



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO

---

---

INSTITUTO DE CIENCIAS BÁSICAS E INGENIERÍA  
ÁREA ACADÉMICA DE MATEMÁTICAS Y FÍSICA

CONSTRUCCIÓN DE LA PRUEBA SMIRNOV-KOLMOGOROV  
A TRAVÉS DE EXPRESIONES EXPLÍCITAS DE LAS  
PROBABILIDADES DE CRUCE DEL PUENTE BROWNIANO

T E S I S

QUE PARA OBTENER EL TÍTULO DE

LICENCIADO EN MATEMÁTICAS APLICADAS

P R E S E N T A

ISAAC VÁZQUEZ MENDOZA

Bajo la dirección de

DRA. LILIANA PERALTA HERNÁNDEZ

MINERAL DE LA REFORMA, HIDALGO  
OCTUBRE 2021



Mineral de la Reforma, Hgo., a 28 de septiembre de 2021

Número de control: ICBI-D/896/2021  
 Asunto: Autorización de impresión.

**MTRO. JULIO CÉSAR LEINES MEDÉCIGO**  
**DIRECTOR DE ADMINISTRACIÓN ESCOLAR DE LA UAEH**

Con fundamento en lo dispuesto en el Título Tercero, Capítulo I, Artículo 18 Fracción IV; Título Quinto, Capítulo II, Capítulo V Fracción IX del Estatuto General de nuestra Institución, por este medio le comunico que el Jurado asignado al Pasante de la Licenciatura en Matemáticas Aplicadas Isaac Vázquez Mendoza, quien presenta el trabajo de titulación **“Construcción de la prueba Smirnov-Kolmogorov a través de expresiones explícitas de las probabilidades de cruce del puente Browniano”**, después de revisar el trabajo en reunión de Sinodales ha decidido autorizar la impresión del mismo, hechas las correcciones que fueron acordadas.

A continuación, firman de conformidad los integrantes del Jurado:

**Presidente:** Dr. José Félix Fernando Barrera Mora

**Secretario:** Dr. Jorge Viveros Rogel

**Vocal:** Dra. Liliana Peralta Hernández

**Suplente:** Dr. Roberto Ávila Pozos

Sin otro particular por el momento, reciba un cordial saludo.

Atentamente  
 “Amor, Orden y Progreso”

Dr. Otilio Arturo Acevedo Sandoval  
 Director del ICBI



Ciudad del Conocimiento  
 Carretera Pachuca-Tulancingo km 4.5 Colonia  
 Carboneras, Mineral de la Reforma, Hidalgo,  
 México. C.P. 42184  
 Teléfono: 771 71 720 00 ext. 2231 Fax 2109  
 direccion\_icbi@uaeh.edu.mx

OAAS/YCC





# Agradecimientos

Estoy profundamente agradecido con la Universidad Autónoma del Estado de Hidalgo (UAEH) por todo el crecimiento personal, profesional y académico que recibí a través de la Licenciatura en Matemáticas Aplicadas (LIMA) de esta institución. Mi paso por la LIMA superó por mucho mis expectativas y me mostró una belleza en las matemáticas que no sabía que existía.

Agradezco a mi asesora, la Dra. Liliana Peralta Hernández, cuyo apoyo incondicional me permitió culminar este trabajo. Colaborar con ella fue sumamente sorprendente y, en extremo, satisfactorio pues durante la realización de este trabajo compartió conmigo su tiempo, conocimiento y experiencias. Recordaré siempre que la probabilidad y la estadística son matemáticas *risueñas*.

Quiero agradecer a mis sinodales por enriquecer este trabajo con sus revisiones, observaciones y comentarios. Asimismo, agradezco a mis profesoras y profesores en la LIMA que con su ayuda se me presentaron experiencias académicas extraordinarias, en algunas de ellas fui aceptado y en algunas otras no, pero lo que viví es invaluable. En particular, mi más sincera gratitud a la M. en C. Margarita Tetlalmatzi Montiel, al Dr. Rubén A. Martínez Avendaño, al Dr. Roberto Ávila Pozos y al Dr. Jorge Viveros Rogel.

Durante mi tiempo en la LIMA conocí muchas personas, que hoy puedo considerar amigas y amigos cuya presencia, apoyo y compañía me salvó más de una vez. Le reitero mi más sincero agradecimiento a mis amigas y amigos: Mariana, Jailene, Mónica, Ceci, Leslie, Paola Jaqueline, Ana, Miriam, Sol, Bibi, Mau, Alex, Orlando, René, Manuel y Arturo. Puede que sea una lista algo extensa, pero hay al menos una razón por la cual cada persona aparece en ella.

En especial, agradezco a Marlene, quien ha hecho tanto por mí, desde acompañarme en la sala de lectura o los cubículos estudiando en las madrugadas hasta acompañarme en cuestiones de salud; las palabras se quedan cortas para expresar cuán agradecido estoy por su amistad. Asimismo, agradezco a Paola, su mente siempre me sorprendió, nuestro carácter nos ha hecho discutir, pero nuestra amistad es muy grande. Además, agradezco a Sitka, quien me ayudó de manera incondicional hasta su último momento aquí, siempre recordaré su sentido del humor, su formalismo para trabajar y su bonita costumbre de endulzar la vida.

Por último, es un placer para mí agradecer a mi familia: mi mamá ha sido un pilar para mi desarrollo, mi papá ha sido flexible para entenderme, mi hermano ha sido crucial para mi acercamiento a la ciencia, mi tía Liz ha sido mi más grande soporte y finalmente

mi amiga Ciomara, resumir estos once años de conocernos no es fácil, pero ha sido un motor en mi vida, impulsándome a lograr cosas que no consideré posibles y es alguien a quien le auguro un futuro muy próspero.

Por si olvidé mencionar a alguien, no lo hice de manera intencional; al contrario, agradezco a todas las personas que me apoyaron en el proceso de la realización de este trabajo, el cual fue toda una aventura: me retó intelectualmente, pero abrió mi panorama de las matemáticas. Aún hoy, para mí, es difícil creer que en efecto pude culminar esta licenciatura y este trabajo, pero todo el tiempo y esfuerzo que invertí valió la pena.

I'm a litte bit unholy, so what?  
So is everyone else.

---

Miley Cyrus

## Resumen

Cuando un experimento aleatorio es analizado estadísticamente es frecuente comenzar averiguando si las variables aleatorias que modelan al experimento tienen alguna distribución en específico, por ejemplo la distribución normal. Esto resulta importante pues al determinar la distribución, también sabemos qué herramientas son las más apropiadas para continuar con el análisis. Para estudiar la distribución de las variables aleatorias, cuando no se conoce a la función de distribución, se emplean las pruebas de bondad de ajuste, las cuales se desprenden de la inferencia estadística no paramétrica. En este trabajo construimos una de las pruebas de bondad de ajuste más ampliamente utilizadas: la prueba Smirnov-Kolmogorov, estudiando las probabilidades de cruce del puente Browniano. Nuestro enfoque constituye una forma nueva y más natural de obtener este resultado clásico.

## Abstract

When a random experiment is statistically analyzed, it is frequent to begin finding out if the random variables modeling the experiment follow a distribution in specific, for example, the normal distribution. This is important since determining the distribution also allows us to define the proper tools to proceed with the analysis. To find out the cumulative probability function of a random variable, when it is unknown, nonparametric statistical inference tools are used: goodness-of-fit tests. The Smirnov-Kolmogorov test, one of the most used goodness-of-fit tests, is constructed in this thesis using boundary crossing probabilities for the Brownian bridge. Our approach establishes a new and more natural way to prove this classical theorem.



# Índice general

<b>Índice general</b>	<b>7</b>
<b>Introducción</b>	<b>9</b>
<b>1. Preliminares</b>	<b>11</b>
1.1. Conceptos generales de Probabilidad . . . . .	11
1.2. Pruebas de hipótesis . . . . .	23
1.2.1. Pruebas de bondad de ajuste . . . . .	29
<b>2. Función de distribución empírica</b>	<b>33</b>
2.1. Teorema de Glivenko-Cantelli . . . . .	39
<b>3. Puente Browniano</b>	<b>43</b>
3.1. Movimiento Browniano . . . . .	44
3.1.1. Propiedades de las trayectorias del movimiento Browniano . . . . .	46
3.2. Puente Browniano . . . . .	50
3.3. Probabilidades de cruce . . . . .	53
<b>4. Estadístico Smirnov-Kolmogorov</b>	<b>65</b>
4.1. Ejemplo de aplicación de la prueba Smirnov-Kolmogorov . . . . .	70
<b>Conclusiones</b>	<b>81</b>
<b>Bibliografía</b>	<b>83</b>
<b>A. Construcción del movimiento Browniano</b>	<b>89</b>



# Introducción

La estadística inferencial nos permite derivar conclusiones sobre el comportamiento probabilístico de un experimento aleatorio a partir de los datos observados. No obstante, es imprescindible verificar la validez de tales conclusiones; así, surgen las pruebas de hipótesis, las cuales son procedimientos que contrastan afirmaciones a fin de decidir si son verdaderas.

Desde la perspectiva práctica, la parte más importante de las pruebas de hipótesis son las decisiones que se concluyen. En cambio, desde la perspectiva teórica, el estadístico de prueba es la pieza clave. Una prueba de hipótesis generalmente se especifica en términos de un estadístico de prueba, es decir, de una función de la muestra aleatoria.

Las pruebas de hipótesis de tipo no paramétrico destacan de entre las pruebas de hipótesis pues son métodos libres de distribución, es decir, no es necesario que las realizaciones de la muestra aleatoria se adapten a una distribución específica para que estas pruebas puedan aplicarse, por lo que representan una alternativa con gran versatilidad. Dentro de las pruebas de hipótesis no paramétricas se encuentran las pruebas de bondad de ajuste, las cuales permiten medir el acoplamiento entre el comportamiento probabilístico de las realizaciones de una muestra aleatoria y una función de distribución dada.

En este trabajo construiremos el estadístico de prueba Smirnov-Kolmogorov, una prueba de bondad de ajuste. Este estadístico surge de los trabajos de Andrei Kolmogorov [37] y Nikolai Smirnov [53], y es una de las herramientas más utilizadas para determinar la distribución de una variable aleatoria continua en contextos donde, por algún motivo, es desconocida (ver por ejemplo [1], [40], [62]).

Para el estadístico Smirnov-Kolmogorov, se ha estudiado principalmente su distribución límite [16], [21]. En este trabajo mantendremos dicha perspectiva y mostraremos cómo las probabilidades de cruce del puente Browniano están relacionadas con el comportamiento límite de este estadístico. Es decir, en esta tesis trataremos el caso en el que las probabilidades

$$P\left(\max_{0 \leq t \leq 1} |W_t| \leq f(t)\right),$$

donde  $W$  es un movimiento Browniano y  $f : [0, 1] \rightarrow \mathbb{R}$  es una función lineal, están condicionadas al valor que toma este proceso en  $t = 1$ . A través de las propiedades de las trayectorias del movimiento Browniano, usadas como en [50], [42], y el resultado clásico

de [16], calcularemos la probabilidad

$$P\left(\max_{0 \leq t \leq 1} |W_t - tW_1| \leq \alpha\right),$$

con  $\alpha \in \mathbb{R}$ . De esta manera, aludiendo al Teorema de Donsker, obtendremos una forma más natural de determinar la distribución límite del estadístico Smirnov-Kolmogorov.

Este trabajo se divide en cuatro capítulos. En el Capítulo 1 realizaremos una revisión general de conceptos de probabilidad necesarios para introducir el concepto de pruebas de hipótesis en la Sección 1.2. Luego, en el Capítulo 2 estudiaremos la función de distribución empírica en la que se basa la prueba Smirnov-Kolmogorov así como su comportamiento asintótico, es decir, estudiaremos el Teorema de Glivenko-Cantelli. El Capítulo 3 estará dedicado al puente Browniano, primero definiremos y construiremos al movimiento Browniano luego, estudiaremos las propiedades de invarianza de sus trayectorias, las cuales, serán de utilidad para estudiar al puente Browniano y abordar el problema central de este trabajo. Finalmente, en el Capítulo 4 se proporcionará la construcción de la distribución límite del estadístico Smirnov-Kolmogorov.

# Preliminares

En este capítulo presentamos las herramientas y resultados básicos necesarios para el desarrollo de este trabajo. La lectora o el lector familiarizado con los conceptos básicos de la Teoría de Probabilidad puede comenzar la lectura en la Sección 1.2. Además, para no desviarnos del objetivo de este trabajo hemos omitido algunos conceptos básicos relativos a Teoría de la Medida; sin embargo, recomendamos consultar el libro de Kolmogorov [38, Cap. 1-2] para todas aquellas personas que deseen profundizar en esta área en el contexto de probabilidad. Los resultados de probabilidad fueron tomados de libro de Resnick [45], mientras que los resultados referentes a estadística del libro de Hogg, McKean y Craig [30].

## 1.1. Conceptos generales de Probabilidad

Para poder definir apropiadamente un espacio de probabilidad es conveniente introducir algunos conceptos de teoría de la medida.

