omega_j,\ j \in \{1, 2, ..., n\}\}$$

y las ***proyecciones*** $\mathfrak{p}_j : \mathop{\times}_{k=1}^{n} \Omega_k \to \Omega_j$ *definidas como*

$$\mathfrak{p}_j(\omega_1, ..., \omega_n) = \omega_j$$

para todo $j \in \{1, 2, ..., n\}$. Definimos el ***$\sigma-$álgebra producto*** *sobre $\Omega_1 \times \cdots \times \Omega_n$ como*

$$\mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_n := \sigma(\mathfrak{p}_1, ..., \mathfrak{p}_n)$$

1.2.3 Medida

Sea Ω un espacio y $\mathcal{A}$ una clase no vacía de subconjuntos de Ω. Si $A \in 2^{\Omega}$ y $\{A_k : k = 1, 2, ..., n\}$ es una subclase disjunta de $\mathcal{A}$ tal que $A = \bigcup_{k=1}^{m} A_k$, entonces se dice que $\{A_k : k = 1, 2, ..., n\}$ forma una **partición finita** de A en $\mathcal{A}$. Para $A \in \mathcal{A}$ diremos que la clase $\{A_k : k = 1, 2, ...\}$ es una **σ-partición** de A en $\mathcal{A}$ si es una subclase disjunta de $\mathcal{A}$ y $\bigcup_{k=1}^{\infty} A_k = A$.

Definición 1.2.14 *Sea $\mathcal{A}$ una clase de conjuntos de Ω. Una función μ de $\mathcal{A}$ en $\overline{\mathbb{R}}$ se llama* ***finitamente aditiva*** *si, y sólo si $\mu(\emptyset) = 0$ y para cada partición finita, $\{A_k : k = 1, 2, ..., n\}$ en $\mathcal{A}$, de $A \in \mathcal{A}$ la suma $\sum_{k=1}^{n} \mu(A_k)$ está bien definida y*

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

Definición 1.2.15 *Sea $\mathcal{A}$ una clase de conjuntos de Ω. Una función μ de $\mathcal{A}$ en $\overline{\mathbb{R}}$ se llama* ***σ-aditiva*** *o* ***numerablemente aditiva*** *si, y sólo si $\mu(\emptyset) = 0$ y para cada σ-partición finita, $\{A_k : k = 1, 2, ..., \} \subset \mathcal{A}$, de $A \in \mathcal{A}$ la suma $\sum_{k=1}^{\infty} \mu(A_k)$ está bien definida y*

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

Cuando en las definiciones anteriores se pide que la suma esté bien definida se quiere decir que para dos conjuntos distintos, A_i y A_j, en la partición de A, no ocurra simultáneamente que $\mu(A_i) = -\infty$ y $\mu(A_j) = +\infty$.

Definición 1.2.16 *Sea $(\Omega, \mathcal{A})$ un espacio medible. Una función σ-aditiva $\mu : \mathcal{A} \to [0, \infty]$ es llamada* **medida** *y $(\Omega, \mathcal{A}, \mu)$ es llamado* **espacio de medida.**

Ejemplo 1.2.10 *Denotemos por $\#(A)$ la cardinalidad de A. Sea Ω un espacio numerable. $\mathcal{A} = 2^{\Omega}$ y $\mu : \mathcal{A} \to [0, \infty]$ definida por*

$$\mu(A) := \begin{cases} \#(A) & \text{si } A \text{ es finito} \\ \infty & \text{si } A \text{ es infinito} \end{cases}$$

para todo $A \subset \Omega$. Entonces es fácil verificar que $(\Omega, \mathcal{A}, \mu)$ es un espacio de medida. La medida μ de este ejemplo es conocida como **medida de conteo**. ▲

Ejemplo 1.2.11 *Sea Ω un espacio, y $\mathcal{A} = 2^{\Omega}$. Tomemos un elemento fijo $a \in \Omega$, y definimos $\mu : \mathcal{A} \to [0, \infty]$ por $\mu(A) := 0$ si $a \notin A$ y $\mu(A) := 1$ si $a \in A$. Es claro que μ es una medida y por lo tanto $(\Omega, \mathcal{A}, \mu)$ un espacio de medida. La medida presentada en este ejemplo se conoce como* **medida de Dirac**. ▲

Sea $(\Omega, \mathcal{A}, \mu)$ es un espacio de medida y sean A_1, $A_2 \in \mathcal{A}$, claramente

$$A_1 = (A_1 \cap A_2) \cup (A_1 \cap A_2^c)$$

y

$$(A_1 \cap A_2) \cap (A_1 \cap A_2^c) = \emptyset,$$

entonces

$$\mu(A_1) = \mu(A_1 \cap A_2) + \mu(A_1 \backslash A_2). \tag{1.5}$$

Si ahora escribimos

$$A_1 \cup A_2 = A_2 \cup (A_1 \cap A_2^c),$$

entonces

$$\mu(A_1 \cup A_2) = \mu(A_2) + \mu(A_1 \backslash A_2)$$

y por (1.5)

$$\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2) - \mu(A_2 \cap A_1)$$

así queda demostrado el

Teorema 1.2.17 *Si $(\Omega, \mathcal{A}, \mu)$ es un espacio de medida y A_1, $A_2 \in \mathcal{A}$, entonces*

1. $\mu(A_1) = \mu(A_1 \cap A_2) + \mu(A_1 \backslash A_2)$.

2. $\mu(A_1) + \mu(A_2) = \mu(A_1 \cup A_2) + \mu(A_2 \cap A_1)$.

Como una consecuencia inmediata de 1 del teorema anterior se tiene el siguiente

Corolario 1.2.18 *Si* $(\Omega, \mathcal{A}, \mu)$ *es un espacio de medida y* $A_2 \subset A_1$ *para* $A_1, A_2 \in \mathcal{A}$

1. $\mu(A_2) \leq \mu(A_1)$

2. Si $\mu(A_2) < \infty$ *entonces* $\mu(A_1 \backslash A_2) = \mu(A_1) - \mu(A_2)$

Ahora supongamos que $A, A_1, A_2, ..., A_n$ están en $\mathcal{A}$, donde $(\Omega, \mathcal{A}, \mu)$ es un espacio de medida, entonces, si $A \subset \bigcup_{k=1}^{n} A_k$,

$$\mu(A) \leq \mu\left(\bigcup_{k=1}^{n} A_k\right);$$

sea

$$B_1 := A_1 \qquad \text{y} \qquad B_n := A_n \backslash \bigcup_{k=1}^{n-1} A_k, \tag{1.6}$$

claramente los B_k son disjuntos y además

$$\bigcup_{k=1}^{n} A_k = \bigcup_{k=1}^{n} B_k,$$

entonces

$$\mu\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} \mu(B_k) \leq \sum_{k=1}^{n} \mu(A_k).$$

Por lo tanto,

$$\mu(A) \leq \sum_{k=1}^{n} \mu(A_k)$$

Supongamos ahora que $\bigcup_{k=1}^{n} A_k \subset A$ y que las A_k son disjuntas, entonces

$$\sum_{k=1}^{n} \mu(A_k) = \mu\left(\bigcup_{k=1}^{n} A_k\right) \leq \mu(A).$$

Así, se tiene el siguiente teorema.

Teorema 1.2.19 *Sea* $(\Omega, \mathcal{A}, \mu)$ *un espacio de medida y* $A, A_1, A_2, ..., A_n$ *elementos de* $\mathcal{A}$. *Entonces*

1. $\mu(A) \leq \sum_{k=1}^{n} \mu(A_k)$ *si* $A \subset \bigcup_{k=1}^{n} A_k$.
2. $\sum_{k=1}^{n} \mu(A_k) \leq \mu(A)$ *si* $\bigcup_{k=1}^{n} A_k \subset A$ *y* $A_1, A_2, ..., A_n$ *son disjuntos.*

Corolario 1.2.20 *Si* $(\Omega, \mathcal{A}, \mu)$ *es un espacio de medida y* $A, A_1, A_2, ...$ *están en* $\mathcal{A}$ *entonces*

1. $\mu(A) \leq \sum_{k=1}^{\infty} \mu(A_k)$ *si* $A \subset \bigcup_{k=1}^{\infty} A_k$.
2. $\sum_{k=1}^{\infty} \mu(A_k) \leq \mu(A)$ *si* $\bigcup_{k=1}^{\infty} A_k \subset A$ *y* $A_1, A_2, ..., A_n$ *son disjuntos.*

Si $\{A_k\}_{k\geq 1}$ es una sucesión tal que $A_1 \subset A_2 \subset ...$ y $A := \bigcup_{k=1}^{\infty} A_k$, decimos que $\{A_k\}_{k\geq 1}$ es una **sucesión creciente con límite** A y lo escribimos como $A_n \uparrow A$. Si $\{A_k\}_{k\geq 1}$ es tal que $A_1 \supset A_2 \supset ...$ y $A := \bigcap_{k=1}^{\infty} A_k$, entonces decimos que $\{A_k\}_{k\geq 1}$ es una **sucesión decreciente** con límite A y en este caso escribimos $A_n \downarrow A$.

Supongamos que $A_n \uparrow A$ y hagamos como en (1.6)

$$B_1 := A_1 \qquad \text{y} \qquad B_n := A_n \backslash \bigcup_{k=1}^{n-1} A_k,$$

entonces, si $\mu(A_m) = \infty$ para alguna $m \geq 1$, $\mu(A_n) = \infty = \mu(A)$ para toda $n \geq m$ y se tiene que $\mu(A_n) \to \mu(A)$. Si ahora suponemos que $\mu(A_m) < \infty$ para toda $m \geq 1$, entonces

$$\begin{aligned}
\mu(A) &= \sum_{i=1}^{\infty} \mu(B_i) = \lim_{k\to\infty} \sum_{i=1}^{k} \mu(B_i) \\
&= \lim_{k\to\infty} [\mu(A_1) + (\mu(A_2) - \mu(A_1)) + ... + (\mu(A_k) - \mu(A_{k-1}))] \\
&= \lim_{k\to\infty} \mu(A_k).
\end{aligned}$$

Por lo tanto,

$$\mu(A) = \lim_{k\to\infty} \mu(A_k).$$

Para el caso en que $A_n \downarrow A$ y $\mu(A_1) < \infty$ si tomamos $D_n := A_n \backslash A_{n+1}$ se tienen que $D_1, D_2, ...$ son disjuntos y $A_1 \setminus A = \bigcup_{i=1}^{\infty} D_i$, entonces

$$
\begin{aligned}
\mu(A_1) - \mu(A) &= \mu(A_1 \setminus A) = \sum_{i=1}^{\infty} \mu(D_i) = \lim_{k\to\infty} \sum_{i=1}^{k-1} \mu(D_i) \\
&= \lim_{k\to\infty} \left[(\mu(A_1) - \mu(A_2)) + (\mu(A_2) - \mu(A_3)) + ... + (\mu(A_{k-1}) - \mu(A_k)) \right] \\
&= \mu(A_1) - \lim_{k\to\infty} \mu(A_k).
\end{aligned}
$$

Por lo tanto,

$$\mu(A) = \lim_{k\to\infty} \mu(A_k).$$

Lo anterior demuestra el

Teorema 1.2.21 *Si $(\Omega, \mathcal{A}, \mu)$ es un espacio de medida y $A, A_1, A_2, ..., A_n$ están en $\mathcal{A}$, entonces*

1. $\mu(A) = \lim_{k\to\infty} \mu(A_k)$ *si* $A_n \uparrow A$.
2. $\mu(A) = \lim_{k\to\infty} \mu(A_k)$ *si* $A_n \downarrow A$ *y* $\mu(A_1) < \infty$.

1.3 Espacios de Probabilidad

Definición 1.3.1 *Se dice que un espacio de medida $(\Omega, \mathcal{A}, P)$ es un* ***espacio de probabilidad*** *si, y sólo si, $P(\Omega) = 1$.*

Definición 1.3.2 *En un espacio de probabilidad $(\Omega, \mathcal{A}, P)$*

1. *El conjunto Ω es llamado* ***espacio muestral.***
2. *P es una* ***medida de probabilidad.***
3. *Los conjuntos $A \in \mathcal{A}$ son conocidos como* ***eventos*** *y el número $P(A)$ es llamado* ***probabilidad*** *del evento A.*

Las propiedades de una medida de probabilidad se enuncian en el siguiente teorema. La demostración es una particularización de los resultados de la sección 1.2.3 y la aplicación de la condición $P(\Omega) = 1$.

Teorema 1.3.1 *Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, y $A, A_1, A_2, \ldots \in \mathcal{A}$ eventos. Entonces*

1. $P(A_1) = P(A_1 \cap A_2) + P(A_1 \cap A_2^c)$.
2. $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$.
3. *Si* $A_1 \subset A_2$, *entonces* $P(A_2 \backslash A_1) = P(A_2) - P(A_1)$ *y* $P(A_1) \leq P(A_2)$.
4. $P(A) \leq \sum_{k=1}^{\infty} P(A_k)$ *si* $A \subset \bigcup_{k=1}^{\infty} A_k$.
5. $\sum_{k=1}^{\infty} P(A_k) \leq P(A)$ *si* $\bigcup_{k=1}^{\infty} A_k \subset A$ *y* $A_1, A_2, \ldots, A_n$ *son disjuntos.*

Teorema 1.3.2 *Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, y $A, A_1, A_2, \ldots \in \mathcal{A}$ eventos. Entonces*

1. $P(A_1^c) = 1 - P(A_1)$.
2. $0 \leq P(A_1) \leq 1$.
3. *Si* $\lim_{n\to\infty} P(A_n) = 1$, *entonces* $\lim_{n\to\infty} P(A \cap A_n) = P(A)$, .

Teorema 1.3.3 *Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y $A, A_1, A_2, \ldots, A_n$ eventos en $\mathcal{A}$. Si $A_n \to A$ entonces $P(A) = \lim_{k\to\infty} P(A_k)$.*

Definición 1.3.3 *Un evento de probabilidad nula es llamado* ***evento nulo****. Si una propiedad se cumple excepto para un evento nulo, entonces se dice que tal propiedad se cumple* ***casi seguramente*** *(abreviado* ***c.s.****).*

El siguiente teorema es un caso particular de un resultado importante de teoría de la medida conocido como teorema de Extensión. La prueba no se presenta aquí pero el lector interesado puede encontrarla en Aliprantis [1981, sec. 12], Billingsley [1995, sec. 11], Clarke [1975, apen. B] o Dudley [1989, sec. 3.1].

Teorema 1.3.4 *Sea $\mathcal{I}$ un semi-álgebra en Ω, y sea $\mu : \mathcal{I} \to [0, \infty]$ una función σ-aditiva tal que $\mu(\Omega) = 1$. Entonces existe una única medida de probabilidad P sobre $\sigma(\mathcal{I})$ la cual es igual a μ sobre $\mathcal{I}$; es decir,*

$$P(A) = \mu(A) \qquad \text{para todo } A \in \mathcal{I}.$$

Consideremos nuevamente el semi-álgebra J definida en el ejemplo 1.2.1 y dos medidas de probabilidad P' y P'' en $\mathcal{B}(\mathbb{R})$ tales que $P' = P''$ sobre J. En el ejemplo 1.2.9 se demostró que $\sigma(J) = \mathcal{B}(\mathbb{R})$. Supongamos ahora que $\mu : J \to [0, \infty]$ es una función $\sigma-$aditiva con la propiedad $\mu(\Omega) = 1$, entonces, por el teorema 1.3.4, existe una única medida de probabilidad P sobre $\sigma(J) = \mathcal{B}(\mathbb{R})$ tal que $P = \mu$ sobre J. Si suponemos que P es precisamente P', es decir que $P'(A) = \mu(A)$ para cada $A \in J$, entonces, como por hipótesis $P' = P''$ en J, se tiene que también $P''(A) = \mu(A)$ para cada $A \in J$, pero, por el teorema 1.3.4, P' es única, entonces se debe cumplir que $P' = P''$ en $\mathcal{B}(\mathbb{R})$, lo que demuestra el siguiente resultado que se presenta como corolario del teorema 1.3.4.

Corolario 1.3.5 *Si P' y P'' son dos medidas de probabilidad en $\mathcal{B}(\mathbb{R})$ tal que son iguales sobre la semi-álgebra J definida en el ejemplo 1.2.1. Entonces $P' = P''$ sobre $\mathcal{B}(\mathbb{R})$.*

1.4 Variables Aleatorias y Funciones de Distribución

Definición 1.4.1 *Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y $X : \Omega \to \overline{\mathbb{R}}$ una función medible. Se dice que X es una* **variable aleatoria** *si X es finita casi seguramente.*

Los siguientes teoremas definen las operaciones entre variables aleatorias y demuestran que límites y ciertas funciones de variables aleatorias también son variables aleatorias. La demostración de cada uno de estos teoremas es un caso particular de las demostraciones realizadas para funciones medibles presentadas en la sección 1.2.2.

Teorema 1.4.1 *Sea X una variable aleatoria en el espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Entonces*

1. *cX es una variable aleatoria.*
2. *$|X|$ es una variable aleatoria.*
3. *X^2 una variable aleatoria.*
4. *Si $X \neq 0$, entonces $\dfrac{1}{X}$ es variable aleatoria.*

Teorema 1.4.2 *Sean X y Y variables aleatorias en el espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Entonces $X + Y$ y XY son variables aleatorias.*

Teorema 1.4.3 *Si X es una variable aleatoria en el espacio de probabilidad $(\Omega, \mathcal{A}, P)$ y $g : \mathbb{R} \to \mathbb{R}$ es una función Borel-medible. entonces $g \circ X$ es una variable aleatoria*

Teorema 1.4.4 *Sean X y Y variables aleatorias en el espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Entonces* $\max(X, Y)$ *y* $\min(X, Y)$ *son variables aleatorias.*

Corolario 1.4.5 *Si X es una variable aleatoria en el espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Entonces la parte positiva y la parte negativa de X definidas como*

$$X^+ := \max(0, X) \qquad y \qquad X^- := \min(0, -X) \tag{1.7}$$

respectivamente son variables aleatorias.

Teorema 1.4.6 *Si $\{X_n\}_{n\geq 1}$ es una sucesión de variables aleatorias en $(\Omega, \mathcal{A}, P)$, entonces*

1. $\sup X_n$ *es una variable aleatoria.*
2. $\inf X_n$ *es una variable aleatoria.*

Corolario 1.4.7 *Si $\{X_n\}_{n\geq 1}$ es una sucesión de variables aleatorias en $(\Omega, \mathcal{A}, P)$ entonces*

1. $\limsup X_n$ *es una variable aleatoria.*
2. $\liminf X_n$ *es una variable aleatoria.*
3. *Si $\lim\limits_{n\to\infty} X_n = X$ entonces X es una variable aleatoria.*

Ahora consideremos una variable aleatoria X, un conjunto $B \in \mathcal{B}(\mathbb{R})$ y la función P_X definida como

$$P_X(B) := P\left\{X^{-1}(B)\right\}$$

Como X es una transformación medible $(\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ y $P : \mathcal{A} \to [0, 1]$ se sigue que $P_X : \mathcal{B}(\mathbb{R}) \to [0, 1]$. También, como $X^{-1}(\mathbb{R}) = \Omega$, se cumple que $P_X(\mathbb{R}) = P(\Omega) = 1$. Además, si consideramos una sucesión $\{B_k\}_{k\geq 1}$ de conjuntos disjuntos en $\mathcal{B}(\mathbb{R})$, entonces

$$\begin{aligned} P_X\left(\bigcup_{k\geq 1} B_k\right) &= P\left\{X^{-1}\left(\bigcup_{k\geq 1} B_k\right)\right\} = P\left\{\bigcup_{k\geq 1} X^{-1}(B_k)\right\} \\ &= \sum_{k\geq 1} P\left\{X^{-1}(B_k)\right\} = \sum_{k\geq 1} P_X(B_k). \end{aligned}$$

Por lo tanto, $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ es un espacio de probabilidad.

Definición 1.4.2 *Sea X una variable aleatoria en el espacio $(\Omega, \mathcal{A}, P)$. La medida de probabilidad*

$$P_X(B) := P\left\{X^{-1}(B)\right\} \text{ donde } B \in \mathcal{B}(\mathbb{R})$$

se conoce como la ***distribución*** *de la variable aleatoria X.*

Definición 1.4.3 *Sean X y Y variables aleatorias con distribución P_X y P_Y respectivamente. Se dice que X y Y son* ***variables aleatorias idénticamente distribuidas*** *si $P_X = P_Y$ sobre $\mathcal{B}(\mathbb{R})$.*

Definición 1.4.4 *Sea X una variable aleatoria en el espacio $(\Omega, \mathcal{A}, P)$ y $x \in \mathbb{R}$. La* ***función de distribución de la variable aleatoria*** *X, denotada por F_X, se define como*

$$F_X(x) := P\{X \leq x\} = P\left\{X^{-1}(]-\infty, x])\right\} = P_X\{]-\infty, x]\}.$$

El siguiente teorema establece que dos variables aleatorias son idénticamente distribuidas si, y sólo si, sus correspondientes funciones de distribución son iguales sobre $\mathbb{R}$. La demostración es una simple aplicación del corolario 1.3.5.

Teorema 1.4.8 *Si P_X y P_Y son medidas de probabilidad con función de distribución F_X y F_Y respectivamente. Entonces $F_X = F_Y$ en $\mathbb{R}$ si, y sólo si, $P_X = P_Y$ en $\mathcal{B}(\mathbb{R})$.*

Para una función F y un número $h > 0$ definimos

$$F(x-) := \lim_{h \to 0} F(x-h) \qquad F(x+) := \lim_{h \to 0} F(x+h)$$

$$\Delta F(x) := F(x) - F(x-)$$

también en algunas ocasiones se utiliza la notación

$$F(x-) = \lim_{u \uparrow x} F(u) \qquad F(x+) = \lim_{u \downarrow x} F(u)$$

Definición 1.4.5 *Una función de variable real $G(x)$ en $\mathbb{R}$ es llamada* ***función de distribución*** *si*

1. $\lim_{x \to -\infty} G(x) = 0$ *y* $\lim_{x \to \infty} G(x) = 1$

2. *G es no decreciente*

3. G es continua por la derecha es decir

$$G(x+) = G(x)$$

para toda $x \in \mathbb{R}$.

El siguiente teorema establece que la función de distribución asociada a una variable aleatoria siempre es una función de distribución en el sentido de la definición 1.4.5.

Teorema 1.4.9 *Sea* X *una variable aleatoria en* $(\Omega, \mathcal{A}, P)$ *con función de distribución* F_X. *Entonces* F_X *tiene las siguientes propiedades*

1. $\lim_{x \to -\infty} F_X(x) = 0$ *y* $\lim_{x \to \infty} F_X(x) = 1$
2. F_X *es no decreciente en* $\mathbb{R}$
3. F_X *es continua por la derecha*

Demostración.

1. Sea $B_n := \{X \leq n\}$ donde $n = 0, 1, 2, \ldots$ Es claro que $B_{-n} \downarrow \emptyset$ y $B_n \uparrow \Omega$ de donde, por el teorema 1.3.3,

$$\lim_{x \to -\infty} F_X(x) = \lim_{n \to -\infty} P(B_n) = P\left(\lim_{n \to -\infty} B_n\right) = P(\emptyset) = 0$$

y

$$\lim_{x \to \infty} F_X(x) = \lim_{n \to \infty} P(B_n) = P\left(\lim_{n \to \infty} B_n\right) = P(\Omega) = 1.$$

2. Sean $a, b \in \mathbb{R}$ donde $a < b$, entonces

$$F_X(b) - F_X(a) = P\left\{X^{-1}(]-\infty, b])\right\} - P\left\{X^{-1}(]-\infty, a])\right\}.$$

Ya que $]-\infty, a] \subset]-\infty, b]$,

$$\begin{aligned} F_X(b) - F_X(a) &= P\left\{X^{-1}(]-\infty, b]) \backslash X^{-1}(]-\infty, a])\right\} \\ &= P\left\{X^{-1}(]a, b])\right\} \geq 0. \end{aligned}$$

Por lo tanto, F_X es no decreciente.

3. Se desea demostrar que , dado $h > 0$, $\lim_{h\to 0}[F_X(x+h) - F_X(x)] = 0$. Hacemos $h = \frac{1}{n}$, así la condición $h \to 0$ es equivalente a que $n \to \infty$. Ahora tomamos $C_n := \left]c, c + \frac{1}{n}\right]$ y observamos que $C_n \downarrow \emptyset$ de donde $P(C_n) \to 0$ cuando $n \to \infty$. Pero,

$$F_X\left(c + \frac{1}{n}\right) - F_X(c) = P\left\{X^{-1}\left(\left]c, c + \frac{1}{n}\right]\right)\right\},$$

por lo que

$$\begin{aligned}\lim_{h\to 0}[F_X(c+h) - F_X(c)] &= \lim_{n\to\infty}\left[F_X\left(c + \frac{1}{n}\right) - F_X(c)\right] \\ &= \lim_{n\to\infty} P\left\{X^{-1}\left(\left]c, c + \frac{1}{n}\right]\right)\right\} \\ &= P\left\{X^{-1}(\emptyset)\right\} = 0\end{aligned}$$

y por lo tanto, F_X es continua por la derecha.

■

El siguiente resultado establece que una función que cumpla con las propiedades listadas en el teorema 1.4.9 es la función de distribución de una medida de probabilidad y que esta medida de probabilidad es única. La demostración puede encontrarse en Clarke [1975, p. 49].

Teorema 1.4.10 *Sea $F : \mathbb{R} \to [0,1]$ una función de distribución, entonces existe una, y sólo una, medida de probabilidad P sobre $\mathcal{B}(\mathbb{R})$ tal que*

$$P\{]a, b]\} = F(b) - F(a) \qquad a, b \in \mathbb{R}.$$

Como corolario del teorema 1.4.10 se tiene el siguiente resultado.

Corolario 1.4.11 *Sea $F : \mathbb{R} \to [0,1]$ una función tal que cumple con las hipótesis del teorema 1.4.10. Entonces existe una variable aleatoria X, en algún espacio de probabilidad adecuado, que tiene como función de distribución a F.*

Definición 1.4.6 *Se dice que una función de distribución F es* ***discreta*** *si existe un conjunto $S = \{x_j : j = 1, 2, ..., n;\ n \leq \infty\}$ tal que*

$$\Delta F(x_j) = p_j > 0 \qquad y \qquad \sum_j p_j = 1$$

En la definición 1.4.6 los elementos de S son conocidos como **puntos de discontinuidad** o **puntos de salto** de F, p_j es la **longitud del j-ésimo salto**; el conjunto $L = \{p_j : j = 1, 2, ...\}$ se conoce como **distribución de probabilidad discreta**. Si la medida de probabilidad P y la variable aleatoria X están asociadas a una función de distribución discreta F, entonces P es una **medida de probabilidad discreta** y X una **variable aleatoria discreta**.

Definición 1.4.7 *Sea F una función de distribución discreta con conjunto de puntos de discontinuidad $S = \{x_j : j = 1, 2, ...\}$ y distribución de probabilidad $L = \{p_j : j = 1, 2, ...\}$. La función f definida por*

$$f(x) := \begin{cases} p_j & si\ x = x_j,\ j \geq 1 \\ 0 & si\ x \neq x_j \end{cases}$$

*es llamada **función densidad de probabilidad** o **función masa de probabilidad**.*

Definición 1.4.8 *Una medida de probabilidad P y su correspondiente función de distribución F, se dicen **absolutamente continuas**, si existe una función de Borel f en $\mathbb{R}$ tal que*

$$F(x) = \int_{-\infty}^{x} f(t)\,dt \qquad x \in \mathbb{R} \tag{1.8}$$

*La función f es conocida como la **función de densidad** de la función de distribución F.*

Claramente cada función f no negativa, integrable en el sentido de Riemann y tal que

$$\int_{-\infty}^{\infty} f(t)\,dt = 1 \qquad x \in \mathbb{R}$$

define una función de distribución que estará dada por (1.8).

Definición 1.4.9 *Si la variable aleatoria X está asociada a una función de distribución absolutamente continua entonces se dirá que X es una **variable aleatoria absolutamente continua**.*

A continuación se extienden las definiciones anteriores al caso de vectores aleatorios.

Definición 1.4.10 *Sean $X_1, X_2, ..., X_n$ variables aleatorias en $(\Omega, \mathcal{A}, P)$. El conjunto ordenado*

$$\mathbf{X}(\omega) = (X_1(\omega), X_2(\omega), ..., X_n(\omega))$$

*es llamado **vector aleatorio n-dimensional** en $(\Omega, \mathcal{A}, P)$.*

Definición 1.4.11 *La función $G : \mathbb{R}^n \to \mathbb{R}$, es una **función de distribución** n**-dimensional** si cumple con cada una de la siguientes condiciones*

1. $G(x_1, x_2, ..., x_n)$ *es no decreciente en cada argumento.*
2. $G(x_1, x_2, ..., x_n)$ *es continua por la derecha en cada argumento.*
3. $G(x_1, x_2, ..., x_n) \to 0$ *si* $x_j \to -\infty$ *para al menos una* $j = 1, 2, ...n$,
   $G(x_1, x_2, ..., x_n) \to 1$ *si* $x_j \to \infty$ *para toda* $j = 1, 2, ...n$.
4. *Si* $a_i \leq b_i$, $i = 1, 2, ..., n$, *y*

   $$\Delta_{a_i, b_i} G(x_1, x_2, ..., x_n) := G(x_1, ..., x_{i-1}, b_i, x_{i+1} ..., x_n) - G(x_1, ..., x_{i-1}, a_i, x_{i+1} ..., x_n),$$

   entonces

   $$\Delta_{a_1, b_1} \Delta_{a_2, b_2} ... \Delta_{a_n, b_n} G(x_1, x_2, ..., x_n) \geq 0.$$

Definición 1.4.12 *Sea $\mathbf{X}(\omega) = (X_1(\omega), X_2(\omega), ..., X_n(\omega))$ un vector aleatorio en $(\Omega, \mathcal{A}, P)$. La medida de probabilidad $P_{\mathbf{X}}$, definida por*

$$P_{\mathbf{X}}(B) := P\left(\left\{\mathbf{X}^{-1}(B)\right\}\right) = P\left(\{\omega : (X_1(\omega), X_2(\omega), ..., X_n(\omega)) \in B\}\right); \qquad B \in \mathcal{B}(\mathbb{R}^n),$$

*es llamada **distribución conjunta** de las variables aleatorias $X_1, X_2, ..., X_n$, o **distribución** del vector aleatorio $\mathbf{X}$.*

Definición 1.4.13 *Sea $\mathbf{X}(\omega) = (X_1(\omega), X_2(\omega), ..., X_n(\omega))$un vector aleatorio en $(\Omega, \mathcal{A}, P)$. La función $F_{\mathbf{X}}$ en $\mathbb{R}^n$, definida por*

$$F_{\mathbf{X}}(x_1, x_2 ..., x_n) := P\left(\{\omega : X_1(\omega) \leq x_1, X_2(\omega) \leq x_2, ..., X_n(\omega) \leq x_n\}\right) \tag{1.9}$$

*es llamada **función de distribución conjunta** de las variables aleatorias $X_1, X_2, ..., X_n$, o **función de distribución** del vector aleatorio $\mathbf{X}$.*

Puede demostrarse que la función definida por la ecuación (1.9) es una función de distribución en el sentido de la definición 1.4.11.

El siguiente teorema es la extensión al caso multidimensional del teorema 1.4.10

Teorema 1.4.12 *Sea $A = \mathop{\times}\limits_{i=1}^{n}]a_i, b_i]$ y $F(x_1, x_2 ..., x_n)$ una función de distribución en $\mathbb{R}^n$. Entonces existe una única medida de probabilidad P sobre $(\mathbb{R}^n, \mathcal{B}^n)$ tal que*

$$P(A) = \Delta_{a_1, b_1} \Delta_{a_2, b_2} ... \Delta_{a_n, b_n} F(x_1, x_2, ..., x_n).$$

Dada una función de distribución conjunta n-dimensional $F_{\mathbf{X}}$ de las variables aleatorias $X_1, ..., X_n$, podemos determinar la función de distribución conjunta k-dimensional, $k \leq n$, $F_{\mathbf{X}'}$ de $X_{i_1}, X_{i_2}, ..., X_{i_k}$, donde $0 < i_1 < i_2 < ... < i_k \leq n$ mediante la relación

$$F_{\mathbf{X}'}\left(x_{i_1}, x_{i_2}, ..., x_{i_k}\right) = F_{\mathbf{X}}\left(x_1, x_2, ..., x_n\right)\Big|_{j \neq i_1, i_2, ..., i_k}^{x_j = +\infty} \tag{1.10}$$

donde $F(+\infty) = \lim\limits_{x \to \infty} F(x)$. En particular, por medio de (1.10), podemos determinar la función de distribución marginal de cada X_k, $1 \leq k \leq n$ por

$$F_{X_k}(x) = F_{\mathbf{X}}\left(\overbrace{+\infty, ..., +\infty}^{k-1}, x, +\infty, ..., +\infty\right)$$

La relación (1.10) se deduce de (1.9) y de que $\{X_j < \infty\}$ es un evento con probabilidad uno.

Definición 1.4.14 *Las funciones de distribución conjuntas obtenidas mediante la ecuación* (1.10) *se denominan* ***funciones de distribución marginales (parciales)****.*

Si cada una de las componentes del vector aleatorio $\mathbf{X} = (X_1, X_2, ..., X_n)$, tiene una distribución discreta, entonces se dice que el vector aleatorio $\mathbf{X} = (X_1, X_2, ..., X_n)$ **tiene distribución discreta** o que $X_1, X_2, ..., X_n$, son **variables aleatorias conjuntamente discretas**.

Sea $F_{\mathbf{X}}$ la función de distribución conjunta del vector aleatorio $\mathbf{X} = (X_1, X_2, ..., X_n)$. Se dice que $X_1, X_2, ..., X_n$ son **variables aleatorias conjuntamente continuas** o que el vector aleatorio $\mathbf{X} = (X_1, X_2, ..., X_n)$ tiene **distribución continua**, si $F_{\mathbf{X}}(x_1, x_2, ..., x_n)$ puede escribirse como

$$F_{\mathbf{X}}\left(x_1, x_2, ..., x_n\right) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\mathbf{X}}\left(t_1, t_2, ..., t_n\right) dt_1 ... dt_n \tag{1.11}$$

para cada $(x_1, x_2, ..., x_n) \in \mathbb{R}^n$, donde $f_{\mathbf{X}}(x_1, x_2, ..., x_n)$ es una función medible no negativa tal que

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}\left(x_1, x_2, ..., x_n\right) dx_1 ... dx_n = 1.$$

La función $f_{\mathbf{X}}(x_1, x_2, ..., x_n)$ es llamada **función de densidad conjunta** de las variables aleatorias $X_1, X_2, ..., X_n$.

1.5 Independencia

Definición 1.5.1 *Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y sean $A_1, ..., A_n$ eventos en $\mathcal{A}$. Decimos que $A_1, ..., A_n$ son eventos independientes si*

$$P\left\{\bigcap_{i\in\alpha} A_i\right\} = \prod_{i\in\alpha} P\{A_i\}$$

para todo α subconjunto no vacío de $\{1, 2, ..., n\}$.

Como $\{1, 2, ..., n\}$ posee $2^n - 1$ subconjuntos no vacíos, podemos ver que para demostrar independencia tendríamos que verificar $2^n - n - 1$ ecuaciones no triviales.

Definición 1.5.2 *Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y sea $\{A_t : t \in T\}$ una familia de eventos en $\mathcal{A}$. Decimos que los eventos $\{A_t : t \in T\}$ son* ***independientes*** *si $A_{t_1}, ..., A_{t_n}$ son eventos independientes siempre que $t_1, ..., t_n$ sean elementos diferentes de T y $n = 1, 2, \cdots$. Es decir, $\{A_t : t \in T\}$ son independientes si, y sólo si,*

$$P\left\{\bigcap_{t\in\alpha} A_t\right\} = \prod_{t\in\alpha} P\{A_t\}$$

para todo $\alpha \subset T$ finito no vacío.

Definición 1.5.3 *Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad. Si $\mathcal{C}$ y $\mathcal{D}$ son clases de eventos contenidas en $\mathcal{A}$. Decimos que $\mathcal{C}$ y $\mathcal{D}$ son* ***independientes*** *si C y D son independientes siempre que $C \in \mathcal{C}$ y $D \in \mathcal{D}$, esto es,*

$$P(C \cap D) = P(C)\, P(D)$$

para todo $C \in \mathcal{C}$ y todo $D \in \mathcal{D}$.

Definición 1.5.4 *Sean X y Y variables aleatorias definidas sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$ decimos que X y Y son* ***variables aleatorias independientes****, y lo denotamos por $X \perp Y$, si $\sigma(X)$ y $\sigma(Y)$ son independientes.*

De las definiciones anteriores se desprende el siguiente teorema.

Teorema 1.5.1 *Las n variables aleatorias $X_1, X_2, ..., X_n$ son independientes si, y sólo si,*

$$P\left\{\bigcap_{i=1}^{n} X_i^{-1}(B_i)\right\} = \prod_{i=1}^{n} P\left\{X_i^{-1}(B_i)\right\}$$

para cada elección de conjuntos de Borel $B_1, B_2, ..., B_n$.

El siguiente teorema proporciona una definición alternativa para la independencia de variables aleatorias. La demostración puede encontrarse en Shiryayev [1984, p. 177].