**Definición 1 ( $\sigma$ -álgebra)** Sea  $\Omega \neq \emptyset$  un conjunto. Una  $\sigma$ -álgebra  $\mathcal{F}$  es una colección de subconjuntos de  $\Omega$  que satisface las siguientes propiedades:

- (a)  $\Omega \in \mathcal{F}$ .
- (b) Si  $A \in \mathcal{F}$ , entonces  $A^c \in \mathcal{F}$ .
- (c) Si  $\{A_n\}_{n \in \mathbb{N}}$  es una colección numerable de conjuntos en  $\mathcal{F}$ , entonces  $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}$ .

La pareja  $(\Omega, \mathcal{F})$  recibe el nombre de **espacio medible**.

Ahora, veamos algunos ejemplos de  $\sigma$ -álgebra.

**Ejemplo 2 ( $\sigma$ -álgebra generada)** Sea  $\mathcal{A}$  una colección de subconjuntos de  $\Omega$ , entonces

$$\sigma(\mathcal{A}) = \bigcap \{ \mathcal{F} \mid \mathcal{F} \text{ es una } \sigma\text{-álgebra en } \Omega \text{ y } \mathcal{A} \subseteq \mathcal{F} \}$$

es la  $\sigma$ -álgebra generada por  $\mathcal{A}$ .

Un caso particular, y muy importante, de una  $\sigma$ -álgebra generada es la  $\sigma$ -álgebra de Borel. Si tomamos

$$\mathcal{A} = \{(a, b) \subseteq \mathbb{R} \mid a, b \in \mathbb{R}, a < b\}.$$

Entonces, la  $\sigma$ -álgebra generada por  $\mathcal{A}$  se conoce como  $\sigma$ -álgebra de Borel y se denota por  $\mathcal{B}(\mathbb{R})$ .

En todo espacio medible, la  $\sigma$ -álgebra nos dice qué conjuntos pueden ser medidos; sin embargo, no se ha fijado una forma de medir, lo que nos lleva a introducir el concepto de medida.

**Definición 3 (Medida)** Sea  $(\Omega, \mathcal{F})$  un espacio medible. Una medida en  $(\Omega, \mathcal{F})$  es una función  $\mu : \mathcal{F} \rightarrow [0, \infty]$  que satisface:

- (a)  $\mu(\emptyset) = 0$ .
- (b)  $\mu(A) \geq 0$ , para todo  $A \in \mathcal{F}$ .
- (c) Si  $\{A_n\}_{n \in \mathbb{N}}$  es una colección numerable de conjuntos disjuntos de  $\mathcal{F}$ , entonces

$$\mu \left( \bigcup_{n \in \mathbb{N}} A_n \right) = \sum_{n \in \mathbb{N}} \mu(A_n).$$

Es importante recordar que las propiedades que satisfacen los conceptos previos serán omitidas; sin embargo, recomendamos a las personas interesadas consultar [45] y [33].

**Nota 4** A la terna  $(\Omega, \mathcal{F}, \mu)$  se le conoce como **espacio de medida** y  $\mu(B)$ , con  $B \in \mathcal{F}$ , como la medida del conjunto  $B$ . Luego, si  $\mu(\Omega) < \infty$  se dice que  $\mu$  es una medida finita y, en particular, si  $\mu(\Omega) = 1$  se dice que  $\mu$  es una medida de probabilidad y se denotará como  $P$ ; por lo que, en este caso, a  $P(B)$  se le denomina como la probabilidad del evento  $B \in \mathcal{F}$ . Además, como en este trabajo nos restringimos a emplear medidas de probabilidad, no es necesario hablar de la incorporación del infinito en la recta real.

Finalmente, al espacio de medida  $(\Omega, \mathcal{F}, P)$  se le conoce como **espacio de probabilidad** y, este contexto, a  $\Omega$  se le conoce como **espacio muestral** pues el conjunto de resultados de un experimento aleatorio.

Los espacios de probabilidad son más interesantes cuando definimos variables aleatorias en ellos, entonces procedamos con su definición.

**Definición 5 (Variable aleatoria)** Se dice que una función  $X : \Omega \rightarrow \mathbb{R}$  es una función medible si  $X^{-1}((-\infty, x]) = \{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F}$ , para toda  $x \in \mathbb{R}$ . Las funciones medibles en un espacio de probabilidad se conocen como variables aleatorias (v.a.).

Consideremos los espacios medibles  $(\Omega, \mathcal{F})$  y  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , con  $\mathcal{B}(\mathbb{R})$  la  $\sigma$ -álgebra de Borel, y sea  $X$  una v.a. Si  $P : \mathcal{F} \rightarrow [0, 1]$  es una medida de probabilidad, diremos que  $P_X(B) := P(\{\omega \in \Omega \mid X(\omega) \in B\})$ , para todo  $B \in \mathcal{B}(\mathbb{R})$ , es la medida de probabilidad inducida por  $X$  (ver [19]).

En particular, la probabilidad del evento  $X^{-1}((-\infty, x])$ , para todo  $x \in \mathbb{R}$ , se llama función de distribución de la v.a.  $X$  y la definición formal es la siguiente.

**Definición 6 (Función de distribución)** Sea  $X$  una variable aleatoria. La función de distribución (f.d.), de  $X$ , es una función,  $F_X : \mathbb{R} \rightarrow [0, 1]$ , que está definida por

$$F_X(x) = P_X(X \leq x), \text{ para toda } x \in \mathbb{R}.$$

Con lo anterior, tenemos que si  $P_X$  denota la medida de probabilidad inducida por la v.a.  $X$ , entonces hemos construido el espacio de medida  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$  y, más fuerte aún, hemos construido un espacio de probabilidad, pues notemos que la f.d. de  $X$  y  $P_X$  están relacionadas de la manera siguiente:

$$F_X(x) = P_X((-\infty, x]), \text{ para toda } x \in \mathbb{R}.$$

**Ejemplo 7 (Variable aleatoria uniforme)** Sea  $(a, b) \subset \mathbb{R}$  un intervalo. Consideremos a la f.d.  $F : \mathbb{R} \rightarrow [0, 1]$ , definida por

$$F(x) = \begin{cases} 0, & \text{si } x \leq a \\ \frac{x-a}{b-a}, & \text{si } x \in (a, b) \\ 1, & \text{si } x \geq b. \end{cases} \quad (1.1)$$

La v.a.  $X$ , inducida por  $F$ , es conocida como variable uniforme continua con parámetros  $a$  y  $b$ .

Diremos que una v.a.  $X$  tiene distribución  $F$ , y en símbolos lo denotamos como  $X \sim F$ , si  $X$  es inducida por  $F$ ; así, si  $F$  es como en la ecuación (1.1) es frecuente llamarla distribución uniforme y para expresar que una v.a. tiene dicha distribución se escribe  $X \sim Unif(a, b)$ .

Una vez aclarada la importancia de las f.d., estudiaremos algunas de sus propiedades, las cuales serán de utilidad en el estudio de la función de distribución empírica (ver Capítulo 2).

**Teorema 8** Sea  $F(x) = P(X \leq x)$  una función de distribución. Entonces se satisfacen las siguientes propiedades:

- (a)  $\lim_{x \rightarrow \infty} F(x) = 1$ .
- (b)  $\lim_{x \rightarrow -\infty} F(x) = 0$ .
- (c) Si  $x \leq y$ , entonces  $F(x) \leq F(y)$ .
- (d) F es continua por la derecha, es decir, si  $h > 0$  y  $F(x^+) := \lim_{h \rightarrow 0} F(x + h)$ , entonces  $F(x^+) = F(x)$ .
- (e) Si  $h > 0$  y  $F(x^-) := \lim_{h \rightarrow 0} F(x - h)$ , entonces  $F(x^-) = P(X < x)$ .

DEMOSTRACIÓN.

- (a) Sea  $a = \{a_n\}_{n \in \mathbb{N}}$  una sucesión creciente de números reales tal que  $\lim_{n \rightarrow \infty} a_n = \infty$ .  
Sea  $A = \{A_n\}_{n \in \mathbb{N}}$  una sucesión de eventos con  $A_n = (X \leq a_n)$ , para toda  $n \in \mathbb{N}$ . Como la sucesión  $a$  es creciente, entonces la sucesión  $A$  lo es; además, como en el límite de la sucesión  $a$  no converge, entonces

$$\lim_{n \rightarrow \infty} A_n = \Omega.$$

Así, por la continuidad de la medida de probabilidad (ver [19], [47]), se tiene que

$$\lim_{n \rightarrow \infty} F(a_n) = \lim_{n \rightarrow \infty} P(A_n) = P(\Omega) = 1.$$

Y dado que de cualquier sucesión  $\{x_n\}_{n \in \mathbb{N}}$  tal que  $\lim_{n \rightarrow \infty} x_n = \infty$ , podemos extraer una subsucesión creciente  $\{x_{n_k}\}_{k \in \mathbb{N}}$  tal que  $\lim_{k \rightarrow \infty} x_{n_k} = \infty$ , pues  $\mathbb{R}$  con la métrica del valor absoluto es un espacio métrico, entonces se sigue el resultado.

- (b) Sea  $a = \{a_n\}_{n \in \mathbb{N}}$  una sucesión decreciente de números reales tal que  $\lim_{n \rightarrow \infty} a_n = -\infty$ .  
Sea  $A = \{A_n\}_{n \in \mathbb{N}}$  una sucesión de eventos con  $A_n = (X \leq a_n)$ , para toda  $n \in \mathbb{N}$ . Como la sucesión  $a$  es decreciente, entonces la sucesión  $A$  también lo es y esto implica que

$$\lim_{n \rightarrow \infty} A_n = \emptyset.$$

Entonces, nuevamente por la continuidad de la medida de probabilidad, se tiene que

$$\lim_{n \rightarrow \infty} F(a_n) = \lim_{n \rightarrow \infty} P(A_n) = P(\emptyset) = 0.$$

Así, sin importar la forma en cómo los números tiendan a  $-\infty$  se cumple que  $\lim_{n \rightarrow \infty} F(x_n) = 0$ .

(c) Como  $x \leq y$ , entonces

$$\begin{aligned} F(x) &\leq F(x) + P(x < X \leq y) \\ &= P([X \leq x] \cup [x < X \leq y]) \\ &= P(X \leq y) \\ &= F(y). \end{aligned}$$

(d) Sea  $\{a_n\}_{n \in \mathbb{N}}$  una sucesión decreciente de números reales positivos tal que  $\lim_{n \rightarrow \infty} a_n = 0$ . Notemos que

$$F(x + a_n) = F(x) + P(x < X \leq x + a_n).$$

Entonces, tomando la sucesión de eventos  $\{A_n\}$  donde  $A_n = (x < X \leq x + a_n)$ , para toda  $n \in \mathbb{N}$ , se tiene que la sucesión converge al conjunto vacío, por lo que usando la continuidad de la probabilidad obtenemos

$$\lim_{n \rightarrow \infty} F(x + a_n) = F(x) + \lim_{n \rightarrow \infty} P(A_n) = F(x),$$

de donde se sigue lo deseado.

(e) Sea  $\{a_n\}_{n \in \mathbb{N}}$  una sucesión decreciente de números reales positivos tal que  $\lim_{n \rightarrow \infty} a_n = 0$ . Tomando la sucesión de eventos  $\{A_n\}_{n \in \mathbb{N}}$  con  $A_n = (X < x - a_n)$ , para toda  $n \in \mathbb{N}$ , se tiene que la sucesión converge al evento  $(X < x)$ , por lo que la continuidad de la probabilidad implica

$$P(X < x) = \lim_{n \rightarrow \infty} P(A_n) = \lim_{n \rightarrow \infty} F(x - a_n) = F(x^-).$$

□

Otra propiedad de la f.d se encuentra descrita en la siguiente definición.

**Definición 9 (Variable aleatoria absolutamente continua)** Una v.a.  $X$ , con función de distribución  $F$ , es absolutamente continua si existe una función no negativa e integrable en el sentido de Riemman,  $f : \mathbb{R} \rightarrow \mathbb{R}$ , que denominaremos **función de densidad**, tal que

$$F(x) = \int_{-\infty}^x f(u) du.$$

En el caso que  $X$  sea una v.a. discreta, es decir, una variable aleatoria cuyo rango es a lo sumo numerable el símbolo de integral se reemplaza por una suma sobre el conjunto de puntos en donde no es cero.

Cabe aclarar que solo en algunos casos, como en el Ejemplo 7, la función de densidad existe, y tanto esta como su f.d son conocidas bajo algún nombre específico, como distribución geométrica, binomial, normal, Poisson, etc., por tal razón, generalmente se trabaja con la f.d.

Ahora presentamos algunos ejemplos que serán relevantes a lo largo de este trabajo.

**Ejemplo 10 (Distribución Bernoulli)** Sea  $p \in (0, 1)$ . Consideremos a la f.d.  $F : \mathbb{R} \rightarrow [0, 1]$  definida por

$$F(x) = \begin{cases} 0, & \text{si } x < 0, \\ 1 - p, & \text{si } 0 \leq x < 1, \\ 1, & \text{en otro caso.} \end{cases}$$

Verificaremos que  $f : \mathbb{R} \rightarrow \mathbb{R}$  definida por

$$f(x) = \begin{cases} p^x(1-p)^{1-x}, & \text{si } x = 0, 1, \\ 0, & \text{en otro caso,} \end{cases}$$

es una función tal que

$$F(x) = \sum_{u \leq x} f(u).$$

Por la definición de  $f$ , para  $x < 0$ , se cumple que  $f(x) = 0$ , por lo que  $F(x) = 0$ . Luego, para  $x = 0$  se tiene que  $f(x) = 1 - p$ ; mientras que, para  $0 < x < 1$  se satisface que  $f(x) = 0$ , en consecuencia, para  $0 \leq x < 1$  se cumple que  $F(x) = 1 - p$ . Finalmente,  $f(1) = p$  y si  $x > 1$ , entonces  $f(x) = 0$ ; por lo tanto  $F(x) = 1$  para  $x \geq 1$ . Verificando la afirmación.

Ahora, si una v.a.  $X$  tiene distribución  $F$ , diremos que  $X$  tiene distribución Bernoulli con parámetro  $p$  y escribiremos  $X \sim Ber(p)$ .

**Ejemplo 11 (Distribución Binomial)** Sean  $n \in \mathbb{N}$  y  $p \in (0, 1)$ . Consideremos la función de densidad  $f : \mathbb{R} \rightarrow \mathbb{R}$  definida por

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{si } x = 1, \dots, n, \\ 0, & \text{otro caso.} \end{cases}$$

Si una v.a.  $X$ , tiene función de densidad  $f$ , diremos que  $X$  tiene distribución Binomial con parámetros  $n$  y  $p$ , y escribiremos  $X \sim Bino(n, p)$ .

**Ejemplo 12 (Distribución normal)** Sean  $\mu, \sigma^2 \in \mathbb{R}$  con  $\sigma^2 > 0$ . Además, sea  $f : \mathbb{R} \rightarrow \mathbb{R}$  una función de densidad definida por

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \text{ para toda } x \in \mathbb{R}.$$

Si una v.a.  $X$  tiene función de densidad  $f$ , diremos que  $X$  tiene distribución normal, o **Gaussiana**, con parámetros  $\mu$  y  $\sigma^2$ , y escribiremos  $X \sim N(\mu, \sigma^2)$ .

Las v.a. poseen características numéricas entre las que destacan el valor esperado, la varianza y la covarianza. A continuación presentamos su definición.

**Definición 13** Sea  $X$  una v.a. absolutamente continua con función de densidad  $f$ .

- (a) El valor esperado, o esperanza, de  $X$  se define como la siguiente integral en el sentido de Riemann

$$E(X) = \int_{\mathbb{R}} xf(x)dx.$$

En el caso de que  $X$  sea una v.a. discreta y absolutamente continua, sustituiremos el símbolo de integral por una suma sobre el conjunto de puntos en donde es distinta de cero. Además, cuando  $E(X) = \pm\infty$  diremos que  $X$  no tiene esperanza finita.

- (b) La varianza de  $X$  se denota y define como

$$\text{Var}(X) = E([X - E(X)]^2).$$

- (c) Si  $Y$  es una v.a., entonces la covarianza de  $X$  y  $Y$  se define como

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

La razón por la cual los tres conceptos previamente definidos destacan, es la interpretación que podemos asociarles. La esperanza es el número alrededor del cual se concentran los eventos con mayor probabilidad de ocurrir; por otro lado, la varianza se refiere a cuán dispersos se encuentran los eventos y, finalmente, la covarianza indica cómo es la relación entre dos variables aleatorias dadas. Para no desviarnos del objetivo de este trabajo sugerimos consultar [57], [45] y [33], para profundizar en las diversas propiedades que satisfacen los conceptos antes mencionados.

Ahora, introduciremos los vectores aleatorios, pues estos relacionan dos o más v.a. a fin de modelar experimentos aleatorios más sofisticados y la covarianza toma un papel fundamental en el estudio de ellos.

**Definición 14 (Vector aleatorio)** Sea  $X : \Omega \rightarrow \mathbb{R}^n$  una función. Diremos que  $X$  es un vector aleatorio si

$$X^{-1}((-\infty, x_1] \times \cdots \times (-\infty, x_n]) \in \mathcal{F}, \forall x_1, \dots, x_n \in \mathbb{R}.$$

De la definición anterior no es del todo claro que estemos hablando de un vector aleatorio; así, presentamos el siguiente resultado.

**Proposición 15** Una función  $(X_1, \dots, X_n)^T : \Omega \rightarrow \mathbb{R}^n$  es un vector aleatorio si, y sólo si, cada coordenada es una variable aleatoria.

Introduciremos ahora a las funciones de distribución conjuntas, las cuales son análogas a las funciones de distribución definidas previamente.

**Definición 16 (Función de distribución conjunta)** Sean  $X_1, \dots, X_n$  variables aleatorias. Diremos que la función de distribución conjunta (f.d.c.) del vector  $X = (X_1, \dots, X_n)^T$ ,  $F_X : \mathbb{R}^n \rightarrow [0, 1]$ , se define por

$$F_X(x_1, \dots, x_n) = P_X(X_1 \leq x_1, \dots, X_n \leq x_n),$$

para cualquier  $(x_1, \dots, x_n) \in \mathbb{R}^n$ .

Para vectores aleatorios existen definiciones y resultados análogos a los de v.a. Recomendamos consultar [56], [45] y [33].

Ahora, extenderemos a vectores aleatorios el concepto de variables aleatorias absolutamente continuas y el concepto de función de densidad.

**Definición 17 (Función de densidad conjunta)** Sea  $X = (X_1, \dots, X_n)^T$  un vector aleatorio con función de distribución  $F_X$ . Se dice que  $X$  es absolutamente continuo si existe una función no negativa e integrable,  $f_X : \mathbb{R}^n \rightarrow \mathbb{R}$ , tal que para todo  $(x_1, \dots, x_n) \in \mathbb{R}^n$ , se cumple que

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

empleando la integral en el sentido de Riemann. Notemos que, similar a variables aleatorias, en el caso de que el vector aleatorio  $X$  esté formado de variables aleatorias discretas, sustituiremos los símbolos de integral por sumas sobre los correspondientes soportes. Además, a la función  $f_X$  la denominaremos función de densidad conjunta.

Ilustraremos la utilidad de la definición anterior mostrando un ejemplo que será relevante en el Capítulo 3.

**Ejemplo 18 (Distribución normal p-variada)** Sea  $X = (X_1, \dots, X_p)^T$  un vector aleatorio. Diremos que  $X$  tiene distribución normal  $p$ -variada si su función de densidad conjunta es

$$f_X(y) = (\sqrt{2\pi})^{-p/2} \det(\Sigma)^{-1/2} \exp\left(-\frac{1}{2}(y - \mu)^T \Sigma^{-1}(y - \mu)\right), \text{ para todo } y \in \mathbb{R}^p,$$

donde  $\mu \in \mathbb{R}^p$  es tal que  $\mu = (E(X_1), \dots, E(X_p))^T$  y  $\Sigma \in \mathcal{M}_{p \times p}(\mathbb{R})$  es tal que  $\Sigma = (\text{Cov}(X_i, X_j))_{i,j}$ .

Dado un vector aleatorio, en ocasiones, es útil conocer el comportamiento estocástico de alguna colección de las variables aleatorias que lo componen, de esto surge el concepto que definimos a continuación.

**Definición 19 (Función de densidad marginal)** Sea  $X = (X_1, \dots, X_n)^T$  un vector aleatorio con función de densidad  $f_X$ . Definimos a la función de densidad marginal  $f_{X_i} : \mathbb{R} \rightarrow \mathbb{R}$ , para cada  $i = 1, \dots, n$ , por

$$f_{X_i}(x_i) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_{i-1}} \int_{-\infty}^{x_{i+1}} \cdots \int_{-\infty}^{x_n} f_X(u_1, \dots, u_n) du_n \cdots du_{i+1} du_{i-1} \cdots du_1,$$

empleando la integral en el sentido de Riemann.

Habiendo definido las funciones de densidad marginales podemos extender el concepto de independencia a las variables aleatorias.

**Definición 20 (Independencia de variables aleatorias)** Sean  $X$  y  $Y$  variables aleatorias con funciones de densidad  $f_X$  y  $f_Y$ , respectivamente. Diremos que  $X$  y  $Y$  son independientes si

$$f_{(X,Y)}(x, y) = f_X(x)f_Y(y), \text{ para todos } x, y \in \mathbb{R}.$$

Donde  $f_{(X,Y)}$  denota la función de densidad conjunta del vector aleatorio  $(X, Y)^T$ .

Enunciaremos a continuación algunas propiedades que satisfacen las variables aleatorias independientes, la demostración de la siguiente proposición puede ser consultada en [57], [10] y [33].

**Proposición 21** Sean  $X, Y$  variables aleatorias independientes, entonces

- (a)  $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ , para cualesquiera eventos  $A$  y  $B$ .
- (b)  $E(XY) = E(X)E(Y)$ .
- (c) Para cualesquiera  $a, b \in \mathbb{R}$  se cumple que  $\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$ .
- (d)  $\text{Cov}(X, Y) = 0$ , y más aún, si  $X$  y  $Y$  tienen distribución normal, entonces son independientes si, y sólo si,  $\text{Cov}(X, Y) = 0$ .

Ahora, mostramos algunos tipos de convergencia para variables aleatorias definidas en un espacio de probabilidad. Las demostraciones pueden consultarse en [10], [45] y [19].

**Definición 22 (Convergencia puntual)** Sea  $\{X_n\}_{n \in \mathbb{N}}$  una sucesión de variables aleatorias definidas en un mismo espacio de probabilidad y  $X$  una v.a. sobre dicho espacio. Diremos que la sucesión  $\{X_n\}_{n \in \mathbb{N}}$  converge puntualmente a  $X$  si para todo  $\omega \in \Omega$  se cumple que

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

En este caso, escribiremos  $X_n \rightarrow X$ .

La convergencia puntual es una herramienta esencial pues a través de ella puede definirse la convergencia casi segura, y para hablar de ella es necesario introducir el concepto siguiente.

**Definición 23** Consideremos un espacio de probabilidad  $(\Omega, \mathcal{F}, P)$ . Diremos que una proposición sobre elementos aleatorios es válida **casi seguramente** (c.s.) si existe un evento  $E \in \mathcal{F}$ , con  $P(E) = 0$ , tal que la proposición es válida si  $\omega \in E^c$ . El evento  $E$  recibe el nombre de conjunto excepción.

En ocasiones, si una proposición se cumple c.s. es usual decir que dicha proposición es válida **con probabilidad uno**, ya que si  $E$  es un conjunto excepción entonces  $P(E^c) = 1 - P(E) = 1$ . Con lo anterior, procedamos a definir la convergencia casi segura.

**Definición 24 (Convergencia casi segura)** Sea  $\{X_n\}_{n \in \mathbb{N}}$  una sucesión de variables independientes e idénticamente distribuidas (i.i.d.) y  $X$  una v.a. Diremos que  $\{X_n\}_{n \in \mathbb{N}}$  converge casi seguramente a  $X$  si

$$P(\{\omega \in \Omega \mid X_n(\omega) \rightarrow X(\omega)\}) = 1.$$

En este caso, diremos que

$$\lim_{n \rightarrow \infty} X_n = X \text{ casi seguramente.}$$

Alternativamente, escribiremos

$$X_n \xrightarrow{\text{c.s.}} X.$$

Para continuar, introduciremos un tipo de convergencia que surge a partir de proponer una forma alterna de medir la cercanía entre una sucesión de variables aleatorias y la v.a. candidata a límite.

**Definición 25 (Convergencia en probabilidad)** Sea  $\{X_n\}_{n \in \mathbb{N}}$  una sucesión de v.a. Diremos que  $X_n$  converge en probabilidad a la v.a.  $X$  si para toda  $\varepsilon > 0$  se cumple que

$$\lim_{n \rightarrow \infty} P(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) = 0.$$

En tal caso, escribimos  $X_n \xrightarrow{P} X$ .

Finalmente, hablaremos sobre un tipo de convergencia la cual mide la cercanía entre una sucesión y la v.a. candidata a límite a través de la función de distribución.

**Definición 26 (Convergencia en distribución)** Sea  $\{X_n\}_{n \in \mathbb{N}}$  una sucesión de variables aleatorias i.i.d. con función de distribución  $F_n$ . Para una función  $g$ , denotemos por  $\mathcal{C}(g)$  al conjunto de puntos en donde  $g$  es continua. De esta manera, decimos que  $X_n$  converge en distribución a la v.a.  $X$ , con f.d.  $F$ , si

$$\lim_{n \rightarrow \infty} F_n(x) = F(x), \text{ para todo } x \in \mathcal{C}(F).$$

En tal caso, escribimos  $X_n \xrightarrow{D} X$ .

A continuación, presentaremos resultados importantes para la Teoría de Probabilidad que emanan del estudio de los tipos de convergencia.

**Teorema 27 (Ley débil de los grandes números)** Sea  $\{X_n\}_{n \in \mathbb{N}}$  una sucesión de variables aleatorias i.i.d. con  $E(X_i) = \mu$  y  $\text{Var}(X_i) < \infty$ . Si definimos  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ , entonces

$$\bar{X}_n \xrightarrow{P} \mu.$$

Ahora, introduciremos un resultado similar al Teorema 27 en donde el tipo de convergencia que se obtiene resulta ser convergencia casi segura.

**Teorema 28 (Ley fuerte de los grandes números)** Sea  $\{X_n\}_{n \in \mathbb{N}}$  una sucesión de variables aleatorias i.i.d. Sea  $\bar{X}_n$  como en el Teorema 27, entonces existe una constante  $c \in \mathbb{R}$  tal que

$$\bar{X}_n \xrightarrow{c.s.} c,$$

si, y sólo si,  $E(|X_i|) < \infty$ . En tal caso,  $c = E(X_i)$ .