Teorema 1.5.2 *Sean $X_1, X_2, ..., X_n$ variables aleatorias en $(\Omega, \mathcal{A}, P)$ con función de distribución conjunta $F_{\mathbf{X}}$. Si denotamos por F_{X_i} la función de distribución de X_i, $1 \leq i \leq n$, entonces una condición necesaria y suficiente para que las variables aleatorias sean independientes es que*

$$F_{\mathbf{X}}(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} F_{X_i}(x_i).$$

Supongamos que X_1 y X_2 variables aleatorias independientes y que $f : \mathbb{R} \to \mathbb{R}$ y $g : \mathbb{R} \to \mathbb{R}$ funciones Borel medibles, entonces, para $A_1, A_2 \in \mathcal{B}(\mathbb{R})$

$$\begin{aligned} P(f(X_1) \in A_1, g(X_2) \in A_2) &= P\left(X_1 \in f^{-1}(A_1), X_2 \in g^{-1}(A_2)\right) \\ &= P\left(X_1 \in f^{-1}(A_1)\right) P\left(X_2 \in g^{-1}(A_2)\right) \\ &= P(f(X_1) \in A_1) P(g(X_2) \in A_2). \end{aligned}$$

Así, se ha demostrado el

Teorema 1.5.3 *Sean X_1 y X_2 variables aleatorias independientes. Si $f : \mathbb{R} \to \mathbb{R}$ y $g : \mathbb{R} \to \mathbb{R}$ son funciones Borel medibles, entonces $f(X_1)$ y $g(X_2)$ son independientes.*

Capítulo 2

Valores Esperados

2.1 Introducción

El concepto de valor esperado fue introducido por primera vez por Christaan Huyghens (1629-1695) en su libro *De Ratiociniis in Ludo Alea.* En *De Ratiociniis in Ludo Alea*, Huyghens toma el valor esperado como el concepto fundamental de teoría de la probabilidad y deduce tres teoremas sobre valor esperado.

Ya que el concepto de valor esperado es fundamental para el desarrollo de la teoría de la funciones características, se requiere un estudio detallado de sus propiedades. Iniciamos nuestro estudio sobre el valor esperado en la sección 2.2, donde primero se da la definición para variables aleatorias simples y después se extiende a variables aleatorias no negativas y variables aleatorias que pueden tomar cualquier signo; aquellos lectores familiarizados con teoría de la integración, podrán notar que en realidad el valor esperado de una variable aleatoria X en $(\Omega, \mathcal{A}, P)$ es la integral de Lebesgue respecto a la medida de probabilidad P. En la sección 2.3 se enuncia y se demuestran el teorema de convergencia monótona y el teorema de convergencia dominada y algunas de sus aplicaciones. La sección 2.4 tiene que ver con el cálculo de valores esperados bajo el supuesto de independencia. El objetivo de la sección 2.5 es extender la definición de valor esperado al caso de variables aleatorias complejas. En la sección 2.6 se presenta un caso particular del teorema de cambio de variable para integrales de Lebesgue. El cálculo de valores esperados de variables aleatorias con distribución discreta o absolutamente continua se discute en la sección 2.7. En la última sección de este capítulo se enuncian, sin demostración, algunas desigualdades de uso común en probabilidad y estadística.

2.2 Definición y Propiedades

2.2.1 Variables Aleatorias Simples

Definición 2.2.1 *Una variable aleatoria X en $(\Omega, \mathcal{A}, P)$, se llama **simple** si su rango es finito.*

Si X es una variable aleatoria simple con rango $H = \{x_i : i = 1, 2, ..., n\}$ entonces, para cada $\omega \in \Omega$, $X(\omega) \in H$ de donde $X^{-1}(H) = \Omega$. Ahora, si $A_i := \{\omega : X(\omega) = x_i\}$, $i \in \{1, 2, ..., n\}$, entonces $X^{-1}(\{x_i\}) = A_i \in \mathcal{A}$ y para cada $i \neq j$, $i, j \in \{1, 2, ..., n\}$, $A_i \cap A_j = \emptyset$; además

$$\bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} X^{-1}(\{x_i\}) = X^{-1}(H) = \Omega$$

por lo tanto, $\{A_i : i = 1, 2, ..., n\} \subset \mathcal{A}$ forma una partición de Ω. Así, X puede representarse como

$$X = \sum_{i=1}^{n} x_i I_{A_i} \tag{2.1}$$

donde $H = \{x_i : i \in \{1, 2, ..., n\}\}$ es el rango de X, $A_i = \{\omega : X(\omega) = x_i\} \in \mathcal{A}$, y

$$I_{A_i}(\omega) = \begin{cases} 1 \text{ si } \omega \in A_i \\ 0 \text{ si } \omega \notin A_i \end{cases}$$

Definición 2.2.2 *Si $X = \sum_{i=1}^{n} x_i I_{A_i}$ es una variable aleatoria simple no negativa, entonces el **valor esperado** de X, denotado por $E(X)$, es*

$$E(X) := \sum_{i=1}^{n} x_i P(A_i)$$

En ocasiones X puede tener otras representaciones de la forma (2.1), debido a que $x_i I_{A_i}$ puede reemplazarse por $\sum_j x_i I_{A_{ij}}$, donde las A_{ij} forman una partición finita de A_i. Por lo tanto debemos demostrar que, para una variable aleatoria X simple y no negativa, $E(X)$ está bien definido, esto es, *el valor de $E(X)$ no depende de la representación de X.*

Para ello, supongamos que tiene la representación $X = \sum_{i=1}^{n} x_i I_{A_i}$ dada por (2.1), y la representación adicional $X = \sum_{j=1}^{m} b_j I_{B_j}$. Como $\{A_i : i = 1, ..., n\} \in \mathcal{A}$ y $\{B_j : j = 1, ..., m\} \in \mathcal{A}$ son particiones de Ω, entonces

$$P(B_j) = \sum_{i=1}^{n} P(A_i \cap B_j) \qquad j \in \{1, ..., m\}$$

y

$$P(A_i) = \sum_{j=1}^{m} P(A_i \cap B_j) \qquad i \in \{1, ..., n\}$$

de donde

$$\sum_{i=1}^{n} x_i P(A_i) = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i P(A_i \cap B_j)$$

y

$$\sum_{j=1}^{m} b_j P(B_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} b_j P(A_i \cap B_j).$$

Pero

$$x_i P(A_i \cap B_j) = b_j P(A_i \cap B_j)$$

ya que, si $A_i \cap B_j \neq \emptyset$, entonces $x_i = b_j = X(\omega_{ij})$, donde $\omega_{ij} \in A_i \cap B_j$. Por lo tanto,

$$\sum_{i=1}^{n} \sum_{j=1}^{m} x_i P(A_i \cap B_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} b_j P(A_i \cap B_j)$$

lo que concluye la demostración.

En el siguiente teorema se establecen las propiedades más elementales del valor esperado de variables aleatorias simples no negativas.

Teorema 2.2.1 *Si*

$$X = \sum_{i=1}^{n} x_i I_{A_i} \qquad y \qquad Y = \sum_{j=1}^{m} y_j I_{B_j}$$

son variables aleatorias simples no negativas, entonces

1. $E(X) \geq 0.$
2. *Si* c *es una constante y* $X \overset{c.s.}{=} c$ *entonces* $E(X) = c.$
3. *Si* $X = I_A$ *entonces* $E(X) = P(A).$
4. *Si* $a, b \in \mathbb{R}$ *entonces* $E(aX + bY) = aE(X) + bE(Y).$
5. *Si* $X \geq Y$ *entonces* $E(X) \geq E(Y).$

Demostración.

1. El rango de X es un conjunto de números no negativos y $P : \Omega \to [0,1]$, entonces $x_i P(A_i) \geq 0$ y por lo tanto,
$$\sum_{i=1}^{n} x_i P(A_i) \geq 0.$$

2. Sea $A \in \mathcal{A}$ el conjunto donde $X \stackrel{c.s.}{=} c$, entonces $P(A) = 1$ y
$$E(X) = cP(A) = c.$$

3. Si $X = I_A$, entonces
$$E(X) = 1 \cdot P(A) + 0 \cdot P(A^c) = P(A).$$

4. $aX + bY = \sum_{i=1}^{n} \sum_{j=1}^{m} (ax_i + by_i) I_{A_i \cap B_j}$, entonces
$$\begin{aligned} E(aX + bY) &= \sum_{i=1}^{n} \sum_{j=1}^{m} (ax_i + by_i) P(A_i \cap B_j) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} (ax_i) P(A_i \cap B_j) + \sum_{i=1}^{n} \sum_{j=1}^{m} (by_i) P(A_i \cap B_j) \\ &= \sum_{i=1}^{n} (ax_i) P(A_i) + \sum_{j=1}^{m} (by_i) P(B_j) \\ &= aE(X) + bE(Y). \end{aligned}$$

5. Si $X \geq Y$, entonces $X - Y$ es una variable aleatoria no negativa entonces, por 1 y 4
$$E(X - Y) = E(X) - E(Y) \geq 0.$$

■

Ejemplo 2.2.1 *Supongamos que X es una variable aleatoria con distribución Bernoulli (p), esto es, X es una variable aleatoria que toma el valor 1 con probabilidad p y el valor 0 con probabilidad $1 - p$. Entonces*
$$E(X) = 0 \cdot P[X = 0] + 1 \cdot P[X = 1] = p.$$

▲

Ejemplo 2.2.2 *Supongamos que X tiene distribución Binomial (n,p), en este caso*

$$P[X=k]=\binom{n}{k}p^k(1-p)^{n-k}; \qquad k=0,1,...,n,$$

entonces

$$\begin{aligned}
E(X) &= \sum_{k=0}^{n} kP[X=k] \\
&= \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} \\
&= \sum_{k=0}^{n} k\frac{n!}{(n-k)!k!}p^k(1-p)^{n-k} \\
&= np\sum_{k=1}^{n} \frac{(n-1)!}{((n-1)-(k-1))!(k-1)!}p^{k-1}(1-p)^{(n-1)-(k-1)} \\
&= np\sum_{m=0}^{n-1} \frac{(n-1)!}{((n-1)-m)!m!}p^m(1-p)^{(n-1)-m} \\
&= np.
\end{aligned}$$

▲

El siguiente teorema será de especial utilidad para extender la definición de valor esperado a variables aleatorias no negativas.

Teorema 2.2.2 *Si $\{X_n\}_{n\geq 1}$ es una sucesión de variables aleatorias simples y no negativas tal que $X_n \uparrow X$ y Y es una variable aleatoria simple no negativa tal que $X \geq Y$, entonces*

$$\lim_{n\to\infty} E(X_n) \geq E(Y).$$

Demostración. Sea $Y=\sum_{j=1}^{m} y_j I_{B_j}$. Para $\varepsilon>0$ definimos $D_n := \{\omega : X_n(\omega) \geq Y(\omega)-\varepsilon\}$. Es claro que $D_n \uparrow \Omega$, por lo tanto, para toda $j \in \{1,2,...,m\}$, $(B_j \cap D_n) \uparrow B_j$, entonces, por el teorema 1.3.3

$$\lim_{n\to\infty} P(D_n)=1$$

y

$$\lim_{n\to\infty} P(B_j \cap D_n) = P(B_j),$$

de donde

$$\lim_{n\to\infty} E\left(YI_{D_n}\right) = \lim_{n\to\infty} \sum_{j=1}^{m} y_j P\left(B_j \cap D_n\right) = E\left(Y\right).$$

Además, para todo n entero positivo,

$$X_n \geq X_n I_{D_n} \geq Y I_{D_n} - \varepsilon I_{D_n}$$

de donde, por 3 y 5 del teorema 2.2.1

$$\lim_{n\to\infty} E\left(X_n\right) \geq E\left(Y\right) - \varepsilon \tag{2.2}$$

y como (2.2) se cumple para cualquier $\varepsilon > 0$, entonces

$$\lim_{n\to\infty} E\left(X_n\right) \geq E\left(Y\right).$$

■

2.2.2 Variables Aleatorias no Negativas

Sea X una variable aleatoria no negativa y considérese la sucesión $\{X_n\}_{n\geq 1}$ con

$$X_n := \sum_{i=1}^{n2^n} 2^{-n}\left(i-1\right) I_{A_{in}}$$

donde $A_{in} := \{\omega \in \Omega : (i-1)\, 2^{-n} \leq X\left(\omega\right) \leq i2^{-n}\}$ para $i \in \{1, 2, ..., n2^n\}$. Nótese que $\{X_n\}_{n\geq 1}$ es una sucesión de variables aleatorias simples. Se puede demostrar que

$$0 \leq X_n\left(\omega\right) \leq X_{n+1}\left(\omega\right) \leq X\left(\omega\right)$$

para toda ω y toda n. Además, si ω es fija, se tiene que

$$0 \leq X\left(\omega\right) - X_n\left(\omega\right) \leq 2^{-n}.$$

Por lo tanto, si $n \longrightarrow \infty$, $X_n \uparrow X$ y tenemos el siguiente resultado

Proposición 2.2.3 *Sea X una variable aleatoria no negativa. Entonces existe una sucesión de variables aleatorias simples no negativas que converge a X.*

Definición 2.2.3 *Si X es una variable aleatoria no negativa y $\{X_n\}_{n\geq 1}$ es una sucesión de variables aleatorias simples tal que $X_n \uparrow X$, entonces definimos*

$$E\left(X\right) := \lim_{n\to\infty} E\left(X_n\right).$$

La proposición 2.2.3 nos garantiza que la definición 2.2.3 tiene sentido para toda variable aleatoria no negativa.

Para demostrar que $E(X)$ está bien definido para X no negativa supongamos que $\{X_n\}_{n\geq 1}$ y $\{Y_m\}_{m\geq 1}$ son sucesiones de variables aleatorias simples tales que $X_n \uparrow X$ y $Y_m \uparrow X$. Claramente

$$\lim_{n\to\infty} X_n \geq Y_m$$

y

$$\lim_{m\to\infty} Y_m \geq X_n$$

entonces, por el teorema 2.2.2, para toda n

$$\lim_{n\to\infty} E(X_n) \geq E(Y_m)$$

y

$$\lim_{m\to\infty} E(Y_m) \geq E(X_n),$$

de donde

$$\lim_{n\to\infty} E(X_n) \geq \lim_{m\to\infty} E(Y_m)$$

y

$$\lim_{m\to\infty} E(Y_m) \geq \lim_{n\to\infty} E(X_n),$$

entonces se sigue que

$$\lim_{n\to\infty} E(X_n) = \lim_{m\to\infty} E(Y_m)$$

y por lo tanto, $E(X)$ está bien definido.

Para demostrar que las definiciones 2.2.2 y 2.2.3 son equivalentes cuando X es una variable aleatoria simple no negativa, basta con tomar la sucesión $\{X_n\}_{n\geq 1}$, donde $X_1 = X_2 = \ldots = X$, entonces claramente ambas definiciones son equivalentes.

Teorema 2.2.4 *Sean X y Y variables aleatorias no negativas entonces*

1. $E(X) \geq 0$.
2. *Si* $c \geq 0$, *entonces* $E(cX) = cE(X)$.
3. $E(X+Y) = E(X) + E(Y)$.
4. *Si* $X \geq Y$ *entonces* $E(X) \geq E(Y)$.

Demostración. A lo largo de la prueba suponemos que $\{X_n\}_{n\geq 1}$ y $\{Y_m\}_{m\geq 1}$ son sucesiones de variables aleatorias simples con $X_n \uparrow X$ y $Y_m \uparrow Y$.

1. Es obvio, ya que $E(X)$ es límite de una suma de valores no negativos.

2. Claramente la sucesión $\{cX_n\}_{n\geq 1}$ es una sucesión de variables aleatorias simples con $cX_n \uparrow cX$ en Ω. Entonces

$$\begin{aligned} E(cX) &= \lim_{n\to\infty} E(cX_n) \\ &= \lim_{n\to\infty} cE(X_n) \\ &= cE(X). \end{aligned}$$

3. La sucesión $\{X_n + Y_n\}_{n\geq 1}$ es una sucesión de variables aleatorias simples tal que

$$X_n + Y_n \uparrow X + Y,$$

entonces

$$\begin{aligned} E(X+Y) &= \lim_{n\to\infty} E(X_n + Y_n) \\ &= \lim_{n\to\infty} E(X_n) + \lim_{n\to\infty} E(Y_n) \\ &= E(X) + E(Y). \end{aligned}$$

4. Si $X \geq Y$, entonces $X - Y$ es una variable aleatoria no negativa. Así, por 2 y 3

$$E(X) = E(Y) + E(X - Y) \geq E(Y).$$

■

2.2.3 Caso General

Si X es una variable aleatoria que puede tomar cualquier valor real , entonces X puede expresarse como

$$X = X^+ - X^-$$

donde X^+ y X^- son las variables aleatorias no negativas definidas en (1.7), entonces el valor esperado de X puede definirse de la siguiente manera.

Definición 2.2.4 *Sea* $X = X^+ - X^-$ *una variable aleatoria tal que* $\min\{E(X^+), E(X^-)\} < \infty$ *entonces*

$$E(X) = E\left(X^+\right) - E\left(X^-\right).$$

Si $E(X^+) = E(X^-) = \infty$, *entonces* $E(X)$ *no está definido.*

Supongamos que X es una variable aleatoria con esperanza finita y que c es una constante real. Si $c \geq 0$, entonces

$$(cX)^+ = cX^+ \qquad \text{y} \qquad (cX)^- = cX^-$$

de donde

$$\begin{aligned} E(cX) &= E(cX^+) - E(cX^-) \\ &= cE(X^+) - cE(X^-) \\ &= cE(X). \end{aligned}$$

Si $c < 0$, entonces

$$(cX)^+ = -cX^- \qquad \text{y} \qquad (cX)^- = -cX^+$$

y

$$\begin{aligned} E(cX) &= E(-cX^-) - E(-cX^+) \\ &= -cE(X^-) + cE(X^+) \\ &= cE(X). \end{aligned}$$

Así, hemos demostrado el

Teorema 2.2.5 *Si X es una variable aleatoria con valor esperado finito y c es una constante real, entonces*

$$E(cX) = cE(X).$$

Ahora, supongamos que X y Y son variables aleatorias . Notamos que

$$\begin{aligned} (X+Y)^+ &= \frac{1}{2}[|X+Y| + (X+Y)] \\ &\leq \frac{1}{2}[|X+Y| + |X| + |Y|] \\ &= X^+ + Y^+ \end{aligned}$$

y

$$\begin{aligned} (X+Y)^- &= \frac{1}{2}[|X+Y| - (X+Y)] \\ &\leq \frac{1}{2}[|X| + |Y| - X - Y] \\ &= \frac{1}{2}[|X| - X] + \frac{1}{2}[|Y| - Y] \\ &= X^- + Y^- \end{aligned}$$

Si $E(X)$ y $E(Y)$ son finitos, entonces

$$\max\left\{X^+, Y^+, X^-, Y^-\right\} < \infty,$$

entonces, por 4 del teorema 2.2.4,

$$E\left((X+Y)^+\right) < \infty \qquad \text{y} \qquad E\left((X+Y)^-\right) < \infty$$

y por lo tanto,

$$E(X+Y) < \infty.$$

Ahora, notamos que

$$(X+Y)^+ + X^- + Y^- = \frac{1}{2}\left[|X+Y| + |X| + |Y|\right] = (X+Y)^- + X^+ + Y^+,$$

entonces

$$E\left((X+Y)^+\right) + E\left(X^-\right) + E\left(Y^-\right) = E\left((X+Y)^-\right) + E\left(X^+\right) + E\left(Y^+\right),$$

de donde

$$E(X+Y) = E(X) + E(Y).$$

Por lo tanto, queda demostrado el

Teorema 2.2.6 *Si X y Y son variables aleatorias con valor esperado finito, entonces el valor esperado de $X+Y$ es finito y*

$$E(X+Y) = E(X) + E(Y).$$

Ahora supongamos que los valores esperados de las variables aleatorias X y Y existen y que $X \geq Y$, entonces

$$X^+ \geq Y^+ \qquad \text{y} \qquad Y^- \geq X^-$$

por lo que

$$E\left(X^+\right) \geq E\left(Y^+\right) \qquad \text{y} \qquad E\left(Y^-\right) \geq E\left(X^-\right),$$

entonces

$$E\left(X^+\right) - E\left(X^-\right) \geq E\left(Y^+\right) - E\left(Y^-\right)$$

y por lo tanto,

$$E(X) \geq E(Y).$$

Así, queda demostrado el

Teorema 2.2.7 *Si X y Y son variables aleatorias para las cuales $E(X)$ y $E(Y)$ existen. Si $X \geq Y$ en Ω, entonces*

$$E(X) \geq E(Y).$$

Teorema 2.2.8 *Sea X una variable aleatoria, $E(X)$ es finito si, y sólo si, $E(|X|)$ es finito.*

Demostración. Supongamos que $E(X)$ es finito, entonces

$$\max\left\{E\left(X^+\right), E\left(X^-\right)\right\} < \infty$$

y como

$$E(|X|) = E\left(X^+ + X^-\right) = E\left(X^+\right) + E\left(X^-\right). \tag{2.3}$$

se tiene que $E(|X|)$ es finito.

Ahora suponemos que $E(|X|)$ es finito. Como X^+ y X^- son variables aleatorias no negativas, se sigue de 1 del teorema 2.2.4 y de (2.3) que

$$E(|X|) \geq E\left(X^+\right)$$

y

$$E(|X|) \geq E\left(X^-\right),$$

entonces nuevamente

$$\max\left\{E\left(X^+\right), E\left(X^-\right)\right\} < \infty.$$

Por lo tanto,

$$E(X) = E\left(X^+\right) - E\left(X^-\right)$$

es finito. ∎

Teorema 2.2.9 *Si X es una variable aleatoria con $E(|X|) < \infty$, entonces*

$$|E(X)| \leq E(|X|)$$

Demostración. Tenemos que

$$\begin{aligned} |E(X)| &= \left|E\left(X^+\right) - E\left(X^-\right)\right| \\ &\leq E\left(X^+\right) + E\left(X^-\right) \\ &= E\left(X^+ + X^-\right); \end{aligned}$$

ya que $|X| = X^+ + X^-$ se tiene

$$|E(X)| \leq E(|X|)$$

que es lo que se quería demostrar. ∎

Teorema 2.2.10 *Supongamos que X es una variable aleatoria con valor esperado finito y Y es una variable aleatoria tal que $X \geq |Y|$ en Ω. Entonces $E(Y) < \infty$.*

Demostración. Por el teorema 2.2.7

$$E(X) \geq E(|Y|)$$

de donde, como $E(X) < \infty$ y por el teorema 2.2.9, $E(Y)$ es finito. ■

2.3 Algunos Teoremas Importantes

Teorema 2.3.1 (Teorema de Convergencia Monótona) *Sea $\{X_n\}_{n\geq 1}$ una sucesión monótona creciente de variables aleatorias no negativas que converge a X, entonces*

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

Demostración. Por 3 del corolario 1.4.7 tenemos que $\lim_{n\to\infty} X_n = X$ es una variable aleatoria. Como $X_n \leq X_{n+1} \leq X$, entonces, por el teorema 2.2.7, para cada entero positivo n

$$E(X_n) \leq E(X_{n+1}) \leq E(X),$$

de donde

$$\lim_{n\to\infty} E(X_n) \leq E(X). \tag{2.4}$$

Además, como cada elemento en la sucesión $\{X_n\}_{n\geq 1}$ es una variable aleatoria no negativa, sabemos, de la proposición 2.2.3, que para cada entero positivo n existe una sucesión monótona creciente $\{X_{km}\}_{m\geq 1}$ de variables aleatorias simples con límite X_k. Ahora, para cada n entero positivo, definimos

$$Y_n := \max\{X_{1n}, X_{2n}, ..., X_{nn}\},$$

entonces es claro que $\{Y_n\}_{n\geq 1}$ es una sucesión de variables aleatorias simples no negativas para las cuales

$$X_{kn} \leq Y_n \leq X_n, \tag{2.5}$$

si $k, n \in \{1, 2, ...\}$, de donde, por el teorema 2.2.7,

$$E(Y_n) \leq E(X_n). \tag{2.6}$$

Si en (2.5) hacemos $n \to \infty$ tenemos que

$$X_k \leq \lim_{n\to\infty} Y_n \leq X, \tag{2.7}$$

de donde, si en (2.7) $k \to \infty$, se sigue que

$$\lim_{n\to\infty} Y_n = X,$$

entonces, como $\{Y_n\}_{n\geq 1}$ es una sucesión de variables aleatorias simples que converge a X,

$$\lim_{n\to\infty} E(Y_n) = E(X).$$

Pero, por (2.6),

$$\lim_{n\to\infty} E(X_n) \geq \lim_{n\to\infty} E(Y_n) = E(X); \tag{2.8}$$

por lo tanto, combinando (2.4) y (2.8),

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

■

Teorema 2.3.2 (Lema de Fatou) *Si $\{X_n\}_{n\geq 1}$ es una sucesión de variables aleatorias no negativas, entonces*

$$E\left(\liminf_{n\longrightarrow\infty} X_n\right) \leq \liminf_{n\longrightarrow\infty} E(X_n).$$

Demostración. Sea $Y_m := \inf\{X_m, X_{m+1}, ...\}$, entonces, si $m \leq n$, $Y_m \leq X_n$, de donde

$$E(Y_m) \leq E(X_n),$$

para $m \leq n$. Entonces

$$E(Y_m) \leq \liminf_{n\longrightarrow\infty} E(X_n).$$

Pero, como $\{Y_m\}_{m\geq 1} \uparrow \liminf_{n\longrightarrow\infty} X_n$, se sigue, del teorema de convergencia monótona,

$$E\left(\liminf_{n\longrightarrow\infty} X_n\right) = \lim_{m\to\infty} E(Y_m) \leq \liminf_{n\longrightarrow\infty} E(X_n)$$

ésto demuestra el teorema. ■

Teorema 2.3.3 (Teorema de Convergencia Dominada de Lebesgue) *Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias no negativas que converge a X en Ω. Si existe una variable aleatoria Y con valor esperado finito tal que, para toda k, $|X_k| \leq Y$, entonces $E(X)$ es finita y*

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

Demostración. Como para todo entero positivo k, $|X_k| \leq Y$ en Ω, se tiene que

$$|X| \leq Y,$$

entonces, por el teorema 2.2.7,

$$E(|X|) \leq E(Y)$$

de donde $E(X)$ es finito. Claramente $Y + X_n$ es una variable aleatoria no negativa, por lo que podemos aplicar el Lema de Fatou

$$\begin{aligned} E(Y) + E(X) &= E(Y + X) \\ &\leq \liminf_{n \longrightarrow \infty} E(Y + X_n) \\ &= \liminf_{n \longrightarrow \infty} [E(Y) + E(X_n)] \\ &= E(Y) + \liminf_{n \longrightarrow \infty} E(X_n), \end{aligned}$$

entonces

$$E(X) \leq \liminf_{n \longrightarrow \infty} E(X_n). \tag{2.9}$$

También es claro que $Y - X_n \geq 0$, de donde, aplicando nuevamente el Lema de Fatou, obtenemos

$$\begin{aligned} E(Y) - E(X) &= E(Y - X) \\ &\leq \liminf_{n \longrightarrow \infty} E(Y - X_n) \\ &= \liminf_{n \longrightarrow \infty} [E(Y) - E(X_n)] \\ &= E(Y) - \limsup_{n \longrightarrow \infty} E(X_n), \end{aligned}$$

y se sigue que

$$\limsup_{n \longrightarrow \infty} E(X_n) \leq E(X). \tag{2.10}$$

Así, combinado (2.9) y (2.10), tenemos que

$$E(X) = \lim_{n \longrightarrow \infty} E(X_n).$$

■

Para el caso particular del teorema de convergencia dominada cuando $Y(\omega) = c$ para todo $\omega \in \Omega$ tenemos el siguiente resultado; el cual se conoce como teorema de convergencia Acotada.

Teorema 2.3.4 (Teorema de Convergencia Acotada) *Si $\{X_n\}_{n\geq 1}$ es una sucesión de variables aleatorias no negativas con límite X en Ω y si existe una constante real c tal que, para toda $n, |X_n| \leq c$, entonces $E(X)$ es finito y*

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

Teorema 2.3.5 *Supongamos que $\{X_n\}_{n\geq 1}$ es una sucesión de variables aleatorias no negativas y que $\sum\limits_{n=1}^{\infty} X_n$ converge en Ω, entonces*

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

Demostración. Sea $Y_k = \sum\limits_{n=1}^{k} X_n$, entonces, como para cada m entero positivo X_m es una variable aleatoria no negativa, $\{Y_k\}_{k\geq 1}$ es una sucesión monótona creciente de variable aleatorias no negativas, además por hipótesis sabemos que existe $Y = \lim\limits_{k\to\infty} Y_k$ de donde, por el teorema de convergencia monótona,

$$E(Y) = \lim_{k\to\infty} E(Y_k);$$

es decir,

$$E\left(\sum_{n=1}^{\infty} X_n\right) = \lim_{k\to\infty} E\left(\sum_{n=1}^{k} X_n\right).$$

Entonces, por el teorema 2.2.6,

$$\begin{aligned} E\left(\sum_{n=1}^{\infty} X_n\right) &= \lim_{k\to\infty} \sum_{n=1}^{k} E(X_n) \\ &= \sum_{n=1}^{\infty} E(X_n) \end{aligned}$$

que es lo que se quería demostrar. ∎

2.3.1 Dependencia de un Parámetro

En el siguiente capítulo necesitaremos considerar valores esperados de variables aleatorias que dependen de un parámetro real. Por ello en esta sección veremos como el teorema de convergencia dominada puede utilizarse en este caso.

En lo que resta de esta sección supondremos que

$$X^{(t)} : \Omega \times [a,b] \to \mathbb{R}$$

es una variable aleatoria para cada $t \in [a,b]$.

Teorema 2.3.6 *Supongamos que para algún $t_0 \in [a,b]$*

$$X^{(t_0)} = \lim_{t \to t_0} X^{(t_0)}$$

y que existe una variable aleatoria Y con esperanza finita tal que $|X^{(t)}| \leq Y$. Entonces

$$E\left(X^{(t_0)}\right) = \lim_{t \to t_0} E\left(X^{(t)}\right).$$

Demostración. Sea $\{t_n\}_{n \geq 1}$ una sucesión en $[a,b]$ que converge a t_0. Definimos $X_n = X^{(t_n)}$, entonces $X_n \to X^{(t_0)}$ y por el teorema de convergencia dominada

$$E\left(X^{(t_0)}\right) = E\left(\lim_{n \to \infty} X_n\right) = \lim_{n \to \infty} E\left(X^{(t_n)}\right) = \lim_{t \to t_0} E\left(X^{(t)}\right).$$

■

Como una consecuencia inmediata del teorema 2.3.6 se tiene el

Corolario 2.3.7 *Si $X^{(t)}$ es continua en $[a,b]$, para cada $\omega \in \Omega$, y existe una variable aleatoria Y con esperanza finita tal que $|X^{(t)}| \leq Y$. Entonces $F(t) = E\left(X^{(t)}\right)$ es continua en $[a,b]$.*

Teorema 2.3.8 *Supongamos que para algún $t_0 \in [a,b]$ la variable aleatoria $X^{(t)}$ tiene valor esperado finito, que $\frac{\partial X^{(t)}}{\partial t}$ existe sobre $\Omega \times [a,b]$, y que existe una variable aleatoria Y con esperanza finita tal que*

$$\left|\frac{\partial X^{(t)}}{\partial t}\right| \leq Y; \qquad t \in [a,b],$$

entonces $E\left[\frac{\partial X^{(t)}}{\partial t}\right]$ es finita sobre $[a,b]$ y

$$\frac{d}{dt} E\left(X^{(t)}\right) = E\left[\frac{\partial X^{(t)}}{\partial t}\right]$$

sobre $]a,b[$.

Demostración. Sea $t \in [a,b]$. Si $\{t_n\}_{n\geq 1}$ una sucesión en $[a,b]$ que converge a t, $t_n \neq t$, entonces por los teoremas 1.4.1, 1.4.2 y el corolario 1.4.7, para cada $\omega \in \Omega$

$$\frac{\partial X^{(t)}}{\partial t} = \lim_{n\to\infty} \frac{X^{(t_n)} - X^{(t)}}{t_n - t}$$

es una variable aleatoria.

Si $\omega \in \Omega$ y $t \in [a,b]$, entonces, según el teorema del valor medio, existe c entre t_0 y t tal que

$$X^{(t)} - X^{(t_0)} = (t - t_0)\frac{\partial X^{(c)}}{\partial t},$$

así,

$$\left|X^{(t)}\right| \leq \left|X^{(t_0)}\right| + |t - t_0| Y$$

y se sigue que $E\left(X^{(t)}\right)$ es finito para cada $t \in [a,b]$.

Ahora sea $F(t) = E\left(X^{(t)}\right)$, entonces, si $t_n \neq t$,

$$\frac{F(t_n) - F(t)}{t_n - t} = E\left[\frac{X^{(t_n)} - X^{(t)}}{t_n - t}\right];$$

pero como $\frac{X^{(t_n)} - X^{(t)}}{t_n - t}$ está dominada por Y, aplicamos el teorema de convergencia dominada y obtenemos

$$\frac{d}{dt}E\left(X^{(t)}\right) = E\left[\frac{\partial X^{(t)}}{\partial t}\right].$$

■

Teorema 2.3.9 *Si, para todo $\omega \in \Omega$, $X(\omega, u)$ es continua en $[a,b]$ y existe una variable aleatoria Y con esperanza finita tal que $|X(\omega,u)| \leq Y(\omega)$. Entonces*

$$\int_a^b E(X(\omega,t))\,dt = E\left(\int_a^b X(\omega,t)\,dt\right)$$

donde la integrales con respecto a t son integrales de Riemann.

Demostración. Sea Z definida sobre $\Omega\times [a,b]$ como

$$Z(\omega,t) = \int_a^t X(\omega,u)\,du,$$

entonces

$$\frac{\partial}{\partial t}Z(\omega,t) = X(\omega,t).$$

Como la integral es el límite de una sucesión de sumas de Riemann entonces Z es una variable aleatoria; además, como $|X(\omega,t)| \leq Y(\omega)$ en Ω, se sigue que $|Z(\omega,t)| \leq Y(\omega)(b-a)$ en Ω de donde, por el teorema de convergencia dominada, $E(Z(\omega,t))$ es finito para cada $t \in [a,b]$.

Ahora, sea

$$W(t) = E(Z(\omega,t)),$$

entonces, por el teorema 2.3.8,

$$\begin{aligned} \frac{d}{dt}W(t) &= E\left(\frac{\partial}{\partial t}Z(\omega,t)\right) \\ &= E(X(\omega,t)). \end{aligned}$$

Por lo tanto,

$$\begin{aligned} \int_a^b E(X(\omega,t))\,dt &= W(b) - W(a) \\ &= E(Z(\omega,b) - Z(\omega,a)) \\ &= E\left(\int_a^b X(\omega,t)\,dt\right). \end{aligned}$$

■

2.4 Independencia y Valor Esperado

Supongamos que X, Y son variables aleatorias independientes con valor esperado finito.

i) Si

$$X = \sum_{i=1}^{n} x_i I_{A_i} \quad \text{y} \quad Y = \sum_{j=1}^{m} y_j I_{B_j}$$

son variables aleatorias no negativas, entonces

$$\begin{aligned} E(XY) &= E\left(\left(\sum_{i=1}^{n} x_i I_{A_i}\right)\left(\sum_{j=1}^{m} y_j I_{B_j}\right)\right) \\ &= E\left(\sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j I_{A_i \cap B_j}\right) \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j P(A_i \cap B_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j P(A_i) P(B_j) \\
&= \left(\sum_{i=1}^{n} x_i P(A_i)\right)\left(\sum_{j=1}^{m} y_j P(B_j)\right) \\
&= E(X) E(Y)
\end{aligned}$$

ii) Ahora, supongamos que X y Y son variables aleatorias no negativas. Consideremos la sucesión $\{X_n\}_{n\geq 1}$, con

$$X_n := \sum_{i=1}^{n2^n} 2^{-n}(i-1) I_{A_{in}},$$

donde $A_{in} := \{\omega \in \Omega : (i-1)2^{-n} \leq X(\omega) \leq i2^{-n}\}$ para $i \in \{1,2,...,n2^n\}$, y la sucesión $\{Y_n\}_{n\geq 1}$, con

$$Y_n := \sum_{i=1}^{n2^n} 2^{-n}(i-1) I_{B_{in}}$$

donde $B_{in} := \{\omega \in \Omega : (i-1)2^{-n} \leq Y(\omega) \leq i2^{-n}\}$, $i \in \{1,2,...,n2^n\}$, entonces, como se vio en la prueba de la proposición 2.2.3, $\{X_n\}_{n\geq 1}$ y $\{Y_n\}_{n\geq 1}$ son sucesiones monótonas crecientes de variables aleatorias simples con límite X y Y respectivamente. Entonces, $\{X_nY_n\}_{n\geq 1}$ es una sucesión monótona creciente de variables aleatorias simples que converge a la variable aleatoria XY. Además, ya que X_n y Y_n son transformaciones Borel medibles de X y Y respectivamente, se sigue del teorema 1.5.3 que $X_1, X_2, ..., Y_1, Y_2, ...$ son variables aleatorias simples independientes. Así

$$E(XY) = \lim_{n\to\infty} E(X_nY_n)$$

pero, como ya se demostró,

$$E(X_nY_n) = E(X_n)E(Y_n),$$

entonces

$$\begin{aligned}
E(XY) &= \lim_{n\to\infty} E(X_n)E(Y_n) \\
&= \lim_{n\to\infty} E(X_n) \lim_{n\to\infty} E(Y_n) \\
&= E(X)E(Y)
\end{aligned}$$

que es finito.

iii) Si $X = X^+ - X^-$ y $Y = Y^+ - Y^-$ son variables aleatorias que pueden tomar cualquier signo, entonces

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

pero como X^+, X^-, Y^+ y Y^- son transformaciones medibles de X y Y, se tiene, por el teorema 1.5.3, que

$$X^+ \perp Y^+,\ X^- \perp Y^+,\ X^+ \perp Y^-,\ X^- \perp Y^-;$$

entonces

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

además, como $E(X)$ y $E(Y)$ son finitos $E(XY) = E(X)E(Y) < \infty$.

Lo anterior se resume en el

Teorema 2.4.1 *Si X y Y son variables aleatorias independientes con valor esperado finito entonces $E(XY)$ es finito y*

$$E(XY) = E(X)E(Y).$$

Como extensión del teorema 2.4.1 al caso de n variables aleatorias tenemos el

Corolario 2.4.2 *Si $X_1, X_2, ..., X_n$ son variables aleatorias independientes con valor esperado finito entonces $E(X_1X_2 \cdots X_n)$ es finito y*

$$E\left(\prod_{1=1}^{n} X_i\right) = \prod_{1=1}^{n} E(X_i)$$

2.5 Variables Aleatorias Complejas

En esta sección se presentan extensiones al caso complejo de la definición de valor esperado y, cuando es posible la analogía, de los resultados probados para variables aleatorias reales. En lo siguiente suponemos que Z con o sin subíndice representa una variable aleatoria de tipo complejo. La mayoría de las pruebas son análogas y por lo tanto las omitimos..

Definición 2.5.1 *Sea $Z = X + iY$ una variable aleatoria compleja, decimos que $E(Z)$ es finito si $E(X)$ y $E(Y)$ son ambos finitos, y en tal caso definimos*

$$E(Z) := E(X) + iE(Y).$$

Teorema 2.5.1 *Si $E(Z)$ es finito y c es un número complejo, entonces $E(cZ)$ es finito y*

$$E(cZ) = cE(Z).$$

Teorema 2.5.2 *Si $E(Z_1)$ y $E(Z_2)$ son finitos, entonces $E(Z_1 + Z_2)$ es finito y*

$$E(Z_1 + Z_2) = E(Z_1) + E(Z_2).$$

Teorema 2.5.3 *$E(Z)$ es finito si, y sólo si, $E(|Z|)$ es finito.*

Teorema 2.5.4 *Si $E(Z)$ es finito, entonces*

$$|E(Z)| \leq E(|Z|).$$

Teorema 2.5.5 *Si W es una variable aleatoria real con valor esperado finito y $|Z| \leq W$ en Ω. Entonces $E(Z)$ es finito.*

Teorema 2.5.6 *Supongamos que $\{Z_n\}_{n\geq 1}$ es una sucesión de variables aleatorias complejas con límite Z en Ω y que existe una variable aleatoria real W con valor esperado finito tal que, para $n = 1, 2, ..., |Z_n| \leq W$. Entonces $E(Z)$ es finito y $E(Z_n) \to E(Z)$, cuando $n \to \infty$.*

Corolario 2.5.7 *Supongamos que $\{Z_n\}_{n\geq 1}$ es una sucesión de variables aleatorias complejas con límite Z en Ω y que existe una constante real c tal que para $n = 1, 2, ..., |Z_n| \leq c$. Entonces $E(Z)$ es finito y*

$$E(Z_n) \to E(Z)$$

cuando $n \to \infty$.

Ahora suponemos que

$$Z^{(t)} : \Omega \times [a, b] \to \mathbb{C}$$

es una variable aleatoria compleja para cada $t \in [a, b]$.

Teorema 2.5.8 *Supongamos que para algún* $t_0 \in [a, b]$

$$Z^{(t_0)} = \lim_{t \to t_0} Z^{(t)}$$

y que existe una variable aleatoria real W *con esperanza finita, tal que* $\left|Z^{(t)}\right| \leq W$. *Entonces*

$$E\left(Z^{(t_0)}\right) = \lim_{t \to t_0} E\left(Z^{(t)}\right).$$

Corolario 2.5.9 *Si, para todo* $\omega \in \Omega$, $Z^{(t)}$ *es continua en* $[a, b]$, *y existe una variable aleatoria real* W *con esperanza finita tal que* $\left|X^{(t)}\right| \leq W$. *Entonces* $F(t) = E\left(Z^{(t)}\right)$ *es continua en* $[a, b]$.

Teorema 2.5.10 *Supongamos que para algún* $t_0 \in [a, b]$ *la variable aleatoria* $Z^{(t)}$ *tiene valor esperado finito, que* $\frac{\partial Z^{(t)}}{\partial t}$ *existe sobre* $\Omega \times [a, b]$, *y que existe una variable aleatoria real* W *con esperanza finita tal que*

$$\left|\frac{\partial Z^{(t)}}{\partial t}\right| \leq W; \qquad t \in [a, b],$$

entonces $E\left[\frac{\partial Z^{(t)}}{\partial t}\right]$ *es finita sobre* $[a, b]$, *y*

$$\frac{d}{dt} E\left(Z^{(t)}\right) = E\left[\frac{\partial Z^{(t)}}{\partial t}\right]$$

sobre $]a, b[$.

Teorema 2.5.11 *Si* $Z(\omega, u)$ *es continua en* $[a, b]$, *para cada* $\omega \in \Omega$, *y existe una variable aleatoria real* W *con esperanza finita tal que* $|Z(\omega, u)| \leq W(\omega)$ *para toda* $\omega \in \Omega$. *Entonces*

$$\int_a^b E(Z(\omega, t))\, dt = E\left(\int_a^b Z(\omega, t)\, dt\right).$$

2.6 Teorema de Cambio de Variable

Como ya se mencionó en la introducción de este capítulo, el valor esperado de una variable aleatoria X es la integral de Lebesgue de X con respecto a la medida de probabilidad asociada. Es decir, si X es una variable aleatoria en $(\Omega, \mathcal{A}, P)$, entonces el valor esperado de X, puede escribirse como

$$E(X) = \int_\Omega X dP$$

donde la integral del lado derecho se interpreta en el sentido de Lebesgue.

Denotemos ahora por $I : \mathbb{R} \to \mathbb{R}$ la función identidad, es decir

$$I(x) = x, \qquad \text{para toda } x \in \mathbb{R}$$

entonces, como $I^{-1}(]-\infty, x]) =]-\infty, x]$ es un elemento de la σ-álgebra de Borel en $\mathbb{R}$, se tiene que I es una variable aleatoria en $(\mathbb{R}, \mathcal{B}, P_X)$, donde P_X es la distribución de X. En este caso

$$E(I) = \int_{\mathbb{R}} x dP_X.$$

El siguiente teorema demuestra que el valor esperado de una variable aleatoria X sólo depende de la distribución de X. La demostración puede encontrarse en Chow y Teicher [1988, p. 170], Clarke [1975, p. 75] o Shiryayev [1984, p. 194].

Teorema 2.6.1 *Sea g una función Borel-medible, y X una variable aleatoria en $(\Omega, \mathcal{A}, P)$, si I denota a la función identidad en los reales, entonces $E(g(X))$ existe si, y sólo si, existe $E(g(I))$ y en tal caso*

$$E(g(X)) = \int_{\Omega} g(X(\omega))\, dP = \int_{\mathbb{R}} g(x)\, dP_X = E(g(I)).$$

Corolario 2.6.2 *Sea X una variable aleatoria en $(\Omega, \mathcal{A}, P)$, si I denota a la función identidad en los reales, entonces $E(X)$ existe si, y sólo si, existe $E(I)$ y en tal caso*

$$E(X) = \int_{\Omega} X dP = \int_{\mathbb{R}} x dP_X = E(I).$$

Como un corolario al teorema 2.6.1 concluimos que podemos calcular el valor esperado de la variable aleatoria X si se conoce su función de distribución.

Corolario 2.6.3 *Sea X una variable aleatoria con función de distribución F_X, y g una función Borel-medible, entonces*

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)\, dF_X(x).$$

Demostración. Se sigue del teorema 1.4.8 ■

2.7 Valor Esperado de Variables Aleatorias Discretas y Absolutamente Continuas

Para hacer notar que la definición y las propiedades del valor esperado son totalmente generales, a lo largo de esta sección se ha trabajado con integrales de Lebesgue. Sin embargo, existen casos de alta importancia práctica (variables aleatorias discretas y absolutamente continuas para ser precisos), para los cuales su valor esperado se puede calcular de una manera más sencilla.

En esta sección se incluyen dos teoremas que permiten calcular el valor esperado de variables aleatorias discretas y absolutamente continuas. La demostración no se incluye pero el lector interesado puede encontrarla en Clarke [1975, pp. 77-81].

Teorema 2.7.1 *Si X es una variable aleatoria discreta con conjunto de puntos de discontinuidad $S = \{x_j : j = 1, 2, ...\}$ y con distribución de probabilidad $L = \{p_j : j = 1, 2, ...\}$. Entonces el valor esperado de X está dado por*

$$E(X) = \sum_{i=1}^{\infty} p_j x_j.$$

Teorema 2.7.2 *Si X es una variable aleatoria absolutamente continua con función de densidad $f_X(\cdot)$, entonces el valor esperado de X está dado por*

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx$$

donde la integral se interpreta en el sentido de Riemann.

2.8 Momentos y Desigualdades

Definición 2.8.1 *Si $r > 0$ y X es una variable aleatoria en $(\Omega, \mathcal{A}, P)$, $E(|X|^r)$ es llamado **r-ésimo momento absoluto** de X.*

Definición 2.8.2 *Para cualquier entero positivo k, $E(X^k)$ es conocido (si este existe) como el **k-ésimo momento** de X.*

Definición 2.8.3 *Si $E(|X|^2) < \infty$, definimos la **varianza** de X como*

$$\sigma_X^2 = Var(X) = E\left[(X - E(X))^2\right];$$

*la raíz cuadrada positiva de σ_X^2 es llamada **desviación estándar**.*

No es difícil demostrar que la varianza de X puede calcularse alternativamente por la fórmula

$$\sigma_X^2 = E\left(X^2\right) - E^2\left(X\right).$$

Definición 2.8.4 *Si X, Y son variables aleatorias tales que $0 < \min\left\{\sigma_X^2, \sigma_Y^2\right\} < \infty$, el **coeficiente de correlación** entre X y Y está dado por*

$$\rho_{X,Y} = Corr\left(X, Y\right) = \frac{E\left[\left(X - E\left(X\right)\right)\left(Y - E\left(Y\right)\right)\right]}{\sigma_X \sigma_Y}.$$

*Si $\rho_{X,Y} = 0$ entonces se dice que X y Y son **no correlacionadas**.*

El coeficiente de correlación puede calcularse por la relación

$$\rho_{X,Y} = \frac{E\left[XY\right] - E\left(X\right)E\left(Y\right)}{\sigma_X \sigma_Y}.$$

Se sigue inmediatamente de la definición 2.8.4 y del teorema 2.4.1 que si X, Y son variables aleatorias independientes con primer momento absoluto finito, entonces X y Y son no correlacionadas. El siguiente contraejemplo aparece en Stoyanov [1988, p. 53] y demuestra que el recíproco no es cierto en general.

Ejemplo 2.8.1 *Sea $\Omega = \{1, 2, 3\}$ y supongamos que cada $\omega \in \Omega$ tiene probabilidad $\frac{1}{3}$. Definamos las variables aleatorias X y Y como*

$$X\left(\omega\right) := \begin{cases} 1 & si\ \omega = 1 \\ 0 & si\ \omega = 2 \\ -1 & si\ \omega = 3 \end{cases} \qquad Y\left(\omega\right) := \begin{cases} 0 & si\ \omega = 1 \\ 1 & si\ \omega = 2 \\ 0 & si\ \omega = 3 \end{cases}$$

Es claro que $E\left(X\right) = 0$ y $E\left(XY\right) = 0$ lo que implica que $\rho_{X,Y} = 0$. Sin embargo,

$$P\left\{X = 1, Y = 1\right\} = 0 \neq \frac{1}{3} \cdot \frac{1}{3} = P\left\{X = 1\right\} P\left\{Y = 1\right\}$$

de donde se sigue que X y Y no son independientes. ▲

A continuación presentamos algunas desigualdades que son de uso común en probabilidad y estadística. Las pruebas no se incluyen, pero el lector interesado puede consultarlas en Shiryayev [1984, pp. 190-192].

Teorema 2.8.1 (Desigualdad de Markov) *Sea X una variable aleatoria no negativa. Entonces para toda $\varepsilon > 0$*

$$P[X \geq \varepsilon] \leq \frac{E(X)}{\varepsilon}.$$

Por una ligera modificación de la desigualdad de Markov obtenemos la siguiente desigualdad.

Teorema 2.8.2 (Desigualdad de Chebyshev) *Sea X una variable aleatoria y $\varepsilon > 0$. Entonces*

$$P[X \geq \varepsilon] \leq \frac{E(X^2)}{\varepsilon^2}$$

y

$$P[|X - E(X)| \geq \varepsilon] \leq \frac{Var(X)}{\varepsilon^2}.$$

Teorema 2.8.3 (Desigualdad de Cauchy-Schwarz) *Sean X y Y variables aleatorias tal que $\max\{E(X^2), E(Y^2)\} < \infty$. Entonces $E(|XY|) < \infty$ y*

$$E^2(|XY|) \leq E(X^2) E(Y^2).$$

Teorema 2.8.4 (Desigualdad de Jensen) *Sea g una función Borel medible convexa y X una variable aleatoria con valor esperado finito. Entonces*

$$g(E(X)) \leq E(g(X)).$$

Teorema 2.8.5 (Desigualdad de Lyapunov) *Si X una variable aleatoria y s, t son constantes reales tales que $0 < s < t$; entonces*

$$[E(|X|^s)]^{\frac{1}{s}} \leq [E(|X|^t)]^{\frac{1}{t}}.$$

Teorema 2.8.6 (Desigualdad de Hölder) *Sean $1 < p < \infty$ y $1 < q < \infty$, tales que $\frac{1}{p} + \frac{1}{q} = 1$. Si X y Y son variables aleatorias tales que*

$$\max\{E(|X|^p), E(|Y|^p)\} < \infty,$$

entonces $E(|XY|) < \infty$ y

$$E(|XY|) \leq [E(|X|^p)]^{\frac{1}{p}} [E(|X|^q)]^{\frac{1}{q}}.$$

Teorema 2.8.7 (Desigualdad Minkowski) *Si $\max\{E(|X|^p), E(|Y|^p)\} < \infty$, para $1 \leq p < \infty$, entonces $E(|X+Y|^p) < \infty$ y*

$$[E(|X+Y|^p)]^{\frac{1}{p}} \leq [E(|X|^p)]^{\frac{1}{p}} + [E(|Y|^p)]^{\frac{1}{p}}.$$

Capítulo 3

Función Característica

3.1 Introducción

La teoría de la probabilidad hace uso de diferentes herramientas de análisis. Una de ellas es la transformada de Fourier. En probabilidad, la transformada de Fourier de una función de distribución F, es conocida como la función característica de la función de distribución F.

Como se verá en el capítulo 4, la función característica tiene diversas aplicaciones en la solución de problemas probabilísticos y estadísticos. Para el desarrollo y comprensión de éstas aplicaciones es necesario tener un buen conocimiento de las propiedades de la función característica.

El objetivo de este capítulo es estudiar las propiedades analíticas de la función característica, comenzando en la sección 3.2 en la que se define la función característica y se demuestran sus propiedades elementales; además, se presentan varios ejemplos del cálculo de funciones características. La relación entre la función característica y los momentos de una variable aleatoria se expone en la sección 3.3. Los teoremas de inversión y unicidad se incluyen en la sección 3.4; también en esta sección se estudia la equivalencia entre convergencia débil y convergencia de funciones características y el uso de funciones características para determinar la distribución de sumas de variables aleatorias independientes. Algunas condiciones para poder determinar si una función es característica se mencionan en la sección 3.5. Por último, en la sección 3.6 se hace la extensión de la definición de función característica y de algunas de sus propiedades al caso de vectores aleatorios.

3.2 Definición y Propiedades Elementales

Definición 3.2.1 *Sea X una variable aleatoria, para todo $t \in \mathbb{R}$ se define la* ***función característica*** *φ_X de X como*

$$\varphi_X(t) := E\left(e^{itX}\right).$$

Ya que $e^{itx} = \cos tx + i \operatorname{sen} tx$, podemos calcular la función característica de la variable aleatoria X por la fórmula

$$\varphi_X(t) = E(\cos tX) + iE(\operatorname{sen} tX).$$

Si X tiene función de distribución F_X entonces la función característica de X se puede escribir como

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} dF_X(x). \tag{3.1}$$

Es costumbre referirse indistintamente a $\varphi_X(t)$ como la función característica de la variable aleatoria X o de la función de distribución F_X.

Si F_X es absolutamente continua y tiene densidad f_X entonces

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x)\, dx.$$

Para el caso en que X es una variable aleatoria discreta con conjunto de puntos de discontinuidad $S = \{x_j : j = 1, 2, ...\}$ y con distribución de probabilidad $L = \{p_j : j = 1, 2, ...\}$ la función característica de X es

$$\varphi_X(t) = \sum_{j=1}^{\infty} p_j e^{itx_j}.$$

Se puede observar que la función característica siempre existe, ya que

$$\left|e^{itx}\right| = \left|\cos tx + i \operatorname{sen} tx\right| = 1$$

es una cota para toda x.

A continuación se presentan las propiedades elementales de la función característica. En lo siguiente se supone que X es una variable aleatoria con función característica φ_X y función de distribución F_X.

Teorema 3.2.1 *La función característica φ_X satisface las siguientes relaciones*

1. $\varphi_X(0) = 1.$
2. $|\varphi_X(t)| \leq 1.$
3. $\varphi_X(-t) = \overline{\varphi_X(t)}$ *donde* $\overline{\varphi_X(t)}$ *denota el conjugado complejo de* $\varphi_X(t)$.

Demostración.

1. $\varphi_X(0) = E(\exp(0 \cdot iX)) = 1.$
2. $|\varphi_X(t)| = \left|E\left(e^{itX}\right)\right| \leq E\left(\left|e^{itX}\right|\right) = 1.$
3. $\varphi_X(-t) = \int\limits_{-\infty}^{\infty} \exp(-itx)\, dF_X(x) = \overline{\int\limits_{-\infty}^{\infty} \exp(itx)\, dF_X(x)} = \overline{\varphi_X(t)}.$

■

Supongamos que $t, h \in \mathbb{R}$, entonces

$$\begin{aligned} |\varphi_X(t+h) - \varphi_X(t)| &= \left|E\left(e^{itX}\left(e^{ihX} - 1\right)\right)\right| \\ &\leq E\left(\left|e^{ihX} - 1\right|\right) \end{aligned}$$

claramente $\left|e^{ihX} - 1\right|$ está acotado, por lo tanto, por el teorema de convergencia acotada (teorema 2.3.4), si $h \to 0$,

$$|\varphi_X(t+h) - \varphi_X(t)| \to 0$$

independientemente de t. Esto demuestra el

Teorema 3.2.2 *La función característica es uniformemente continua en* $\mathbb{R}$.

Ejemplo 3.2.1 *Sea X una variable aleatoria con función de distribución*

$$F_X(x) = \begin{cases} 0 & si\ x < c \\ 1 & si\ x \geq c \end{cases}$$

entonces la función característica de X es

$$\varphi_X(t) = e^{ict}.$$

▲

Ejemplo 3.2.2 (Distribución Bernoulli) *Si* $X \sim Bernoulli\,(p)$

$$p_k = p^k (1-p)^{1-k} I_{\{0,1\}}(k); \qquad p \in]0,1[,$$

entonces la función característica de X *es*

$$\varphi_X(t) = (1-p) + p\exp(it).$$

▲

Ejemplo 3.2.3 *Si* $X \sim Binomial\,(n,p)$

$$p_k = \binom{n}{k} p^k (1-p)^{n-k} I_{\{0,1,\ldots,n\}}(k); \qquad p \in]0,1[,$$

entonces la función característica de X *es*

$$\begin{aligned}
\varphi_X(t) &= \sum_{k=0}^{n} \exp(itk) \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} (p\exp(it))^k (1-p)^{n-k} \\
&= \left[(1-p) + pe^{it}\right]^n.
\end{aligned}$$

▲

Ejemplo 3.2.4 *Si* $X \sim Geométrica\,(p)$

$$p_k = p^k (1-p) I_{\{0,1,2,\ldots\}}(k); \qquad p \in]0,1[,$$

entonces la función característica de X *está dada por la serie*

$$\begin{aligned}
\varphi_X(t) &= \sum_{k=1}^{\infty} \exp(itk)\, p^k (1-p) \\
&= (1-p) \sum_{k=1}^{\infty} (p\exp(it))^k;
\end{aligned}$$

como $|p\exp(it)| = p < 1$ *tenemos que*

$$\sum_{k=1}^{\infty} (p\exp(it))^k = \frac{1}{1-p\exp(it)}.$$

Por lo tanto la función característica de X es

$$\varphi_X(t) = \frac{1-p}{1-p\exp(it)}.$$

▲

Ejemplo 3.2.5 *Si $X \sim Poisson(\lambda)$*

$$p_k = \frac{e^{-\lambda}\lambda^k}{k!} I_{\{0,1,2,\ldots\}}(k); \qquad \lambda > 0,$$

entonces la función característica de X es

$$\begin{aligned}
\varphi_X(t) &= \sum_{k=1}^{\infty} e^{itk}\frac{e^{-\lambda}\lambda^k}{k!} \\
&= e^{-\lambda}\sum_{k=1}^{\infty}\frac{[\lambda\exp(it)]^k}{k!} \\
&= e^{-\lambda}\exp\left(e^{it\lambda}\right) \\
&= \exp\left(\lambda\left(e^{it}-1\right)\right).
\end{aligned}$$

▲

Ejemplo 3.2.6 *Si $X \sim Normal(0,1)$ entonces su función de distribución es*

$$f_X(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left(-\frac{y^2}{2}\right)dy$$

y su función característica está dada por

$$\begin{aligned}
\varphi_X(t) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(itx)\exp\left(-\frac{x^2}{2}\right)dx \\
&= \frac{1}{\sqrt{2\pi}}\left[\int_{-\infty}^{\infty}\cos tx\exp\left(-\frac{x^2}{2}\right)dx + i\int_{-\infty}^{\infty}\operatorname{sen} tx\exp\left(-\frac{x^2}{2}\right)dx\right] \\
&= \sqrt{\frac{2}{\pi}}\int_{0}^{\infty}\cos tx\exp\left(-\frac{x^2}{2}\right)dx.
\end{aligned}$$

Observamos que

$$\varphi'_X(t) = \sqrt{\frac{2}{\pi}} \int_0^\infty x \exp\left(-\frac{x^2}{2}\right) \operatorname{sen} tx \, dx,$$

entonces, integrando por partes con $u = \operatorname{sen} tx$ *y* $dv = xe^{-\frac{x^2}{2}}$,

$$\begin{aligned}
\varphi'_X(t) &= \sqrt{\frac{2}{\pi}} \left[e^{-\frac{x^2}{2}} \operatorname{sen} tx \Big|_{x=0}^{\infty} - t \int_0^\infty \cos tx \exp\left(-\frac{x^2}{2}\right) dx \right] \\
&= -t\varphi_X(t).
\end{aligned}$$

Así, llegamos a la ecuación diferencial

$$\varphi'_X(t) + t\varphi_X(t) = 0,$$

multiplicando ambos miembros por $e^{\frac{t^2}{2}}$

$$\frac{d}{dt}\left[e^{\frac{t^2}{2}}\varphi_X(t)\right] = e^{\frac{t^2}{2}}\varphi'_X(t) - te^{\frac{t^2}{2}}\varphi_X(t) = 0,$$

entonces, para c constante,

$$e^{\frac{t^2}{2}}\varphi_X(t) = c$$

lo que implica que la función característica de X tiene la forma

$$\varphi_X(t) = cc^{-\frac{t^2}{2}}$$

sabemos, por 3.2.1 del teorema 3.2.1, que $\varphi_X(0) = 1$ *de donde* $c = 1$. *Por lo tanto, la función característica de X es*

$$\varphi_X(t) = e^{-\frac{t^2}{2}}.$$

▲

Ejemplo 3.2.7 *Sea* $X \sim Exponencial(\theta)$. *En este caso X es absolutamente continua con función de densidad*

$$f_X(x) = \theta \exp(-\theta x) I_{]0,\infty[}(x); \qquad \theta > 0,$$

entonces la función característica de X es

$$\begin{aligned}
\varphi_X(t) &= \int_0^\infty \exp(itx)\,\theta \exp(-\theta x)\, dx \\
&= \theta \left[\int_0^\infty \cos tx \exp(-\theta x)\, dx + i \int_0^\infty \operatorname{sen} tx \exp(-\theta x)\, dx \right].
\end{aligned}$$

Integrando por partes tenemos

$$\begin{aligned}\int_0^\infty \cos tx \exp(-\theta x)\,dx &= e^{-\theta x}\frac{\operatorname{sen} tx}{t}\Big|_{x=0}^{\infty} - \int_0^\infty \left(-\theta e^{-\theta x}\frac{\operatorname{sen} tx}{t}\right)dx \\ &= \frac{\theta}{t}\int_0^\infty e^{-\theta x}\operatorname{sen} tx dx\end{aligned}$$

y

$$\begin{aligned}\int_0^\infty e^{-\theta x}\operatorname{sen} tx dx &= -e^{-\theta x}\frac{\cos tx}{t}\Big|_{x=0}^{\infty} - \int_0^\infty \theta e^{-\theta x}\frac{\cos tx}{t}\,dx \\ &= \frac{1}{t} - \frac{\theta}{t}\int_0^\infty e^{-\theta x}\cos tx\,dx\end{aligned}$$

de donde se obtiene que

$$\int_0^\infty \cos tx \exp(-\theta x)\,dx = \frac{\theta}{\theta^2 + t^2}$$

y

$$\int_0^\infty e^{-\theta x}\operatorname{sen} tx dx = \frac{t}{\theta^2 + t^2},$$

entonces

$$\begin{aligned}\varphi_X(t) &= \theta\frac{\theta + it}{\theta^2 + t^2} \\ &= \theta\frac{\theta + it}{(\theta + it)(\theta - it)} \\ &= \frac{\theta}{\theta - it}.\end{aligned}$$

▲

Ejemplo 3.2.8 *Supongamos que $X \sim U[-a, a]$, entonces su densidad es*

$$f_X(x) = \frac{1}{2a}I_{[-a,a]}(x).$$

Entonces la función característica de X es

$$\begin{aligned}\varphi_X(t) &= \int_{-\infty}^{\infty} \exp(itx)\, f_X(x)\, dx \\ &= \frac{1}{2a}\int_{-a}^{a} \exp(itx)\, dx \\ &= \frac{\operatorname{sen} at}{at}.\end{aligned}$$

▲

Ejemplo 3.2.9 *Sea X una variable aleatoria con distribución $Triangular([-a,a])$. Entonces*

$$f_X(x) = \frac{a-|x|}{a^2} I_{[-a,a]}(x).$$

y la función característica de X está dada por

$$\begin{aligned}\varphi_X(t) &= \int_{-\infty}^{\infty} \exp(itx)\, f_X(x)\, dx \\ &= \int_{-a}^{a} \exp(itx) \frac{a-|x|}{a^2} dx \\ &= \int_{-a}^{a} \frac{a-|x|}{a^2} \cos tx\, dx + i\int_{-a}^{a} \frac{a-|x|}{a^2} \operatorname{sen} tx\, dx \\ &= 2\int_{0}^{a} \frac{a-x}{a^2} \cos tx\, dx \\ &= \frac{2}{a^2}\left[\left.\frac{a \operatorname{sen} tx}{t}\right|_{x=0}^{a} - \int_{0}^{a} x \cos tx\, dx\right] \\ &= \frac{2}{a^2}\left[\frac{a \operatorname{sen} at}{t} - \int_{0}^{a} x \cos tx\, dx\right].\end{aligned}$$

Integrando por partes, con $u = x$ y $dv = \cos tx\, dx$, obtenemos

$$\varphi_X(t) = \frac{2}{a^2}\left[\frac{a \operatorname{sen} at}{t} - \left.\frac{x \operatorname{sen} tx}{t}\right|_{x=0}^{a} + \int_{0}^{a} \frac{\operatorname{sen} tx}{t} dx\right]$$

$$\begin{aligned}
&= \frac{2}{a^2}\left[\frac{a \operatorname{sen} at}{t} - \frac{a \operatorname{sen} at}{t} - \left.\frac{\cos tx}{t^2}\right|_{x=0}^{a}\right] \\
&= \frac{2}{a^2t^2}\left[1 - \cos at\right] \\
&= \left(\frac{\operatorname{sen}\frac{at}{2}}{\frac{at}{2}}\right)^2.
\end{aligned}$$

Por lo tanto, la función característica de X es

$$\varphi_X(t) = \left(\frac{\operatorname{sen}\frac{at}{2}}{\frac{at}{2}}\right)^2.$$

▲

Ahora, supongamos que $a, b \in \mathbb{R}$ y consideremos la variable aleatoria Y definida por la forma lineal $Y = aX + b$, entonces la función característica de Y es

$$\begin{aligned}
\varphi_Y(t) &= E\left(e^{itY}\right) \\
&= E\left(e^{it(aX+b)}\right) \\
&= e^{itb}E\left(e^{aitX}\right) \\
&= e^{itb}\varphi_X(at);
\end{aligned}$$

así, tenemos el

Teorema 3.2.3 *Si $Y = aX + b$, donde X es una variable aleatoria con función característica φ_X y $a, b \in \mathbb{R}$, entonces, para toda $t \in \mathbb{R}$, la función característica φ_Y de Y es*

$$\varphi_Y(t) = e^{itb}\varphi_X(at).$$

Ejemplo 3.2.10 *Supongamos que $X \sim Normal(\mu, \sigma^2)$, y definamos $Z = \frac{X-\mu}{\sigma}$ entonces*

$$\begin{aligned}
F_Z(z) &= P[Z \leq z] \\
&= P[X \leq z\sigma + \mu] \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{z\sigma+\mu} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left(-\frac{t^2}{2}\right) dt.
\end{aligned}$$

por lo tanto, $Z \sim Normal\,(0,1)$. En el ejemplo 3.2.6 se demostró que

$$\varphi_Z(t) = \exp\left(-\frac{t^2}{2}\right),$$

entonces, por el teorema 3.2.3 con $a = \sigma$, $b = \mu$, tenemos que $X = \sigma Z + \mu$ tiene función característica

$$\begin{aligned} \varphi_Y(t) &= e^{it\mu}\varphi_X(\sigma t) \\ &= e^{it\mu}e^{-\frac{\sigma^2 t^2}{2}} \\ &= \exp\left(it\mu - \frac{\sigma^2 t^2}{2}\right). \end{aligned}$$

Así, la función característica de una variable aleatoria con distribución $Normal\,(\mu, \sigma^2)$ es

$$\varphi(t) = \exp\left(it\mu - \frac{\sigma^2 t^2}{2}\right).$$

▲

Como una consecuencia del teorema 2.6.1 tenemos el

Teorema 3.2.4 *Sea X una variable aleatoria con función de distribución F_X y supongamos que g es una función Borel-medible. Entonces la función característica φ_Y de la variable aleatoria $Y = g(X)$ está dada por*

$$\varphi_Y(t) = \int_{-\infty}^{\infty} \exp(itg(x))\, dF_X(x).$$

Ejemplo 3.2.11 *Sea $X \sim Normal\,(0,1)$ y se Y la variable aleatoria definida por la relación $Y = X^2$. Entonces la función característica de Y es*

$$\begin{aligned} \varphi_Y(t) &= \int_{-\infty}^{\infty} \exp\left(itx^2\right) dF_X(x) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(itx^2 - \frac{x^2}{2}\right) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}(1 - 2it)\right) dx \\ &= (1 - 2it)^{-\frac{1}{2}}. \end{aligned}$$

▲

Definición 3.2.2 *Sea ψ una función definida sobre $\mathbb{R}$ que toma valores complejos. Se dice que ψ es* **positiva definida** *si para todo entero positivo N, y para cada elección de $t_1, t_2, ..., t_N \in \mathbb{R}$ y de $\omega_1, \omega_2, ..., \omega_N \in \mathbb{C}$ se satisface la condición*

$$\sum_{j=1}^{N}\sum_{k=1}^{N}\psi\left(t_j - t_k\right)\omega_j\overline{\omega}_k \geq 0$$

Consideremos una función característica φ con función de distribución F, es decir

$$\varphi\left(t\right) = \int_{-\infty}^{\infty} e^{itx} dF\left(x\right).$$

Si $t_1, t_2, ..., t_N$ son números reales y $\omega_1, \omega_2, ..., \omega_N$ son números complejos, entonces

$$\begin{aligned}
\sum_{j=1}^{N}\sum_{k=1}^{N}\varphi\left(t_j - t_k\right)\omega_j\overline{\omega_k} &= \sum_{j=1}^{N}\sum_{k=1}^{N}\int_{-\infty}^{\infty} e^{ix(t_j - t_k)} dF\left(x\right)\omega_j\overline{\omega}_k \\
&= \int_{-\infty}^{\infty}\sum_{j=1}^{N}\sum_{k=1}^{N} e^{ix(t_j - t_k)}\omega_j\overline{\omega}_k dF\left(x\right) \\
&= \int_{-\infty}^{\infty}\left(\sum_{j=1}^{N} e^{ixt_j}\omega_j\right)\left(\sum_{k=1}^{N} e^{-ixt_k}\overline{\omega}_k\right) dF\left(x\right) \\
&= \int_{-\infty}^{\infty}\left|\sum_{j=1}^{N} e^{ixt_j}\omega_j\right|^2 dF\left(x\right) \geq 0.
\end{aligned}$$

Así, queda demostrado el

Teorema 3.2.5 *Si φ es función característica, entonces φ es definida positiva.*

3.3 Función Característica y Momentos

La función característica se usa frecuentemente en el cálculo de los momentos de la distribución.

Teorema 3.3.1 *Sea X una variable aleatoria con función característica φ_X. Si $E(|X|^n) < \infty$, entonces φ_X tiene derivadas continuas hasta de orden n y*

$$\varphi_X^{(k)}(0) = i^k E\left(X^k\right)$$

para $k = 1, 2, ..., n$.