La siguiente proposición es una herramienta fundamental en la teoría de probabilidad y estadística por su versatilidad.

**Teorema 29 (Teorema del límite central)** Sea  $\{X_n\}_{n \in \mathbb{N}}$  una sucesión de variables aleatorias i.i.d. con esperanza  $\mu$  y varianza  $\sigma^2 < \infty$ . Si  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  y  $Z$  es una v.a. con distribución normal estándar, entonces

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{D} Z.$$

**Nota 30** Observemos que el Teorema del límite central implica la ley débil de los grandes números pues

$$\text{Var}(\bar{X}_n) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\sigma^2}{n} < \infty.$$

Luego, la desigualdad de Chebyshev (ver [7]) implica que para toda  $\varepsilon > 0$  se cumple que

$$P(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

Finalmente, por el teorema de estricción (ver [54]) se tiene que

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0,$$

es decir,  $\bar{X}_n \xrightarrow{P} \mu$ .

## 1.2. Pruebas de hipótesis

Recordemos que el objetivo central de este trabajo es construir el estadístico de prueba conocido como Smirnov-Kolmogorov, por tal razón en esta sección introducimos los conceptos básicos de pruebas de hipótesis.

En estadística es usual encontrarnos con problemas que involucran incertidumbre, pues al considerar experimentos aleatorios desconocemos parcial o totalmente elementos tales como la distribución de las variables que los modelan y los parámetros que las caracterizan. Para lidiar con dichas situaciones se han desarrollado herramientas de inferencia [30].

El campo de la inferencia estadística se divide en dos: estimación de parámetros y pruebas de hipótesis. Respecto a la estimación de parámetros se tienen a los estimadores puntuales, lo cuales proporcionan un único valor que pretende aproximar el verdadero comportamiento del parámetro de interés, y para analizar qué tan buena es la aproximación se estudian propiedades como el sesgo, la consistencia y la suficiencia del estimador puntual. Además, dentro de la estimación de parámetros podemos encontrar la estimación por intervalo, la cual nos permite generar todo un conjunto de la forma  $(a, b) \subset \mathbb{R}$  en el cual, posiblemente, puede hallarse al verdadero valor del parámetro de estudio. Notemos que con este enfoque, tenemos una cantidad infinita no numerable de posibilidades, como el uso de intervalos de confianza y de credibilidad, lo que puede resultar conveniente pues se tiene un área de trabajo que no resulta tan restrictiva [22].

Por otra parte, respecto a las pruebas de hipótesis, éstas se dividen en dos tipos: las pruebas de hipótesis sobre parámetros, las cuales estudian la validez de afirmaciones hechas sobre el comportamiento del parámetro de interés, por ejemplo verificar si el parámetro es menor o igual a 5; mientras que, el segundo tipo de pruebas de hipótesis se refiere a las pruebas de bondad de ajuste que veremos más adelante. Para comenzar a hablar formalmente sobre las pruebas de hipótesis es necesario introducir los siguientes términos.

**Definición 31 (Muestra aleatoria)** Sea  $X$  una v.a. con f.d.  $F_X$ . Una colección de  $n$  variables aleatorias i.i.d.  $X_1, \dots, X_n$  es denominada una muestra aleatoria, de tamaño  $n$ , de la v.a.  $X$  si

$$X_i \sim F_X, \text{ para toda } i = 1, \dots, n.$$

Es decir,  $X$  y  $X_i$  son iguales en distribución,  $i = 1, \dots, n$ ; o, equivalentemente en símbolos escribimos  $X_i \stackrel{D}{=} X$ .

Además, diremos que  $x_1, \dots, x_n$  es una realización, u observación, de la muestra si

$$x_i = X_i(\omega),$$

para algún  $\omega \in \Omega$ .

**Definición 32 (Estadístico)** Sean  $X$  una variable aleatoria,  $X_1, X_2, \dots, X_n$  una muestra aleatoria de  $X$  y  $A$  es un conjunto no vacío. Un estadístico  $T$  es una función  $T : \Omega \rightarrow A$  tal que  $\Omega$  es el espacio muestral asociado al vector aleatorio  $(X_1, X_2, \dots, X_n)$ .

**Nota 33** Note que un estadístico de prueba es una variable aleatoria.

**Definición 34 (Hipótesis estadística)** Una hipótesis estadística ( $H$ ) es una suposición hecha sobre la f.d.  $F(x, \theta)$  de una variable aleatoria o vector aleatorio. Una hipótesis estadística es llamada *simple* si determina completamente a la función de distribución, en otro caso es llamada *compuesta*.

**Ejemplo 35** Consideremos una v.a.  $X \sim N(\mu, \sigma^2)$  con  $\mu$  y  $\sigma^2$  parámetros desconocidos. Una hipótesis simple sería:

$$H: \mu = 30 \text{ y } \sigma^2 = 10.$$

Mientras que, una hipótesis compuesta sería:

$$H: \mu \neq 30, \mu < 0.1 \text{ y } \sigma^2 > 1.$$

**Definición 36 (Prueba de hipótesis)** Dado un par de hipótesis estadísticas complementarias, es decir, tales que exactamente una de ellas es verdadera, diremos que una prueba de hipótesis es un procedimiento para decidir qué hipótesis aceptar.

Coloquialmente, una prueba de hipótesis es un procedimiento que nos ayuda a tomar la decisión de aceptar o no una hipótesis dada [29], la cual puede ser cuantitativa o cualitativa. Para este fin se sigue el siguiente procedimiento:

1. Se recolectan datos de muestra.
2. Se calculan los estadísticos descritos en la prueba.
3. Se usa la información obtenida en los pasos anteriores para conocer qué tan probable es que la hipótesis sea correcta.

Un estadístico de prueba particiona al espacio muestral  $\Omega$ , generando así un subconjunto  $C \subset \Omega$ , llamado región crítica, el cual determina la aceptación o rechazo de la hipótesis. Para fijar ideas veamos el siguiente ejemplo.

**Ejemplo 37** Sea  $X$  una v.a. con f.d.  $F(x, \theta)$ , donde  $\theta \in \Theta$  es una propiedad desconocida de  $F$  y  $\Theta \neq \emptyset$ . Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de  $X$ . Supongamos que existe un par de conjuntos disjuntos distintos del vacío  $\Theta_0, \Theta_1 \subset \Theta$  tales que

$$\Theta_0 \cup \Theta_1 = \Theta.$$

Además, supongamos que  $\theta$  es elemento de uno, y solo uno de los conjuntos  $\Theta_0$  y  $\Theta_1$ .

Realizar una prueba de hipótesis ayuda a determinar si

$$\theta \in \Theta_0 \quad \text{ó} \quad \theta \in \Theta_1,$$

a través de su estadístico de prueba  $T : \Omega \rightarrow A \subseteq \mathbb{R}$ .

Sea  $C \subset \Omega$  la región crítica que induce el estadístico  $T$ . Si la observación

$$\{\omega \in \Omega \mid X_1(\omega) = x_1, \dots, X_n(\omega) = x_n\} \subset C \quad (1.2)$$

diremos que no hay suficiente información para aceptar que  $\theta \in \Theta_0$  y, en consecuencia, aceptamos que  $\theta \in \Theta_1$ . En otro caso aceptamos que  $\theta \in \Theta_0$ .

Es importante notar que, en la práctica, es usual trabajar con la imagen de  $C$  bajo el estadístico  $T$  para tomar la decisión [6], verificando si  $T(x_1, \dots, x_n)$  es, o no, elemento de  $T(C)$ , en vez de trabajar directamente con la región crítica  $C$ .

Tomando el Ejemplo 37 como base enunciaremos las definiciones siguientes.

**Definición 38** En un problema de prueba de hipótesis, a la suposición que  $\theta \in \Theta_0$  la llamaremos **hipótesis nula** ( $H_0$ ); mientras que, a la suposición que  $\theta \in \Theta_1$  la denominaremos **hipótesis alternativa** ( $H_1$ ).

Veamos el siguiente ejemplo.

**Ejemplo 39** Una empresa automotriz determinó, a través de pruebas de choque a  $10\text{km/h}$ , que solo el 25 % de las veces sus autos no sufren daños visibles. Para incrementar dicho porcentaje se ha diseñado un parachoques reforzado, el cuál será probado bajo las mismas condiciones. De esta manera, si  $p$  denota la proporción de autos con parachoques reforzado que chocaron a  $10\text{km/h}$  y no sufrieron daños visibles, entonces se presenta el siguiente problema de prueba de hipótesis: como hipótesis nula se tiene que el reforzado del parachoques no mejoró la resistencia de los autos a los choques, pero al ser un reforzamiento tampoco empeoró su anterior resistencia, por lo que  $p$  sería igual a la proporción previamente observada, lo que significa que

$$H_0: p = 0.25.$$

Por otro lado, como hipótesis alternativa se tiene que el reforzado cumplió las expectativas de mejorar la resistencia de los autos a los choques lo que implicaría que  $p$  fue mayor a la proporción antes reportada, por lo que

$$H_1: p > 0.25.$$

En el contexto de este ejemplo es intuitivamente claro que solo una de las hipótesis es cierta pues, por la ley de tricotomía, no es posible que  $p$  sea igual a la proporción de autos sin parachoques reforzado que chocaron a  $10km/h$  y no sufrieron daños visibles y que a la vez sea mayor que ella.

**Definición 40 (Región crítica)** Sea  $X$  una variable aleatoria y  $X_1, \dots, X_n$  es una muestra aleatoria de  $X$ . En una prueba de hipótesis, la región crítica de dicha prueba es el subconjunto del espacio muestral asociado al vector aleatorio  $(X_1, \dots, X_n)$ , donde la hipótesis nula,  $H_0$ , se rechaza.

La validez de las conclusiones que se obtienen de una prueba de hipótesis depende de la evidencia recolectada al observar el experimento, por lo que no es posible tener total certeza de la conclusión, pues la incompletitud de evidencia quizá nos haga cometer errores, como aceptar  $H_1$  cuando la hipótesis  $H_0$  en realidad era cierta, o viceversa, lo anterior es conocido como error del tipo I y error del tipo II, respectivamente [30]. De lo anterior, se desprende la Tabla 1.1.

Decisión	Realidad del experimento	
	$H_0$ cierta	$H_1$ cierta
Rechazar $H_0$	Error del tipo I	Decisión correcta
Aceptar $H_0$	Decisión correcta	Error del tipo II

Tabla 1.1: Conclusiones de una prueba de hipótesis.

Dentro de la inferencia estadística es crucial calcular la probabilidad de cometer los errores de tipo I y de tipo II, pues, como mencionábamos antes, al no contar con la totalidad de la información, es fundamental cuantificar la probabilidad de error. Así, presentamos los siguientes conceptos.

**Definición 41** Sea  $\mathcal{X}^T = (X_1, \dots, X_n)$  una muestra aleatoria de la v.a.  $X$ . Consideremos un problema de prueba de hipótesis, con  $C$  la región crítica de dicha prueba. Definimos

$$\alpha(\theta) = P(\mathcal{X} \in C \mid \theta), \quad \forall \theta \in \Theta_0$$

y

$$\beta(\theta) = P(\mathcal{X} \notin C \mid \theta), \quad \forall \theta \in \Theta_1.$$

La probabilidad de cometer el error de tipo I, conocida como *el nivel de significancia de la prueba*, se denota y define por

$$\alpha = P(\text{“Cometer el error de tipo I”}) = \sup_{\theta \in \Theta_0} \alpha(\theta). \quad (1.3)$$

Luego, la probabilidad de cometer el error de tipo II se denota y define por

$$\beta = P(\text{“Cometer el error de tipo II”}) = \sup_{\theta \in \Theta_1} \beta(\theta), \quad (1.4)$$

donde  $\Theta_0$  y  $\Theta_1$  son como en el Ejemplo 37.

Veamos el siguiente ejemplo.

**Ejemplo 42** En adición a lo discutido en el Ejemplo 39 supongamos ahora que se procederá a realizar un prueba de hipótesis, con estadístico de prueba  $T$ , considerando una muestra aleatoria  $X_1, \dots, X_n$  de tamaño  $n = 20$ .

Para llevar a cabo la prueba de hipótesis se requiere la realización de la muestra aleatoria  $x_i = X_i(\omega)$ ,  $\omega \in \Omega$ , la cual en este caso se obtendría de construir 20 autos que cuenten con el prototipo de parachoques reforzado. Luego, para cada auto ejecutar un choque independiente a  $10km/h$  y evaluar si hay daños visibles o no, lo que nos habla de una v.a. con distribución Bernoulli.

Luego, puesto que se desea analizar la cantidad de autos sin daños visibles, se propone que el estadístico  $T$  sea la v.a. que cuenta el número de autos, con parachoques reforzado, que tras chocar no presentan daños visibles, por lo que

$$T = \sum_{i=1}^{20} X_i.$$

En consecuencia  $T \sim Bino(20, p)$ , donde recordemos que  $p$  denota la proporción de autos con parachoques reforzado que chocaron a  $10km/h$  y no sufrieron daños visibles

Sabemos que si los autos hubiesen sido contruidos con el parachoques original se esperaría que el 25 % de los autos, es decir 5 autos, no presentaran daños visibles. Sin embargo, como la empresa quiere asegurarse que los parachoques reforzados proveen mejores resultados, esperan que al menos 8 autos, el 40 % de la muestra, no presenten daños visibles. De esta manera, la región crítica es

$$C = \{8, 9, \dots, 20\}.$$

Además, en este contexto, el nivel de significancia o la probabilidad de cometer el error de tipo I,  $\alpha$ , es la probabilidad de asumir que los parachoques reforzados tuvieron mejor rendimiento que los parachoques originales, cuando en realidad su rendimiento fue igual. Así,

$$\alpha = P(T \geq 8) \approx 0.1018,$$

donde  $T \sim \text{Bino}(20, 0.25)$ , pues suponemos que  $H_0$  es cierta.

En cambio, la probabilidad de cometer el error de tipo II,  $\beta$ , es la probabilidad de asumir que los parachoques reforzados resultaron igual de resistentes que los parachoques originales, cuando en realidad proporcionaron mejores resultados, digamos  $p = 0.32 > 0.25$ . Así,

$$\beta = P(T \leq 7) \approx 0.7077,$$

donde  $T \sim \text{Bino}(20, 0.32)$ , pues suponemos que  $H_1$  es cierta.

Además de preguntarnos sobre la probabilidad de errar, también es de gran interés estudiar qué tan buena es la prueba de hipótesis para tomar la decisión correcta, rechazando la hipótesis alternativa cuando es falsa. De esto, se desprende el siguiente concepto.

**Definición 43 (Poder de una prueba de hipótesis)** Sea  $\mathcal{X}^T = (X_1, \dots, X_n)$  una muestra aleatoria de la v.a.  $X$ . Consideremos un problema de prueba de hipótesis, con  $C$  la región crítica de dicha prueba. Definimos el poder de la prueba

$$\gamma(\theta) = P(\mathcal{X} \in C \mid \theta), \quad (1.5)$$

para cada  $\theta \in \Theta_1$ .

La razón de estudiar el nivel de significancia de una prueba de hipótesis es refinar la aceptación o rechazo de  $H_0$  dando un nivel de certeza a la conclusión. Notemos que entre más cercano a cero sea el nivel de significancia, tendremos mayor certeza sobre que hemos tomado la decisión correcta.

En general, el procedimiento de aceptación-rechazo de  $H_0$  implementando el nivel de significancia depende de la f.d. del estadístico de prueba; sin embargo, a continuación presentamos un ejemplo que ilustra la idea subyacente.

**Ejemplo 44** Sea  $X$  una v.a.,  $X_1, \dots, X_n$  una muestra aleatoria y  $x_1, \dots, x_n$  una realización de ella. Dada una prueba de hipótesis con estadístico de prueba  $T$ , con f.d.  $F_T$ , y un nivel de significancia  $\alpha$ . Aceptaremos la hipótesis  $H_0$  con nivel de significancia  $\alpha$ , si

$$T(x_1, \dots, x_n) \leq t_\alpha,$$

donde  $t_\alpha$  es tal que  $F_T(t_\alpha) = \alpha$ .

Ahora introduciremos una alternativa a los procedimientos de toma de decisión similares al discutido en el ejemplo previo. El  $p$ -valor tiene grandes implicaciones en la toma de decisión asociada a un problema de prueba de hipótesis; para este trabajo, proporcionaremos una breve introducción al tema. Recomendamos consultar [6] y [22] para las personas interesadas en ampliar esta información.

**Definición 45 ( $p$ -valor)** Sea  $X$  una v.a.,  $X_1, \dots, X_n$  una muestra aleatoria de la v.a. y  $x_1, \dots, x_n$  una realización de ella. Dada una prueba de hipótesis con estadístico de prueba  $T : \Omega \rightarrow \mathbb{R}$ , con  $\Omega$  el espacio muestral asociado al vector aleatorio  $(X_1, \dots, X_n)^T$ , definimos el  $p$ -valor asociado a la prueba como

$$P(T \geq t \mid \theta \in \Theta_0), \quad (1.6)$$

donde  $t = T(x_1, \dots, x_n)$ .

El  $p$ -valor, es a veces denominado *nivel de significancia observado* [30], pues el cálculo del  $p$ -valor depende directamente de los datos observados como vemos en su definición. Además, de acuerdo con [22], si  $t$  es el cálculo del estadístico usando los datos observados, el  $p$ -valor puede entenderse como la probabilidad de obtener valores próximos a  $t$  dado que  $H_0$  es cierta. Es por ello que entre más cercano al valor uno sea el del  $p$ -valor, mayor certeza tendremos de que, en efecto,  $H_0$  es cierta.

Recordemos que para introducir el Ejemplo 44 mencionamos que los procedimientos de prueba de hipótesis dependen de la f.d. del estadístico de prueba. No obstante, el  $p$ -valor nos proporciona una alternativa para tomar una decisión en las pruebas de hipótesis en donde tal dependencia es nula, pues dado un nivel de significancia  $\alpha$ , aceptamos la hipótesis  $H_0$  si

$$p \geq \alpha, \quad (1.7)$$

donde  $p$  denota el  $p$ -valor.

### 1.2.1. Pruebas de bondad de ajuste

Ahora, usaremos lo discutido en la sección previa para estudiar las pruebas de hipótesis no paramétricas. Estas son un caso particular de las pruebas de hipótesis que se desarrollaron como parte de la inferencia no paramétrica, la cual tiene como idea subyacente hacer pocas suposiciones y usar a los datos para inferir propiedades cuantitativas y cualitativas desconocidas [60].

Una prueba de hipótesis no paramétrica es un *método libre de distribución* [29], es decir, no pide como hipótesis que los datos sigan una distribución específica, ofreciendo

una alternativa con mayor flexibilidad, que no sacrifica la información que provee la distribución de los datos [31].

Recordemos que al observar un experimento aleatorio suponemos que las v.a.,  $X_1, \dots, X_n$ , que componen una muestra aleatoria son independientes e idénticamente distribuidas, digamos con f.d. común  $F$ ; sin embargo, en estadística, al tener realizaciones de dicha muestra, es usual desconocer a  $F$  y para tratar de hallar una solución a este problema surgen las técnicas no paramétricas de bondad de ajuste [11], [22]. Ahora, veamos el siguiente ejemplo para familiarizarnos con el contexto de un problema de bondad de ajuste.

**Ejemplo 46** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una v.a.  $X$  con función de distribución desconocida. Ya que, por la definición de muestra aleatoria (ver Definición 31),  $X_1, \dots, X_n$  comparten una misma f.d., una realización de dicha muestra  $x_1, \dots, x_n$ , preserva tal información entonces cuando, por alguna razón, se sospecha que las realizaciones se adecúan a alguna f.d.  $F^*$  es indispensable verificar la validez de nuestra afirmación [11], pues esto nos ayudaría a develar información sustancial sobre el experimento de interés.

Con el ejemplo anterior, introducimos las pruebas de bondad de ajuste, las cuales pretenden medir la discrepancia, o equivalentemente el acoplamiento, entre el comportamiento probabilístico de las realizaciones y  $F^*$ .

Formalmente, las pruebas de bondad de ajuste son pruebas de hipótesis no paramétricas cuya hipótesis nula, usualmente, es que la variable aleatoria, o el vector aleatorio,  $X$  que modela el experimento de interés tiene distribución  $F^*$  [11]. Además, la hipótesis nula suele refinarse con el objetivo de conocer los parámetros que caracterizan a  $F^*$ ; sin embargo, cuando los parámetros consisten en propiedades de localización y escala, no resulta necesario refinar la hipótesis nula, puesto que la aceptación o rechazo depende de la forma de  $F^*$  y no de los parámetros [8]. Por otra parte, la hipótesis alternativa no resulta informativa [11] puesto que, en general, no nos hace sugerencia alguna sobre el verdadero comportamiento aleatorio.

Dentro las pruebas de bondad de ajuste podemos encontrar pruebas para distribuciones discretas y también continuas. A continuación discutiremos algunas de las pruebas más conocidas.

### i) Pruebas basadas en la distribución $\chi^2$

En el caso de distribuciones discretas tenemos la prueba de  $\chi^2$  propuesta por Karl Pearson en 1900 [22]. Dada una muestra aleatoria  $X_1, \dots, X_n$  proveniente de una v.a. discreta, esta prueba supone que los datos pueden clasificarse en  $k$  diferentes categorías. Para cada  $i = 1, \dots, k$  definimos  $f_i$  como el número de observaciones

que entran dentro de la categoría  $i$  y  $e_i$  el número esperado de observaciones para la categoría  $i$ . Entonces el estadístico de la prueba  $\chi^2$  es

$$Q_{n,k} = \sum_{i=1}^k \frac{(f_i - e_i)^2}{e_i}.$$

Esta prueba de hipótesis recibe el nombre de  $\chi^2$  pues

$$Q_{n,k} \xrightarrow{D} Q,$$

con  $Q \sim \chi_{k-1}^2$ .

## ii) Pruebas basadas en la función de distribución empírica

En el caso de las v.a. continuas las pruebas de bondad de ajuste se basan en una función aleatoria conocida como Función de Distribución Empírica. Esta función es una aproximación de la f.d  $F$  común a las v.a de la muestra y será definida y estudiada con mayor profundidad en el siguiente capítulo.

De manera general, las pruebas de bondad de ajuste para v.a. continuas miden la distancia entre la función de distribución empírica  $F_n$  y la función de distribución  $F$  propuesta. Esto es, miden la diferencia entre ellas, el valor absoluto de la diferencia o el cuadrado de la diferencia. Esto puede apreciarse en los estadísticos de prueba siguientes

$$D_n^+ = \sup_{x \in \mathbb{R}} \{F_n(x) - F(x)\}. \quad (1.8)$$

$$D_n^- = \sup_{x \in \mathbb{R}} \{F(x) - F_n(x)\}. \quad (1.9)$$

$$\begin{aligned} D_n &= \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \\ &= \text{máx}\{D_n^+, D_n^-\}. \end{aligned} \quad (1.10)$$

$$\omega_n^2 = n \int_{\mathbb{R}} [F_n(x) - F(x)]^2 \psi(F(x)) dF(x), \quad (1.11)$$

con  $\psi : [0, 1] \rightarrow [0, \infty)$  una función de peso e integrando en el sentido de Lebesgue.

Por sí solos, los estadísticos en las ecuaciones (1.8) y (1.9) son de gran utilidad computacionalmente [36]; sin embargo, para ser empleados en pruebas de hipótesis es usual proponer un estadístico de prueba en donde ambos estén relacionados, como en el estadístico en la ecuación (1.10) conocido como el estadístico **Smirnov-Kolmogorov**, del cual hablaremos a lo largo de este trabajo.

Un ejemplo en donde los estadísticos  $D_n^+$  y  $D_n^-$  se relacionan para generar una prueba de hipótesis puede observarse en [39], donde derivado de estudiar la migración de aves se preguntan si la dinámica observada en dos lugares y momentos del año distintos, tiene el mismo comportamiento estocástico; para ello, toman de referencia una brújula y estudian el ángulo en el que se dirigen los grupos de aves. De esta manera, si  $z = (\sin(2\pi x), \cos(2\pi x))$  es el vector aleatorio que modela el ángulo  $x \in [0, 1)$ , es intuitivamente claro que podemos hacer el análisis estadístico sobre  $x$ ; así, para abordar la prueba de hipótesis proponen el estadístico

$$V_n = D_n^+ + D_n^-, \quad (1.12)$$

para comparar con los datos encontrados en la literatura.

Finalmente, en la ecuación (1.11) se observa la familia de estadísticos Cramér-von Mises, la cual es una alternativa a la prueba de bondad de ajuste Smirnov-Kolmogorov, que a través de la función de peso permite modular la sensibilidad de la prueba de hipótesis de acuerdo a los requerimientos acordados [12].

**Nota 47** Las pruebas de bondad de ajuste se aplican con la intención de no rechazar la hipótesis nula, porque en caso contrario no se obtiene ninguna información útil sobre la muestra estudiada.

# Función de distribución empírica

Como ya se ha señalado en la Sección 1.2.1, las pruebas de bondad de ajuste como la prueba Smirnov-Kolmogorov se basan en la función de distribución empírica. En consecuencia, en este capítulo presentamos la definición y algunas propiedades importantes de esta función. El resultado más relevante del capítulo es el Teorema de Glivenko-Cantelli, el cual nos permitirá definir al estadístico de prueba Smirnov-Kolmogorov.

Los resultados relacionados con la función de distribución empírica fueron tomados de [52], mientras que la demostración del Teorema de Glivenko-Cantelli fue tomada de [44].

La función de distribución empírica, propuesta por Francis Galton en 1875 [23], es una función asociada a la medida empírica [58] y sirve para estimar la función de distribución desconocida  $F$  generada por una muestra aleatoria. Esta función se define de la siguiente manera.

**Definición 48 (Función de distribución empírica)** Sean  $X_1, X_2, \dots, X_n$  una muestra aleatoria con f.d.  $F$ . La función aleatoria definida como

$$F_n(x, \omega) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}}, \quad \text{para } x \in \mathbb{R}, \quad (2.1)$$

donde

$$\mathbf{1}_{\{X_i \leq x\}} = \begin{cases} 1, & \text{si } \omega \in \{X_i \leq x\}, \\ 0, & \text{otro caso,} \end{cases}$$

se conoce como función de distribución empírica (f.d.e.) de la muestra aleatoria.