Demostración. Sea F_X la función de distribución de X, entonces

$$\frac{\varphi_X(t+h) - \varphi_X(t)}{h} = \int_{-\infty}^{\infty} e^{itx} \frac{e^{ihx} - 1}{h} dF_X(x).$$

Mediante el desarrollo en series de Taylor se puede demostrar que $\left|e^{ihx} - 1\right| \leq |hx|$, entonces

$$\left|\frac{\varphi_X(t+h) - \varphi_X(t)}{h}\right| \leq \int_{-\infty}^{\infty} |x| \, dF_X(x) < \infty;$$

así, por el teorema 2.5.10,

$$\frac{d}{dt}\varphi_X(t) = i \int_{-\infty}^{\infty} x e^{itx} dF_X(x),$$

Por lo tanto,

$$\left.\frac{d}{dt}\varphi_X(t)\right|_{t=0} = iE(X);$$

el resto se sigue por inducción. ∎

Ejemplo 3.3.1 *En el ejemplo 3.2.7 se demostró que si $X \sim Exponencial(\theta)$ su función característica es*

$$\varphi_X(t) = \frac{\theta}{\theta - it}.$$

Es bien sabido que para $X \sim Exponencial(\theta)$, $E\left(|X|^2\right) < \infty$, además

$$\varphi_X'(t) = \frac{i\theta}{(\theta - it)^2}, \qquad \varphi_X''(t) = -\frac{2\theta}{(\theta - it)^3};$$

entonces por el teorema 3.3.1 tenemos que

$$E(X) = \frac{1}{\theta}, \qquad E\left(X^2\right) = \frac{2}{\theta^2}.$$

▲

Corolario 3.3.2 *Sea X una variable aleatoria tal que $E(|X|^n) < \infty$. Si φ_X denota la función característica de X, entonces*

$$\varphi_X(t) = 1 + \sum_{k=1}^{n} \varphi_X^{(k)}(0)\frac{(it)^k}{k!} + o(t^n), \qquad t \to 0.$$

Demostración. El desarrollo de $\varphi_X(t)$ en series de Maclaurin en una vecindad de $t = 0$ está dado por

$$\varphi_X(t) = 1 + \sum_{k=1}^{n} \varphi_X^{(k)}(0)\frac{t^k}{k!} + R_n(t)$$

donde

$$R_n(t) = \frac{t^n}{n!}\left[\varphi_X^{(n)}(\theta t) - \varphi_X^{(n)}(0)\right], \qquad 0 < \theta < 1.$$

Es claro que

$$|R_n(t)| \le \frac{|t|^n}{n!}\int_{-\infty}^{\infty} |x|^n \left|e^{it\theta x} - 1\right| dF_X(x).$$

Entonces por el teorema de convergencia dominada se sigue que

$$R_n(t) = o(t^n), \qquad t \to 0.$$

■

Teorema 3.3.3 *Si $\varphi_X^{(2n)}(0)$ existe y es finita, entonces $E\left(X^k\right) < \infty$ para $k \le 2n$.*

Demostración. Observamos que

$$\begin{aligned}\left|\varphi_X^{(2n)}(0)\right| &= \left|\lim_{h\to 0}\int_{-\infty}^{\infty}\left(\frac{e^{ihx} - e^{-ihx}}{2h}\right)^{2n} dF_X(x)\right| \\ &= \lim_{h\to 0}\int_{-\infty}^{\infty}\left(\frac{\operatorname{sen} xh}{xh}\right)^{2n} x^{2n} dF_X(x),\end{aligned}$$

de donde, usando el Lema de Fatou (teorema 2.3.2),

$$\left|\varphi_X^{(2n)}(0)\right| \ge \int_{-\infty}^{\infty} x^{2n} dF_X(x).$$

El resultado se sigue de la desigualdad de Lyapunov. ■

3.4 Teoremas Importantes

3.4.1 Teorema de Inversión

Teorema 3.4.1 *Supongamos que X es una variable aleatoria con función de distribución F_X y característica φ_X. Entonces, para $a < b$,*

$$\lim_{c\to\infty} \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t)\, dt = P\{a < X < b\} + \frac{P\{X = a\} + P\{X = b\}}{2}.$$

Demostración. Sea $c > 0$ y definamos

$$I(c) := \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t)\, dt,$$

entonces

$$\begin{aligned} I(c) &= \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} E\left(e^{itX}\right) dt \\ &= \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} \left(\int_{-\infty}^{\infty} e^{itx} dF_X(x) \right) dt. \end{aligned}$$

Como

$$\begin{aligned} \left| \frac{e^{-ita} - e^{-itb}}{it} e^{itx} \right| &= \left| \frac{e^{-ita} - e^{-itb}}{it} \right| \\ &= \left| \int_a^b e^{-itx} dx \right| \\ &\leq \int_a^b \left| e^{-itx} \right| dx \\ &= b - a \end{aligned}$$

y

$$\int_{-c}^{c} \int_{-\infty}^{\infty} (b - a)\, dF_X(x)\, dt \leq 2c(b - a) < \infty$$

podemos intercambiar el orden de integración para obtener

$$I(c) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-c}^{c} \frac{e^{it(x-a)} - e^{it(x-b)}}{it} dt dF_X(x).$$

Hagamos

$$J_c(x) := \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{it(x-a)} - e^{it(x-b)}}{it} dt,$$

entonces, como

$$e^{it} = \cos t + i \operatorname{sen} t,$$

se tiene

$$J_c(x) = \frac{1}{2\pi} \int_{-c}^{c} \frac{\cos(t(x-a)) - \cos(t(x-b))}{it} dt + \frac{1}{2\pi} \int_{-c}^{c} \frac{\operatorname{sen}(t(x-a)) - \operatorname{sen}(t(x-b))}{t} dt. \tag{3.2}$$

Pero la función coseno es una función par, por lo tanto, el primer sumando de (3.2) se anula, y tenemos que

$$\begin{aligned} J_c(x) &= \frac{1}{2\pi} \int_{-c}^{c} \frac{\operatorname{sen}(t(x-a)) - \operatorname{sen}(t(x-b))}{t} dt \\ &= \frac{1}{2\pi} \int_{-c(x-a)}^{c(x-a)} \frac{\operatorname{sen} t}{t} dt - \frac{1}{2\pi} \int_{-c(x-b)}^{c(x-b)} \frac{\operatorname{sen} t}{t} dt. \end{aligned}$$

Se puede demostrar que

$$g(k,s) = \int_{k}^{s} \frac{\operatorname{sen} t}{t} dt$$

es uniformemente continua en k y s, y que cuando $k \to -\infty$ y $s \to \infty$

$$g(k,s) \to \pi.$$

Entonces, cuando $c \to \infty$,

$$J_c(x) \to \begin{cases} 0 & x < a \text{ ó } x > b \\ \frac{1}{2} & x = a \text{ ó } x = b \\ 1 & a < x < b \end{cases}$$

Ahora, por el teorema de convergencia dominada,

$$\begin{aligned}
\lim_{c\to\infty} I(c) &= \lim_{c\to\infty} E[J_c(X)] \\
&= E\left[\lim_{c\to\infty} J_c(X)\right] \\
&= E\left[I_{]a,b[}(X) + \frac{1}{2} I_{\{a,b\}}(X)\right] \\
&= P\{a < X < b\} + \frac{P\{X=a\} + P\{X=b\}}{2}
\end{aligned}$$

que es lo que se quería demostrar. ∎

Corolario 3.4.2 *En particular, si a, b son puntos de continuidad de F_X,*

$$F_X(b) - F_X(a) = \lim_{c\to\infty} \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t)\, dt.$$

Corolario 3.4.3 (Teorema de Inversión de Fourier) *Sea φ_X la función característica de la variable aleatoria X. Si φ_X es absolutamente integrable en $\mathbb{R}$ entonces X es absolutamente continua. Además, la función de densidad f_X de X está dada por*

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi_X(t)\, dt, \qquad x \in \mathbb{R}.$$

Demostración. Ver Laha y Rohatgi [1971] ∎

Ejemplo 3.4.1 *Consideremos la función característica*[1] $\varphi_X(t) = \exp(-|t|)$. *Claramente φ_X es absolutamente integrable, entonces, por el corolario 3.4.3, X es absolutamente continua y su función de densidad está dada por*

$$\begin{aligned}
f_X(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} e^{-|t|} dt \\
&= \frac{1}{2\pi} \left[\int_{0}^{\infty} e^{-t-itx} dt + \int_{-\infty}^{0} e^{t-itx} dt \right] \\
&= \frac{1}{2\pi} \int_{0}^{\infty} e^{-t} \left(e^{itx} + e^{-itx}\right) dt,
\end{aligned}$$

[1] En el ejemplo 3.5.1, p. 84, se dan argumentos de porque es función característica.

pero

$$\begin{aligned} e^{itx} + e^{-itx} &= \cos tx + i \operatorname{sen} tx + \cos tx - i \operatorname{sen} tx \\ &= 2\cos tx, \end{aligned}$$

entonces

$$f_X(x) = \frac{1}{\pi} \int_0^\infty e^{-t} \cos tx dt.$$

Ahora, por el teorema de integración por partes con $u = \cos tx$ *y* $dv = e^{-t}$,

$$\begin{aligned} f_X(x) &= \frac{1}{\pi} \left[-e^{-t} \cos tx \right]_{t=0}^{\infty} - \frac{x}{\pi} \int_0^\infty e^{-t} \operatorname{sen} tx dt \\ &= \frac{1}{\pi} - \frac{x}{\pi} \int_0^\infty e^{-t} \operatorname{sen} tx dt, \end{aligned}$$

integrando por partes una vez más, ahora con $u = \operatorname{sen} tx$ *y* $dv = e^{-t}$,

$$\begin{aligned} f_X(x) &= \frac{1}{\pi} + \left[-e^{-t} \operatorname{sen} tx \right]_{t=0}^{\infty} - \frac{x^2}{\pi} \int_0^\infty e^{-t} \cos tx dt \\ &= \frac{1}{\pi} - x^2 f_X(x). \end{aligned}$$

Despejando encontramos que

$$f_X(x) = \frac{1}{\pi(1+x^2)}$$

▲

Ejemplo 3.4.2 *Consideremos la función característica*

$$\varphi(t) = \begin{cases} 1 - |t| & \text{si } |t| \leq 1 \\ 0 & \text{en otro caso.} \end{cases}$$

claramente $\varphi(t)$ *es absolutamente integrable, se sigue del teorema 3.4.3 que* $\varphi(t)$ *es la función característica de una variable aleatoria absolutamente continua, digamos* X. *Encontremos la función de densidad de* X

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t) dt$$

$$
\begin{aligned}
&= \frac{1}{2\pi}\int_{-1}^{1} e^{-itx}(1-|t|)\,dt \\
&= \frac{1}{\pi}\left[\int_{0}^{1}(1-t)\cos tx dt\right] \\
&= \frac{1}{\pi}\left[\frac{\operatorname{sen} x}{x} - \int_{0}^{1} t\cos tx dt\right]
\end{aligned}
$$

integrando por partes con $u = t$ y $dv = \cos tx$ se tiene que

$$
\begin{aligned}
f_X(x) &= \frac{1}{\pi}\left[\frac{\operatorname{sen} x}{x} - t\frac{\operatorname{sen} tx}{x}\bigg|_{t=0}^{1} + \int_{0}^{1}\frac{\operatorname{sen} tx}{x}dt\right] \\
&= \frac{1}{\pi}\left[\int_{0}^{1}\frac{\operatorname{sen} tx}{x}dt\right] \\
&= \frac{1}{\pi}\left[-\frac{\cos tx}{x^2}\right]_{t=0}^{1} \\
&= \frac{1-\cos x}{\pi x^2}.
\end{aligned}
$$

Por lo tanto,

$$
f_X(x) = \frac{1-\cos x}{\pi x^2}.
$$

▲

En el ejemplo 3.5.2, se demostrará que la condición de integrabilidad absoluta es suficiente pero no necesaria para que una función característica tenga asociada una distribución absolutamente continua.

3.4.2 Teorema de Unicidad

El teorema de unicidad es uno de los resultados más importantes para funciones características. Aunque aquí veremos el teorema de unicidad como un corolario del teorema de inversión cabe señalar que este puede ser probado sin hacer referencia al teorema de inversión (ver por ejemplo Shiryayev [1984, p. 280], Tucker [1967, p. 51]).

Teorema 3.4.4 (Teorema de Unicidad) *Existe una correspondencia uno a uno entre el conjunto de funciones de distribución y el conjunto de sus funciones características.*

Demostración. Sea F_X la función de distribución de X y F_Y la función de distribución de Y.

Supongamos que $F_X = F_Y$, entonces, de (3.1), se sigue que $\varphi_X(t) = \varphi_Y(t)$.

Por otro lado, para todo b punto de continuidad de F_X y F_Y, el teorema de inversión implica que

$$F_X(b) = \lim_{a\to-\infty}\lim_{c\to\infty}\frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-ita}-e^{-itb}}{it}\varphi_X(t)\,dt$$

y

$$F_Y(b) = \lim_{a\to-\infty}\lim_{c\to\infty}\frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-ita}-e^{-itb}}{it}\varphi_Y(t)\,dt$$

de donde si suponemos que $\varphi_X(t) = \varphi_Y(t)$ para toda $t \in \mathbb{R}$, entonces $F_X(b) = F_Y(b)$. ■

Como una aplicación del teorema de unicidad tenemos el siguiente ejemplo.

Ejemplo 3.4.3 *Supongamos que X es una variable aleatoria con distribución simétrica, entonces, si F_X es la función de distribución asociada a X.*

$$F_X(x) = 1 - F_X(-x).$$

No es difícil ver que esto ocurre si y sólo si $-X$ y X tienen la misma distribución, entonces, por el teorema 3.4.4, $\varphi_X(t) = \varphi_{-X}(t)$, pero, por el teorema 3.2.3,

$$\varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi_X(t)}$$

de donde $\varphi_X(t)$ es real.
Si ahora suponemos que X y Y son variables aleatorias tales que $X = -Y$ y la función característica de X es real tenemos que

$$\varphi_X(t) = \varphi_Y(-t)$$

pero como $\varphi_X(t)$ es real debe ocurrir que $\varphi_Y(t) = \varphi_{-Y}(t)$ por lo tanto, por el teorema 3.4.4, X tiene distribución simétrica. ▲

El ejemplo 3.4.3 demuestra que *la función característica es real si, y sólo si, la función de distribución asociada es simétrica.*

En teoría de variable compleja se demuestra que si dos funciones analíticas coinciden sobre algún intervalo, entonces coinciden en todos lados; por ello parece natural hacer la siguiente pregunta: si φ_1 y φ_2 son funciones características tales que $\varphi_1(t) = \varphi_2(t)$ para todo $t \in [-k, k]$, con $k > 0$, ¿se sigue que φ_1 y φ_2 coinciden en $\mathbb{R}$? En 1937, Gnedenko, mediante un contraejemplo[2] demostró que esto no es cierto. El contraejemplo se pudo construir debido a que no toda función característica es analítica.

Ejemplo 3.4.4 *Consideremos las funciones características*[3] φ_1 *y* φ_2 *definidas como*

$$\varphi_1(t) = \begin{cases} 1 - |t| & si\ |t| \leq \frac{1}{2} \\ \frac{1}{4|t|} & si\ |t| > \frac{1}{2} \end{cases}$$

y

$$\varphi_2(t) = \begin{cases} 1 - |t| & si\ |t| \leq 1 \\ 0 & si\ |t| > 1 \end{cases}$$

claramente φ_1 *y* φ_2 *coinciden en* $\left[-\frac{1}{2}, \frac{1}{2}\right]$ *pero son distintas sobre* $\mathbb{R}$. ▲

Ejemplo 3.4.5 *Consideremos nuevamente la función característica*

$$\varphi_X(t) = \begin{cases} 1 - |t| & si\ |t| \leq 1 \\ 0 & en\ otro\ caso. \end{cases}$$

En el ejemplo 3.4.2 se demostró que $\varphi_X(t)$ *es la función característica asociada a la variable aleatoria absolutamente continua con densidad*

$$f_X(x) = \frac{1 - \cos x}{\pi x^2}.$$

Consideremos ahora la variable aleatoria discreta Y *tal que*

$$P[Y = 0] = \frac{1}{2}, \quad P[Y = (2k-1)\pi] = \frac{1}{\pi^2}\frac{2}{(2k-1)^2}, \quad k = \pm 1, \pm 2, \ldots.$$

Entonces, para $-1 \leq t \leq 1$, *la función característica de* Y *es*[4]

$$\varphi_Y(t) = \frac{1}{2} + \frac{4}{\pi^2}\sum_{k=1}^{\infty}\frac{\cos([2k-1]\pi t)}{(2k-1)^2}$$

[2]El lector interesado puede encontrar el contraejemplo en Gnedenko [1937] o en Stoyanov [1988, p. 59].

[3]En el ejemplo 3.5.2, p. 85, se dan argumentos de porque son funciones características.

[4]En Gradshteyn y Ryzhik [1965, p. 39, 1.444.6] aparece la fórmula
$\sum_{k=1}^{\infty}\frac{\cos([2k-1]x)}{(2k-1)^2} = \frac{\pi}{4}\left(\frac{\pi}{2} - |x|\right), \quad -\pi \leq x \leq \pi$

$$
\begin{aligned}
&= \frac{1}{2}+\frac{4}{\pi^2}\left[\frac{\pi}{4}\left(\frac{\pi}{2}-|\pi t|\right)\right] \\
&= 1-|t|
\end{aligned}
$$

se puede ver que $\varphi_X(t)=\varphi_Y(t)$ *sobre* $[-1,1]$, *sin embargo* $\varphi_X(t)$ *es la función característica de una variable aleatoria absolutamente continua mientras que* $\varphi_Y(t)$ *es la función característica asociada a una variable aleatoria discreta.* ▲

3.4.3 Suma de Variables Aleatorias Independientes

Una de las ventajas del uso de funciones características en probabilidad y estadística es la facilidad con la que se puede encontrar la distribución de sumas de variables aleatorias independientes.

Supongamos que X_1 y X_2 son variables aleatorias independientes con función característica φ_{X_1} y φ_{X_2} respectivamente. La función característica de $S_2 := X_1+X_2$ está dada por

$$\varphi_{S_2}(t)=E\left(e^{it(X_1+X_2)}\right)=E\left(e^{itX_1}e^{itX_2}\right).$$

ya que

$$e^{itX}=\cos tX+i\operatorname{sen} tX,$$

entonces

$$
\begin{aligned}
\varphi_{S_2}(t) &= E\left(e^{itX_1}e^{itX_2}\right) \\
&= E\left\{(\cos tX_1+i\operatorname{sen} tX_1)\cdot(\cos tX_2+i\operatorname{sen} tX_2)\right\} \\
&= E(\cos tX_1\cos tX_2)-E(\operatorname{sen} tX_1\operatorname{sen} tX_2) \\
&\quad +iE(\cos tX_1\operatorname{sen} tX_2)+iE(\operatorname{sen} tX_1\cos tX_2).
\end{aligned}
$$

Es fácil demostrar que la función coseno y la función seno son transformaciones Borel-medibles. Entonces se sigue del teorema 1.5.3 que

$$\cos tX_1\perp\cos tX_2,\quad \operatorname{sen} tX_1\perp\operatorname{sen} tX_2$$

y

$$\operatorname{sen} tX_1\perp\cos tX_2,\quad \cos tX_1\perp\operatorname{sen} tX_2.$$

Ahora, por el teorema 2.4.1

$$
\begin{aligned}
\varphi_{S_2}(t) &= E(\cos tX_1)E(\cos tX_2)-E(\operatorname{sen} tX_1)E(\operatorname{sen} tX_2) \\
&\quad +iE(\cos tX_1)E(\operatorname{sen} tX_2)+iE(\operatorname{sen} tX_1)E(\cos tX_2) \\
&= \{E(\cos tX_1)+iE(\operatorname{sen} tX_1)\}\{E(\cos tX_2)+iE(\operatorname{sen} tX_2)\} \\
&= E\left(e^{itX_1}\right)E\left(e^{itX_2}\right)=\varphi_{X_1}(t)\varphi_{X_2}(t)
\end{aligned}
$$

lo que demuestra el

Teorema 3.4.5 *Sean X_1 y X_2 variables aleatorias independientes con función característica φ_{X_1} y φ_{X_2} respectivamente. Entonces la función característica de $S_2 := X_1 + X_2$ está dada por*

$$\varphi_{S_2}(t) = \varphi_{X_1}(t)\,\varphi_{X_2}(t)$$

El siguiente corolario extiende el teorema 3.4.5 al caso de n variables aleatorias. La demostración se sigue fácilmente por inducción.

Corolario 3.4.6 *Sean $X_1, ..., X_n$ variables aleatorias independientes con función característica $\varphi_{X_1}, ..., \varphi_{X_n}$ respectivamente. Entonces*

$$S_n := X_1 + X_2 + \cdots + X_n$$

tiene función característica

$$\varphi_{S_n}(t) = \prod_{i=1}^{n} \varphi_{X_i}(t).$$

Ejemplo 3.4.6 *Supongamos que $X_1, X_2, ..., X_n$ son variables aleatorias independientes con distribución Geométrica (p), entonces la función característica común es*

$$\varphi_X(t) = \frac{1-p}{1-p\exp(it)}.$$

En este caso, según el teorema 3.4.5, la función característica de $S_n := X_1 + ... + X_n$ es

$$\varphi_{S_n}(t) = \left[\frac{1-p}{1-p\exp(it)}\right]^n;$$

que es la función característica de la distribución Binomial Negativa con parámetro p. Así, por el teorema 3.4.4 se concluye que $S_n \sim BN(p)$. ▲

Ejemplo 3.4.7 *Supongamos que $X_1, X_2, ..., X_n$ son variables aleatorias independientes tales que $X_k \sim$ Poisson (λ_k), para $k = 1, 2..., n$, entonces la función característica de la variable aleatoria X_k es*

$$\varphi_{X_k}(t) = \exp\left(\lambda_k\left(e^{it} - 1\right)\right),$$

de donde $S_n := X_1 + ... + X_n$ tiene función característica

$$\begin{aligned}
\varphi_{S_n}(t) &= \prod_{k=1}^{n} \exp\left(\lambda_k \left(e^{it} - 1\right)\right) \\
&= \exp\left(\left(e^{it} - 1\right) \sum_{k=1}^{n} \lambda_k\right)
\end{aligned}$$

que podemos reconocer como la función característica de una distribución Poisson $(\sum_{k=1}^{n} \lambda_k)$. Por lo tanto, por el teorema 3.4.4, $S_n \sim$ Poisson $(\sum_{k=1}^{n} \lambda_k)$. En particular cuando $\lambda_k = \lambda$ para toda $k \in \{1, 2, ..., n\}$ se tiene que $S_n \sim$ Poisson $(n\lambda)$. ▲

3.4.4 Teorema de Continuidad

Probablemente el más importante de los resultados que hacen de la función característica una herramienta poderosa en las aplicaciones es el teorema de continuidad de Lévy.

Definición 3.4.1 *Supongamos que X_n tiene función de distribución F_n. La sucesión de variables aleatorias $\{X_n\}_{n \geq 1}$* ***converge débilmente (o en distribución)*** *a la variable aleatoria X, y escribimos $X_n \xrightarrow{d} X$ o $F_n \xrightarrow{d} X$, si cuando $n \to \infty$*

$$F_n(x) \to F(x)$$

para cada x punto de continuidad de F.

Teorema 3.4.7 *Sea $\{X_n\}_{n \geq 1}$ una sucesión de variables aleatorias con correspondientes funciones características $\{\varphi_n\}_{n \geq 1}$.* Si *$X_n \xrightarrow{d} X$, entonces* $\lim \varphi_n(t) = \varphi(t)$*; donde $\varphi(t)$ es la función característica de X.*

Demostración. Supongamos que $X_n \xrightarrow{d} X$. Ya que $\cos tX$ y $\operatorname{sen} tX$ son funciones continuas y acotadas, entonces por el teorema de Helly (teorema A.4.1) tenemos que

$$\lim_{n \to \infty} E[\cos tX_n] = E[\cos tX]$$

y

$$\lim_{n \to \infty} E[\operatorname{sen} tX_n] = E[\operatorname{sen} tX]$$

entonces

$$\begin{aligned}
\lim_{n\to\infty} \varphi_n(t) &= \lim_{n\to\infty} E\left[e^{itX_n}\right] \\
&= \lim_{n\to\infty} E\left[\cos tX_n + i \operatorname{sen} tX_n\right] \\
&= \lim_{n\to\infty} E\left[\cos tX_n\right] + i \lim_{n\to\infty} E\left[\operatorname{sen} tX_n\right] \\
&= E\left[\cos tX\right] + iE\left[\operatorname{sen} tX\right] \\
&= E\left[e^{itX}\right] = \varphi(t)
\end{aligned}$$

■

Para la demostración del teorema de continuidad necesitamos demostrar el siguiente

Lema 3.4.8 *Sea X una variable aleatoria en $(\Omega, \mathcal{A}, P)$, con función característica $\varphi(t)$. Entonces, para todo $c > 0$,*

$$P\left\{|X| \geq \frac{2}{c}\right\} \leq \frac{1}{c}\int_{-c}^{c} \{1 - \varphi(t)\}\, dt.$$

Demostración. Sea $c > 0$, entonces

$$\begin{aligned}
\frac{1}{c}\int_{-c}^{c} \{1 - \varphi(t)\}\, dt &= \frac{1}{c}\int_{-c}^{c} E\left[1 - e^{itX}\right] dt \\
&= \frac{1}{c} E\left\{\int_{-c}^{c} \left[1 - e^{itX}\right] dt\right\}.
\end{aligned}$$

Ahora, sí $X \neq 0$

$$\begin{aligned}
\int_{-c}^{c} \left[1 - e^{itX}\right] dt &= \int_{-c}^{c} \left[1 - \cos tX - i \operatorname{sen} tX\right] dt \\
&= 2\int_{0}^{c} \left[1 - \cos tX\right] dt \\
&= 2c\left[1 - \frac{\operatorname{sen} cX}{cX}\right] \\
&\geq 2c\left[1 - \frac{\operatorname{sen} cX}{cX}\right] I_A
\end{aligned}$$

donde $A := \left\{\omega \in \Omega : |X(\omega)| \geq \frac{2}{c}\right\}$.

Se puede demostrar que si $|X| \geq \frac{2}{c}$, entonces $\frac{\operatorname{sen} cX}{cX} \leq \frac{1}{2}$; de donde

$$\left[1 - \frac{\operatorname{sen} cX}{cX}\right] I_A \geq \frac{1}{2} I_A$$

y por lo tanto,

$$\frac{1}{c} \int_{-c}^{c} \{1 - \varphi(t)\}\, dt \geq \frac{1}{c} E(cI_A) = P(A).$$

▲

Teorema 3.4.9 (Teorema de Continuidad de Lévy) *Si $\varphi_n(t)$ converge sobre $]-\infty, \infty[$ a una función $g(t)$ continua en $t = 0$, entonces g es la función característica de alguna variable aleatoria X y $X_n \xrightarrow{d} X$.*

Demostración. Sea $\varepsilon > 0$. Ya que

$$g(0) = \lim_{n \to \infty} \varphi_n(0) = 1$$

y $g(t)$ es continua en $t = 0$, existe un número $c > 0$ tal que

$$|1 - g(t)| < \frac{1}{2}\varepsilon$$

para todo $t \in [-c, c]$.

Como $|\varphi_n(t)| \leq 1$, para todo $t \in \mathbb{R}$, podemos aplicar el teorema de convergencia acotada y obtenemos

$$\int_{-c}^{c} \varphi_n(t)\, dt \to \int_{-c}^{c} g(t)\, dt \tag{3.3}$$

además

$$\begin{aligned} \left|\frac{1}{c}\int_{-c}^{c} [1 - g(t)]\, dt\right| &\leq \frac{1}{c}\int_{-c}^{c} |1 - g(t)|\, dt \\ &< \frac{1}{c}\int_{-c}^{c} \frac{1}{2}\varepsilon dt \\ &= \varepsilon \end{aligned}$$

entonces, por (3.3), para n suficientemente grande

$$\left|\frac{1}{c}\int_{-c}^{c} [1 - \varphi_n(t)]\, dt\right| < \varepsilon.$$

Así, por el lema 3.4.8, si X_n es la variable aleatoria asociada a $\varphi_n(t)$

$$P\left\{|X_n| \geq \frac{2}{c}\right\} \leq \left|\frac{1}{c}\int_{-c}^{c}[1-\varphi_n(t)]\,dt\right| < \varepsilon$$

para n suficientemente grande, y también, para tal n,

$$\begin{aligned} F_n\left(\frac{2}{c}\right) - F_n\left(-\frac{2}{c}\right) &= P\left[-\frac{2}{c} < X_n \leq \frac{2}{c}\right] \\ &\geq P\left[|X_n| < \frac{2}{c}\right] \\ &> 1-\varepsilon. \end{aligned}$$

Por el teorema A.4.2 existe una subsucesión $F_{n_1}, F_{n_2}, \ldots$ y una función de distribución F tal que $F_{n_p} \xrightarrow{d} F$. Sea φ^* la función característica de F. Entonces, según el teorema 3.4.7

$$\varphi_{n_p}(t) \longrightarrow \varphi^*(t)$$

cuando $n \longrightarrow \infty$. Entonces, por hipótesis,

$$\varphi^*(t) = g(t)$$

y por lo tanto $\varphi(t)$, es la función característica de F.

Falta demostrar que $F_n \xrightarrow{d} F$. Para ello supongamos que F_n no converge en distribución a F, es decir que existe al menos un b perteneciente al conjunto de puntos de continuidad de F tal que

$$F_n(b) \stackrel{d}{\nrightarrow} F(b).$$

Ya que la sucesión $\{F_k(b)\}_{k\geq 1}$ es acotada, contiene una subsucesión $F_{m_1}, F_{m_2}, \ldots$ que converge a algún límite $k \neq F(b)$. Entonces por el lema A.4.2 existe una subsucesión $F_{m_{r_1}}, F_{m_{r_2}}, \ldots$ y una función de distribución G tal que

$$F_{m_{r_s}} \xrightarrow{d} G$$

cuando $s \rightarrow \infty$.

Así, si $\psi(t)$ es la función característica de G, el teorema 3.4.7 implica que, para todo $t \in \mathbb{R}$

$$\psi(t) = \lim_{n\to\infty} \varphi_{m_{r_s}}(t) = g(t),$$

entonces, por el teorema de unicidad de las funciones características,

$$G(x) = F(x), \quad \text{para todo } x \in \mathbb{R}$$

así, si $s \to \infty$, $F_{m_{r_s}} \xrightarrow{d} F$. Pero $b \in C_F$, entonces

$$F(b) = \lim_{s\to\infty} F_{m_{r_s}}(b) = \lim_{r\to\infty} F_{m_r}(b) = k \neq F(b)$$

que es una contradicción.

Por lo tanto,

$$F_n \xrightarrow{d} F.$$

■

El siguiente ejemplo aparece en Stoyanov [1988, p.157] y sirve para demostrar que la condición de que g sea continua en cero no se puede eliminar.

Ejemplo 3.4.8 *Consideremos la sucesión de funciones de distribución $\{F_n\}_{n\geq 1}$ donde*

$$F_n(x) = \begin{cases} 0 & si\ x < -n \\ \dfrac{n+x}{2n} & si\ -n \leq x \leq n \\ 1 & si\ x \geq n \end{cases}$$

Es fácil ver que las funciones características de la sucesión $\{F_n\}_{n\geq 1}$ están dadas por

$$\varphi_n(t) = \frac{\operatorname{sen} nt}{nt};$$

de donde $\varphi_n(t)$ converge a la función $\widetilde{\varphi}(t)$ definida como

$$\widetilde{\varphi}(t) = \begin{cases} 1 & si\ t = 0 \\ 0 & si\ t \neq 0. \end{cases}$$

Es claro que la función $\widetilde{\varphi}(t)$ es discontinua en el punto $t = 0$; por lo tanto, en este caso el teorema de continuidad no puede ser aplicado. Además, para toda $x \in \mathbb{R}$.

$$\lim_{n\to\infty} F_n(x) = \frac{1}{2}.$$

Esto implica que $\{F_n\}_{n\geq 1}$ converge a la función constante $\frac{1}{2}$, la cual no es función de distribución. ▲

El siguiente corolario proporciona una caracterización de la convergencia débil.

Corolario 3.4.10 *Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias con correspondientes funciones características $\{\varphi_n\}_{n\geq 1}$ y X una variable aleatoria con función característica φ. Entonces $X_n \xrightarrow{d} X$ si, y sólo si $\lim \varphi_n = \varphi$.*

3.5 Condiciones Necesarias y Suficientes

Los teoremas 3.2.1 y 3.2.2 establecen condiciones necesarias para que una función sea función característica; entonces si una función no cumple alguna de las propiedades que aparecen en estos teoremas entonces no puede ser función característica. El recíproco no es cierto, es decir, podemos encontrar funciones que no son funciones características y, sin embargo, cumplan las propiedades de los teoremas 3.2.1 y 3.2.2. Por ello es deseable contar con condiciones suficientes para que una función sea función característica.

En esta sección presentamos algunas condiciones necesarias y suficientes para poder determinar si una función es característica o no. Las pruebas dependen de resultados auxiliares (que no enunciamos) y por lo tanto no se incluyen. Sin embargo se indica donde pueden encontrarse.

3.5.1 Condiciones Suficientes

Uno de los resultados más conocidos sobre condiciones suficientes es el teorema de Polya.

Teorema 3.5.1 (Teorema de Polya) *Sea $\varphi(t)$ una función continua en $\mathbb{R}$. Para que $\varphi(t)$ sea la función característica, de una variable aleatoria absolutamente continua, es suficiente que cumpla las siguientes condiciones:*

1. $\varphi(0) = 1$.
2. $\varphi(t)$ *es par.*
3. $\varphi(t) \to 0$ *si* $t \to \infty$.
4. $\varphi(t)$ *es convexa en* $]0, \infty[$.

Demostración. Ver Laha y Rohatgi [1971, pp. 168-171]. ■

El teorema de Polya es muy usual en la construcción de funciones características para distribuciones simétricas.

Ejemplo 3.5.1 *Consideremos las funciones*

$$\begin{aligned}
\varphi_1(t) &= \exp\left(-\frac{t^2}{2}\right) \\
\varphi_2(t) &= \exp(-|t|) \\
\varphi_2(t) &= \begin{cases} (1-|t|) & si\ |t| \leq 1 \\ 0 & si\ |t| > 1 \end{cases}
\end{aligned}$$

Cada una de ellas cumple las condiciones del teorema 3.5.1 por lo tanto se sigue que son funciones características de distribuciones absolutamente continuas. ▲

El corolario 3.4.3 establece que para que una función característica φ tenga asociada una variable aleatoria absolutamente continua, es suficiente que φ sea absolutamente integrable, sin embargo, como demuestra el ejemplo 3.5.2 esta no es una condición necesaria. En tal caso el teorema 3.5.1 nos ofrece un criterio más efectivo.

Ejemplo 3.5.2 *Consideremos la funciones*

$$\begin{aligned} \varphi_3(t) &= \frac{1}{1+|t|} \\ \varphi_4(t) &= \begin{cases} (1-|t|) & si\ |t| \leq \frac{1}{2} \\ \frac{1}{4|t|} & si\ |t| > \frac{1}{2}. \end{cases} \end{aligned}$$

Obsérvese que ninguna de φ_3 y φ_4 es absolutamente integrable, aún así, cada una de ellas es la función característica de una variable aleatoria absolutamente continua pues cumplen las condiciones del teorema 3.5.1. ▲

3.5.2 Condiciones Necesarias y Suficientes

Probablemente el resultado más conocido, que establece condiciones necesarias y suficientes para que una función sea función característica, es el teorema de Bochner-Khinchine . Bochner y Khinchine encontraron este resultado al mismo tiempo, pero fue publicado primero por Bochner.

Teorema 3.5.2 (Bochner-Khinchine) *Una condición necesaria y suficiente para que una función continua φ, tal que $\varphi(0) = 1$, sea función característica es que sea positiva definida*[5].

Demostración. En la sección 3.2 se probó la necesidad. La demostración de la suficiencia puede consultarse en Laha y Rohatgi [1971, pp. 161-165] o Lukacs [1970, pp. 71-73]. ■

Otro resultado conocido que proporciona condiciones necesarias y suficientes para que una función sea característica es el Criterio de Cramér. La prueba se puede ver en Laha y Rohatgi [1971, pp. 166-168], Lukacs [1970, pp. 73-75].

Teorema 3.5.3 (Criterio de Cramér) *Una función $\varphi(t)$ continua y acotada es función característica si, y sólo si,*

[5] Ver definición 3.2.2.

1. $\varphi(0) = 1$

2. La función $\psi(x, A) := \int_0^A \int_0^A \varphi(t-u) \exp[ix(t-u)]\, dtdu$ *es real y no-negativa para toda* $x \in \mathbb{R}$ *y toda* $A > 0$.

Teorema 3.5.4 (Marcienkiewicz) *Si una función característica es de la forma* $\exp[\zeta(t)]$, *donde* $\zeta(t)$ *es un polinomio, entonces* $\zeta(t)$ *es, a lo más, de grado dos.*

Demostración. Laha y Rohatgi [1971, pp. 258-259], Lukacs [1970, p. 213] ∎

Ejemplo 3.5.3 *Se sigue del teorema 3.5.4 que* e^{-t^3} *no puede ser función característica.* ▲

3.6 Caso Multivariado

3.6.1 Definición y Propiedades Básicas

Definición 3.6.1 *Si* $\mathbf{X} = (X_1, ..., X_n)$ *es un vector aleatorio* n*-dimensional, entonces para todo vector* $\mathbf{t} = (t_1, ..., t_n) \in \mathbb{R}^n$, *la* ***función característica***, $\varphi_{\mathbf{X}}(\mathbf{t}) = \varphi_{\mathbf{X}}(t_1, ..., t_n)$, *del vector aleatorio* $\mathbf{X}$ *se define como*

$$\varphi_{\mathbf{X}}(\mathbf{t}) := E\left(e^{i\langle \mathbf{t}, \mathbf{X}\rangle}\right);$$

donde

$$\langle \mathbf{t}, \mathbf{X}\rangle := \sum_{i=1}^{n} t_i X_i.$$

Si el vector aleatorio $\mathbf{X} = (X_1, ..., X_n)$ tiene función de distribución $F_{\mathbf{X}}(x_1, ..., x_n) = F_{\mathbf{X}}(\mathbf{x})$, entonces, para todo vector $\mathbf{t} = (t_1, ..., t_n) \in \mathbb{R}^n$, podemos escribir la función característica de $\mathbf{X}$ como

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \int_{\mathbb{R}^n} \exp[i\langle \mathbf{t}, \mathbf{x}\rangle]\, dF_{\mathbf{X}}(\mathbf{x}).$$

Teorema 3.6.1 *Si* $\varphi_{\mathbf{X}}(\cdot)$ *es la función característica del vector aleatorio* n*-dimensional* $\mathbf{X}$, *entonces*

1. $\varphi_{\mathbf{X}}(\mathbf{0}) = 1$, *donde* $\mathbf{0} = (0, ..., 0)$.

2. $|\varphi_{\mathbf{X}}(\mathbf{t})| \leq 1$, *para todo* $\mathbf{t} \in \mathbb{R}^n$.

3. $\varphi_{\mathbf{X}}(\cdot)$ *es uniformemente continua en* $\mathbb{R}^n$.

Demostración.