Notemos que la f.d.e.,  $F_n$ , es una v.a. por lo que el siguiente resultado nos muestra cuál es su distribución.

**Teorema 49** La función de distribución empírica es una v.a. cuya distribución es binomial con parámetros  $n$  y  $\frac{F}{n}$ , es decir,  $F_n \sim \text{Bino} \left( n, \frac{F}{n} \right)$ .

DEMOSTRACIÓN. Sea  $x \in \mathbb{R}$  fija. Notemos que  $E(\mathbf{1}_{\{X_i \leq x\}}) = F(x)$ , pues

$$\begin{aligned} E(\mathbf{1}_{\{X_i \leq x\}}) &= 0 \cdot P(\omega \notin \{X_i \leq x\}) + 1 \cdot P(\omega \in \{X_i \leq x\}) \\ &= F(x). \end{aligned}$$

Desde que  $\mathbf{1}_{\{X_i \leq x\}} \sim \text{Ber}(F(x))$ , para toda  $i = 1, \dots, n$  y  $x \in \mathbb{R}$  es arbitraria se satisface que  $F_n(x) \sim \text{Bino}\left(n, \frac{F(x)}{n}\right)$ , lo cual prueba el resultado.  $\square$

El siguiente ejemplo muestras como la f.d.e. estima a la f.d. a través de una muestra aleatoria.

**Ejemplo 50** Sea  $X$  una v.a. tal que  $X \sim N(0, 1)$  y sean  $X_1, \dots, X_n$  una muestra aleatoria de  $X$ . La Figura 2.1, obtenida con ayuda de R, muestra en color negro la gráfica de la f.d.e. calculada a partir de una realización de una muestra aleatoria de tamaño  $n = 30$  y en color rojo se muestra la gráfica de la f.d. de una v.a. con distribución normal estándar.

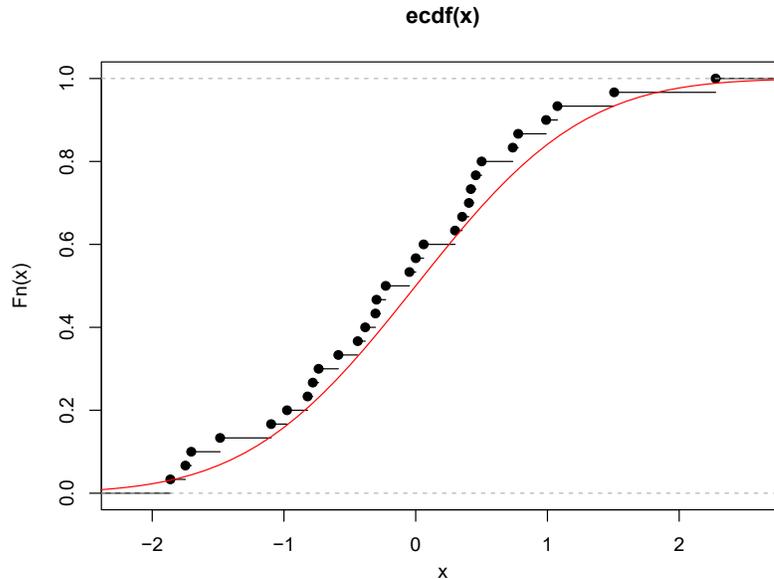


Figura 2.1: Aproximación a la función de distribución a través la función de distribución empírica.

Del ejemplo anterior y tras observar la Figura 2.1 se infiere que, similar a la función de distribución, la función de distribución empírica es no decreciente; así, tenemos el siguiente resultado.

**Teorema 51** Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria. Si  $F_n$  denota la f.d.e. de la muestra, entonces  $F_n$  es no decreciente.

DEMOSTRACIÓN. Sean  $x, y \in \mathbb{R}$  arbitrarios y fijos; además, supongamos que  $x \leq y$ . Notemos que si  $\mathbf{1}_{\{X_i \leq x\}} = 1$ , para algún  $i = 1, \dots, n$ , esto implica que  $\mathbf{1}_{\{X_i \leq y\}} = 1$ , puesto que si  $X_i \leq x$  entonces  $X_i \leq y$ , por transitividad, ya que  $x \leq y$ . De esta manera,

$$\mathbf{1}_{\{X_i \leq x\}} \leq \mathbf{1}_{\{X_i \leq y\}}, \quad \forall i = 1, \dots, n.$$

En consecuencia, por la definición de la f.d.e.,  $F_n(x) \leq F_n(y)$ .  $\square$

Los siguientes resultados representan herramientas relevantes para los fines de este trabajo de tesis, pues junto con el Teorema de Donsker, que será enunciado en el Capítulo 4, nos proporcionará una conexión entre el puente Browniano y el estadístico Smirnov-Kolmogorov.

**Definición 52 (Normalización de la función de distribución empírica)** Sean  $X_1, \dots, X_n$  variables aleatorias i.i.d., con  $F$  la f.d. de  $X_i$ , para toda  $i = 1, \dots, n$ . Si  $F_n$  denota la f.d.e., diremos que la normalización de  $F_n$  es la siguiente v.a.

$$\sqrt{n}(F_n(x) - F(x)), \text{ para toda } x \in \mathbb{R}.$$

**Nota 53** Notemos que  $F_n$  esencialmente asigna una masa  $1/n$  a cada valor  $X_i$ , con  $i = 1, \dots, n$ ; además  $F_n$  es un estimador natural de la f.d.  $F$ , pues si  $Y_i = \mathbf{1}_{\{X_i \leq x\}}$ , entonces la ley fuerte de los grandes números (ver Teorema 28) implica

$$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i \xrightarrow{\text{c.s.}} \mu,$$

donde  $\mu = E(\mathbf{1}_{\{X_i \leq x\}}) = F(x)$ .

Además, por el teorema del límite central (ver Teorema 29), se observa la cercana conexión entre la normalización de la f.d.e. y la distribución normal, además de la versatilidad que hereda de las propiedades de la convergencia en distribución.

De los trabajos de Galton sobre la función de distribución empírica emana la función cuantil [24], también conocida como la transformación inversa de una función de distribución y su definición es la siguiente.

**Definición 54 (Transformación inversa)** Sea  $F$  una función de distribución dada. La transformación inversa de la función de distribución  $F$  se define como

$$F^{-1}(t) := \inf\{x \in \mathbb{R} \mid F(x) \geq t\},$$

para  $0 < t < 1$ .

**Nota 55** Para emplear la prueba Smirnov-Kolmogorov se requiere que la f.d.  $F$  sea continua. Además, como en contextos de estadística aplicada principalmente se trabaja con f.d. estrictamente crecientes se cumple que la inversa de la función  $F$  existe y coincide con la transformación inversa de  $F$  (ver [30, pág. 294]). De esta manera, en  $t = 1$  ya no se depende de si  $F$  alcanza la recta  $y = 1$  o si dicha recta es una asíntota para la gráfica de  $F$ , entonces se puede convenir que  $F^{-1}(1) \in \mathbb{R}$  es el número tal que hace verdadera la igualdad

$$F(F^{-1}(1)) = 1.$$

Similarmente, en  $t = 0$  puede convenirse que  $F^{-1}(0) \in \mathbb{R}$  es el número tal que hace verdadera la igualdad

$$F(F^{-1}(0)) = 0,$$

donde recordamos que  $F^{-1}$  es la transformación inversa de la f.d.  $F$  y, en este caso coincide con la función inversa de  $F$ .

De la Definición 54 surge una pregunta natural, sobre si  $F^{-1}$  está bien definida para todo  $t \in (0, 1)$  por lo que se tiene el siguiente teorema.

**Teorema 56** Si  $F$  es una función de distribución dada, entonces su transformación inversa,  $F^{-1}$ , está bien definida.

**DEMOSTRACIÓN.** Sea  $u \in (0, 1)$  y definamos  $A = \{x \in \mathbb{R} \mid F(x) \geq u\}$ . Defina  $\varepsilon := 1 - u > 0$ , por el inciso (a) del Teorema 8 existe  $N_0 \in \mathbb{N}$  tal que si  $x > N_0$  entonces  $|F(x) - 1| = 1 - F(x) < \varepsilon$ , es decir,  $u < F(x)$ , por tanto como  $N_0 + 1 > N_0$  se tiene que  $F(N_0 + 1) > u$  y se sigue que  $N_0 + 1 \in A$ , por lo que  $A \neq \emptyset$ .

Por el inciso (b) del Teorema 8, para  $\varepsilon_0 = u$  existe  $M_0 \in \mathbb{N}$  tal que si  $x < -M_0$  entonces  $|F(x)| = F(x) < \varepsilon_0$ . Por tanto  $-(M_0 + 1)$  es una cota inferior de  $A$  ya que de no ser así existiría  $y \in A$  tal que  $y < -(M_0 + 1)$  y como  $F$  es no decreciente entonces  $u \leq F(y) \leq F(-(M_0 + 1))$ , una contradicción. De esta manera, como  $A$  es un conjunto no vacío y acotado inferiormente, el  $\inf A$  existe. Además, como  $u$  fue arbitraria, se tiene que  $\inf A$  existe para cada  $0 < u < 1$ .  $\square$

Para familiarizarnos con la transformación inversa de una función de distribución enunciaremos algunas propiedades que se utilizarán posteriormente. Sugerimos consultar el libro de Shorak & Wellmer [52] para profundizar en el tema y el libro de Rincón [46] para obtener una perspectiva computacional.

**Teorema 57** Sea  $F$  una f.d. dada. Si  $F^{-1}$  denota la transformación inversa de  $F$ ,

entonces  $F^{-1}$  satisface las siguientes propiedades.

- (a)  $F(F^{-1}(u)) \geq u$ , para toda  $u \in (0, 1)$ .
- (b)  $F(F^{-1}(u)^-) \leq u$ , para toda  $u \in (0, 1)$ . Donde  $F(x^-)$  es el límite por la izquierda (vea inciso (e) del Teorema 8).

DEMOSTRACIÓN.

- (a) Sea  $u \in (0, 1)$  arbitraria y fija; como antes, definamos  $A := \{x \in \mathbb{R} \mid F(x) \geq u\}$  y sea  $a := F^{-1}(u) = \inf A$ , el cual existe por el Teorema 56. Dada  $\varepsilon > 0$  se tiene que  $a < a + \varepsilon$  y, más aún,  $a + \varepsilon \in A$  pues:

si  $a \in A$ , necesariamente se cumple que  $F(a) \geq u$  y dado que  $F$  es no decreciente se tiene que

$$F(a + \varepsilon) \geq F(a) \geq u.$$

Por otro lado, si  $a \notin A$  y suponemos que  $a + \varepsilon \notin A$ , entonces  $F(a) < u$  y  $F(a + \varepsilon) < u$ , por la definición de  $A$ ; no obstante, por la definición de ínfimo, se cumple que para la  $\varepsilon$  dada existe  $\alpha \in A$  tal que  $\alpha < a + \varepsilon$  lo que implica que  $u \leq F(\alpha) \leq F(a + \varepsilon)$ , una contradicción.

De esta manera, como para cada  $\varepsilon > 0$  se cumple que  $F(a + \varepsilon) \geq u$ , entonces se tiene que

$$\lim_{\varepsilon \rightarrow 0} F(a + \varepsilon) = F(a^+) \geq u.$$

Y puesto que  $F$  es continua por la derecha se tiene que  $F(a) = F(a^+)$ , de donde se sigue que

$$F(a) = F(F^{-1}(u)) \geq u.$$

Luego, como  $u$  fue arbitraria se muestra que el resultado es válido para cada  $u \in (0, 1)$ .

- (b) Sea  $u \in (0, 1)$  arbitraria y fija. Definamos  $A := \{x \in \mathbb{R} \mid F(x) \geq u\}$  y sea  $a := \inf A = F^{-1}(u)$ , el cual existe por el Teorema 56. Dada  $\varepsilon > 0$  se cumple que  $a - \varepsilon < a$  lo que implica que  $a - \varepsilon \notin A$ , pues si  $a - \varepsilon$  fuese un elemento de  $A$  necesariamente se cumpliría, por la definición de ínfimo, que  $a \leq a - \varepsilon$ , lo cual es una contradicción.

Así, se sigue que  $F(a - \varepsilon) < u$ , para toda  $\varepsilon > 0$ , por lo que al tomar límite cuando  $\varepsilon \rightarrow 0$  se tiene que

$$\lim_{\varepsilon \rightarrow 0} F(a - \varepsilon) = F(a^-) = F(F^{-1}(u)^-) \leq u.$$

Notemos que como  $u$  fue arbitraria, se tiene que el resultado es válido para toda  $u \in (0, 1)$ , lo cual concluye la prueba.  $\square$

Del teorema anterior se desprende el siguiente corolario.

**Corolario 58** Sea  $F$  es una función de distribución dada. Si  $u \in (0, 1)$  y  $A = \{x \in \mathbb{R} \mid F(x) \geq u\}$ , entonces  $A$  es un intervalo de la forma  $(a, \infty)$  o  $[a, \infty)$ .

**DEMOSTRACIÓN.** Sea  $u \in (0, 1)$  arbitraria y fija. Del Teorema 56 se tiene que  $A$  es acotado inferiormente; luego, usando un argumento similar a la demostración del inciso (a) del Teorema 57 sabemos que para todo  $\varepsilon > 0$  se tiene que  $\inf A + \varepsilon \in A$  por lo que  $(\inf A, \infty) \subseteq A$ . Por otro lado, si  $\inf A \notin A$  para  $x \in A$  entonces  $x > \inf A$  y se sigue que  $x \in (\inf A, \infty)$ , por lo que  $A = (\inf A, \infty)$ .