1. $\varphi_{\mathbf{X}}(0) = \int\limits_{\mathbb{R}^n} \exp[i\langle 0, \mathbf{x}\rangle] \, dF_{\mathbf{X}}(\mathbf{x}) = \int\limits_{\mathbb{R}^n} \exp[0] \, dF_{\mathbf{X}}(\mathbf{x}) = 1.$

2. Sea $\mathbf{t} \in \mathbb{R}^n$entonces

$$\begin{aligned} |\varphi_{\mathbf{X}}(\mathbf{t})| &= \left| \int\limits_{\mathbb{R}^n} \exp[i\langle \mathbf{t}, \mathbf{x}\rangle] \, dF_{\mathbf{X}}(\mathbf{x}) \right| \\ &\leq \int\limits_{\mathbb{R}^n} |\exp[i\langle \mathbf{t}, \mathbf{x}\rangle]| \, dF_{\mathbf{X}}(\mathbf{x}) \\ &= \int\limits_{\mathbb{R}^n} dF_{\mathbf{X}}(\mathbf{x}) = 1. \end{aligned}$$

3. Para cada $\mathbf{t}, \mathbf{h} \in \mathbb{R}^n$ observamos que

$$\begin{aligned} |\varphi_{\mathbf{X}}(\mathbf{t}+\mathbf{h}) - \varphi_{\mathbf{X}}(\mathbf{t})| &= \left| \int\limits_{\mathbb{R}^n} (\exp[i\langle \mathbf{t}+\mathbf{h}, \mathbf{x}\rangle] - \exp[i\langle \mathbf{t}, \mathbf{x}\rangle]) \, dF_{\mathbf{X}}(\mathbf{x}) \right| \\ &\leq \int\limits_{\mathbb{R}^n} |\exp[i\langle \mathbf{h}, \mathbf{x}\rangle] - 1| \, dF_{\mathbf{X}}(\mathbf{x}). \end{aligned}$$

Como, para todo $\mathbf{x} \in \mathbb{R}^n$,

$$|\exp[i\langle \mathbf{h}, \mathbf{x}\rangle] - 1| \leq |\exp[i\langle \mathbf{h}, \mathbf{x}\rangle]| + |-1| \leq 2$$

e independientemente de $\mathbf{t}$

$$|\exp[i\langle \mathbf{h}, \mathbf{x}\rangle] - 1| \to 0$$

cuando $\mathbf{h} \to \mathbf{0}$, se sigue, del teorema de convergencia acotada, que

$$\int\limits_{\mathbb{R}^n} |\exp[i\langle \mathbf{h}, \mathbf{x}\rangle] - 1| \, dF_{\mathbf{X}}(\mathbf{x}) \to 0$$

si $\mathbf{h} \to \mathbf{0}$. Por lo tanto $\varphi_{\mathbf{X}}(\cdot)$ es uniformemente continua en $\mathbb{R}^n$. ∎

Teorema 3.6.2 *Si* $\mathbf{X}$ *es un vector aleatorio* n*-dimensional con función característica* $\varphi_{\mathbf{X}}(\cdot)$, *y* $\mathbf{Y}$ *es un vector aleatorio* n*-dimensional con* n*-ésima componente definida por la relación*

$$Y_i = a_i X_i + b_i, \qquad 1 \leq i \leq n$$

donde $a_i, b_i \in \mathbb{R}$, *entonces la función característica de* $\mathbf{Y}$, *está dada por*

$$\varphi_{\mathbf{Y}}(\mathbf{t}) = \exp\left[i\langle \mathbf{t}, \mathbf{b}\rangle\right] \varphi_{\mathbf{X}}\left(\widetilde{\mathbf{t}}\right);$$

donde

$$\widetilde{\mathbf{t}} = (a_1 t_1, ..., a_n t_n) \quad y \quad \mathbf{b} = (b_1, ..., b_n).$$

Demostración. Sea $\widetilde{\mathbf{t}} = (a_1 t_1, ..., a_n t_n)$ y $\mathbf{b} = (b_1, ..., b_n)$, entonces

$$\begin{aligned}
\varphi_{\mathbf{Y}}(\mathbf{t}) &= E\left(e^{i\langle \mathbf{t}, \mathbf{Y}\rangle}\right) \\
&= E\left(\exp\left[i\sum_{i=1}^{n} t_i Y_i\right]\right) \\
&= E\left(\exp\left[i\sum_{i=1}^{n} t_i a_i X_i + i\sum_{i=1}^{n} t_i b_i\right]\right) \\
&= \exp\left[i\langle \mathbf{t}, \mathbf{b}\rangle\right] E\left(e^{i\langle \widetilde{\mathbf{t}}, \mathbf{X}\rangle}\right) \\
&= \exp\left[i\langle \mathbf{t}, \mathbf{b}\rangle\right] \varphi_{\mathbf{X}}\left(\widetilde{\mathbf{t}}\right).
\end{aligned}$$

■

3.6.2 Función Característica y Momentos

Sea $\mathbf{X} = (X_1, ..., X_n)$ un vector aleatorio n-dimensional y $\varphi_{\mathbf{X}}(\mathbf{t})$ su función característica. Supongamos que $E\left(|X_i|^k\right) < \infty$ para algún $i = 1, 2, ..., n$ y $k \geq 1$. De las desigualdades de Hölder y Lyapunov (teoremas 2.8.6 y 2.8.5) se sigue que los momentos $E\left(X_1^{v_1} \cdots X_n^{v_n}\right)$ existen para todos los enteros no negativos tales que $\sum_{i=1}^{n} v_i \leq k$. Como en el teorema 3.3.1, esto implica la existencia y continuidad de las derivadas parciales

$$\frac{\partial^{v_1+v_2+\cdots+v_n}}{\partial t_1^{v_1} \partial t_2^{v_2} \cdots \partial t_n^{v_n}} \varphi_{\mathbf{X}}(\mathbf{t})$$

para $\sum_{i=1}^{n} v_i \leq k$. Además

$$\left.\frac{\partial^{v_1+v_2+\cdots+v_n}}{\partial t_1^{v_1} \partial t_2^{v_2} \cdots \partial t_n^{v_n}} \varphi_{\mathbf{X}}(\mathbf{t})\right|_{\mathbf{t}=0} = i^{v_1+v_2+\cdots+v_n} E\left(X_1^{v_1} \cdots X_n^{v_n}\right).$$

Entonces el desarrollo en series de $\varphi_{\mathbf{X}}(\mathbf{t})$ es

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \sum_{v_1+v_2+\cdots+v_n \leq k} \frac{i^{v_1+v_2+\cdots+v_n}}{\prod_{i=1}^n v_i!} E\left(\prod_{i=1}^n X_i^{v_i}\right) \prod_{i=1}^n t_i^{v_i} + o(|t|).$$

donde $|t| = |t_1| + |t_2| + \cdots + |t_n|$.

3.6.3 Teorema de Inversión y Unicidad

Teorema 3.6.3 *Sea* $\mathbf{X} = (X_1, ..., X_n)$ *un vector aleatorio* n*-dimensional con función de distribución* $F_{\mathbf{X}}(\cdot)$ *y función característica* $\varphi_{\mathbf{X}}(\cdot)$. *Sean* $\mathbf{a} = (a_1, ..., a_n)$ *y* $\mathbf{b} = (b_1, ..., b_n)$ *dos puntos en* $\mathbb{R}^n$ *tales que, para toda* $j \in \{1, 2, ..., n\}$, $a_j < b_j$ *y* a_j, b_j *son puntos de continuidad de la función de distribución marginal*

$$F_{X_j}(x_j) = F_{\mathbf{X}}\left(\overbrace{+\infty, ..., +\infty}^{j-1}, x, +\infty, ..., +\infty\right), \qquad j = 1, 2, ..., n.$$

Entonces

$$F_{\mathbf{X}}(\mathbf{b}) - F_{\mathbf{X}}(\mathbf{a}) = \frac{1}{\pi^n} \lim_{\substack{T_j \to \infty \\ j=1,2,...,n}} \int_{-T_n}^{T_n} \cdots \int_{-T_1}^{T_1} \prod_{j=1}^n \frac{e^{-it_j a_j} - e^{-it_j b_j}}{it_j} \varphi_{\mathbf{X}}(\mathbf{t})\, dt_1 dt_2 \cdots dt_n.$$

Corolario 3.6.4 *Si* $\varphi_{\mathbf{X}}(\cdot)$ *es absolutamente integrable en* $\mathbb{R}^n$. *Entonces la correspondiente función de distribución es absolutamente continua en* $\mathbb{R}^n$. *Además, la función de densidad*

$$f_{\mathbf{X}} = \frac{\partial^n F}{\partial x_1 \partial x_2 \cdots \partial x_n}$$

está dada por

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \exp\left[i\langle \mathbf{t}, \mathbf{x}\rangle\right] \varphi_{\mathbf{X}}(\mathbf{t})\, d\mathbf{t}$$

donde $\mathbf{t} = (t_1, ..., t_n)$ *y* $\mathbf{x} = (x_1, ..., x_n)$.

Corolario 3.6.5 *La función característica determina de manera única a la función de distribución.*

3.6.4 Función Característica e Independencia

El siguiente teorema proporciona una definición alternativa de independencia de variables aleatorias, la demostración puede consultarse en Loève [1960, p. 227].

Teorema 3.6.6 *Sea* $\mathbf{X} = (X_1, ..., X_n)$ *un vector aleatorio* n*-dimensional con función de distribución* $F_{\mathbf{X}}(\cdot)$ *y función característica* $\varphi_{\mathbf{X}}(\cdot)$*. Una condición necesaria y suficiente para que* $X_1, ..., X_n$ *sean variables aleatorias independientes es que, para cualesquiera* $t_1, ..., t_n \in \mathbb{R}$,

$$\varphi_{\mathbf{X}}(t_1, ..., t_n) = \prod_{j=1}^{n} \varphi_{X_j}(t_j).$$

Ejemplo 3.6.1 *Supongamos que* X_1 *y* X_2 *son variables aleatorias independientes cada una de ellas con distribución* $Normal\,(0,1)$*. Definamos*

$$Y := \frac{X_1 + X_2}{\sqrt{2}} \qquad y \qquad Z := \frac{X_1 - X_2}{\sqrt{2}}.$$

Haciendo uso de los teoremas 3.4.5 y 3.2.3, es fácil ver que la función característica de Y *es*

$$\varphi_Y(t) = e^{-\frac{t^2}{2}}$$

y la función característica de Z *es*

$$\varphi_Z(t) = e^{-\frac{t^2}{2}}.$$

A simple vista no es claro si Y *y* Z *son independientes. En este caso el teorema 3.6.6 puede ser de utilidad. La función característica conjunta de* Y *y* Z *es*

$$\begin{aligned}
\varphi_{Y,Z}(t_1,t_2) &= E\left(\exp\left[it_1\left(\frac{X_1+X_2}{\sqrt{2}}\right) + it_2\left(\frac{X_1-X_2}{\sqrt{2}}\right)\right]\right)\\
&= E\left(\exp\left[iX_1\left(\frac{t_1+t_2}{\sqrt{2}}\right) + iX_2\left(\frac{t_1-t_2}{\sqrt{2}}\right)\right]\right)\\
&= E\left(\exp\left[iX_1\left(\frac{t_1+t_2}{\sqrt{2}}\right)\right]\right) E\left(\exp\left[iX_2\left(\frac{t_1-t_2}{\sqrt{2}}\right)\right]\right)\\
&= \exp\left[-\frac{(t_1+t_2)^2}{4}\right]\exp\left[-\frac{(t_1-t_2)^2}{4}\right]\\
&= \exp\left[-\frac{t_1^2+t_2^2}{2}\right]\\
&= \exp\left[-\frac{t_1^2}{2}\right]\exp\left[-\frac{t_2^2}{2}\right]\\
&= \varphi_Y(t_1)\,\varphi_Z(t_2).
\end{aligned}$$

Por lo tanto, según el teorema 3.6.6, las variables aleatorias Y y Z son independientes. ▲

3.6.5 Otras Propiedades

Ocurre frecuentemente que, a partir de la distribución conjunta de las n variables aleatorias $X_1, ..., X_n$, se requiere conocer la distribución de un subconjunto del conjunto formado por las variables aleatorias $X_1, ..., X_n$. En la sección 1.4 se mencionó un método para dar solución a este problema; sin embargo, este método puede presentar gran dificultad cuando la función de distribución conjunta se desconoce y es difícil de calcular. A continuación presentaremos algunos resultados que pueden usarse para determinar de una manera más sencilla la distribución marginal parcial cuando se conoce la función característica conjunta o es más fácil de calcular que la función de distribución marginal parcial.

Sea $\varphi_{\mathbf{X}}(\mathbf{t}) = \varphi_{\mathbf{X}}(t_1, ..., t_n)$ la función característica del vector aleatorio $\mathbf{X} = (X_1, ..., X_n)$. Entonces

$$\begin{aligned}\varphi_{\mathbf{X}}(t_1, ..., t_{n-1}, 0) &= E\left(\exp\left[i\sum_{k=1}^{n} t_k X_k\right]\right)\\ &= E\left(\exp\left[i\sum_{k=1}^{n-1} t_k X_k\right]\right)\\ &= \varphi_{\mathbf{X}'}(\mathbf{t}');\end{aligned}$$

donde $\mathbf{t}' = (t_1, ..., t_{n-1})$ y $\mathbf{X}' = (X_1, ..., X_{n-1})$. Es decir, si conocemos la función característica del vector $\mathbf{X} = (X_1, ..., X_n)$ podemos conocer la función característica (y por consiguiente cualquier otra cosa) del subvector $\mathbf{X}' = (X_1, ..., X_{n-1})$. Este resultado se generaliza en el siguiente teorema.

Teorema 3.6.7 *Sea $\mathbf{X} = (X_1, ..., X_n)$ un vector aleatorio n-dimensional con función característica $\varphi_{\mathbf{X}}$. Para $0 < j_1 < j_2 < ... < j_k \leq n$, la función característica del vector aleatorio k-dimensional $\mathbf{X} = (X_{j_1}, ..., X_{j_n})$, $k \leq n$, se puede calcular mediante la relación*

$$\varphi_{\mathbf{X}'}(t_{j_1}, t_{j_2}, ..., t_{j_n}) = \varphi_{\mathbf{X}}(t_1, t_2, ..., t_n)\Big|_{j \neq j_1, j_2, ..., j_k}^{t_j = 0}. \tag{3.4}$$

En particular, por medio de 3.4 podemos determinar la función característica de cada una de las componentes del vector.

Corolario 3.6.8 *Si $\varphi_{\mathbf{X}}$ es la función característica del vector aleatorio $\mathbf{X} = (X_1, ..., X_n)$, entonces la función característica de X_m, $1 \leq m \leq n$ está dada por*

$$\varphi_{X_m}(t) = \varphi_{\mathbf{X}}\left(\overbrace{0, ..., 0}^{m-1}, t, 0, ..., 0\right).$$

La función característica conjunta de $\mathbf{X} = (X_1, ..., X_n)$ también puede ser utilizada para encontrar la función característica de una combinación lineal de las componentes del vector $\mathbf{X}$. Para ello empleamos el teorema 3.6.9 y el corolario 3.6.10.

Teorema 3.6.9 *Si $\varphi_{\mathbf{X}}(\mathbf{t}) = \varphi_{\mathbf{X}}(t_1, t_2, ..., t_n)$ es la función característica asociada al vector aleatorio n-dimensional $\mathbf{X} = (X_1, ..., X_n)$, entonces la función característica de $S_n = \sum_{j=1}^{n} X_j$ está dada por*

$$\varphi_{S_n}(t) = \varphi_{\mathbf{X}}(t_1, t_2, ..., t_n)|_{t_j = t} \quad \textit{para toda } j = 1, 2, ..., n.$$

Corolario 3.6.10 *Si $\varphi_{\mathbf{X}}(\mathbf{t}) = \varphi_{\mathbf{X}}(t_1, t_2, ..., t_n)$ es la función característica asociada al vector aleatorio n-dimensional $\mathbf{X} = (X_1, ..., X_n)$, entonces la función característica de $S_n = \sum_{j=1}^{n} t_j X_j$ está dada por*

$$\varphi_{S_n}(t) = \varphi_{\mathbf{X}}(tt_1, tt_2, ..., tt_n).$$

Capítulo 4

Aplicaciones de la Función Característica

4.1 Introducción

La función característica tiene diversas aplicaciones en la teoría de la probabilidad y en soluciones a problemas, de naturaleza teórica, en la estadística matemática. En problemas de la teoría de la probabilidad el uso de funciones características simplifica notablemente la solución de problemas de convergencia, además, el análisis de ciertas familias de distribuciones, como son las distribuciones estables e infinitamente divisibles, se facilita al establecer sus propiedades en términos de funciones características. En estadística es una herramienta valiosa cuando es necesario determinar la distribución de una estadística para la construcción de pruebas de hipótesis o intervalos de confianza o determinar las propiedades de un estimador. También, se pueden proponer métodos de estimación, útiles cuando los métodos tradicionales fallan o no se pueden aplicar. Además, se pueden establecer pruebas de bondad de ajuste basadas en su análogo empírico.

La aplicación de la función característica en problemas de distribución, se ilustra en la sección 4.2, en la sección 4.3 se muestra como se puede usar la función característica en problemas de convergencia estocástica. Las secciones 4.4 y 4.5 presentan las distribuciones infinitamente divisibles y las distribuciones estables; en estas secciones se pueden encontrar claros ejemplos de que muchas propiedades se pueden establecer y probar más sencillamente en términos de funciones características. En la sección 4.6 se introduce la función característica empírica y algunas de sus propiedades. Algunos de los métodos de estimación y de bondad de ajuste basados en la función característica empírica se describen en la sección 4.7 y 4.8 respectivamente.

4.2 Distribución de Estadísticas

En estadística uno de los problemas fundamentales es determinar la distribución de ciertas estadísticas de prueba o estimadores. En esta sección se mostrará como las funciones características pueden ser de utilidad en la solución de problemas de esta naturaleza.

4.2.1 Distribución de la Media Muestral

Definición 4.2.1 *Se dice que las variables aleatorias $X_1, X_2, ..., X_n$ forman una* ***muestra aleatoria*** *de una población con distribución $F(\cdot)$ si $X_1, X_2, ..., X_n$ son variables aleatorias independientes e idénticamente distribuidas con función de distribución común $F(\cdot)$.*

Frecuentemente ante problemas de tipo inferencial resulta necesario conocer la distribución de la **media muestral**

$$\overline{X}_n := \frac{1}{n}\sum_{j=1}^{n} X_j.$$

Ya que $\overline{X}_n$ es una suma de variables aleatorias independientes, se puede calcular la función característica $\varphi_{\overline{X}_n}$ de $\overline{X}_n$ de manera sencilla aplicando los resultados de la sección 3.4.3.

Como $X_1, X_2, ..., X_n$ son variables aleatorias independientes e idénticamente distribuidas tenemos que, si φ_X es la función característica común a $X_1, X_2, ..., X_n$, entonces $S_n := \sum_{j=1}^{n} X_j$ tiene función característica

$$\varphi_{S_n}(t) = [\varphi_X(t)]^n$$

entonces por el teorema 3.2.3

$$\varphi_{\overline{X}_n}(t) = \varphi_{S_n}\left(\frac{t}{n}\right) = \left[\varphi_X\left(\frac{t}{n}\right)\right]^n \tag{4.1}$$

Ejemplo 4.2.1 *Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una distribución $Normal(\mu, \sigma^2)$. En el ejemplo 3.2.10 se demostró que*

$$\varphi_X(t) = \exp\left(i\mu t - \frac{\sigma^2 t^2}{2}\right).$$

Por lo tanto, aplicando la ecuación (4.1), *tenemos que*

$$\varphi_{\overline{X}_n}(t) = \exp\left(i\mu t - \frac{\sigma^2 t^2}{2n}\right)$$

que reconocemos como la función característica de una variable aleatoria Normal $\left(\mu, \frac{\sigma^2}{n}\right)$. *Así, por el teorema 3.4.4, concluimos que* $\overline{X}_n \sim Normal\left(\mu, \frac{\sigma^2}{n}\right)$. ▲

Ejemplo 4.2.2 *Consideremos ahora el problema de encontrar la distribución de* $\overline{X}_n$ *calculada a partir de una muestra aleatoria de tamaño n de una población con distribución* $Cauchy\,(\alpha, \beta)$. *En este caso la función característica común de los elementos de la muestra es*

$$\varphi_X(t) = \exp\left(i\alpha t - \beta\,|t|\right),$$

entonces claramente la función característica de $\overline{X}_n$ *es*

$$\varphi_{\overline{X}_n}(t) = \exp\left(i\alpha t - \beta\,|t|\right).$$

Por lo tanto, $\overline{X}_n \sim Cauchy\,(\alpha, \beta)$. ▲

4.2.2 Cantidades Pivotales

Uno de los métodos más usados para la construcción de intervalos de confianza es el método de la cantidad pivotal (ver Mood, et. al. [1974, cap. 8]).

Definición 4.2.2 *Sea* $X_1, X_2, ..., X_n$ *una muestra aleatoria de la distribución* $F(\cdot;\theta)$. *Supongamos que* Q *es una función de* $X_1, X_2, ..., X_n$ *y de* θ. *Si la distribución de* Q *no depende de* θ, *entonces se dice que* Q *es una* ***cantidad pivotal***.

En muchos casos, con ayuda del teorema 3.2.3, podemos emplear funciones características para determinar si alguna función es cantidad pivotal.

Ejemplo 4.2.3 *Sea* $X_1, X_2, ..., X_n$ *una muestra aleatoria de una población* $Normal\,(\mu, 3)$, *en este caso*

$$\varphi_{\overline{X}_n}(t) = \exp\left(i\mu t - \frac{3}{2n}t^2\right),$$

entonces, por el teorema 3.2.3, tenemos que la función característica de $Q = \overline{X}_n - \mu$ *es*

$$\begin{aligned} \varphi_Q(t) &= \exp(-i\mu t)\exp\left(i\mu t - \frac{3}{2n}t^2\right) \\ &= \exp\left(-\frac{3}{2n}t^2\right). \end{aligned}$$

Por lo tanto $Q \sim Normal\left(0, \frac{3}{n}\right)$. *Así, hemos demostrado que la distribución de* Q *no depende de* μ *y por lo tanto* Q *es una cantidad pivotal.* ▲

Ejemplo 4.2.4 *Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una población Cauchy $(0, \beta)$. En este caso la función característica de $\overline{X}_n$ es*

$$\varphi_{\overline{X}_n}(t) = \exp(-\beta |t|).$$

Si definimos $Q := \frac{\overline{X}_n}{\beta}$ entonces, por el teorema 3.2.3 tenemos que la función característica de Q es

$$\begin{aligned}\varphi_Q(t) &= \varphi_{\overline{X}_n}\left(\frac{t}{\beta}\right) \\ &= \exp(-|t|)\end{aligned}$$

y se sigue que $Q \sim$ Cauchy $(0, 1)$ y por lo tanto Q es una cantidad pivotal. ▲

4.2.3 Muestreo de la Distribución Normal

En la inferencia estadística la distribución Normal es de especial importancia, por ello a continuación ilustraremos la aplicación de la función característica en la solución de problemas de distribución de estadísticas cuando la muestra proviene de una distribución Normal.

Teorema 4.2.1 *Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una distribución Normal $(0, 1)$. La estadística*

$$\chi^2 := \sum_{j=1}^{n} X_j^2$$

tiene distribución $\chi^2_{(n)}$.

Demostración. Por el teorema 1.5.3 tenemos que $X_1^2, X_2^2, ..., X_n^2$ son variables aleatorias independientes. En el ejemplo 3.2.11 se demostró que la función característica de cada X_j^2 es

$$\varphi(t) = (1 - 2it)^{-\frac{1}{2}}.$$

Entonces, por el teorema 3.4.5, se tiene que

$$\varphi_{\chi^2}(t) = [\varphi(t)]^n = (1 - 2it)^{-\frac{n}{2}};$$

esta es la función característica asociada a la distribución $\chi^2_{(n)}$. Por lo tanto, según el teorema de unicidad de las funciones características $\chi^2 \sim \chi^2_{(n)}$. ■

Definición 4.2.3 *Sea $X_1, X_2, ..., X_n$ una muestra aleatoria. de tamaño n, de una distribución $F(\cdot)$. La estadística*

$$S^2 := \frac{1}{n}\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2; \qquad n > 1$$

es llamada ***varianza muestral.***

No es difícil demostrar que

$$S^2 = \frac{1}{n}\sum_{j=1}^{n} X_j^2 - \overline{X}^2$$

Teorema 4.2.2 *Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una distribución Normal (μ, σ^2). Entonces la media muestral, $\overline{X}$, y la varianza muestral, S^2, son independientes. Además, $\frac{nS^2}{\sigma^2}$ tiene distribución $\chi^2_{(n-1)}$.*

Demostración. La función característica conjunta de $\overline{X}$ y S^2 es

$$\varphi(t_1, t_2) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left[it_1\overline{X} + \frac{it_2}{n}\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2 - \frac{1}{2\sigma^2}\sum_{j=1}^{n}\left(X_j - \mu\right)^2\right] dx_1 \cdots dx_n$$

hagamos la transformación[1]

$$\begin{cases} Y_1 = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} X_j \\ Y_k = \frac{1}{\sqrt{k(k-1)}}\left[\sum\limits_{j=1}^{k-1} X_j - (k-1)X_k\right] & \text{si } k > 1 \end{cases}$$

puede demostrarse (ver Harris [1966, pp. 184-185]) que esta es una transformación ortogonal, es decir,

$$\sum_{j=1}^{n} X_j^2 = \sum_{j=1}^{n} Y_j^2.$$

Por lo tanto, el valor absoluto del Jacobiano de la transformación es $|J| = 1$. Además, se tiene que

$$\frac{1}{n}\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2 \;=\; \frac{1}{n}\sum_{j=1}^{n} X_j^2 - \overline{X}^2$$

[1] Esta transformación es conocida como la transformación de Helmert.

$$
\begin{aligned}
&= \frac{1}{n}\sum_{j=1}^{n} Y_j^2 - \overline{X}^2 \\
&= \frac{1}{n}\sum_{j=2}^{n} Y_j^2 + \overline{X}^2 - \overline{X}^2 \\
&= \frac{1}{n}\sum_{j=2}^{n} Y_j^2
\end{aligned}
$$

y

$$
\begin{aligned}
\sum_{j=1}^{n}(X_j-\mu)^2 &= \sum_{j=1}^{n} X_j^2 - 2\mu\sum_{j=1}^{n} X_j + n\mu^2 \\
&= \sum_{j=1}^{n} Y_j^2 - 2\mu n\overline{X} + n\mu^2 \\
&= \sum_{j=2}^{n} Y_j^2 + n\overline{X}^2 + 2\mu n\overline{X} + n\mu^2 \\
&= \sum_{j=2}^{n} Y_j^2 + \left(\sqrt{n}\,\overline{X} + \sqrt{n}\mu\right)^2 \\
&= \sum_{j=2}^{n} Y_j^2 + \left(Y_1 + \sqrt{n}\mu\right)^2
\end{aligned}
$$

de donde

$$
\begin{aligned}
\varphi(t_1,t_2) &= (2\pi\sigma^2)^{-\frac{n}{2}} \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \left[\exp\left[\frac{it_1y_1}{\sqrt{n}} + \frac{it_2}{n}\sum_{j=2}^{n} y_j^2\right]\right. \\
&\quad \left.\times \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{j=2}^{n} y_j^2 + \left(y_1+\sqrt{n}\mu\right)^2\right)\right]\right] dy_1\cdots dy_n \\
&= \left\{\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\exp\left[\frac{it_1y_1}{\sqrt{n}} - \frac{1}{2\sigma^2}\left(y_1+\sqrt{n}\mu\right)^2\right]dy_1\right\} \\
&\quad \times\left\{\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\exp\left[\frac{it_2}{n}y^2 - \frac{1}{2\sigma^2}y^2\right]dy\right\}^{n-1} \\
&= \exp\left[it_1\mu - \frac{\sigma^2t_1^2}{2n}\right]\left(1-2it_2\frac{\sigma^2}{n}\right)^{-\frac{n-1}{2}} .
\end{aligned}
$$

Entonces, por el teorema 3.6.7,

$$\varphi(t_1, 0) = \exp\left[it_1\mu - \frac{\sigma^2 t_1^2}{2n}\right]$$

$$\varphi(0, t_2) = \left(1 - 2it_2\frac{\sigma^2}{n}\right)^{-\frac{n-1}{2}}$$

y se sigue que

$$\overline{X} \sim Normal\left(\mu, \frac{\sigma^2}{n}\right) \qquad \text{y} \qquad S^2 \sim \chi^2_{(n-1)}.$$

Finalmente observamos que

$$\varphi(t_1, t_2) = \varphi(t_1, 0) \cdot \varphi(0, t_2),$$

entonces, según el teorema 3.6.6, se concluye que $\overline{X}$ y S^2 son independientes. ∎

Teorema 4.2.3 *Sean X_1, X_2 variables aleatorias independientes, donde $X_j \sim Gamma(\lambda_j, \theta)$; $j = 1, 2$. La densidad de $W = \frac{X_1}{X_2}$ es*

$$f_W(x) = \frac{1}{B(\lambda_1, \lambda_2)} x^{\lambda_1 - 1}(1 + x)^{-(\lambda_1+\lambda_2)} I_{(0,\infty)}(x),$$

donde $B(\cdot, \cdot)$ denota la función Beta.

Demostración. Sea $Y_1 = \ln X_1$, entonces, por el teorema 3.2.4, la función característica de Y_1 es

$$\begin{aligned}\varphi_{Y_1}(t) &= \int_0^\infty \frac{1}{\Gamma(\lambda_1)} \theta^{\lambda_1} e^{-\theta x} x^{\lambda_1 + it - 1} dx \\ &= \theta^{-it} \frac{\Gamma(\lambda_1 + it)}{\Gamma(\lambda_1)}.\end{aligned}$$

De la misma forma para $Y_2 = \ln X_2$

$$\begin{aligned}\varphi_{Y_2}(t) &= \int_0^\infty \frac{1}{\Gamma(\lambda_2)} \theta^{\lambda_2} e^{-\theta x} x^{\lambda_2 + it - 1} dx \\ &= \theta^{-it} \frac{\Gamma(\lambda_2 + it)}{\Gamma(\lambda_2)}.\end{aligned}$$

Entonces, por el teorema 3.4.5, tenemos que la función característica de $Y = \ln W$ es

$$\varphi_Y(t) = \frac{\Gamma(\lambda_1 + it)\,\Gamma(\lambda_2 - it)}{\Gamma(\lambda_1)\,\Gamma(\lambda_2)}.$$

Ahora, según el teorema 3.4.3, la densidad de Y está dada por

$$\begin{aligned} f_Y(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\varphi_Y(t)\,dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\frac{\Gamma(\lambda_1 + it)\,\Gamma(\lambda_2 - it)}{\Gamma(\lambda_1)\,\Gamma(\lambda_2)}dt; \end{aligned}$$

tomando $-s = \lambda_2 - it$ tenemos

$$f_Y(x) = \frac{1}{2\pi i}\int_{\lambda_2 - i\infty}^{\lambda_2 + i\infty} e^{-x(s+\lambda_2)}\frac{\Gamma(\lambda_1 + \lambda_2 + s)\,\Gamma(-s)}{\Gamma(\lambda_1)\,\Gamma(\lambda_2)}ds,$$

entonces la densidad de $W = e^Y$ es

$$f_W(x) = \frac{x^{-\lambda_2 - 1}}{\Gamma(\lambda_1)\,\Gamma(\lambda_2)}\frac{1}{2\pi i}\int_{\lambda_2 - i\infty}^{\lambda_2 + i\infty} x^{-s}\Gamma(\lambda_1 + \lambda_2 + s)\,\Gamma(-s)\,ds$$

pero, para $\operatorname{Re}(1-\beta) < \gamma < 0$ y $|\arg t| < \pi$, se tiene que[2]

$$\int_{\lambda_2 - i\infty}^{\lambda_2 + i\infty} t^{-s}\Gamma(\beta + s)\,\Gamma(-s)\,ds = 2\pi i\Gamma(\beta)\,(1+t)^{-\beta}. \tag{4.2}$$

Por lo tanto,

$$\begin{aligned} f_W(x) &= \frac{\Gamma(\lambda_1 + \lambda_2)}{\Gamma(\lambda_1)\,\Gamma(\lambda_2)}x^{\lambda_1 - 1}(1+x)\,I_{(0,\infty)}(x) \\ &= \frac{1}{B(\lambda_1, \lambda_2)}x^{\lambda_1 - 1}(1+x)^{-(\lambda_1 + \lambda_2)}\,I_{(0,\infty)}(x). \end{aligned}$$

■

Del teorema 4.2.3 se desprenden los siguientes corolarios

[2] La integral 4.2 se encuentra en Gradshteyn and Ryzhik [1965, p. 657, 6.422-3]

Corolario 4.2.4 *Sean χ_1^2 y χ_2^2 variables aleatorias independientes tales que*

$$\chi_k^2 \sim Ji - Cuadrada\,(n_k)$$

para $k = 1, 2$. Entonces la función de densidad de $F = \frac{\chi_1^2/n_1}{\chi_2^2/n_2}$ es[3]

$$f_F(x) = \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} x^{\frac{n_1}{2}-1}\left(1 + x\frac{n_1}{n_2}\right)^{-\left(\frac{n_1+n_2}{2}\right)} I_{(0,\infty)}(x).$$

Corolario 4.2.5 *Sean $X \sim Normal\,(0, 1)$ y $Y \sim \chi_{(n)}^2$ variables aleatorias independientes. La variable aleatoria $T = \frac{X}{\sqrt{Y/n}}$ tiene función de densidad*

$$f_T(x) = \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}}\left(1 + \frac{x^2}{n}\right)^{-\left(\frac{n+1}{2}\right)}.$$

Como una aplicación de los corolarios anteriores tenemos los siguientes teoremas.

Teorema 4.2.6 *Sea S_1^2 y S_2^2 las varianzas muestrales de dos muestras independientes, de tamaño n_1 y n_2 respectivamente, tomadas de poblaciones Normales con la misma varianza. Entonces*

$$F = \frac{n_1 S_1^2/(n_1 - 1)}{n_2 S_2^2/(n_2 - 1)}$$

tiene distribución $F(n_1 - 1, n_2 - 1)$.

Teorema 4.2.7 *Sea $X_1, X_2, ..., X_n$ muestra aleatoria de una distribución $Normal\,(\mu, \sigma^2)$. Entonces*

$$T = \sqrt{n-1}\,\frac{\overline{X} - \mu}{S}$$

tiene distribución $t - Student\,(n - 1)$.

[3] Se dice que una variable aleatoria Y con función de densidad dada por 4.2.4 tiene distribución F con n_1, n_2 grados de libertad y se denota por $Y \sim F(n_1, n_2)$.

4.3 Convergencia Estocástica

Muchos resultados importantes de la teoría de la probabilidad son formulados como teoremas límites. Los primeros teoremas límites en probabilidad fueron el teorema de De Moivre-Laplace y el teorema de Poisson bajo el esquema de Bernoulli. Estos teoremas se probaron mediante un análisis directo de las funciones límites, las cuales se pueden expresar de manera sencilla en términos de probabilidades binomiales. Sin embargo, puede resultar difícil, o hasta imposible, analizar sistemas de variables aleatorias más complicadas mediante este procedimiento.

La primera idea para probar teoremas límites para sumas de variables aleatorias con distribución arbitraria fue dada por Chebyshev. La desigualdad que el descubrió y que en la actualidad lleva su nombre, facilita la demostración de algunos teoremas límite, como la ley de los grandes números de Bernoulli, y establece condiciones muy generales bajo las cuales se cumplen.

Un tiempo después, Lyapunov propuso un método alternativo para probar el teorema del límite central, basado en la función característica. Desarrollos posteriores demostraron que el método de las funciones características de Lyapunov es muy efectivo para probar diversos teoremas límites.

Basándose en la correspondencia uno a uno que existe entre las funciones características y las funciones de distribución, el método de las funciones características de Lyapunov, propone estudiar las propiedades de la función de distribución mediante su función característica. Es el teorema de continuidad de Lévy (teorema 3.4.9), el que hace posible el análisis del comportamiento límite de una función de distribución mediante el análisis de su función característica.

4.3.1 Ley Débil de los Grandes Números

Para nuestro estudio de las leyes de los grandes números será necesario introducir otro tipo de convergencia, la convergencia en probabilidad.

Definición 4.3.1 *Decimos que la sucesión de variables aleatorias $\{X_n\}_{n\geq 1}$ converge en probabilidad a la variable aleatoria X, y lo denotamos como $X_n \xrightarrow{P} X$, si*

$$\lim_{n\to\infty} P\left[|X_n - X| \geq \varepsilon\right] = 0$$

para toda $\varepsilon > 0$.

Puede demostrarse (ver apéndice A) que si $X_n \xrightarrow{P} X$, entonces $X_n \xrightarrow{d} X$. El recíproco no es cierto en general, sin embargo si $X_n \xrightarrow{d} c$, donde c es una constante, es cierto que $X_n \xrightarrow{P} c$.

Para demostrar lo anterior supongamos que $X_n \xrightarrow{d} c$ donde c es una constante. La función de distribución F de la variable aleatoria constante c es

$$F(c) = \begin{cases} 0 & \text{si } x < c \\ 1 & \text{si } x \geq c \end{cases}$$

claramente F es continua en todos los puntos diferentes de c. Sea $\varepsilon > 0$ y F_n la función de distribución de X_n entonces

$$\begin{aligned} P[|X_n - c| \geq \varepsilon] &= P[X_n - c \geq \varepsilon] + P[X_n - c \leq -\varepsilon] \\ &= P[X_n \geq c + \varepsilon] + P[X_n \leq c - \varepsilon] \\ &\leq P\left[X_n > c + \frac{1}{2}\varepsilon\right] + P[X_n \leq c - \varepsilon] \\ &= 1 - F_n\left(c + \frac{1}{2}\varepsilon\right) + F_n(c - \varepsilon) \end{aligned}$$

entonces

$$\lim_{n\to\infty} P[|X_n - c| \geq \varepsilon] \leq 1 - F\left(c + \frac{1}{2}\varepsilon\right) + F(c - \varepsilon) = 0.$$

Por lo tanto, según la definición 4.3.1, $X_n \xrightarrow{P} c$. Así, hemos demostrado el

Teorema 4.3.1 *Si $X_n \xrightarrow{d} c$ donde c es una constante, entonces $X_n \xrightarrow{P} c$.*

Definición 4.3.2 *Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias tales que $E(X_n) < \infty$ para todo entero positivo n. Decimos que $\{X_n\}_{n\geq 1}$ obedece **la ley débil de los grandes números** si*

$$\frac{S_n}{n} \xrightarrow{P} \frac{E(S_n)}{n}$$

donde $S_n := \sum_{i=1}^{n} X_i$.

Teorema 4.3.2 *Supongamos que $X_1, X_2, \ldots$ son observaciones independientes de la variable aleatoria X. Si $E(X) = \mu < \infty$, entonces $\overline{X}_n \xrightarrow{P} \mu$. Es decir, $X_1, X_2, \ldots$ obedece la ley débil de los grandes números.*

Demostración. Sea φ_X la función característica de X, entonces, para $t \in \mathbb{R}$, la función característica de $\overline{X}_n$ es

$$\varphi_{\overline{X}_n}(t) = \left[\varphi_X\left(\frac{t}{n}\right)\right]^n$$

pero según el corolario 3.3.2

$$\varphi_X(t) = 1 + it\left(\frac{\mu}{n} + o(1)\right)$$

entonces, para todo $t \in \mathbb{R}$,

$$\varphi_{\overline{X}}(t) = \left[1 + \frac{it}{n}(\mu + \varepsilon_n)\right]^n;$$

donde $\varepsilon_n \to 0$ cuando $n \to \infty$. Por lo tanto, si $n \to \infty$

$$\varphi_{\overline{X}_n}(t) \to e^{i\mu t}.$$

Además, por el ejemplo 3.2.1 sabemos que $e^{i\mu t}$ es la función característica de la variable aleatoria constante μ, entonces, por el corolario 3.4.10, $\overline{X}_n \xrightarrow{d} \mu$ cuando $n \to \infty$. Finalmente, por el teorema 4.3.1, se tiene que

$$\overline{X}_n \xrightarrow{P} \mu$$

cuando $n \to \infty$. ■

4.3.2 Teorema del Límite Central

Existen diferentes versiones del teorema de límite central; la más sencilla de éstas versiones es el teorema del límite central de Lévy que involucra sumas de variables aleatorias independientes e idénticamente distribuidas. A continuación, se presenta y se demuestra, por el método de las funciones características, este caso particular. Versiones más fuertes pueden encontrarse en Serfling [1980, p. 28-35] o en Chung [1968, cap. 7]. En Billingsley [1968, p. 42] y Feller [1971, sec. 8.4] se pueden encontrar demostraciones sin el uso de funciones características.

Teorema 4.3.3 (Teorema del Límite Central de Lévy) *Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas (no degeneradas) tales que $E(X_1) = \mu$ y $Var(X_1) = \sigma^2 < \infty$. Sea $S_n := X_1 + X_2 + \cdots + X_n$, entonces, cuando $n \to \infty$,*

$$\frac{S_n - E(S_n)}{\sqrt{Var(S_n)}} \xrightarrow{d} \Phi(x);$$

donde

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$$

Demostración. Para $n = 1, 2, \ldots$, denotemos por φ_n la función característica de

$$\frac{S_n - E(S_n)}{\sqrt{Var(S_n)}}$$

y por $\varphi(t)$ la función característica de $X_1 - \mu$, entonces, ya que $X_1 - \mu, X_2 - \mu, ..., X_n - \mu$ son variables aleatorias independientes, la función característica de $S_n - E(S_n)$ es $[\varphi(t)]^n$ y por el teorema 3.2.3

$$\varphi_n(t) = \left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n.$$

Pero, por el teorema 3.3.2,

$$\varphi(t) = 1 - \frac{1}{2}\sigma^2 t^2 (1 + o(1)); \qquad t \to \infty.$$

entonces,

$$\varphi_n(t) = \left[1 - \frac{1}{2\sigma^2 n}\sigma^2 t^2 \left(1 + o\left(\frac{1}{n}\right)\right)\right]^n \to e^{-\frac{t^2}{2}}$$

cuando $n \to \infty$.

Por lo tanto, según el teorema de continuidad ,

$$\frac{S_n - E(S_n)}{\sqrt{Var(S_n)}} \xrightarrow{d} Normal(0,1).$$

■

4.3.3 Teorema de Poisson

Teorema 4.3.4 *Para cada $n \geq 1$ supongamos que $X_{n,1}, X_{n,2}, ..., X_{n,n}$ son variables aleatorias independientes tales que*

$$\begin{cases} p_{nk} & si\ X_{n,k} = 1 \\ q_{nk} & si\ X_{n,k} = 0 \\ 0 & en\ otro\ caso \end{cases}$$

y $S_n = \sum_{k=1}^{n} X_{n,k}$. Si, cuando $n \to \infty$,

$$\max_{1 \leq k \leq n} p_{nk} \to 0 \qquad y \qquad \sum_{k=1}^{n} p_{nk} \to \lambda$$

entonces, para cada $m = 0, 1, ...,$

$$P[S_n = m] \to \frac{e^{-\lambda}\lambda^m}{m!}.$$

Demostración. Para $1 \leq k \leq n$, la función característica de $X_{n,k}$ es

$$\varphi_{X_{n,k}}(t) = p_{nk}e^{it} + q_{nk},$$

entonces

$$\begin{aligned}
\varphi_{S_n}(t) &= \prod_{k=1}^{n} \left[p_{nk}e^{it} + q_{nk}\right] \\
&= \prod_{k=1}^{n} \left[1 + p_{nk}\left(e^{it} - 1\right)\right],
\end{aligned}$$

de donde, si $n \to \infty$,

$$\varphi_{S_n}(t) \to \exp\left[\lambda\left(e^{it} - 1\right)\right].$$

Como $\exp\left[\lambda\left(e^{it} - 1\right)\right]$ es la función característica de la distribución $Poisson\,(\lambda)$, entonces, por el teorema 3.4.9, tenemos que

$$S_n \xrightarrow{d} Poisson\,(\lambda).$$

■

4.4 Distribuciones Infinitamente Divisibles

Definición 4.4.1 *Una función de distribución F se llama* ***infinitamente divisible*** *si existe un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ tal que para cada entero positivo n existen n variables aleatorias independientes e idénticamente distribuidas, $X_{n,1}, X_{n,2}, \ldots, X_{n,n}$, tales que $S_n := \sum_{k=1}^{n} X_{n,k}$ tiene función de distribución F.*

Claramente podemos establecer una definición alternativa en términos de funciones características como sigue.

Definición 4.4.2 *La función de distribución F es* ***infinitamente divisible*** *si para todo n entero positivo su función característica puede expresarse como la n-ésima potencia de alguna otra función característica. Si F es infinitamente divisible diremos que la correspondiente* ***función característica es infinitamente divisible****.*

Ejemplo 4.4.1 *Sea φ la función característica de una distribución de $Poisson\,(\lambda)$, es decir*

$$\varphi(t) = \exp\left[\lambda\left(e^{it} - 1\right)\right],$$

entonces la distribución de Poisson es infinitamente divisible pues

$$\varphi_n(t) = \exp\left[\frac{\lambda}{n}\left(e^{it} - 1\right)\right]$$

es la función característica de la distribución Poisson $\left(\frac{\lambda}{n}\right)$. ▲

Ejemplo 4.4.2 *Para la distribución Normal* (μ, σ^2) *la función característica es*

$$\varphi(t) = \exp\left[it\mu - \frac{\sigma^2 t^2}{2}\right].$$

Como

$$\varphi_n(t) = \exp\left[it\frac{\mu}{n} - \frac{\sigma^2 t^2}{2n}\right]$$

es la función característica de la distribución Normal $\left(\frac{\mu}{n}, \frac{\sigma^2}{n}\right)$, *entonces la distribución Normal es infinitamente divisible.* ▲

Ejemplo 4.4.3 *La función característica de la distribución Cauchy* $\left(\frac{\alpha}{n}, \frac{\beta}{n}\right)$ *es*

$$\varphi_n(t) = \exp\left[it\frac{\alpha}{n} - \frac{\beta}{n}|t|\right]$$

ya que

$$[\varphi_n(t)]^n = \exp\left[it\alpha - \beta|t|\right]$$

es la función característica de la distribución Cauchy (α, β) *se concluye que la distribución Cauchy* (α, β) *es infinitamente divisible.* ▲

Ejemplo 4.4.4 *Para la distribución Gamma* (α, β) *la función característica está dada por*

$$\varphi(t) = \left[\frac{\beta}{\beta - it}\right]^{\alpha},$$

entonces, tomando

$$\varphi_n(t) = \left[\frac{\beta}{\beta - it}\right]^{\frac{\alpha}{n}}$$

es claro que la distribución Gamma (α, β) *es infinitamente divisible.* ▲

Teorema 4.4.1 *Si F es infinitamente divisible también lo es $G = 1 - F$.*

Demostración. Sea φ la función característica de F y ψ la función característica de G. entonces claramente φ y ψ se relacionan mediante la relación $\psi(t) = \overline{\varphi}(t)$. Si $\varphi_n(t)$ es una función característica tal que $\varphi(-t) = [\varphi_n(t)]^n$, se tiene que $\psi(t) = [\overline{\varphi}_n(t)]^n$. Ya que $\overline{\varphi}_n$ es una función característica se sigue que G es infinitamente divisible. ∎

En términos de la definición 4.4.2 el teorema 4.4.1 establece que *si la función característica φ es infinitamente divisible entonces también lo es $\overline{\varphi}$.*

Teorema 4.4.2 *Si $X_1, X_2, ..., X_n$ son variables aleatorias independientes con función de distribución infinitamente divisibles entonces también $X_1 + X_2 + \cdots + X_n$ tiene función de distribución infinitamente divisible. En otras palabras, el producto de funciones características infinitamente divisibles es infinitamente divisible.*

Demostración. Basta probar el resultado para $n = 2$, el caso general es inmediato por inducción. Sea φ y ψ la función característica de X_1 y X_2 respectivamente. Ya que φ y ψ son infinitamente divisibles, para cada entero positivo n, existen φ_n y ψ_n tales que

$$\varphi(t) = [\varphi_n(t)]^n \qquad \text{y} \qquad \psi(t) = [\psi_n(t)]^n,$$

entonces

$$\begin{aligned} \varphi(t)\,\psi(t) &= [\varphi_n(t)]^n\,[\psi_n(t)]^n \\ &= [\varphi_n(t)\,\psi_n(t)]^n. \end{aligned}$$

de donde, como $\varphi_n(t)\,\psi_n(t)$ es función característica, $\varphi(t)\,\psi(t)$ es infinitamente divisible. ∎

Teorema 4.4.3 *Una función característica infinitamente divisible no tiene ceros reales.*

Demostración. Sea φ infinitamente divisible, entonces para todo entero positivo n existe φ_n tal que

$$\varphi(t) = [\varphi_n(t)]^n,$$

entonces

$$|\varphi(t)|^{\frac{2}{n}} = |\varphi_n(t)|^2$$

Nótese que, por el teorema 4.4.1 y 4.4.2, $|\varphi_n(t)|^2 = \varphi_n(t)\,\overline{\varphi}_n(t)$ es una función característica infinitamente divisible. Ya que para cualquier función característica γ se tiene que $|\gamma(t)| \leq 1$. entonces

$$\lim_{n\to\infty} |\varphi(t)|^{\frac{2}{n}} = \lim_{n\to\infty} |\varphi_n(t)|^2 = \begin{cases} 1 & \text{en } \{t : \varphi(t) \neq 0\} \\ 0 & \text{en } \{t : \varphi(t) = 0\} \end{cases}$$

pero, por los teoremas 3.2.1 y 3.2.2, $\varphi(t)$ es continua y $\varphi(0) = 1$ entonces $\varphi(t) \neq 0$ en alguna vecindad de 0. Entonces $\lim_{n\to\infty} |\varphi(t)|^{\frac{2}{n}}$ es continua en alguna vecindad de 0. Así, por el teorema de continuidad (teorema 3.4.9), $\lim_{n\to\infty} |\varphi(t)|^{\frac{2}{n}}$ es una función característica y como, por el teorema 3.2.2, toda función característica es continua se tiene que $\varphi(t) \neq 0$ para todo $t \in \mathbb{R}$. ∎

Ejemplo 4.4.5 *Por el ejemplo 3.2.8 sabemos que la función característica de una variable aleatoria con distribución* $Uniforme(-1,1)$ *es*

$$\varphi(t) = \frac{\operatorname{sen} t}{t}, \quad t \in \mathbb{R}.$$

Ya que para algún valor de t*, por ejemplo* π*,* $\varphi(t) = 0$ *se concluye que la distribución* $Uniforme(-1,1)$ *no es infinitamente divisible.* ▲

Una de las propiedades que dan importancia a las distribuciones infinitamente divisibles está dada por el teorema 4.4.4. La demostración puede encontrarse en Shiryayev [1984, pp. 335-336].

Teorema 4.4.4 *Una variable aleatoria* X *puede ser límite de sumas* $S_n = \sum_{k=1}^{n} X_{n,k}$ *si, y sólo si,* X *es infinitamente divisible.*

En 1932, Kolmogorov encontró una representación que describe la función característica de cualquier variable aleatoria infinitamente divisible con varianza finita (ver Gnedenko [1973, p. 270]), después, en 1934, Lévy encontró una representación para la función de cualquier función característica infinitamente divisible. Finalmente, en 1937, Khinchine encontró un resultado ligeramente distinto al de Lévy; este resultado hoy en día se conoce como el teorema de representación de Lévy-Khinchine. A continuación se menciona este resultado, el lector interesado en la demostración puede encontrarla en Lukacs [1970. sec. 5.5], Laha y Rohatgi [1971, pp. 239-244], Chow y Teicher [1988, p. 431].

Teorema 4.4.5 (Lévy-Khinchine) *Una función característica* φ *es infinitamente divisible si, y sólo si, admite la representación*

$$\varphi(t) = \exp\left\{it\alpha + \int_{-\infty}^{\infty} \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) \frac{1+x^2}{x^2} dG(x)\right\} \tag{4.3}$$

donde $\alpha \in \mathbb{R}$ *y* G *es una función acotada no decreciente tal que* $G(-\infty) = 0$*. El integrando se define por continuidad para* $x = 0$ *como* $\frac{-t^2}{2}$*. Además la representación* (4.3) *es única.*

El lector interesado puede encontrar un resumen de otras propiedades de las distribuciones infinitamente divisibles en Bondesson [1995] y para aquel que desee profundizar más es recomendable consultar Gnedenko y Kolmogorov [1954] y Lukacs [1970].

4.5 Distribuciones Estables

Definición 4.5.1 *Se dice que la variable aleatoria X tiene una* **distribución estable** *o bien que la función de distribución asociada a X es estable si, para cada $n \geq 1$ existen constantes $a_n > 0$, b_n y variables aleatorias independientes $X_1, X_2, ..., X_n$, con la misma distribución que X, tales que* [4]

$$a_n X + b_n \stackrel{d}{=} X_1 + \cdots + X_n.$$

Se puede dar una definición en términos de funciones características como sigue

Definición 4.5.2 *Una variable aleatoria X, su función de distribución F y su función característica son* **estables** *si*

$$[\varphi(t)]^n = \varphi(a_n t) e^{itb_n} \tag{4.4}$$

donde $a_n > 0$ y b_n son constantes reales.

De la expresión 4.4 tenemos que

$$\varphi(t) = \left[\varphi\left(\frac{t}{a_n}\right) e^{-\frac{itb_n}{na_n}}\right]^n,$$

de donde se sigue el

Teorema 4.5.1 *Una función característica estable siempre es infinitamente divisible.*

A continuación damos sin prueba un teorema sobre la expresión general de una función característica estable. La demostración puede encontrarse en Laha y Rohatgi [1971, pp. 332-335].

Teorema 4.5.2 (Lévy-Khinchine) *Una variable aleatoria X es estable si, y sólo si, su función característica φ admite la representación*

$$\varphi(t) = \exp\left\{iat - \gamma |t|^\alpha \left[1 + i\beta \frac{t}{|t|} \omega(t;\alpha)\right]\right\} \tag{4.5}$$

donde $0 < \alpha \leq 2, |\beta| \leq 1$, $a \in \mathbb{R}$, $\gamma \geq 0$ *y*

$$\omega(t;\alpha) = \begin{cases} \tan \frac{\alpha\pi}{2} & si\ \alpha \neq 1 \\ \frac{2}{\pi} \ln |t| & si\ \alpha = 1. \end{cases}$$

[4] $\stackrel{d}{=}$ indica igualdad en distribución

Definición 4.5.3 *En* (4.5) , α *es conocido como el* ***exponente característico*** *de la distribución estable,* γ *es el* ***parámetro de escala****,* a *es llamado* ***parámetro de localización*** *y* β *es el* ***parámetro de simetría****.*

Ejemplo 4.5.1 *Si en* (4.5) , $\gamma = 0$ *la función característica resultante corresponde a una distribución degenerado, por lo tanto la distribución degenerada es estable.* ▲

Ejemplo 4.5.2 *Para el caso en que* $\alpha = 2$, (4.5) *es la función característica de una distribución Normal, entonces la distribución Normal es estable.* ▲

Ejemplo 4.5.3 *Si tomamos* $\beta = 0$ *y* $\alpha = 1$ *se sigue que la distribución de Cauchy es estable.* ▲

Ejemplo 4.5.4 *Cuando* $\alpha = \frac{1}{2}$, $\beta = 1$, $a = 0$ *y* $\gamma = 1$ *la función característica resultante pertenece a la densidad*

$$f(x) = \frac{1}{\sqrt{2\pi}} x^{-\frac{3}{2}} e^{-\frac{1}{2}x} I_{(0,\infty)}(x).$$

▲

Además de las mencionadas en los ejemplos 4.5.1-4.5.4, no se conoce ninguna otra distribución estable que tenga una expresión, en términos de funciones elementales, para su densidad. En Lucaks [1970, pp. 141-142] aparece una representación para las densidades de distribuciones estables en términos de series convergentes.

Teorema 4.5.3 *Toda distribución estable es absolutamente continua.*

Demostración. Del teorema 4.5.2, tenemos que si φ es una función característica estable entonces

$$|\varphi(t)| = \exp\{-\gamma |t|^{\alpha}\}.$$

Puede demostrarse que $\exp\{-\gamma |t|^{\alpha}\}$ es integrable sobre $(-\infty, \infty)$, de donde, por el teorema de inversión de Fourier, la distribución asociada a φ es absolutamente continua. ■

Teorema 4.5.4 *Toda distribución estable es unimodal.*

Demostración. Ver Lukacs[1970, p. 158]. ■

A continuación enunciamos lo que podría considerarse la propiedad más importante de las distribuciones estables. La demostración puede encontrarse en Shiryayev [1984, p. 338].

Teorema 4.5.5 *Sean $X_1, X_2, ..., X_n$ variables aleatorias independientes e idénticamente distribuidas y sea $S_n = X_1 + \cdots + X_n$. Una condición necesaria y suficiente para que la variable aleatoria X sea límite, en el sentido de convergencia débil, de las variables aleatorias $\frac{S_n - b_n}{a_n}$, $a_n > 0$, es que X sea estable.*

La importancia de las distribuciones estables no sólo se debe a que son las únicas que pueden ser límites para sumas de variables aleatorias independientes ni a sus propiedades analíticas, sino porque tienen aplicaciones en diversas disciplinas como son física, economía, finanzas y ciencias sociales en general. Por ejemplo, Mandelbroot [1960] y Mandelbroot [1963] sugiere a las distribuciones estables como posibles modelos para la distribución del ingreso[5] y precios especulativos; Press [1972a] desarrolló análisis de portafolios suponiendo que la distribución de los precios de los activos en el portafolio obedecen una distribución estable; Holtsmark [1919], y Chandrasekhar [1943] dan aplicaciones en astronomía y física; aplicaciones en procesos estocásticos pueden encontrarse en Feller [1971, p. 178].

El uso de distribuciones estables en aplicaciones se puede justificar por distintas razones. Antes, se creía que muchos procesos tenían un comportamiento aproximadamente Normal porque las observaciones extremas eran tratadas como "outliers". Después, al estudiar el conjunto completo de observaciones, se encontró que muchos procesos tenían colas más pesadas que la distribución Normal. Tal comportamiento de colas pesadas es típico de distribuciones estables no Normales. Aunque las distribuciones estables no son las únicas que tienen colas pesadas, para algunos procesos, el análisis del comportamiento de combinaciones lineales se simplifica bajo el supuesto de una distribución estable.

4.6 La Función Característica Empírica

Definición 4.6.1 *Sea $X_1 X_2, ..., X_n$ una muestra aleatoria de tamaño n de la función de distribución F. La **función de distribución empírica** $\widehat{F}_n$ se define como*

$$\widehat{F}_n(x) := \frac{1}{n} \sum_{k=1}^{n} I_{[X_k \leq x]}.$$

Muchos procedimientos de inferencia estadística están basados en la función de distribución empírica. Debido a que existe una correspondencia uno a uno entre las funciones características y las funciones de distribución, parece natural tratar de desarrollar procedimientos de inferencia basados en el análogo empírico de la función característica.

[5] Por esta razón las distribuciones estables también son conocidas como distribuciones de Levy-Pareto.

Definición 4.6.2 *Sea $X_1 X_2, ..., X_n$ una muestra aleatoria de tamaño n de una población con función característica φ. Definimos la* **función característica empírica** $\widehat{\varphi}_n$ *como*

$$\widehat{\varphi}_n(t) := \frac{1}{n}\sum_{k=1}^{n} e^{itX_k}.$$

La función característica empírica fue introducida por Parzen [1962] y sus primeras aplicaciones en problemas de inferencia estadística fueron propuestas por Press [1972b], Heathcote [1972], Press [1975], Heathcote [1977], Feigen y Heathcote [1976]. En la actualidad existen bastantes propuestas de aplicación. En Feuverger y McDunnough [1984] aparece una extensa lista de referencias.

Aunque una función de distribución posee diferentes características funcionales con las que mantiene una relación unívoca, los procedimientos inferenciales basados en la función característica empírica son los que se proponen más frecuentemente. La preferencia por el uso de la función característica empírica radíca en el hecho de que existe una extensa teoría sobre la función característica teórica; además, propiedades como la independencia y la simetría, se escriben y manejan más fácilmente en términos de funciones características.

Sea $\widehat{\varphi}_n(t)$ la función característica empírica calculada con la muestra aleatoria $X_1 X_2, ..., X_n$ de una población con función característica $\varphi(t) = C(t) + iS(t)$. Como $\left|e^{itX_j}\right| \leq 1$, entonces $|\widehat{\varphi}_n(t)| \leq 1$; esto implica que *todos los momentos de $\widehat{\varphi}_n(t)$, vista como variable aleatoria, son finitos.*

Si fijamos t, $\widehat{\varphi}_n(t)$ es un promedio muestral de variables aleatorias independientes e idénticamente distribuidas por lo que, según el teorema A.2.1, $\widehat{\varphi}_n(t)$ *obedece las leyes de los grandes números, y por lo tanto es un estimador consistente de $\varphi(t)$.* También, es claro que $\widehat{\varphi}_n(t)$ *es un estimador insesgado de $\varphi(t)$.*

Denotemos por $C(t)$ y $S(t)$ la parte real y la parte compleja de $\varphi(t)$ respectivamente y definamos z_n como el vector $2k$-dimensional

$$z_n^t := (\cos t_1 X_n, ..., \cos t_k X_n, \operatorname{sen} t_1 X_n, ..., \operatorname{sen} t_k X_n)$$

es claro que

$$E[z_n] = (C(t_1), ..., C(t_k), S(t_1), ..., S(t_k))^t := \xi^t$$

además, de las identidades trigonométricas elementales

$$\begin{aligned}
\cos A \cos B &= \frac{1}{2}\{\cos(A-B) + \cos(A+B)\}, \\
\operatorname{sen} A \operatorname{sen} B &= \frac{1}{2}\{\cos(A-B) - \cos(A+B)\}, \\
\operatorname{sen} A \cos B &= \frac{1}{2}\{\operatorname{sen}(A-B) + \operatorname{sen}(A+B)\};
\end{aligned}$$

se tiene que

$$\begin{aligned}
cov\{\cos t_jX, \cos t_kX\} &= E[\cos t_jX \cdot \cos t_kX] - E[\cos t_jX]\,E[\cos t_kX] \\
&= \frac{1}{2}E[\cos((t_j - t_k)X) + \cos((t_j + t_k)X)] - C(t_j)\,C(t_k) \\
&= \frac{1}{2}[C(t_j - t_k) + C(t_j + t_k)] - C(t_j)\,C(t_k)
\end{aligned}$$

$$\begin{aligned}
cov\{\operatorname{sen} t_jX, \operatorname{sen} t_kX\} &= E[\operatorname{sen} t_jX \cdot \operatorname{sen} t_kX] - E[\operatorname{sen} t_jX]\,E[\operatorname{sen} t_kX] \\
&= \frac{1}{2}E[\cos((t_j - t_k)X) - \cos((t_j + t_k)X)] - S(t_j)\,S(t_k) \\
&= \frac{1}{2}[C(t_j - t_k) - C(t_j + t_k)] - S(t_j)\,S(t_k)
\end{aligned}$$

$$\begin{aligned}
cov\{\cos t_jX, \operatorname{sen} t_kX\} &= E[\cos t_jX \cdot \operatorname{sen} t_kX] - E[\cos t_jX]\,E[\operatorname{sen} t_kX] \\
&= \frac{1}{2}E[\operatorname{sen}((t_k - t_j)X) + \operatorname{sen}((t_j + t_k)X)] - C(t_j)\,S(t_k) \\
&= \frac{1}{2}[S(t_j + t_k) - S(t_j - t_k)] - C(t_j)\,S(t_k)
\end{aligned}$$

por lo que la matriz de covarianza $\Sigma := (\sigma_{jk})$ para z_n tiene como elemento jk-ésimo

$$\sigma_{jk} := \begin{cases} \frac{1}{2}[C(t_j - t_k) + C(t_j + t_k) - 2C(t_j)\,C(t_k)] & \text{si } 1 \le j,k \le m \\ \frac{1}{2}[C(t_j - t_k) - C(t_j + t_k) - 2S(t_j)\,S(t_k)] & \text{si } m < j,k \le 2m \\ \frac{1}{2}[S(t_j + t_k) - S(t_j - t_k) - 2C(t_j)\,S(t_k)] & \text{si } 1 \le j \le m \text{ y } m < k \le 2m \end{cases}$$

donde $t_j = t_{j-m}$ para $m + 1 \le j \le 2m$.

Denotemos por $C_n(t)$ la parte real de $\widehat{\varphi}_n(t)$ y por $S_n(t)$ su parte imaginaria, es decir

$$C_n(t) := \frac{1}{n}\sum_{j=1}^{n}\cos tX_j \qquad \text{y} \qquad S_n(t) := \frac{1}{n}\sum_{j=1}^{n}\operatorname{sen} tX_j.$$

Sean $t_1, ..., t_k$ números reales y definamos $\xi_n = \xi_n(t_1, ..., t_k)$ como el vector

$$\xi_n^t := (C_n(t_1), ..., C_n(t_k), S_n(t_1), ..., S_n(t_k)),$$

entonces por el teorema del límite central multivariado (ver Rao [1973, p. 128]) se obtiene que

$$\sqrt{n}(\xi_n - \xi) \xrightarrow{d} N_{2k}(0, \Sigma).$$

Investigaciones sobre el comportamiento límite de la función característica empírica se pueden hallar en Kent [1975], Csörgo [1981], Marcus [1981]. Feuverger y Mureika [1977] realizaron estudios sobre las propiedades de la función característica empírica orientados hacia sus aplicaciones en estadística. Para obtener una idea más intuitiva de la función característica empírica, incluyendo su interpretación geométrica, se puede consultar Epps [1993].

4.7 Estimación de Parámetros

Existen situaciones en las que los métodos de estimación tradicionales no pueden aplicarse. Un ejemplo de ello son las distribuciones estables que, como ya se mencionó, la mayoría de las veces carecen de una expresión en forma cerrada para su densidad, lo que ocasiona que el método de máxima verosimilitud no pueden ser aplicado. Por ello surge la necesidad de elaborar métodos de inferencia alternativos. Entre estos métodos alternativos se encuentran algunos basados en la función característica empírica.

Aunque en la literatura existen diferentes propuestas sobre la forma de utilizar la función característica empírica en problemas de estimación, no existe un consenso sobre cual es el mejor método. Aquí presentamos los métodos más conocidos.

4.7.1 Método de la Mínima Distancia

Supongamos que $\varphi(t)$ es la función característica de una distribución estable con parámetros a, γ, α y β como en la definición 4.5.3. El método de la mínima distancia propone como estimador de a, γ, α y β aquellos valores de estos parámetros que minimicen la función

$$g(a, \gamma, \alpha, \beta) := \sup_t |\varphi(t) - \widehat{\varphi}_n(t)|$$

donde $\widehat{\varphi}_n(t)$ denota la función característica empírica calculada con una muestra aleatoria de tamaño n proveniente de una población con función característica $\varphi(t)$.

El método de la mínima distancia fue propuesto por Press [1972b] y es el más sencillo de los métodos de estimación basados en la función característica empírica.

4.7.2 Método de los Momentos

Este método fue propuesto por Press [1972b] como una alternativa para estimar los parámetros de una distribución estable. A continuación se describe el método para el caso univariado; la generalización para distribuciones estables multivariadas puede encontrarse en Press[1972a, p. 364] o en Press[1972b]. Supongamos que deseamos estimar los parámetros de una distribución estable con función característica $\varphi(t)$ y que para ello contamos con una muestra aleatoria $X_1, ..., X_n$ a partir de la cual calculamos la función característica empírica $\widehat{\varphi}_n(t)$ que sirve como estimador de $\varphi(t)$. Para obtener estimadores $\widehat{\alpha}$ y $\widehat{\gamma}$, del índice característico y el parámetro de escala respectivamente, observamos que

$$|\varphi(t)|^2 = \exp(-2\gamma |t|^\alpha);$$

o equivalentemente

$$-\ln|\varphi(t)| = \gamma |t|^{\alpha}. \tag{4.6}$$

Suponiendo que $\alpha \neq 1$ y que t_1 y t_2 son dos valores distintos de t tales que $t_1 t_2 \neq 0$ se tienen las ecuaciones

$$-\ln|\varphi(t_1)| = \gamma |t_1|^{\alpha}, \qquad -\ln|\varphi(t_2)| = \gamma |t_2|^{\alpha}$$

que al resolverse simultáneamente para α y γ y sustituyendo $\varphi(t)$ por su estimador $\widehat{\varphi}(t)$ resulta

$$\widehat{\alpha} = \frac{\ln\left|\frac{\ln|\widehat{\varphi}(t_1)|}{\ln|\widehat{\varphi}(t_2)|}\right|}{\ln\left|\frac{t_1}{t_2}\right|}$$

y

$$\widehat{\gamma} = \frac{\ln|t_1| \ln[-\ln|\widehat{\varphi}(t_2)|] - \ln|t_2| \ln[-\ln|\widehat{\varphi}(t_1)|]}{\ln\left|\frac{t_1}{t_2}\right|}.$$

Para estimar a y β, consideramos dos casos, cuando $\alpha \neq 1$ y cuando $\alpha = 1$.

Bajo el supuesto de que $\alpha \neq 1$, definimos $S(t) := \operatorname{Im}[\ln \varphi(t)]$, es decir

$$S(t) := at - \gamma |t|^{\alpha-1} t\beta\omega(t;\alpha)$$

y elegimos dos valores distintos t_3 y t_4 para t tales que $t_3 t_4 \neq 0$. Entonces

$$a - \left[\gamma |t_3|^{\alpha-1} \tan\frac{\pi\alpha}{2}\right]\beta = \frac{S(t_3)}{t_3} \tag{4.7}$$

y

$$a - \left[\gamma |t_4|^{\alpha-1} \tan\frac{\pi\alpha}{2}\right]\beta = \frac{S(t_4)}{t_4}. \tag{4.8}$$

Tomando

$$\rho^2(t) = \left(\frac{1}{n}\sum_{j=1}^{n} \cos tX_j\right)^2 + \left(\frac{1}{n}\sum_{j=1}^{n} \operatorname{sen} tX_j\right)^2$$

y

$$\tan\theta(t) = \left(\frac{\sum_{j=1}^{n} \operatorname{sen} tX_j}{\sum_{j=1}^{n} \cos tX_j}\right)$$

expresamos $\widehat{\varphi}(t)$ en coordenadas polares como

$$\widehat{\varphi}(t) = \rho(t) \exp[i\theta(t)]$$

y obtenemos la expresión

$$\ln \widehat{\varphi}(t) = \ln \rho(t) + i\theta(t),$$

de donde se sigue que un estimador para $S(t)$ es

$$\widehat{S}(t) = \theta(t) = \arctan \left(\frac{\sum_{j=1}^{n} \operatorname{sen} tX_j}{\sum_{j=1}^{n} \cos tX_j} \right).$$

Finalmente, si en las ecuaciones (4.7) y (4.8) reemplazamos $S(t)$, α y γ por su valor estimado y resolvemos estas ecuaciones simultáneamente para a y β tenemos que

$$\widehat{\beta} = \frac{\frac{\widehat{S}(t_3)}{t_3} - \frac{\widehat{S}(t_4)}{t_4}}{\left[|t_4|^{\widehat{\alpha}-1} - |t_3|^{\widehat{\alpha}-1}\right] \widehat{\gamma} \tan \frac{\pi\widehat{\alpha}}{2}}$$

y

$$\widehat{a} = \frac{|t_4|^{\widehat{\alpha}-1} \frac{\widehat{S}(t_3)}{t_3} - |t_3|^{\widehat{\alpha}-1} \frac{\widehat{S}(t_4)}{t_4}}{|t_4|^{\widehat{\alpha}-1} - |t_3|^{\widehat{\alpha}-1}}.$$

Para el caso en que $\alpha = 1$, de la ecuación (4.6) tenemos que

$$\widehat{\gamma} = -\frac{\ln \widehat{\varphi}(t_1)}{|t_1|}$$

y $S(t)$ se reduce a

$$S(t) := at - \frac{2\gamma\beta t}{\pi} \ln |t|,$$

entonces para dos valores distintos t_3 y t_4 de t no nulos

$$a - \left[\frac{2\gamma}{\pi} \ln |t_3|\right] \beta = \frac{S(t_3)}{t_3} \tag{4.9}$$

y

$$a - \left[\frac{2\gamma}{\pi} \ln |t_4|\right] \beta = \frac{S(t_4)}{t_4}. \tag{4.10}$$

Resolviendo (4.9) y (4.10) simultáneamente para a y β y sustituyendo los estimadores de $\varphi(t_1)$, $S(t)$ y γ se tiene que

$$\widehat{\beta} = \frac{\frac{\widehat{S}(t_3)}{t_3} - \frac{\widehat{S}(t_4)}{t_4}}{\frac{2\widehat{\gamma}}{\pi} \ln \left|\frac{t_4}{t_3}\right|}$$

y

$$\widehat{a} = \frac{\ln|t_4| \frac{\widehat{S}(t_3)}{t_3} - \ln|t_3| \frac{\widehat{S}(t_4)}{t_4}}{\ln\left|\frac{t_4}{t_3}\right|}.$$

Ya se vio (sec. 4.6) que $\widehat{\varphi}_n(t)$ es un estimador consistente de $\varphi(t)$,entonces, según el teorema A.1.5, los estimadores obtenidos por el método de los momentos son consistentes, pues se basan en $\widehat{\varphi}_n(t)$. Además, obtener estimadores por el método de los momentos presenta la ventaja de que no se requieren procedimientos de cálculo complicados. En Press [1972b], se construyen intervalos de confianza, basados en la distribución asintótica de los estimadores obtenidos por este método.

La elección óptima de los puntos donde se debe evaluar $\widehat{\varphi}_n(t)$ para calcular los estimadores es un problema abierto.

4.7.3 Estimadores de Error Cuadrático Integrado

En Press [1972b], aparece el método de la mínima distancia en r-media, como una alternativa para estimar parámetros de distribuciones estables. Este método sugiere usar como estimadores de a, γ, α y β las estadísticas que minimizan la función

$$h(a,\gamma,\alpha,\beta) := \int_{-\infty}^{\infty} |\varphi(t) - \widehat{\varphi}_n(t)|^r W(t)\, dt;$$

donde $\widehat{\varphi}_n(t)$ es la función característica empírica y $W(t)$ es un factor de convergencia adecuado. Este método tiene la desventaja de requerir métodos númericos complicados. Paulson, Holcomb y Leitch [1975], desarrollaron un procedimiento de minimización numérica para el caso particular en que $r = 2$ y $W(t) = e^{-\frac{t^2}{2}}$ y dieron una aplicación en finanzas. Este método fue generalizado por Heathcote [1977] quien estudio las propiedades de la estadística $\widetilde{\theta}_n$ que minimiza la función

$$I_n(\theta) := \int_{-\infty}^{\infty} |\widehat{\varphi}_n(t) - \varphi(t;\theta)|^2\, dG(t);$$

donde G es una función de ponderación no decreciente y de variación acotada. La estadística $\widetilde{\theta}_n$ es llamada **estimador de error cuadrático integrado**. A continuación presentamos algunas de las propiedades de éstos estimadores.

Sea $X_1, ..., X_n$ una muestra aleatoria de tamaño n de una variable aleatoria X con función de distribución $F(x;\theta)$ y función característica $\varphi(t;\theta) = C(t;\theta) + iS(t;\theta)$. Sea $\widehat{\varphi}_n(t)$ la función característica empírica calculada con $X_1, ..., X_n$. Consideremos la función

$$\begin{aligned} I_n(\theta) &:= \int_{-\infty}^{\infty} |\widehat{\varphi}_n(t) - \varphi(t;\theta)|^2 \, dG(t) \\ &= \int_{-\infty}^{\infty} \left[\{C_n(t) - C(t;\theta)\}^2 + \{S_n(t) - S(t;\theta)\}^2\right] dG(t) \end{aligned}$$

donde $C_n(t)$ y $S_n(t)$ denotan, respectivamente, la parte real y la parte compleja de $\widehat{\varphi}_n(t)$.