Note que si  $\inf A \in A$  entonces  $x \geq \inf A$ , es decir,  $x \in [\inf A, \infty)$ . Además, si  $y \in [\inf A, \infty)$  se cumple que  $y = \inf A + (y - \inf A)$  de donde se sigue que  $y \in A$ , pues si  $y > \inf A$  entonces tomando  $\varepsilon_0 = y - \inf A > 0$  se tiene que  $\inf A + \varepsilon_0 \in A$ , mientras que si  $y = \inf A$  entonces trivialmente  $y \in A$ . De esta manera se tiene que  $A = [\inf A, \infty)$ .

Con lo anterior se ha mostrado que  $A$  es intervalo no acotado superiormente y dado que  $u$  fue arbitraria el resultado es válido para toda  $u \in (0, 1)$ .  $\square$

**Nota 59** Desde que  $u \in (0, 1)$ , los incisos (a)-(c) del Teorema 8 implican que si  $A = \{x \in \mathbb{R} \mid F(x) \geq u\}$ , entonces  $\inf A \in A$ , por lo que

$$A = [\inf A, \infty).$$

O, equivalentemente, en términos de la transformación inversa

$$A = [F^{-1}(u), \infty).$$

A continuación presentaremos un resultado que será crucial en para la construcción del estadístico de prueba de Smirnov-Kolmogorov.

**Teorema 60** Sea  $F$  una f.d. dada y  $\xi$  una v.a. con distribución uniforme, es decir,  $\xi \sim Unif(0, 1)$ . Si  $X = F^{-1}(\xi)$ , con  $F^{-1}$  la transformación inversa de la f.d.  $F$ , entonces  $X \sim F$ .

**DEMOSTRACIÓN.** Sea  $x \in \mathbb{R}$  arbitraria y fija. Demostraremos a continuación que los eventos  $E_1 := \{\omega \in \Omega \mid F(x) \geq \xi(\omega)\}$  y  $E_2 := \{\omega \in \Omega \mid x \geq F^{-1}(\xi(\omega))\}$  son equivalentes.

Sea  $\omega^* \in E_1$ . Como  $\xi(\omega^*) \in (0, 1)$ , pues  $\xi \sim Unif(0, 1)$ , entonces de la definición de la transformación inversa se tiene que

$$F^{-1}(\xi(\omega^*)) = \inf\{y \in \mathbb{R} \mid F(y) \geq \xi(\omega^*)\}.$$

Notemos que se cumple que  $F(x) \geq \xi(\omega^*)$ , puesto que  $\omega^* \in E_1$ , entonces  $x \in \{y \in \mathbb{R} \mid F(y) \geq \xi(\omega^*)\}$ ; así, por la definición de ínfimo, se tiene que  $F^{-1}(\xi(\omega^*)) \leq x$ , probando que  $E_1 \subseteq E_2$ .

Ahora, sea  $\omega^* \in E_2$ , entonces  $x \geq F^{-1}(\xi(\omega^*))$ ; de esta manera, aplicando el Corolario 58 y la Nota 59, se tiene que  $x \in \{y \in \mathbb{R} \mid F(y) \geq \xi(\omega^*)\}$ , por lo que  $F(x) \geq \xi(\omega^*)$ , mostrando que  $E_2 \subseteq E_1$ .

De lo anterior se tiene que  $E_1 = E_2$  y ya que  $x$  fue arbitraria se tiene que el resultado es válido para toda  $x \in \mathbb{R}$ . De esta manera, se sigue que

$$P(X \leq x) = P(F^{-1}(\xi) \leq x) = P(\xi \leq F(x)) = \int_0^{F(x)} \xi d\xi = F(x).$$

Es decir,  $X \sim F$ . □

## 2.1. Teorema de Glivenko-Cantelli

En esta sección profundizaremos en la conexión de las funciones de distribución con la función de distribución empírica correspondiente estudiando la convergencia entre estas. El Teorema de Glivenko-Cantelli (ver [25] y [5]), a veces denominado “Teorema fundamental de la estadística” [52], establece que la f.d.e.  $F_n$  definida en ecuación (2.1) converge uniformemente y en distribución a la f.d.  $F$ . Esto nos permite remarcar que  $F_n$  es, en efecto, un buen estimador de  $F$ . De hecho, por el Teorema 49 podemos concluir que  $F_n$  es un estimador insesgado de  $F$ , es decir, satisface que

$$E(F_n) = F.$$

**Teorema 61 (Glivenko-Cantelli)** Sea  $\{X_n\}_{n \in \mathbb{N}}$  una sucesión de variables aleatorias i.i.d. con f.d. común  $F$ . Si  $F_n$  denota la f.d.e. correspondiente definida en la ecuación (2.1) y

$$D_n := \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|, \tag{2.2}$$

entonces

$$D_n \xrightarrow{c.s.} 0.$$

**DEMOSTRACIÓN.** Sea  $F^{-1}$  la transformación inversa de  $F$  y  $k \in \mathbb{N}$  arbitraria. Para cada  $v = 1, \dots, k$  definamos

$$x_{v,k} := F^{-1}(v/k).$$

Sea  $x \in [x_{v,k}, x_{v+1,k})$ , para  $v = 1, \dots, k-1$  arbitraria y fija, el caso  $v = k$  será estudiado posteriormente. Por las propiedades de la f.d. (ver Teorema 8) se satisface que

$$F(x_{v,k}) \leq F(x) \leq F(x_{v+1,k}^-). \quad (2.3)$$

Además, para cada  $n \in \mathbb{N}$

$$F_n(x_{v,k}) \leq F_n(x) \leq F_n(x_{v+1,k}^-), \quad (2.4)$$

ya que  $F_n$  es no decreciente (ver Teorema 51).

Combinando las desigualdades (2.3) y (2.4) se obtiene

$$F_n(x_{v,k}) - F(x_{v+1,k}^-) \leq F_n(x) - F(x) \leq F_n(x_{v+1,k}^-) - F(x_{v,k}). \quad (2.5)$$

La definición de las  $x_{v,k}$  y los incisos (a) y (b) del Teorema 57 implican que

$$F(x_{v+1,k}^-) - F(x_{v,k}) \leq \frac{v+1}{k} - \frac{v}{k} = \frac{1}{k},$$

para toda  $v = 1, \dots, k-1$ .

Lo anterior, nos permite reescribir la desigualdad (2.5) de la manera siguiente

$$F_n(x_{v,k}) - F(x_{v,k}) - \frac{1}{k} \leq F_n(x) - F(x) \leq F_n(x_{v+1,k}^-) - F(x_{v+1,k}^-) + \frac{1}{k}. \quad (2.6)$$

Usando la desigualdad (2.6) junto con la desigualdad del triángulo obtenemos

$$|F_n(x) - F(x)| \leq \max\{|F_n(x_{v,k}) - F(x_{v,k})|, |F_n(x_{v+1,k}^-) - F(x_{v+1,k}^-)|\} + \frac{1}{k},$$

por lo tanto

$$\sup_{x \in [x_{v,k}, x_{v+1,k})} |F_n(x) - F(x)| \leq \max\{|F_n(x_{v,k}) - F(x_{v,k})|, |F_n(x_{v+1,k}^-) - F(x_{v+1,k}^-)|\} + \frac{1}{k},$$

para toda  $v = 1, \dots, k-1$ . En consecuencia

$$\sup_{x \in [x_{1,k}, x_{k,k})} |F_n(x) - F(x)| \leq C, \quad (2.7)$$

donde

$$C := \max_{v \in \{1, \dots, k-1\}} \{\max\{|F_n(x_{v,k}) - F(x_{v,k})|, |F_n(x_{v+1,k}^-) - F(x_{v+1,k}^-)|\}\} + \frac{1}{k}.$$

Mostraremos que

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| < C.$$

Supongamos que  $x < x_{1,k}$  y  $F(x) \leq F_n(x)$ . Como  $F_n$  es una función no decreciente, entonces

$$\begin{aligned} |F_n(x) - F(x)| &= F_n(x) - F(x) \\ &\leq F_n(x_{1,k}^-) - F(x) \\ &= F_n(x_{1,k}^-) - F(x) + F(x_{1,k}^-) - F(x_{1,k}^-) \\ &\leq |F_n(x_{1,k}^-) - F(x) + F(x_{1,k}^-) - F(x_{1,k}^-)| \end{aligned}$$

Aplicando la desigualdad del triángulo se tiene que

$$|F_n(x) - F(x)| \leq |F_n(x_{1,k}^-) - F(x_{1,k}^-)| + |F(x_{1,k}^-) - F(x)|. \quad (2.8)$$

Como  $x < x_{1,k}$  entonces  $F(x) \leq F(x_{1,k}^-)$ ; así, usando el inciso (a) del Teorema 57 se cumple que

$$\begin{aligned} -\frac{1}{k} \leq -F(x_{1,k}^-) &\leq -F(x_{1,k}^-) + F(x) \\ &\leq F(x) \leq F(x_{1,k}^-) \leq \frac{1}{k}. \end{aligned}$$

Así, se tiene que

$$|F(x) - F(x_{1,k}^-)| \leq \frac{1}{k}. \quad (2.9)$$

Empleando lo anterior en la desigualdad (2.8) se tiene lo siguiente.

$$\begin{aligned} |F_n(x) - F(x)| &\leq |F_n(x_{1,k}^-) - F(x_{1,k}^-)| + \frac{1}{k} \\ &\leq C, \end{aligned}$$

observemos que la última desigualdad se obtiene de la definición de  $C$  y se satisface para toda  $x < x_{1,k}$  tal que  $F(x) \leq F_n(x)$ .

Ahora si suponemos que  $x < x_{1,k}$  y  $F(x) > F_n(x)$ , entonces

$$\begin{aligned} |F_n(x) - F(x)| &= F_n(x) - F(x) \leq F_n(x) \\ &\leq F_n(x_{1,k}^-) \leq \frac{1}{k} \\ &\leq C, \end{aligned}$$

y en consecuencia, podemos extender la desigualdad (2.7) de la siguiente manera

$$\sup_{x \in (-\infty, x_{k,k})} |F_n(x) - F(x)| \leq C. \quad (2.10)$$

Ahora abordaremos el caso  $v = k$ , notemos que basta analizar cuando  $x \geq x_{k,k}$ , entonces  $F(x) \geq F(x_{k,k}) \geq \frac{k}{k} = 1$  y ya que  $F(x) \leq 1$ , para toda  $x \in \mathbb{R}$  se tiene que

$F(x) = 1$ . Similarmente, como  $F_n(x) \geq F_n(x_{k,k}) = F_n(1)$ , entonces usando la Nota 55 se tiene que  $F_n(1) = 1$ , lo que implica que  $F_n(x) = 1$ . Por tanto, para  $x > x_{k,k}$  se cumple que  $|F_n(x) - F(x)| = 0 \leq C$ ; así, podemos extender la desigualdad (2.10) a todos los números reales, es decir,

$$D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq C. \quad (2.11)$$

Como  $E(F_n(x)) = F(x) \leq \infty$ , entonces la ley fuerte de los grandes números (ver Teorema 28) implica que

$$F_n(x) \xrightarrow{c.s.} F(x).$$

Análogamente, como  $E(F_n(x^-)) = F(x^-)$ , entonces

$$F_n(x^-) \xrightarrow{c.s.} F(x^-).$$

De esta manera, se cumple que

$$F_n(x) - F(x) \xrightarrow{c.s.} 0 \quad \text{y} \quad F_n(x^-) - F(x^-) \xrightarrow{c.s.} 0,$$

esto implica que

$$\max_{v \in \{1, \dots, k-1\}} \left\{ \max\{|F_n(x_{v,k}) - F(x_{v,k})|, |F_n(x_{v+1,k}^-) - F(x_{v+1,k}^-)|\} \right\} \xrightarrow{c.s.} 0,$$

por lo que  $C = \frac{1}{k}$  casi seguramente.

Así, se sigue que

$$\limsup_{n \rightarrow \infty} D_n \leq \frac{1}{k} \quad \text{c.s.}$$

Y ya que  $k$  fue arbitraria se concluye que

$$\limsup_{n \rightarrow \infty} D_n \leq 0 \quad \text{c.s.}$$

Finalmente, como  $0 \leq D_n$  entonces  $0 \leq \liminf_{n \rightarrow \infty} D_n$ , por lo que

$$0 \leq \liminf_{n \rightarrow \infty} D_n \leq \limsup_{n \rightarrow \infty} D_n \leq 0,$$

esto implica que  $\lim_{n \rightarrow \infty} D_n = 0$  casi seguramente, lo que concluye la prueba.  $\square$

A continuación presentaremos un par de resultados inmediatos del teorema anterior.

**Corolario 62** La f.d.e.  $F_n$  converge en distribución a la f.d.  $F$ .

**Corolario 63** La f.d.e.  $F_n$  converge uniformemente a la f.d.  $F$ , en  $\mathbb{R}$ , casi seguramente.

# Puente Browniano

En este capítulo introduciremos un proceso estocástico conocido como puente Browniano. Este es un proceso Gaussiano construido a través del movimiento Browniano y resulta ser un objeto matemático con aplicaciones muy interesantes en estadística debido a que es el proceso límite de aquel estadístico definido en el Teorema de Glivenko-Cantelli.