Supongamos que se tiene el caso regular, esto es, $I_n(\theta)$ puede diferenciarse bajo el signo de integración. Si denotamos con una prima la derivada con respecto a θ tenemos que

$$\begin{aligned} I_n'(\theta) &= -2\int_{-\infty}^{\infty} \left[\{C_n(t) - C(t;\theta)\} C'(t;\theta) + \{S_n(t) - S(t;\theta)\} S'(t;\theta)\right] dG(t) \\ &= -2\int_{-\infty}^{\infty} \left[\left\{\sum_{j=1}^{n} \frac{\cos tX_j}{n} - C(t;\theta)\right\} C'(t;\theta) + \left\{\sum_{j=1}^{n} \frac{\operatorname{sen} tX_j}{n} - S(t;\theta)\right\} S'(t;\theta)\right] dG(t) \\ &= -\frac{2}{n}\sum_{j=1}^{n}\int_{-\infty}^{\infty} \left[\{\cos tX_j - C(t;\theta)\} C'(t;\theta) + \{\operatorname{sen} tX_j - S(t;\theta)\} S'(t;\theta)\right] dG(t). \end{aligned}$$

Supongamos que $\widetilde{\theta}$ denota el valor de θ que minimiza $I_n(\theta)$ entonces debe ocurrir que

$$I_n'\left(\widetilde{\theta}\right) = 0 \qquad \text{e} \qquad I_n''\left(\widetilde{\theta}\right) > 0.$$

Definamos la variable aleatoria $K(\theta)$ como

$$K(\theta) := \int_{-\infty}^{\infty} \left[\{\cos tX - C(t;\theta)\} C'(t;\theta) + \{\operatorname{sen} tX - S(t;\theta)\} S'(t;\theta)\right] dG(t)$$

entonces, si $f(x;\theta)$ denota la densidad de X,

$$E[K(\theta)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[\{\cos tx - C(t;\theta)\} C'(t;\theta) + \{\operatorname{sen} tx - S(t;\theta)\} S'(t;\theta)\right] dG(t) \, f(x;\theta) \, dx$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[\{\cos tx - C(t;\theta)\} C'(t;\theta) + \{\operatorname{sen} tx - S(t;\theta)\} S'(t;\theta) f(x;\theta)\right] dx dG(t) \\
&= \int_{-\infty}^{\infty} \left[\{C(t;\theta) - C(t;\theta)\} C'(t;\theta) + \{S(t;\theta) - S(t;\theta)\} S'(t;\theta)\right] dG(t) \\
&= 0
\end{aligned}
$$

y

$$
\begin{aligned}
\operatorname{var}[K(\theta)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[(\{\cos tx - C(t;\theta)\} C'(t;\theta) + \{\operatorname{sen} tx - S(t;\theta)\} S'(t;\theta))\right. \\
&\quad \times (\{\cos ux - C(u;\theta)\} C'(u;\theta) \\
&\quad \left. + \{\operatorname{sen} ux - S(u;\theta)\} S'(u;\theta))\right] dG(u)\, dG(t)\, f(x;\theta)\, dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[\operatorname{cov}[\cos tX, \cos uX] C'(t;\theta) C'(u;\theta)\right. \\
&\quad + 2\operatorname{cov}[\cos tX, \operatorname{sen} uX] C'(t;\theta) S'(u;\theta) \\
&\quad \left. + \operatorname{cov}[\operatorname{sen} tX, \operatorname{sen} uX] S'(t;\theta) S'(u;\theta)\right] dG(u)\, dG(t);
\end{aligned}
$$

donde[6]

$$
\begin{aligned}
\operatorname{cov}\{\cos tX, \cos uX\} &= \frac{1}{2}[C(t-u;\theta) + C(t+u;\theta)] - C(t;\theta) C(u;\theta), \\
\operatorname{cov}\{\operatorname{sen} tX, \operatorname{sen} uX\} &= \frac{1}{2}[C(t-u;\theta) - C(t+u;\theta)] - S(t;\theta) S(u;\theta), \\
\operatorname{cov}\{\cos tX, \operatorname{sen} uX\} &= \frac{1}{2}[S(t+u;\theta) - S(t-u;\theta)] - C(t;\theta) S(u;\theta).
\end{aligned}
$$

Denotemos por $\mathcal{I}(\theta)$ la información esperada de Fisher por unidad muestral, es decir

$$
\mathcal{I}(\theta) := E\left[\left(\frac{\partial}{\partial\theta} \ln f(x;\theta)\right)^2\right].
$$

Se puede demostrar que si $\mathcal{I}(\theta)$ es finita

$$
\operatorname{var}[K(\theta)] \leq 2\mathcal{I}(\theta).
$$

[6] Ver p. 113

Entonces $\{K_j(\theta)\}_{j\geq 1}$ es una sucesión de variables aleatorias independientes con la misma distribución que $K(\theta)$ y varianza finita, por lo tanto, la sucesión $\{K_j(\theta)\}_{j\geq 1}$ obedece la ley fuerte de los grandes números [7] y el teorema del límite central. En particular, si θ_0 denota el verdadero valor de θ tenemos que

$$\sqrt{n} I_n'(\theta_0) \xrightarrow{d} N(0, 4\operatorname{var}[K(\theta)]),$$

cuando $n \to \infty$.

Ahora, observamos que, para n suficientemente grande, $I_n(\theta)$ tiene un valor mínimo en θ_0 casi seguramente. Para ello supongamos que δ es un número positivo arbitrario, y consideremos la diferencia

$$\begin{aligned}
I_n(\theta_0 \pm \delta) - I_n(\theta_0) &= \int_{-\infty}^{\infty} \left[\{C_n(t) - C(t;\theta_0 \pm \delta)\}^2 + \{S_n(t) - S(t;\theta_0 \pm \delta)\}^2\right.\\
&\quad \left. - \{C_n(t) - C(t;\theta_0)\}^2 - \{S_n(t) - S(t;\theta_0)\}^2\right] dG(t)\\
&= \int_{-\infty}^{\infty} \left[C_n^2(t) - 2C_n(t)C(t;\theta_0 \pm \delta) + C^2(t;\theta_0 \pm \delta)\right.\\
&\quad + S_n^2(t) - 2S_n(t)S(t;\theta_0 \pm \delta) + S^2(t;\theta_0 \pm \delta)\\
&\quad - C_n^2(t) + 2C_n(t)C(t;\theta_0) - C^2(t;\theta_0)\\
&\quad \left. - S_n^2(t) + 2S_n(t)S(t;\theta_0) - S^2(t;\theta_0)\right] dG(t)\\
&= \int_{-\infty}^{\infty} \left[\{C(t;\theta_0 \pm \delta) - C(t;\theta_0)\}\{C(t;\theta_0 \pm \delta) + C(t;\theta_0) - 2C_n(t)\}\right.\\
&\quad \left. + \{S(t;\theta_0 \pm \delta) - S(t;\theta_0)\}\{S(t;\theta_0 \pm \delta) + S(t;\theta_0) + 2S_n(t)\}\right] dG(t).
\end{aligned}$$

Ya que $E[\cos tX_j] \leq 1$ y $E[\operatorname{sen} tX_j] \leq 1$, por la ley fuerte de los grandes números (ver apéndice A) se tiene que $C_n(t) \xrightarrow{c.s.} C(t;\theta_0)$ y $S_n(t) \xrightarrow{c.s.} S(t;\theta_0)$. Así, por el teorema de convergencia dominada de Lebesgue,

$$E[I_n(\theta_0 \pm \delta) - I_n(\theta_0)] = \int_{-\infty}^{\infty} \left[\{C(t;\theta_0 \pm \delta) - C(t;\theta_0)\}^2 + \{S(t;\theta_0 \pm \delta) - S(t;\theta_0)\}^2\right] dG(t)$$

de donde, nuevamente por la ley fuerte de los grandes números, resulta que $I_n(\theta_0 \pm \delta) > I_n(\theta_0)$ casi seguramente, y por lo tanto que θ_0 es un mínimo para $I_n(\theta)$ con probabilidad uno. Así, hemos demostrado que existe un valor $\widetilde{\theta}_n$ tal que $I_n\left(\widetilde{\theta}_n\right) = 0$ y $\widetilde{\theta}_n \xrightarrow{c.s.} \theta_0$.

[7] Ver teorema A.2.1, p. 140

Ahora supongamos que $C(t;\theta)$ y $S(t;\theta)$ son tales que $C''(t;\theta)$ y $S''(t;\theta)$ están uniformemente acotadas por funciones integrables respecto a G. Entonces

$$\begin{aligned} I_n''(\theta) &= 2\int_{-\infty}^{\infty}\Big[\{C'(t;\theta)\}^2+\{S'(t;\theta)\}^2-\{C_n(t)-C(t;\theta)\}C''(t;\theta) \\ &\quad +\{S_n(t)-S(t;\theta)\}S''(t;\theta)\Big]\,dG(t) \end{aligned}$$

y se sigue que

$$E\left[I_n''(\theta_0)\right]=2\lambda(\theta_0);$$

donde

$$\lambda(\theta_0)=\int_{-\infty}^{\infty}|\varphi(t;\theta_0)|^2\,dG(t).$$

Por lo tanto, $I_n''(\theta_0)\xrightarrow{c.s.}2\lambda(\theta_0)$. Finalmente, si $\widetilde{\theta}_n$ es una raíz consistente para

$$I_n'\left(\widetilde{\theta}_n\right)=I_n'(\theta_0)+\left(\widetilde{\theta}_n-\theta_0\right)I_n''\left(\theta_0+\varepsilon\left(\widetilde{\theta}_n-\theta_0\right)\right),$$

donde $|\varepsilon|\le 1$, se concluye que

$$\sqrt{n}\left(\widetilde{\theta}_n-\theta_0\right)=\frac{\sqrt{n}I_n'(\theta_0)}{I_n''\left(\theta_0+\varepsilon\left(\widetilde{\theta}_n-\theta_0\right)\right)}\xrightarrow{d}N\left(0,\frac{\operatorname{var}\left[K(\theta_0)\right]}{\lambda^2(\theta_0)}\right)$$

cuando $n\longrightarrow\infty$. Así, queda demostrado el

Teorema 4.7.1 (Heathcote (1977)) *Bajo condiciones de regularidad adecuadas*

1. *Existe una raíz $\widetilde{\theta}_n$ de $I_n'(\theta)$ que es un estimador fuertemente consistente de θ.*
2. *Además, si $C''(t;\theta)$ y $S''(t;\theta)$ están uniformemente acotadas por funciones G-integrables, entonces*

$$\sqrt{n}\left(\widetilde{\theta}_n-\theta_0\right)\xrightarrow{d}N\left(0,\frac{\operatorname{var}\left[K(\theta_0)\right]}{\lambda^2(\theta_0)}\right)$$

donde θ_0 es el verdadero valor de θ.

En Heathcote [1977] aparece la extensión de los estimadores de error cuadrático integrado al caso multivariado.

4.7.4 Procedimiento $k - L$

El procedimiento $k - L$ obtiene estimadores aplicando el método de máxima verosimilitud a la distribución asintótica de k puntos de la función característica empírica. El número k de puntos y su localización son fijos y no dependen del tamaño de muestra n.

Este procedimiento fue propuesto por Feueverger y McDunnough [1981], quienes además demostraron que los estimadores obtenidos por este procedimiento son consistentes, altamente eficientes y tienen como distribución asintótica a la distribución Normal.

Sea $X_1, ..., X_n$ una muestra aleatoria de una población con función característica $\varphi(t;\theta)$ donde $\theta \in \Theta \subset \mathbb{R}^p$. Denotemos por $C(t;\theta)$ y $S(t;\theta)$ la parte real y compleja, respectivamente, de $\varphi(t;\theta)$. Análogamente, $C_n(t)$ y $S_n(t)$, representan la parte real y compleja de la función característica empírica $\widehat{\varphi}_n(t)$.

Sean $t_1, ..., t_k$ números reales y definamos a $\xi(\theta)$ como el vector $2k$-dimensional

$$\xi^t(\theta) := (C(t_1;\theta), ..., C(t_k;\theta), S(t_1;\theta), ..., S(t_k;\theta))$$

y a ξ_n como el vector

$$\xi_n^t := (C_n(t_1), ..., C_n(t_k), S_n(t_1), ..., S_n(t_k)).$$

Es claro que cada una de las componentes de ξ_n es un promedio de variables aleatorias independientes e idénticamente distribuidas con varianza finita y que $E[\xi_n] = \xi(\theta)$. Entonces, por el teorema del límite central multivariado, (apéndice A, p. 143), se sigue que

$$\sqrt{n}(\xi_n - \xi(\theta)) \xrightarrow{d} N_{2k}(0, n\Sigma)$$

donde $\Sigma : = \operatorname{var}(\xi_n)$.

Entonces la verosimilitud asintótica es proporcional a

$$\exp\left[-\frac{n}{2}(\xi_n - \xi(\theta))' \Sigma^{-1} (\xi_n - \xi(\theta))\right]$$

de donde la log-verosimilitud es

$$-\frac{n}{2}(\xi_n - \xi(\theta))' \Sigma^{-1} (\xi_n - \xi(\theta)).$$

Esto sugiere como estimador de θ a la estadística $\widehat{\theta}_n$ que minimiza la forma cuadrática

$$L_n(\theta) := (\xi_n - \xi(\theta))' \Sigma^{-1} (\xi_n - \xi(\theta)).$$

Nótese que $L_n(\theta)$ depende únicamente de la función característica y no de la función de densidad.

4.7.5 Aproximación Mediante Polinomios de Chebyshev

Este método, propuesto por Madan y Senata [1987a], consiste en aproximar la función de densidad de una transformación de la variable aleatoria mediante polinomios de Chebyshev y entonces aplicar el método de máxima verosimilitud a esta aproximación para estimar los parámetros de la distribución de la variable aleatoria de interés.

Sea X una variable aleatoria con distribución simétrica y función característica $\varphi_X(t)$. Definamos la variable aleatoria Y como $Y = \cos tX$ para $t \in \mathbb{R}$. Ya que $e^{itX} = \cos tX + i \operatorname{sen} tX$ tenemos que

$$Y = \frac{e^{itX} + e^{-itX}}{2}$$

de donde

$$\begin{aligned} E[Y^n] &= \frac{1}{2^n} E\left[\left(e^{itX} + e^{-itX}\right)^n\right] \\ &= \frac{1}{2^n} E\left[\sum_{j=0}^{n} \binom{n}{j} e^{itjX} e^{-it(n-j)X}\right] \\ &= \frac{1}{2^n} \sum_{j=0}^{n} \binom{n}{j} E\left[e^{it(2j-n)X}\right] \\ &= \frac{1}{2^n} \sum_{j=0}^{n} \binom{n}{j} \varphi_X((2j-n)t) \end{aligned}$$

y por lo tanto, se sigue que todos los momentos de la variable aleatoria Y pueden ser expresados en términos de la función característica de X.

Definamos

$$q_n(y) := \frac{T_n(y)}{2^{n-1}};$$

donde T_n es el n-ésimo polinomio de Chebyshev del primer tipo[8]. Entonces, por $(B.3)$ del apéndice B, es claro que

$$q_0(y) = 2, \quad q_1(y) = y, \quad q_{n+1}(y) = y q_n(y) - \frac{1}{4} q_{n-1}(y) \quad n \geq 1.$$

Si suponemos que Y tiene densidad g continua en $[-1, 1]$. entonces podemos expresar a g como

$$g(y) = \sum_{n=0}^{\infty} \frac{a_n q_n(y)}{\pi \sqrt{1 - y^2}}$$

[8] El lector que no esté familiarizado con los polinomios de Chebyshev, puede encontrar las definiciones y propiedades básicas en el apéndice B.

por lo que la densidad de Y quedará determinada si encontramos los valores de las a_n.

Para determinar las a_n primero observamos que, por la ortogonalidad de los polinomios de Chebyshev,

$$E\left[q_n(Y)\right] = a_n \int_{-1}^{1} \frac{q_n^2(y)}{\pi\sqrt{1-y^2}} dy,$$

entonces, por $(B.2)$ del apéndice B, se sigue que

$$E\left[q_0(Y)\right] = 4a_0 \tag{4.11}$$

y, para $n \geq 1$,

$$E\left[q_n(Y)\right] = \frac{a_n}{2^{2n-1}}. \tag{4.12}$$

El siguiente teorema será de utilidad para determinar los valores de las a_n para $n \geq 1$.

Teorema 4.7.2 *Si $\varphi_X(t)$ es la función característica de una variable aleatoria simétrica X y $Y = \cos tX$, entonces, para $m \geq 0$ y $n \geq 0$*

$$E\left[Y^m q_n(Y)\right] = \frac{1}{2^{n+m-1}} \sum_{j=0}^{m} \binom{m}{j} \varphi_X\left((n+m-2j)\,t\right)$$

Demostración. Ver Madan y Senata [1987a] ■

Haciendo $m = 0$ en el teorema 4.7.2, obtenemos

$$E\left[q_0(Y)\right] = \frac{\varphi_X(0)}{2^{-1}} = 2$$

y

$$E\left[q_n(Y)\right] = \frac{1}{2^{n-1}}\varphi_X(nt); \quad n \geq 1,$$

de donde, por las ecuaciones (4.11) y (4.12), se sigue que

$$a_0 = \frac{1}{2}, \qquad a_n = 2^n \varphi_X(nt), \quad n \geq 1.$$

Por lo tanto, se concluye que podemos escribir la densidad de Y como

$$g(y) = \frac{1}{\pi\sqrt{1-y^2}} + \sum_{n=1}^{\infty} 2^n \frac{\varphi_X(nt)\, q_n(y)}{\pi\sqrt{1-y^2}}. \tag{4.13}$$

Entonces los parámetros de la función característica pueden estimarse por el método de máxima verosimilitud. Para que el procedimiento sea numéricamente posible es necesario truncar la serie (4.13). En Madan y Senata [1987a] aparecen resultados de un estudio de simulación y una aplicación en finanzas de este método.

4.8 Pruebas de Bondad de Ajuste

Supongamos que $X_1, X_2, ..., X_n$ es una muestra aleatoria de una variable aleatoria con función de distribución F y función característica φ desconocidas. Supongamos que deseamos probar la hipótesis

$$H_0 : F_X(x) = F_0(x) \tag{4.14}$$

donde $F_0(x)$ es alguna distribución conocida particular con función característica $\varphi_0(t)$. El problema de probar (4.14) es conocido como **problema de bondad de ajuste** y cualquier prueba de (4.14) es llamada **prueba de bondad de ajuste**. Si el lector no está familiarizado con los métodos de bondad de ajuste puede consultar D'agostino y Stephens [1986] o Kendall y Stuart [1987, c. 30].

Gran parte de la pruebas de bondad de ajuste comunes en literatura requieren evaluar $F_0(x)$ para varios valores de x. Esta tarea se dificulta cuando $F_0(x)$ carece de una expresión cerrada y no se cuenta con tablas o algoritmos adecuados.

Debido a la correspondencia unívoca existente entre una función de distribución y su función característica es razonable tratar de construir pruebas de bondad de ajuste basadas en la función característica. Las pruebas basadas en la función característica serán útiles cuando φ_0 sea más fácil de evaluar que F_0.

En está sección describimos dos pruebas de bondad de ajuste basadas en la función característica empírica. La primera fue propuesta por Koutrouvelis [1980a] y la segunda por Koutrouvelis y Kellermeier [1981].

4.8.1 Prueba de Bondad de Ajuste Simple

Sea $X_1, X_2, ..., X_n$ es una muestra aleatoria de tamaño n de una variable aleatoria con función de distribución F y función característica φ desconocidas. Supongamos que F_0 es una función de distribución conocida con función característica φ_0. Denotemos por φ_n la función característica empírica calculada con la muestra.

Denotemos por C_l la parte real y por S_l la parte compleja de φ_l para $l = 0, n$. Dados k

números reales $t_1, ..., t_k$ definimos $\xi_l = \xi_l(t_1, ..., t_k)$ como el vector $2k$−dimensional

$$\xi_l^t := (C_l(t_1), ..., C_l(t_k), S_l(t_1), ..., S_l(t_k));$$

además, para $l = 0, n$, definimos $\Omega_l = \left(\omega_{jm}^l\right)$ como la matriz simétrica de $2k \times 2k$ con elementos

$$\omega_{jk}^l := \begin{cases} C_l(t_j - t_m) + C_l(t_j + t_m) - 2C_l(t_j)C_l(t_m) & \text{si } 1 \leq j, m \leq k \\ C_l(t_j - t_m) - C_l(t_j + t_m) - 2S_l(t_j)S_l(t_m) & \text{si } m < j, m \leq 2k \\ S_l(t_j + t_m) - S_l(t_j - t_m) - 2C_l(t_j)S_l(t_m) & \text{si } 1 \leq j \leq k \text{ y } k < m \leq 2k. \end{cases}$$

Se sigue del teorema del límite central para el caso multivariado, (143), que

$$\sqrt{2n}(\xi_n - \xi_0) \xrightarrow{d} N_{2k}(\mathbf{0}, \Omega_0).$$

Si ahora suponemos que $t_1, ..., t_m$ han sido elegidos de tal manera que Ω_0 es no singular, entonces, por el teorema A.1.5, p.139,

$$2n(\xi_n - \xi_0)^t \Omega_0^{-1}(\xi_n - \xi_0) \xrightarrow{d} \chi_{2m}^2.$$

Esto demuestra el

Teorema 4.8.1 *Si $t_1, ..., t_m$ son números reales tales que Ω_0 es no singular, entonces la forma cuadrática*

$$\Omega_n^0 = 2n(\xi_n - \xi_0)^t \Omega_0^{-1}(\xi_n - \xi_0)$$

tiene, bajo H_0, como distribución asintótica la distribución $Ji - Cuadrada\,(2m)$.

Basándose en este resultado Koutrouvelis [1980a] propuso, para n grande, probar la hipótesis

$$H_0 : \varphi(t) = \varphi_0(t), \qquad \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$$

con tamaño aproximado de α tomando como regla de decisión

$$\text{Rechazar } H_0 \text{ si } \Omega_n^0 > \chi_{\alpha,2m}^2$$

donde $\chi_{\alpha,2m}^2$ es el cuantíl $1 - \alpha$ de la distribución $Ji - Cuadrada\,(2m)$.

La desventaja de esta prueba es que depende fuertemente de la elección del número m de puntos así como de su localización, ya que, para valores grandes de m, el procedimiento resulta computacionalmente intratable debido al número de cálculos necesarios para invertir Ω_0, así como problemas de precisión numérica en el calculo de Ω_0^{-1}; además, Ω_n^0 converge muy lentamente a χ_{2m}^2 y entonces sólo que n sea muy grande la prueba será útil. También la matriz Ω_0 es singular si $t_j = 0$ para algún índice j, o si para al menos dos distintos valores del índice, digamos j y k, se tiene que $t_j = |t_k|$.

Comparada con la prueba de Kolmogorov-Smirnov esta prueba tiene la ventaja de ser aplicable a distribuciones de tipo continuo y discreto, mientras que la prueba de Kolmogorov-Smirnov en su forma original sólo es aplicable a distribuciones continuas. A comparación de la prueba Ji-cuadrada de Pearson, cuando la distribución hipotética es continua, la prueba presentada no pierde información al agrupar los datos en intervalos.

4.8.2 Prueba de Bondad de Ajuste Compuestas

En la sección 4.8.1 se desarrolló la prueba propuesta por Koutrouvelis [1980a] para probar la hipótesis simple

$$H_0 : \varphi(t) = \varphi_0(t)$$

tomando como estadística de prueba la forma cuadrática

$$Q_n^0 = 2n(\xi_n - \xi_0)^t \Omega_0^{-1} (\xi_n - \xi_0)$$

donde ξ_l es el vector $2m$-dimensional

$$\xi_l^t = (C_l(t_1), ..., C_l(t_m), S_l(t_1), ..., S_l(t_m))$$

para $l = 0, n$ y $\frac{1}{2n}\Omega_0$ es la matriz de covarianza de ξ_n bajo H_0.

Koutrouvelis y Kellermeier [1981] extendieron esta prueba para probar la hipótesis compuesta

$$H_0 : \varphi(t) = \varphi(t;\theta), \qquad \theta \in \Theta \subset \mathbb{R}^p$$

donde $\varphi(t;\theta)$ es un miembro de una familia específica de funciones características, pero el parámetro θ es desconocido y es necesario estimarlo. A continuación damos una descripción de esta prueba.

Sea $X_1, ..., X_n$ una muestra aleatoria de una población con función característica $\varphi(t) = C(t) + iS(t)$ y sea

$$\widehat{\varphi}_n(t) := \frac{1}{n}\sum_{j=1}^{n} \exp(itX_j)$$

la función característica empírica.

Sea $\varphi(t;\theta)$ una función característica con parte real $C(t;\theta)$ y parte imaginaria $S(t;\theta)$. Definimos a $\xi(\theta)$ como el vector $2m$-dimensional

$$\xi^t(\theta) := (C(t_1;\theta), ..., C(t_m;\theta), S(t_1;\theta), ..., S(t_m;\theta))$$

y $\Omega(\theta) = \omega_{jk}(\theta)$ como la matriz simétrica de dimensión $2m \times 2m$ cuyo jk-ésimo elemento está definido por

$$\omega_{jk}(\theta) := \begin{cases} C(t_j + t_k; \theta) + C(t_j - t_k; \theta) - 2C(t_j; \theta) C(t_k; \theta) & \text{si } 1 \leq j, k \leq m \\ C(t_j - t_k; \theta) - C(t_j + t_k; \theta) - 2S(t_j; \theta) S(t_k; \theta) & \text{si } m < j, k \leq 2m \\ S(t_j + t_k; \theta) - S(t_j - t_k; \theta) - 2C(t_j; \theta) S(t_k; \theta) & \text{si } 1 \leq j \leq m \text{ y } m < k \leq 2m \end{cases}$$

donde $t_j = t_{j-m}$ para $m + 1 \leq j \leq 2m$.

Deseamos probar la hipótesis compuesta

$$H_0 : \varphi(t) = \varphi(t; \theta), \qquad \theta \in \Theta \subset \mathbb{R}^p.$$

Debido a que el parámetro θ es desconocido, resulta necesario estimarlo. Ya que bajo H_0, $\Omega(\theta) = 2n \operatorname{var}(\xi_n)$, Koutrouvelis y Kellermeier [1981] proponen estimar θ mediante el procedimiento $k - L$, que se describió en la sección 4.7.4, es decir proponen usar como estimador de θ la estadística $\widehat{\theta}_n$ que minimiza la forma cuadrática

$$L_n(\theta) := (\xi_n - \xi(\theta))^t \, \Omega^{-1}(\theta) (\xi_n - \xi(\theta)).$$

Entonces, siguiendo a Koutrouvelis [1980a], la elección de $\widehat{\theta}_n$ como estimador de θ, sugiere que la estadística de prueba adecuada es la forma cuadrática Q_n definida como

$$Q_n := 2n \left(\xi_n - \xi\left(\widehat{\theta}_n\right)\right)^t \Omega^{-1}\left(\widehat{\theta}_n\right) \left(\xi_n - \xi\left(\widehat{\theta}_n\right)\right).$$

Desafortunadamente, para determinar el máximo de $L_n(\theta)$ se requiere encontrar la inversa de $\Omega(\theta)$; la cual depende del parámetro desconocido θ. Por ello Koutrouvelis y Kellermeier [1981] sugieren que puede resultar más simple reemplazar $\Omega(\theta)$ por la matriz $\Omega_n = \left(\omega_{jk}^n\right)$ definida como

$$\omega_{jk}^n := \begin{cases} C_n(t_j + t_k;) + C_n(t_j - t_k) - 2C_n(t_j) C_n(t_k) & \text{si } 1 \leq j, k \leq m \\ C_n(t_j - t_k) - C_n(t_j + t_k) - 2S_n(t_j) S_n(t_k) & \text{si } m < j, k \leq 2m \\ S_n(t_j + t_k) - S_n(t_j - t_k) - 2C_n(t_j) S_n(t_k) & \text{si } 1 \leq j \leq m \text{ y } m < k \leq 2m \end{cases}$$

donde $t_j = t_{j-m}$ para $m + 1 \leq j \leq 2m$, $C_n(t) := \frac{1}{n}\sum_{j=1}^n \cos(tX_j)$, $S_n(t) := \frac{1}{n}\sum_{j=1}^n \operatorname{sen}(tX_j)$ y en lugar de $\widehat{\theta}_n$, tomar como estimador de θ a la estadística $\widetilde{\theta}_n$ que minimiza la forma cuadrática

$$L_n'(\theta) := (\xi_n - \xi(\theta))^t \, \Omega_n^{-1} (\xi_n - \xi(\theta)).$$

Entonces la estadística de prueba adecuada es

$$Q_n' := 2n \left(\xi_n - \xi\left(\widetilde{\theta}_n\right)\right)^t \Omega^{-1}\left(\widetilde{\theta}_n\right) \left(\xi_n - \xi\left(\widetilde{\theta}_n\right)\right).$$

Ahora es necesario encontrar la distribución, al menos asintótica, de las estadísticas de prueba. Para ello supongamos que $\theta_0 = (\theta_1, ..., \theta_p)$ es el verdadero valor de θ bajo H_0 y definamos Z como la matriz

$$Z := \begin{pmatrix} \frac{\partial C(t_1;\theta_0)}{\partial \theta_1} & \cdots & \frac{\partial C(t_1;\theta_0)}{\partial \theta_p} \\ \vdots & & \vdots \\ \frac{\partial C(t_m;\theta_0)}{\partial \theta_1} & \cdots & \frac{\partial C(t_m;\theta_0)}{\partial \theta_p} \\ \frac{\partial S(t_1;\theta_0)}{\partial \theta_1} & \cdots & \frac{\partial S(t_1;\theta_0)}{\partial \theta_p} \\ \vdots & & \vdots \\ \frac{\partial S(t_m;\theta_0)}{\partial \theta_1} & \cdots & \frac{\partial S(t_m;\theta_0)}{\partial \theta_p} \end{pmatrix}$$

la distribución asintótica de Q_n y Q'_n la da el siguiente

Teorema 4.8.2 *Si los m valores $t_1, ..., t_m$ son elegidos de manera tal que $\Omega(\theta_0)$ sea invertible y la matriz Z tenga rango $p < 2m$, entonces, bajo H_0, Q_n y Q'_n tienen distribución asintótica $Ji - cuadrada\,(2m - p)$.*

Demostración. Ver Koutrouvelis y Kellermeier [1981] ■

El teorema 4.8.2 sugiere que la hipótesis

$$H_0 : \varphi(t) = \varphi(t; \theta), \qquad \theta \in \Theta \subset \mathbb{R}^p$$

puede probarse con tamaño aproximado α tomando como regla de decisión

$$\text{Rechazar } H_0 \text{ si } Q_n > \chi^2_{\alpha,2m-p}$$

donde $\chi^2_{\alpha,2m-p}$ es el cuantil $1 - \alpha$ de la distribución $Ji - cuadrada\,(2m - p)$.

Capítulo 5

Conclusiones

La función característica tiene diversas aplicaciones en la probabilidad y en la estadística matemática. En el capítulo 4 se han mencionado aquellas que pueden ser analizadas con un nivel matemático intermedio.

En la estadística matemática a menudo es de interés (cuando se necesitan estudiar las propiedades de un estimador, establecer límites de confianza o encontrar regiones de rechazo) conocer la distribución de una estadística. Las funciones características pueden ser usadas para dar solución a este tipo de problemas, en especial cuando la estadística es una suma de variables aleatorias independientes. Este tipo de aplicaciones fue ilustrada en la sección 4.2 . Debido a su importancia en estadística, se estudió el caso particular cuando la distribución de la cual se ha obtenido la muestra es una distribución Normal. Otras aplicaciones de la función característica a problemas de distribución pueden encontrarse en Lukacs y Laha [1964].

En la sección 4.2 se vió como las funciones características pueden ser de utilidad al estudiar el comportamiento límite de variables aleatorias. Para ello se demostraron teoremas límites bien conocidos (ley débil de los grandes números, teorema del límite central y teorema de Poisson) por el método de las funciones características. El método de las funciones características para probar teoremas límite fue propuesto por Lyapunov a principios de siglo XX y es considerada una de las primeras aplicaciones sistemáticas de las funciones características. Pruebas de otros teoremas límites usando funciones características, pueden encontrarse en Kawata [1972].

En la sección 4.4 se dió la definición y se demostraron algunas propiedades de las distribuciones infinitamente divisibles. Las distribuciones infinitamente divisibles han sido ampliamente estudiadas por varios autores y son de importancia no sólo por sus propiedades analíticas sino también por sus aplicaciones en áreas como teoría del riesgo. Estas distribuciones son un claro ejemplo de como las funciones características pueden ser de utilidad, o hasta indispensables, para estudiar ciertas familias de distribuciones con bastante facilidad. Por esta última razón es que las distribuciones infinitamente divisibles han tenido un lugar dentro de la presente tesis.

Otro ejemplo de distribuciones que se han estudiado haciendo amplio uso de las funciones características son las distribuciones estables, que fueron estudiadas en la sección 4.5. Para ambas familias de distribuciones, estables e infinitamente divisibles , existen caracterizaciones en términos de su función característica, me refiero a la factorización de Lévy-Khinchine, que, principalmente para las distribuciones estables, ha conseguido que estas tengan un lugar dentro de las aplicaciones de la estadística y probabilidad a diferentes disciplinas como finanzas, economía, astronomía y física entre otras.

En la sección 4.6 se definió la función característica empírica y se demostraron sus propiedades más elementales. La función característica empírica fue introducida por primera vez por Parzen[1] [1962] y a la fecha muchos investigadores como son Feueverger y McDunnough [1977], Press [1972b] , Heathcote [1972] han realizado diferentes propuestas para su uso en estadística. Otros investigadores, por ejemplo Kent [1975], Csörgo[1981] y Marcus [1981], se han encargado de hacer estudios sobre el comportamiento límite de esta función.

También dentro de la literatura existen diferentes métodos de estimación de parámetros que hacen uso de la función característica. Estos métodos son de utilidad cuando la función característica de la distribución de interés es más fácil de evaluar que su función de distribución o su densidad, como es el caso cuando la distribución es estable. No existe un consenso sobre cual de estos métodos es el mejor y por ello en la sección 4.7 sólo se mencionaron los más conocidos. Otras propuestas sobre el uso de la función característica en problemas de estimación pueden encontrarse en Koutrouvelis [1980b], Koutrouvelis [1981] Koutrouvelis [1982].

El primero de los métodos de estimación enunciado en la sección 4.7 es el método de la mínima distancia, que fue propuesto por Press [1972b]. Este método se usa para estimar parámetros de distribuciones estables y propone como estimador a la estadística que minimiza la función

$$g(a,\gamma,\alpha,\beta) := \sup_t |\widehat{\varphi}_n(t) - \varphi(t)|;$$

donde $\varphi(t)$ es la función característica teórica y $\widehat{\varphi}_n(t)$ su contraparte empírica.

El segundo método de estimación estudiado fue el método de los momentos también propuesto por Press [1972b]. Este método obtiene estimadores fuertemente consistentes. Sin embargo, la elección óptima de los puntos en los que se debe evaluar $\widehat{\varphi}_n(t)$ sigue siendo un problema abierto.

[1] Según Feuerverger y Mureika [1977]

Los estimadores de error cuadrático integrado fueron definidos en el apartado 4.7.3. Estos estimadores fueron introducidos por Heathcote[1977] y son un caso especial de otro método de estimación propuesto previamente por Press [1972b]. Además en este apartado se incluye la demostración de un teorema debido a Heathcote que establece que estos estimadores son fuertemente consistentes y, bajo condiciones de regularidad adecuadas, asintóticamente Normales.

En el apartado 4.7.4 se introdujo el procedimiento $k - L$ propuesto por Feueverger y McDunnough [1981]. El procedimiento $k - L$ obtiene estimadores aplicando el método de máxima verosimilitud a la distribución asintótica de k puntos de la función característica empírica; de ahí el nombre de procedimiento $k - L$. Feueverger y McDunnough [1981] también demostraron que los estimadores obtenidos mediante el procedimiento $k - L$ son altamente eficientes, consistentes y asintóticamente Normales.

Otro método de estimación, propuesto por Madan y Senata [1987a], aparece en el apartado 4.7.5. Este método consiste en aproximar la función de densidad de una transformación de la variable aleatoria mediante polinomios de Chebyshev, y entonces aplicar el método de máxima verosimilitud a esta aproximación.

Finalmente, en la sección 4.8 se presentaron dos pruebas de Bondad de Ajuste basadas en la función característica empírica. La primera, debida a Koutrouvelis[1980a], prueba la hipótesis simple

$$H_0 : \varphi(t) = \varphi_0(t) .$$

La segunda, debida a Koutrouvelis y Kellermeier [1981], prueba la hipótesis compuesta

$$H_0 : \varphi(t) = \varphi(t;\theta) ; \qquad \theta \in \Theta \subset \mathbb{R}^p,$$

donde el parámetro θ es desconocido y por lo tanto es necesario estimarlo.

Además de las aplicaciones que se presentaron en esta tesis se pueden encontrar muchas más aplicaciones de las funciones características. A continuación se mencionan algunas de ellas.

Aplicaciones en problemas de Regresión se pueden encontrar en Lukacs y Laha [1964]. Algunas aplicaciones en procesos estocásticos se pueden consultar en Press [1968] y Lukacs y Laha [1964]. Baglivo et al. [1992], usan la función característica en problemas de bondad de ajuste y análisis de tablas de contingencia. En Epps [1987] y Epps [1988] se usa la función característica para probar que una serie de tiempo es Gausiana y estacionaria. Baringhaus y Henze [1988], Epps y Pulley [1983] y Hall y Welsh [1983], hacen propuestas sobre el uso de la función característica empírica para probar normalidad; además en Baringhaus, Danschke y Henze [1989] aparece un estudio comparativo de pruebas de normalidad en el cual se argumenta que los mejores métodos disponibles para probar normalidad, univariada y multivariada, son aquellos que hacen uso de la función característica. También para probar exponencialidad existe una propuesta por Epps y Pulley [1986]. Feuerverger y Mureika [1977], usan la función característica empírica para probar simetría. Jones y Lotwick [1983], utilizan la función característica para hacer estimaciones tipo kernel. Epps y Singleton [1986] han estudiado el problema de probar la homogeneidad de dos muestras usando la función característica.

Si el lector esta interesado en realizar estudios de simulación de los métodos presentados en las secciones 4.7 y 4.8 probablemente le sea de utilidad consultar Chambers, et al. [1976] y Madan y Senata [1987b].

Apéndice A

Convergencia Estocástica

En la teoría de la probabilidad existen diferentes tipos de convergencia. La finalidad de este apéndice es proporcionar las definiciones básicas y los resultados más elementales relacionados con el comportamiento límite de variables aleatorias. Existen muchos libros sobre el tema. Para el lector interesado en un estudio más detallado se le recomienda consultar Lukacs [1968] y Billingsley [1968].

A.1 Tipos de Convergencia y sus Relaciones

Definición A.1.1 *Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias definidas en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. La sucesión $\{X_n\}_{n\geq 1}$* ***converge en probabilidad*** *a X cuando $n \longrightarrow \infty$, si*

$$\lim_{n\to\infty} P\{|X_n - X| \geq \varepsilon\} = 0$$

para toda $\varepsilon > 0$. Denotamos esto por $X_n \xrightarrow{P} X$.

La noción intuitiva de la convergencia en probabilidad es que, para n suficientemente grande, X_n está muy cerca de X con probabilidad muy alta.

Un tipo de convergencia más fuerte que la convergencia en probabilidad es la convergencia casi segura, también llamada convergencia con probabilidad uno.

Definición A.1.2 *Sea X una variable aleatoria y $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias definidas en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Decimos que la sucesión $\{X_n\}_{n\geq 1}$*

converge ***casi seguramente*** *a* X *cuando* $n \longrightarrow \infty$, *y lo denotamos como* $X_n \xrightarrow{c.s.} X$, *si existe un conjunto* $A \in \mathcal{A}$, *tal que* $P(A) = 0$ *y* $X_n \longrightarrow X$ *en* A^c. *Es decir*

$$P\left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\} = 1.$$

La relación entre los dos tipos de convergencia vistos se establece en el siguiente teorema.

Teorema A.1.1 *La convergencia casi segura implica convergencia en probabilidad.*

Demostración. Sea $\varepsilon > 0$. Supongamos que X_n converge casi seguramente a X. Entonces existe un conjunto A tal que $P(A) = 0$ y $X_n \longrightarrow X$ sobre A^c. Definimos D_n como el conjunto

$$D_n := \{|X_m - X| < \varepsilon,\ m \geq n\},$$

entonces claramente los D_n forma una sucesión creciente con límite $D := \bigcup_{n\geq 1} D_n$ y $A^c \subset D$. Entonces $P(D) = 1$ y $\lim_{n\to\infty} P(D_n) = 1$. Además

$$\{|X_m - X| \geq \varepsilon\} \subset D_n^c,$$

entonces

$$P\{|X_m - X| \geq \varepsilon\} \leq P(D_n^c)$$

de donde se sigue el resultado. ∎

Otro tipo de convergencia es la convergencia en r-media.

Definición A.1.3 *Sea* X *una variable aleatoria y* $\{X_n\}_{n\geq 1}$ *una sucesión de variables aleatorias definidas en un espacio de probabilidad* $(\Omega, \mathcal{A}, P)$. *Se dice que* $\{X_n\}_{n\geq 1}$ *converge en* $r-$***media*** *o en* L_r *a* X *cuando* $n \longrightarrow \infty$, *y se denota por* $X_n \xrightarrow{r} X$, *si para* $r > 0$ *se tiene que*

$$\lim_{n\to\infty} E\left[|X_n - X|^r\right] = 0.$$

Teorema A.1.2 *Si* $X_n \xrightarrow{r} X$, *entonces* $X_n \xrightarrow{P} X$.

Demostración. Sea $r > 0$ y $A_n = \{|X_n - X| \geq \varepsilon\}$. El resultado es trivial observando que

$$|X_n - X|^r \geq |X_n - X|^r I_{A_n} \geq \varepsilon^r I_{A_n}.$$

∎

Definición A.1.4 *Sean $\{F_n\}_{n\geq 1}$ las funciones de distribución de $\{X_n\}_{n\geq 1}$ y F la función de distribución de X. Decimos que X_n* ***converge en distribución****, o* ***débilmente****, a X cuando $n \longrightarrow \infty$, y lo expresamos como $X_n \xrightarrow{d} X$, si para todo x punto de continuidad de F*

$$\lim_{n\to\infty} F_n(x) = F(x).$$

Del siguiente resultado se concluye que la convergencia en distribución es el tipo de convergencia más débil.

Teorema A.1.3 *Supongamos que $X_n \xrightarrow{P} X$ entonces $X_n \xrightarrow{d} X$.*

Demostración. Sean F y F_n la función de distribución de X y de X_n respectivamente. Supongamos que x es punto de continuidad de F, entonces existe $\varepsilon > 0$ y dos números, digamos a y b, tales que $a < x < b$,

$$F(a) > F(x) - \varepsilon \qquad \text{y} \qquad F(b) < F(x) + \varepsilon . \tag{A.1}$$

Entonces

$$\begin{aligned} F_n(x) &= P(X_n \leq x) \\ &= P(X_n \leq x, X \leq b) + P(X_n \leq x, X > b) \\ &\leq P(X \leq b) + P(|X_n - X| \geq b - x); \end{aligned}$$

pero por hipótesis $X_n \xrightarrow{P} X$, entonces

$$\lim_{n\to\infty} F_n(x) \leq P(X \leq b) = F(b). \tag{A.2}$$

Así mismo, se puede demostrar que

$$\lim_{n\to\infty} F_n(x) \geq F(a). \tag{A.3}$$

Relacionando las ecuaciones $(A.1)$, $(A.2)$ y $(A.3)$ se sigue que, para n suficientemente grande,

$$F(x) - \varepsilon < F(a) \leq F_n(x) \leq F(b) < F(x) + \varepsilon ;$$

por lo tanto, haciendo $\varepsilon \downarrow 0$ obtenemos

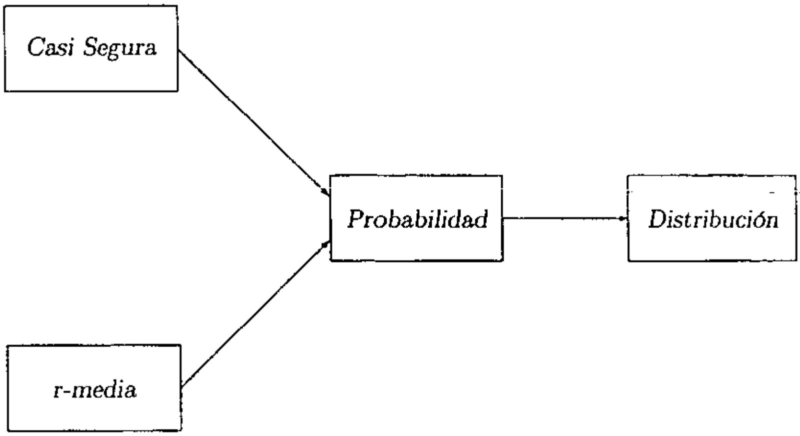
$$\lim_{n\to\infty} F_n(x) = F(x)$$

que es lo que se quería demostrar. ∎

De los teoremas A.1.1, A.1.2 y A.1.3 se sigue el

Corolario A.1.4 *Convergencia casi segura y convergencia en r-media implican convergencia en distribución.*

Los resultados anteriores se resumen en el siguiente diagrama.



Los recíprocos de los teoremas A.1.1, A.1.2 y A.1.3 no se cumplen en general; además la convergencia casi-segura no implica convergencia en r-media ni viceversa. Algunos contraejemplos pueden encontrarse en Romano y Siegel [1986, cap. 5] o Stoyanov [1988, cap. 3]. Condiciones para la validez de los resultados recíprocos aparecen en Clarke [1975, cap. 7] y Serfling [1980, sec. 1.3].

Definición A.1.5 *Sea $\{\mathbf{X}_n\}_{n\geq 1}$ una sucesión de vectores aleatorios k-dimensionales definidos en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Decimos que la sucesión $\{\mathbf{X}_n\}_{n\geq 1}$ **converge en probabilidad** al vector aleatorio $\mathbf{X}$ cuando $n \longrightarrow \infty$, y lo denotamos como $\mathbf{X}_n \xrightarrow{P} \mathbf{X}$, si*

$$\|\mathbf{X}_n - \mathbf{X}\| \xrightarrow{P} 0,$$

en el sentido de la definición A.1.1, donde, para un vector k-dimensional $\mathbf{z}$,

$$\|\mathbf{z}\| = \left(\sum_{j=1}^{k} z_j^2\right)^{\frac{1}{2}}.$$

La definición A.1.5, extiende la definición de convergencia en probabilidad al caso de vectores aleatorios. La extensión de la definición de los otros tipos de convergencia al caso multivariado es inmediata a partir de la definición para el caso univariado.

El siguiente teorema es de mucha utilidad cuando se desea encontrar la distribución asintótica para algunas funciones de variables aleatorias.

Teorema A.1.5 *Sean* $\mathbf{X}, \mathbf{X}_1, \mathbf{X}_2, \ldots,$ *vectores aleatorios k-dimensionales definidos en algún espacio de probabilidad y sea g una función de Borel continua definida en* $\mathbb{R}^k$.

1. *Si* $\mathbf{X}_n \xrightarrow{c.s.} \mathbf{X}$, *entonces* $g(\mathbf{X}_n) \xrightarrow{c.s.} g(\mathbf{X})$.
2. *Si* $\mathbf{X}_n \xrightarrow{P} \mathbf{X}$, *entonces* $g(\mathbf{X}_n) \xrightarrow{P} g(\mathbf{X})$.
3. *Si* $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$, *entonces* $g(\mathbf{X}_n) \xrightarrow{d} g(\mathbf{X})$.

Demostración. Ver Serfling [1980, p. 24]. ■

En particular, si $\mathbf{A}$ y $\mathbf{B}$ son matrices de $m \times k$ y $k \times k$ respectivamente, como

$$\mathbf{A}\mathbf{x}' = \left(\sum_{j=1}^{k} a_{1j}x_1, \ldots, \sum_{j=1}^{k} a_{mj}x_m \right)$$

y

$$\mathbf{x}\mathbf{B}\mathbf{x}' = \sum_{r=1}^{k} \sum_{j=1}^{k} b_{rj}x_r x_j$$

son funciones continuas de $\mathbf{x}$, tenemos el

Corolario A.1.6 *Sean* $\mathbf{A}$ *y* $\mathbf{B}$ *son matrices de* $m \times k$ *y* $k \times k$ *respectivamente y* $\mathbf{X}, \mathbf{X}_1, \mathbf{X}_2, \ldots,$ *vectores aleatorios k-dimensionales. Si la sucesión* $\{\mathbf{X}_n\}_{n\geq 1}$ *converge a* $\mathbf{X}$ *en distribución, en probabilidad o casi seguramente, entonces* $\mathbf{A}\mathbf{X}_n'$ *y* $\mathbf{X}\mathbf{B}\mathbf{X}_n'$ *convergen a* $\mathbf{A}\mathbf{X}'$, $\mathbf{X}\mathbf{B}\mathbf{X}'$ *respectivamente en el mismo tipo de convergencia que* $\{\mathbf{X}_n\}_{n\geq 1}$ *converge a* $\mathbf{X}$.

A.2 Ley de los Grandes Números

Existen dos distintos tipos de leyes de los grandes números, "fuerte" (c.s.) y "débil" (en probabilidad), dependiendo del tipo de convergencia.

Definición A.2.1 *Sea* $\{X_n\}_{n\geq 1}$ *una sucesión de variables aleatorias, tales que* $E(X_n) < \infty$ *para toda n entero positivo y sea* $S_n := \sum_{n=1}^{\infty} X_n$. *Decimos que* $\{X_n\}_{n\geq 1}$ *obedece la* ***ley débil de los grandes números*** *si*

$$\frac{S_n - E(S_n)}{n} \xrightarrow{P} 0.$$

Si se tiene que

$$\frac{S_n - E(S_n)}{n} \xrightarrow{c.s.} 0,$$

entonces decimos que $\{X_n\}_{n\geq 1}$ obedece la ***ley fuerte de los grandes números****.*

Se sigue inmediatamente del teorema A.1.1 que la ley fuerte implica la ley débil.

A continuación mencionamos un caso particular de sucesiones que obedecen las leyes de los grandes números. Condiciones más generales pueden encontrarse en Chow y Teicher [1988, sec. 5.2].

Teorema A.2.1 (Kolmogorov) *Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con valor esperado finito. Entonces $\{X_n\}_{n\geq 1}$ obedece la ley fuerte de los grandes números.*

Demostración. Ver Chow y Teicher [1988, p.125]. ■

Corolario A.2.2 *Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con valor esperado finito. Entonces $\{X_n\}_{n\geq 1}$ obedece la ley débil de los grandes números.*

A.3 Teorema del Límite Central

Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias con valor esperado $E(X_n) = \mu_n$ y varianza $\operatorname{var}(X_n) = \sigma_n^2 < \infty$. Definamos Z_n como la variable aleatoria

$$Z_n := \frac{\sum_{j=1}^{n}(X_n - \mu_n)}{\sqrt{\sum_{j=1}^{n}\sigma_n^2}}.$$

El **problema del límite central** consiste en encontrar condiciones sobre $\{X_n\}_{n\geq 1}$, $\{\mu_n\}_{n\geq 1}$ y $\{\sigma_n^2\}_{n\geq 1}$ para que

$$Z_n \xrightarrow{d} Z$$

donde Z es una variable aleatoria distribuida Normalmente con media cero y varianza unitaria.

La primera solución a este problema está dada por el teorema de De Moivre- Laplace.

Teorema A.3.1 (De Moivre-Laplace) *Si $\{X_n\}_{n\geq 1}$ es una sucesión de variables aleatorias independientes con distribución Bernoulli (p), entonces*

$$\frac{\sum_{j=1}^{n} X_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} Z,$$

donde $Z \sim Normal\,(0,1)$.

Actualmente existen muchas formulaciones del teorema del límite central, aunque la más conocida es el teorema del límite central de Lindeberg, probablemente la más usada es la formulación de Lévy que se demostró en la sección 4.3.

Definición A.3.1 *Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias con esperanza $E(X_n) = \mu_n$ y varianza $\operatorname{var}(X_n) = \sigma_n^2 < \infty$ para todo n entero positivo. Si F_n denota a la función de distribución de X_n y $V_n^2 := \sum_{j=1}^{n} \sigma_j^2$, se dirá que $\{X_n\}_{n\geq 1}$ cumple la* **condición de Lindeberg** *si, para toda $\varepsilon > 0$,*

$$\frac{1}{V_n^2} \sum_{j=1}^{n} \int_{\{X:|X-\mu_j|\geq \varepsilon V_n\}} |X-\mu_j|^2 \, dF_j(x) \longrightarrow 0$$

cuando $n \longrightarrow \infty$.

Teorema A.3.2 (Lindeberg) *Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias con esperanza $E(X_n) = \mu_n$ y varianza $\operatorname{var}(X_n) = \sigma_n^2 < \infty$ para todo n entero positivo. Si $\{X_n\}_{n\geq 1}$ cumple la condición de Lindeberg, entonces*

$$\frac{S_n - E(S_n)}{\sqrt{\operatorname{var}(S_n)}} \xrightarrow{d} Z$$

donde $Z \sim Normal\,(0,1)$.

Una Condición que frecuentemente resulta más sencilla de verificar es la condición de Lyapunov.

Definición A.3.2 *Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias con valor esperado $E(X_n) = \mu_n$ y varianza $\operatorname{var}(X_n) = \sigma_n^2 < \infty$ para todo n entero positivo. Si F_n denota a la función de distribución de X_n y $V_n^2 := \sum_{j=1}^{n} \sigma_j^2$, se dirá que $\{X_n\}_{n\geq 1}$ cumple la* **condición de Lyapunov** *si, para toda $\varepsilon > 0$ y $\delta > 0$,*

$$\frac{1}{V_n^{2+\delta}} \sum_{j=1}^{n} E|X_j - \mu_j|^{2+\delta} \longrightarrow 0$$

cuando $n \longrightarrow \infty$.

Teorema A.3.3 *La condición de Lyapunov implica la condición de Lindeberg.*

Demostración.

$$\begin{aligned} E\left|X_j-\mu_j\right|^{2+\delta} &= \int_{-\infty}^{\infty}\left|x-\mu_j\right|^{2+\delta}dF_j(x) \\ &\geq \int_{\{X:|X-\mu_j|\geq\epsilon V_n\}}\left|x-\mu_j\right|^{\delta}\left(x-\mu_j\right)^2 dF_j(x) \\ &\geq \epsilon^{\delta}V_n^{\delta}\int_{\{X:|X-\mu_j|\geq\epsilon V_n\}}\left(x-\mu_j\right)^2 dF_j(x). \end{aligned}$$

Por lo tanto,

$$\frac{1}{V_n^2}\sum_{j=1}^{n}\int_{\{X:|X-\mu_j|\geq\epsilon V_n\}}\left|X-\mu_j\right|^2 dF_j(x)\leq\frac{1}{\epsilon^{\delta}V_n^{2+\delta}}\sum_{j=1}^{n}E\left|X_j-\mu_j\right|^{2+\delta}.$$

■

Corolario A.3.4 (Lyapunov) *Sea* $\{X_n\}_{n\geq 1}$ *una sucesión de variables aleatorias con valor esperado* $E(X_n)=\mu_n$ *y varianza* $\operatorname{var}(X_n)=\sigma_n^2<\infty$ *para todo n entero positivo. Si* $\{X_n\}_{n\geq 1}$ *cumple la condición de Lyapunov, entonces*

$$\frac{S_n-E(S_n)}{\sqrt{\operatorname{var}(S_n)}}\xrightarrow{d}Z$$

donde $Z\sim Normal(0,1)$.

A continuación se menciona el teorema del límite central multivariado; el lector interesado en la demostración puede consultarla en Varadarajan [1958].

Teorema A.3.5 *Sea* $\{\mathbf{X}_n\}_{n\geq 1}$ *una sucesión de vectores aleatorios k-dimensionales con funciones de distribución* $\{F_{\mathbf{X}_n}(\mathbf{x})\}_{n\geq 1}$. *Sea* $F_{Z_n}(\mathbf{x})$ *la función de distribución de* $Z_n=\boldsymbol{\lambda}'\mathbf{X}_n$, *donde* $\boldsymbol{\lambda}$ *es un vector fijo. Una condición necesaria y suficiente para que* $F_{\mathbf{X}_n}(\mathbf{x})$ *converja a la función de distribución k-dimensional* $F_{\mathbf{Z}}(\mathbf{z})$ *es que* $F_{Z_n}(\mathbf{x})$ *converja a algún límite para cada* $\boldsymbol{\lambda}$.

El siguiente resultado es una versión más débil del teorema A.3.5 que es de mucha utilidad.

Teorema A.3.6 *Sea $\{\mathbf{X}_n\}_{n\geq 1}$ una sucesión de vectores aleatorios k-dimensionales independientes e idénticamente distribuidos con valor esperado $E(\mathbf{X}_j) = \boldsymbol{\mu}$, y matriz de covarianza Σ. Entonces, si Σ es positiva definida,*

$$\frac{\sum_{j=1}^{n}(\mathbf{X}_j - \boldsymbol{\mu})}{\sqrt{n}} \xrightarrow{d} \mathbf{Z}$$

donde $\mathbf{Z} \sim N_k(\mathbf{0}, \Sigma)$.

Demostración. Cramér [1974, sec. 21.11]. ■

A.4 Teoremas de Helly y Prohorov

En esta sección presentamos dos teoremas importantes relacionados con convergencia débil. El primero es un caso particular del teorema de Helly que proporciona una caracterización de la convergencia en distribución. El segundo es un caso particular del teorema de Prohorov, este establece condiciones necesarias y suficientes bajo las cuales, dada una sucesión de funciones de distribución, podemos extraer una subsucesión que converge débilmente a una función de distribución. Las demostraciones pueden encontrase en Clarke [1975, p. 117] y Shiryayev [1984, p. 315] respectivamente.

Teorema A.4.1 (Teorema de Helly) *Supongamos que $X, X_1, X_2, \ldots$ son variables aleatorias. Entonces*

$$X_n \xrightarrow{d} X$$

si, y sólo si,

$$\lim_{n\longrightarrow\infty} E[f(Xn)] = E[f(X)]$$

para toda función continua y acotada $f : \mathbb{R} \longrightarrow \mathbb{R}$.

Teorema A.4.2 (Teorema de Prohorov) *Sea $\{F_n\}_{n\geq 1}$ una sucesión de funciones de distribución. Una condición necesaria y suficiente para que exista una subsucesión $\{F_{n_k}\}_{k\geq 1} \subset \{F_n\}_{n\geq 1}$ tal que*

$$\lim_{k\longrightarrow\infty} F_{n_k} = F,$$

donde F es una función de distribución, es que, para cada $\varepsilon > 0$ exista un número $t = t(\varepsilon) > 0$ tal que

$$F_n(t) - F_n(-t) > 1 - \varepsilon,$$

para toda n suficientemente grande.

Apéndice B

Polinomios de Chebyshev

En este apéndice se presentan sólo las propiedades más elementales de los polinomios de Chebyshev. El lector interesado en conocer más sobre los polinomios de Chebyshev puede consultar Hamming [1973].

Existen diferentes maneras de definir los polinomios de Chebyshev $T_n(x)$. Probablemente la manera más sencilla es definiendo $T_n(x)$ como

$$T_n(x) := \cos(n \arccos x).$$

Consideremos la integral

$$\begin{aligned} I \quad : \quad &= \int_{-1}^{1} \frac{T_n(x)\, T_m(x)}{\sqrt{1-x^2}} dx \\ &= \int_{-1}^{1} \frac{\cos(n \arccos x)\cos(m \arccos x)}{\sqrt{1-x^2}} dx \end{aligned}$$

haciendo el cambio de variable $\theta = \arccos x$ y por la identidad

$$\cos A \cos B = \frac{1}{2}\left[\cos(A+B) + \cos(A-B)\right] \tag{B.1}$$

obtenemos

$$I = \int_{0}^{\pi} \cos(n\theta)\cos(m\theta)\, dx$$

$$
\begin{aligned}
&= \frac{1}{2}\int_0^{\pi} \left[\cos\left((m+n)\,\theta\right) + \cos\left((m-n)\,\theta\right)\right] \\
&= \frac{1}{2}\left[\frac{\operatorname{sen}\left((m+n)\,\theta\right)}{m+n} + \frac{\operatorname{sen}\left((m-n)\,\theta\right)}{m-n}\right]_{\theta=0}^{\pi} \qquad m \neq n \\
&= \begin{cases} 0 & \text{si } m \neq n \\ \frac{\pi}{2} & \text{si } m = n \neq 0 \\ \pi & \text{si } m = n = 0. \end{cases}
\end{aligned}
$$

Por lo tanto,

$$
\int_{-1}^{1} \frac{T_n(x)\,T_m(x)}{\sqrt{1-x^2}}\,dx = \begin{cases} 0 & \text{si } m \neq n \\ \frac{\pi}{2} & \text{si } m = n \neq 0 \\ \pi & \text{si } m = n = 0. \end{cases} \tag{B.2}
$$

Es decir los polinomios de Chebyshev forman un conjunto ortogonal.

Consideremos ahora la suma $T_{n+1}(x) + T_{n-1}(x)$. Nuevamente haciendo la sustitución $\theta = \arccos x$ y aplicando la identidad $(B.1)$ tenemos que

$$
\begin{aligned}
T_{n+1}(x) + T_{n-1}(x) &= \cos\left((n+1)\,\theta\right) + \cos\left((n+1)\,\theta\right) \\
&= 2\left[\cos n\theta\right]\left[\cos\theta\right] \\
&= 2x\cos\left(n\arccos x\right) \\
&= 2xT_n(x),
\end{aligned}
$$

entonces despejando $T_{n+1}(x)$ obtenemos la fórmula recursiva

$$
T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \tag{B.3}
$$

Con la ayuda de la fórmula $(B.3)$, podemos calcular de manera sencilla los valores de $T_n(x)$. Por ejemplo, los primeros cinco valores de $T_n(x)$ son

$$
\begin{aligned}
T_0(x) &= 1 \\
T_1(x) &= x \\
T_2(x) &= 2x^2 - 1 \\
T_3(x) &= 4x^3 - 1 \\
T_4(x) &= 8x^4 - 8x^2 + 1.
\end{aligned}
$$

También, a partir de la fórmula recursiva $(B.3)$, es fácil ver que, para $n \geq 1$,

$$
T_n(x) = 2^{n-1}x^n + \ldots
$$

y que $T_n(x)$ es un polinomio de grado par si n es par y si n es impar entonces $T_n(x)$ es un polinomio de grado impar.

Definamos

$$q_n(y) := \frac{T_n(y)}{2^{n-1}},$$

entonces es fácil ver de $(B.3)$ que

$$q_0(y) = 2, \quad q_1(y) = y, \quad q_{n+1}(y) = yq_n(y) - \frac{1}{4}q_{n-1}(y) \quad n \geq 1.$$

Apéndice C

Distribuciones de Probabilidad

En este apéndice se resumen las distribuciones de probabilidad más usuales y sus funciones características. Sólo se incluyen aquellas distribuciones que tienen una expresión cerrada para su función característica.

C.1 Distribuciones Discretas

Nombre	Función de Densidad	Parámetros	Función Característica
Degenerada	$I_{\{a\}}(x)$	$a \in \mathbb{R}$	e^{iat}
Bernoulli	$p^x (1-p)^{1-x} I_{\{0,1\}}(x)$	$p \in \,]0,1[$	$(1-p) + pe^{it}$
Binomial	$\binom{n}{x} p^x (1-p)^{n-x} I_{\{0,\ldots,n\}}(x)$	$p \in \,]0,1[$ $n = 1, 2, \ldots$	$[(1-p) + pe^{it}]^n$
Poisson	$\frac{e^{-\lambda}\lambda^x}{x!} I_{\{0,1,\ldots\}}(x)$	$\lambda > 0$	$\exp[\lambda(e^{it} - 1)]$
Geométrica	$p(1-p)^x I_{\{0,1,\ldots\}}(x)$	$p \in \,]0,1[$	$\frac{1-p}{1-p\exp(it)}$
Binomial Negativa	$\binom{r+x-1}{x} p^r (1-p)^x I_{\{0,1,\ldots\}}(x)$	$p \in \,]0,1[$ $r > 0$	$\left[\frac{1-p}{1-p\exp(it)}\right]^r$

C.2 Distribuciones Continuas

Nombre	Función de densidad	Parámetros	Función Característica
Uniforme	$\frac{1}{b-a} I_{[a,b]}(x)$	$a, b \in \mathbb{R}$ $a < b$	$\frac{e^{ibt} - e^{iat}}{(b-a)it}$
Triangular	$\frac{a - \lvert x \rvert}{a^2} I_{[-a,a]}(x)$	$a > 0$	$\left(\frac{\operatorname{sen} \frac{at}{2}}{\frac{at}{2}} \right)^2$
Exponencial	$\theta \exp(-\theta x) I_{]0,\infty[}(x)$	$\theta > 0$	$\frac{\theta}{\theta - it}$
Normal	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$\mu \in \mathbb{R}$ $\sigma > 0$	$e^{i\mu t - \frac{\sigma^2 t^2}{2}}$
Gamma	$\frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} I_{]0,\infty[}(x)$	$\lambda > 0$ $r > 0$	$\left[\frac{\lambda}{\lambda - it}\right]^r$
Cauchy	$\frac{1}{\pi\beta \left\{1 + [(x-\alpha)/\beta]^2\right\}}$	$\alpha \in \mathbb{R}$ $\beta > 0$	$e^{i\alpha t - \beta\lvert t \rvert}$
Laplace	$\frac{1}{2\beta} \exp\left(-\frac{\lvert x-\alpha \rvert}{\beta}\right)$	$\alpha \in \mathbb{R}$ $\beta > 0$	$\frac{e^{i\alpha t}}{1 + \beta^2 t^2}$
Ji-Cuadrada	$\frac{1}{\Gamma\left(\frac{k}{2}\right)} \left(\frac{1}{2}\right)^{\frac{k}{2}} x^{\frac{k}{2}-1} e^{-\frac{1}{2}x} I_{]0,\infty[}(x)$	$k = 1, 2, \ldots$	$\left[\frac{1}{1-2it}\right]^{\frac{k}{2}}$

Referencias

[1981] C. D. ALIPRANTIS, *Principles of Real Analysis,* Edward Arnold, London.

[1992] J. BAGLIVO, D. OLIVIER Y M. PAGANO, "Methods for exact goodness-of-fit test", Amer. Statis. Assoc., 87, 464-469.

[1989] L. BARINGHAUS, R. DANSCHKE Y N. HENZE, "Recent and Classical Test for Normality: a Comparative Study", *Communications in Statistics,* 18, 363-379.

[1988] L. BARINGHAUS Y N. HENZE, "A Consistent Test for Multivariate Normality Based on the Empirical Characteristic Function", *Metrika,* 35, 339-348.

[1968] P. BILLINGSLEY, *Convergence of Probability Measures,* Wiley, New York.

[1995] P. BILLINGSLEY, *Probability and Measure,* Wiley, New York..

[1995] L. BONDESSON, "Factorization theory for probability distributions", *Scand. Actuarial J.* 1:44-53.

[1976] J. M. CHAMBERS, C. L. MALLOWS Y B. W. STUCK, "A method for simulating stable random variables", *J. Amer. Statist. Assoc.*, 71, 340-344.

[1943] S. CHANDRASEKHAR, "Stochastic Problems in Physics and Astronomy", *Review of Modern Physics*, 15, 1-89.

[1988] Y. CHOW Y H. TEICHER, *Probability Theory,* Springer Verlag, New York.

[1968] K. L. CHUNG, *A Course in Probability Theory.* Academic Press, New York.

[1975] L. E. CLARKE, *Random Variables,* Longman, London.

[1974] H. CRAMÉR, *Mathematical Methods of Statistics,* Princeton University Press, Princeton.

[1981] S. CSÖRGO, "Limit behaviour of the empirical characteristic function based estimators", *Sankhya Ser. A*, 38, 109-325.

[1986] R. D'AGOSTINO Y M. STEPHENS, *Goodness of fit Techniques,* Marcel Dekker, New York.

[1989] R. M. DUDLEY, *Real Analysis and Probability,* Wadsworth, California.

[1993] T. W. EPPS, "Characteristic Functions and their Empirical Counterparts: Geometrical Interpretations and Applications to Statistical Inference", *Amer. Statist. Assoc.*, 47, 33-38.

[1987] T. W. EPPS, "Testing That a Stationary Time Series is Gaussian", *Annals of Stat.*, 15, 1683-1698.

[1988] T. W. EPPS, "Testing That a Gaussian Process is Stationary", *Annals of Stat.*, 16, 1667-1683.

[1983] T. W. EPPS Y L. B. PULLEY, "A Test for Normality Based on the Empirical Characteristic Function", *Biometrika*, 70, 3, 723-726.

[1986] T. W. EPPS Y L. B. PULLEY, "A Test of Exponentiality vs. Monotone-Hazard Alternatives Derived From the Empirical Characteristic Function", *J. Royal Statist. Soc.,* Ser. B, 48, 206-213.

[1986] T. W. EPPS Y K. J. SINGLETON, "An Omnibus Test for the Two-Sample Problem Using the Empirical Characteristic Function", *J. Statis. Computation and Simulation,* 26, 177-203.

[1976] P. D. FEIGEN Y C. R. HEATHCOTE, "The empirical characteristic function and the Cramér-von Mises statistic". *Sankhya Ser. A*, 38, 309-325.

[1971] W. FELLER, *An Introduction to Probability Theory and its Applications,* Vol. II, Wiley, New York.

[1977] A. FEUERVERGER Y R. A. MUREIKA, "The empirical characteristic function and its applications", *Ann. Statist.*, 5, 88-97.

[1981] A. FEUERVERGER Y P. MCDUNNOUGH, "On the efficiency of empirical characteristic function procedures", *J. Roy, Statist. Soc*, 76, No. 374, pp. 379-387.

[1984] A. FEUERVERGER Y R. A. MUREIKA, "On statistical transform methods and their efficiency", *The Can. J. of Statist.*, 12, No. 4, 303-317.

[1937] B. V. GNEDENKO, "Sur les functions caracterisques". *Bull. Univ. Moscou, Ser. Internat., Sect. A* **1**, 16-17.

[1954] B. V. GNEDENKO Y A.N. KOLMOGOROV, *Limit Distributions for Sums of Independent Random Variables,* Addison Wesley, Moscow.

[1973] B. V. GNEDENKO, *The Theory of Probability,* Mir Publishers, Moscow.

[1965] I. S. GRADSHTEYN Y I. M. RYZHIK, *Table of Integrals Series and Products,* Academic Press, New York.

[1983] P. HALL Y A. H. WELSH, "A test for normality based on the empirical characteristic function", *Biometrika,* 70, 2, 485-489.

[1950] P. HALMOS, *Measure Theory,* Van Nostrand, New York.

[1973] R. W. HAMMING, *Numerical Methods for Scientists and Engineers,* McGraw-Hill, New York.

[1966] B. HARRIS, *Theory of Probability,* Adisson Wesley, New York.

[1972] C. R. HEATHCOTE, "A test of goodness of fit for symmetric random variables", *Austral. J. Statist.,* 14, 172-181.

[1977] C. R. HEATHCOTE, "The integrated squared error estimation of parameters", *Biometrika,* 64, 255-264.

[1919] J. HOLTSMARK, *Ann. d. Physik,* 58, 577.

[1983] M. C. JONE Y H. W. LOTWICK, "On the errors involved in computing the empirical characteristic function", *J. Statist. Comput. Simul.,* 17, 133-149.

[1972] T. KAWATA, *Fourier Analysis in Probability Theory,* Academic Press, New York.

[1987] M. G. KENDALL Y A. STUART, *The Advanced Theory of Statistics,* Vol. II, Oxford University Press, New York.

[1975] J. T. KENT, "A weak convergence theorem for the empirical characteristic function", *J. Appl. Probab.,* 12, 515-523.

[1981] V. S. KOROLIUK, *Manual de la teoría de probabilidades y estadística matemática.* Mir Moscú, URSS.

[1980a] I. A. KOUTROUVELIS, "A goodness-of-fit test of simple hypotheses based on the empirical characteristic function", *Biometrika,* **67**, pp. 238-240.

[1980b] I. A. KOUTROUVELIS, "Regression-type estimation of the parameters of stable laws", *J. Amer. Statist. Assoc.*, 75, 918-928.

[1981] I. A. KOUTROUVELIS, "An iterative procedure for the estimation of the parameters of stable laws", *Commun. Statist.*, B 10, 17-28.

[1982] I. A. KOUTROUVELIS, "Estimation of location and scale in Cauchy distributions using the empirical characteristic function", *Biometrika*, 69, 1, 205-213.

[1981] I. A. KOUTROUVELIS Y J. KELLERMEIER, "A goodness-of-fit test based on the Empirical Characteristic Function when Parameters must be Estimated", *J. R. Statist. Soc.* B, 42, 173-176.

[1971] R. G. LAHA Y V. K.ROHATGI, *Probability Theory,* Wiley, New York.

[1960] M. LOÈVE, *Probability Theory,* Van Nostrand, New York.

[1968] E. LUKACS, *Stochastic Convergence*, D. C. Heath and Company, Washington.

[1970] E. LUKACS, *Characteristic Functions*, Hafner, New York.

[1964] E. LUKACS Y R. G. LAHA, *Applications of Characteristic Functions*, Charles Griffin, London.

[1987a] D. B. MADAN Y E. SENATA, "Chebyshev Polynomial Approximations and Characteristic Function Estimation", *J. R. Statist. Soc. B*, 42, No. 2, pp. 163-169.

[1987b] D. B. MADAN Y E. SENATA, "Simulation of Estimates Using the Empirical Characteristic Function", *Int. Statist. Rev.*, 55, 2, pp. 153-161.

[1960] B. MANDELBROOT, "The Pareto -Lévy law and the distribution of income", *Int. Econ. Rev.* 1, 76-106.

[1963] B. MANDELBROOT, "The variation of certain speculative prices", *J. Business* 36, 394-419.

[1981] M. B. MARCUS, "Weak convergence of the empirical characteristic function", *Ann. Probab.*, 9, 194-201.

[1974] A. M. MOOD, F. A. GRAYBILL Y D. C. BOES, *Introduction to the Theory of Statistics*, McGraw-Hill, New York.

[1968] P. A. MORAN, *An Introduction to Probability Theory,* Oxford University Press, New York.

[1962] E. PARZEN, "On estimation of a probability density function and mode", *Ann. Math. Statist.* 33 1065-1076.

[1975] A. S. PAULSON, E. W. HALCOMB Y R. A. LEITCH, "The estimation of the parameters of the stable laws", *Biometrika*, 62, 163-170.

[1968] S. J. PRESS, "A modified compound Poisson Process with Normal compounding", *Amer. Statist. Assoc. J.*, June, 607-613.

[1972a] S. J. PRESS, *Applied Multivariate Analysis,* Holt, Rinehart and Winston, New York.

[1972b] S. J. PRESS, "Estimation in Univariate and Multivariate Stable Distributions", *Journal of the American Statistical Association,* December 1972, volume 67, Number 340, 842-846.

[1975] S. J. PRESS, "Stable distributions: Probability inference and applications in finance". *Statistical Distributions in Scientific Work, Volume I,* Reidel, Dordrecht.

[1973] C. R. RAO, *Linear Statistical Inference and Its Applications,* Wiley, New York.

[1986] J. P. ROMANO Y A. F. SIEGEL, *Counterexamples in Probability and Statistics,* Wadsworth & Brooks, California.

[1980] R. J. SERFLING, *Approximation Theorems of Mathematical Statistics,* Wiley. New York.

[1984] A. N. SHIRYAYEV, *Probability,* Springer Verlag, New York.

[1988] J. STOYANOV, *Counterexamples in probability,* Wiley, New York.

[1967] H. G. TUCKER, *A Graduate Course in Probability,* Academic Press, London.

[1958] V. S. VARADARAJAN, "A useful convergence theorem", *Sankhya,* **20,** 221-222.

Indice