Debido a que el movimiento y el puente Browniano son procesos Gaussianos comenzaremos presentando algunas propiedades elementales de dicha familia de procesos. Los resultados de esta sección fueron obtenidos del libro de Steel [55].

Para comenzar, hace falta introducir el siguiente concepto.

**Definición 64 (Proceso Gaussiano)** Sea  $X$  un proceso estocástico, es decir, una colección de v.a. definidas en un mismo espacio de probabilidad indexadas por algún conjunto  $T \subseteq \mathbb{R}$ . Definiremos a las distribuciones finito dimensionales del proceso  $X$  como el conjunto de las funciones de distribución conjunta (ver Definición 16) de los vectores  $(X_{t_1}, \dots, X_{t_n})$ , para toda  $n \in \mathbb{N}$ . Diremos que  $X$  es un proceso Gaussiano si sus distribuciones finito dimensionales son Gaussianas.

Es claro que el trabajar con las distribuciones finito dimensionales implica lidiar con distribuciones multivariadas; sin embargo, para los procesos Gaussianos se tiene la siguiente alternativa.

**Teorema 65** Sea  $X$  un proceso estocástico. Diremos que  $X$  es un proceso Gaussiano si, y sólo si, cualquier combinación lineal de sus componentes tiene distribución normal univariada.

El comportamiento de los procesos Gaussianos es similar al comportamiento de la distribución normal en el siguiente sentido: quedan determinados por un par de parámetros, su función media y su función de covarianza [55], que definimos a continuación.

**Definición 66 (Función media)** Sea  $X$  un proceso estocástico. La función media de

$X$  se denota y define para cada  $t \in T$  como

$$\mu_X(t) = E(X_t).$$

**Definición 67 (Función de covarianza)** Sea  $X$  un proceso estocástico. Para cada  $s, t \in T$  definimos a la función de covarianza de  $X$  por

$$\kappa_X(s, t) = E([X_s - E(X_s)][X_t - E(X_t)]).$$

Con lo anterior hemos desarrollado herramientas suficientes para comenzar el estudio del movimiento Browniano y el puente Browniano.

### 3.1. Movimiento Browniano

Descrito en un artículo científico por primera vez en 1828 por el botánico Robert Brown [51] como trayectorias incesantes, impredecibles e irregulares tras observar el desplazamiento que seguían partículas de polen al ser inmersas en una gota de agua, el movimiento Browniano revolucionó la ciencia por sus diversas propiedades y las aplicaciones que tiene en la actualidad.

La historia del movimiento Browniano va más allá de Robert Brown ya que derivado de las observaciones realizadas por el biólogo alemán Wilhelm Friedrich von Gleichen-Rußwurm sobre el movimiento de las partículas de polen fue que Brown comenzó sus propias investigaciones [51]. Asimismo, durante la primera década del siglo XX, Marian von Smoluchowski y Albert Einstein se interesaron, de manera independiente, en el movimiento Browniano como un fenómeno físico, a partir de las publicaciones de Brown.

De forma casi simultánea a las investigaciones de Einstein y Smoluchowski, el matemático Louis Bachelier en su tesis doctoral estudió al movimiento Browniano en el contexto de las fluctuaciones del mercado de acciones [35]; sin embargo, al ser un trabajo orientado completamente a la economía no fue notado de inmediato, sino hasta la década de 1960 [51].

Poco a poco, el movimiento Browniano se volvió más afín a las matemáticas y fue así como Norbert Wiener en 1923 [51], debido a los trabajos de Einstein y Smoluchowski, formaliza una construcción matemática del movimiento Browniano y es por eso que en ocasiones suele llamarse proceso de Wiener o movimiento Browniano de manera indistinta.

A partir de Wiener muchos matemáticos estudiaron el movimiento Browniano, entre ellos Andrei Kolmogorov, Paul Lévy, Zbigniew Ciesielski, Monroe Donsker, entre otros y es que tanto por parte de las matemáticas puras como de las aplicadas el movimiento Browniano generó, y sigue generando, gran curiosidad.

Los resultados que ahora presentamos pueden consultarse en [44].

**Definición 68 (Movimiento Browniano estándar)** Sea  $W = \{W_t \mid t \geq 0\}$  un proceso estocástico. Si  $W$  satisface:

- (a) Tiene incrementos independientes; esto es, para cualquier  $n \in \mathbb{N}$  y cualesquiera tiempos  $0 \leq t_1 < t_2 < \dots < t_{n-1} < t_n$ , las variables aleatorias

$$W_{t_1}, W_{t_2} - W_{t_1}, \dots, W_{t_n} - W_{t_{n-1}}$$

son independientes.

- (b) Para  $0 \leq s < t$ , la variable aleatoria  $W_t - W_s$  tiene distribución normal con esperanza 0 y varianza  $t - s$ .
- (c) Tiene trayectorias continuas casi seguramente.
- (d)  $W_0 = 0$  casi seguramente.

Diremos que  $W$  es un movimiento Browniano estándar.

Para familiarizarnos con el movimiento Browniano, observemos la figura siguiente, en la que se muestra la gráfica de una trayectoria de dicho proceso estocástico.

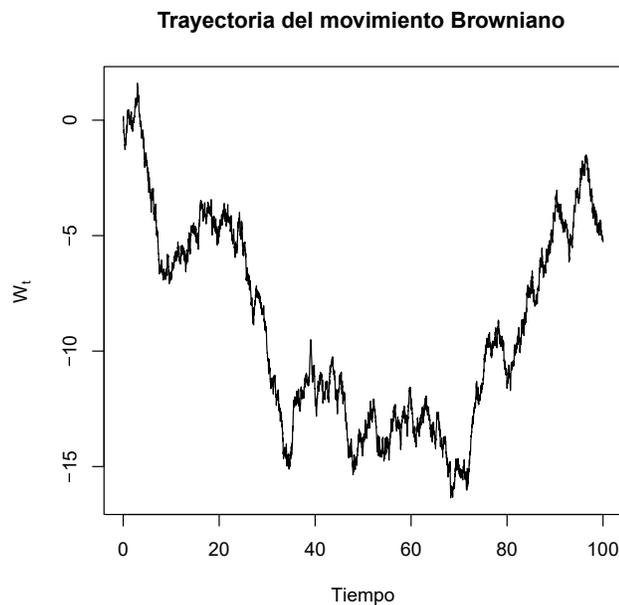


Figura 3.1: Una posible trayectoria del movimiento Browniano.

Antes de seguir, haremos la observación siguiente. Ya que trabajaremos exclusivamente con el movimiento Browniano estándar nos referiremos a él, a partir de ahora, solo como movimiento Browniano.

**Nota 69** En ocasiones, al enunciar la definición de movimiento Browniano se hace mención que es un proceso con incrementos estacionarios, es decir, satisface que dados  $0 \leq s < t$  entonces

$$W_t - W_s \stackrel{D}{=} W_{t+h} - W_{s+h}, \forall h \geq 0,$$

de donde se sigue que  $W_t - W_s \stackrel{D}{=} W_{t-s}$ . Sugerimos consultar el libro de Resnick [44] en donde se muestra que dicha propiedad se sigue de la Definición 68.

Estudiar al movimiento Browniano a partir de la perspectiva de la Definición 68 resulta útil en ciertos contextos; sin embargo, en ocasiones es conveniente verlo desde la perspectiva de los procesos Gaussianos.

**Definición 70 (Movimiento Browniano como un proceso Gaussiano)** Consideremos  $W = \{W_t \mid t \in T\}$  un proceso Gaussiano. Si  $W$  satisface:

- (a) Tiene incrementos independientes,
- (b) Para cualesquiera tiempos  $t, s \geq 0$ , se cumple que  $\mu_W(t) = 0$  y  $\kappa_W(s, t) = \min\{s, t\}$ ,
- (c) Tiene trayectorias continuas casi seguramente, y
- (d)  $W_0 = 0$  casi seguramente.

Diremos que  $W$  es un movimiento Browniano.

La demostración del resultado anterior puede consultarse en el libro de Steel [55, pp. 33]. En adición, para tener un trabajo más completo, incluimos la demostración de la existencia de este proceso estocástico en el Apéndice A.

### 3.1.1. Propiedades de las trayectorias del movimiento Browniano

A lo largo de esta sección estudiaremos algunas propiedades de las trayectorias del movimiento Browniano. El objetivo es que a través de estas, podamos obtener expresiones explícitas para las probabilidades de cruce del puente Browniano.

Los siguientes resultados provienen de [44] y [51].

**Teorema 71 (Propiedad de diferencia)** Sea  $W$  un movimiento Browniano. Definamos el proceso  $W^s = \{W_{t+s} - W_s \mid t \geq 0\}$ , para  $s > 0$  dada, entonces  $W^s$  es un movimiento Browniano independiente de  $W$ .

DEMOSTRACIÓN. Sean  $0 \leq t_1 < t_2 < \dots < t_n$  y  $\alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{R}$  para  $n \in \mathbb{N}$ . La Definición 70 y el Teorema 65 implican que la combinación lineal

$$\sum_{i=1}^n \alpha_i W_{t_i}^s = \sum_{i=1}^n \alpha_i (W_{t_{i+s}} - W_s),$$

tiene distribución normal, y, se sigue que  $W^s$  es un proceso Gaussiano.

Ahora, sean  $r, t \geq 0$  tiempos arbitrarios tales que  $0 \leq r < t$ . Como la función media de  $W$  es  $\mu_W(\tau) = 0$ , entonces  $\mu_{W^s}(t) = E(W_t^s) = E(W_{t+s} - W_s) = E(W_{t+s}) - E(W_s) = \mu_W(t+s) - \mu_W(s) = 0$ , para toda  $t \geq 0$ .

Con lo anterior, se tiene que

$$\begin{aligned} \kappa_{W^s}(r, t) &= E(W_r^s W_t^s) \\ &= E([W_{r+s} - W_s][W_{t+s} - W_s]) \\ &= E(W_{r+s} W_{t+s}) - E(W_{r+s} W_s) - E(W_s W_{t+s}) + E(W_s^2) \\ &= \kappa_W(r+s, t+s) - \kappa_W(r+s, s) - \kappa_W(s, t+s) + \kappa_W(s, s) \\ &= \min\{r+s, t+s\} - \min\{r+s, s\} - \min\{s, t+s\} + \min\{s, s\} \\ &= r \\ &= \min\{r, t\}. \end{aligned}$$

Luego, sean  $t_1, t_2, \dots, t_n$  tiempos arbitrarios. Notemos que podemos escribir los incrementos de  $W^s$  como incrementos de  $W$  de la forma siguiente

$$(W_{t_2}^s - W_{t_1}^s, \dots, W_{t_n}^s - W_{t_{n-1}}^s) = (W_{t_2+s} - W_{t_1+s}, \dots, W_{t_n+s} - W_{t_{n-1}+s}),$$

y como los incrementos de  $W$  son independientes, se sigue que los incrementos de  $W^s$  también lo son.

Por último, las propiedades (c) y (d) de la Definición 70 son inmediatas, por lo tanto,  $W^s$  es un movimiento Browniano.

Ahora, basta probar la independencia entre los procesos  $W$  y  $W^s$ , lo cual se sigue del hecho de que  $W$  es un proceso con incrementos independientes. Si  $t_i, \tau_j \geq 0$  son tiempos tales que  $t_i \leq s \leq \tau_j + s$  para  $s > 0$ , entonces las variables aleatorias  $W_{t_i}, W_s - W_{t_i}, W_{\tau_j+s} - W_s$  son independientes. Así, considerando todo par de vectores finitos  $X$  y  $Y$  con entradas cuyas componentes son incrementos de  $W$  y  $W^s$  obteniendo la propiedad de independencia.  $\square$

**Teorema 72 (Propiedad de escalamiento)** Sea  $W$  un movimiento Browniano. Para  $c \in \mathbb{R}^+$ , definimos  $\widehat{W} = \{\sqrt{c}W_{\frac{t}{c}} \mid t \geq 0\}$ . Entonces  $\widehat{W}$  es un movimiento Browniano.

DEMOSTRACIÓN. Es claro que  $\widehat{W}_0 = 0$  casi seguramente, pues  $\widehat{W}_0 = \sqrt{c}W_0$ . De igual manera, como las trayectorias de  $\widehat{W}$  son una contracción, o dilatación, de las

trayectorias de  $W$  entonces se preservan continuas casi seguramente. Además, la propiedad de incrementos independientes se preserva pues la independencia es invariante bajo transformaciones invertibles.

Falta mostrar que  $\widehat{W}$  es un proceso Gaussiano, calcular su función media y su función de covarianza. Sean  $0 \leq t_1 < t_2 < \dots < t_n$  y  $\alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{R}$  para  $n \in \mathbb{N}$ . Consideremos la combinación lineal

$$\sum_{i=1}^n \alpha_i \widehat{W}_{t_i} = \sum_{i=1}^n \alpha_i \sqrt{c} W_{t_i/c}. \quad (3.1)$$

Por definir  $\beta_i := \sqrt{c} \alpha_i, i = 1, \dots, n$  es claro que la ecuación (3.1) es una combinación lineal de las componentes de  $W, W_{t_1/c}, \dots, W_{t_n/c}$ . En consecuencia, se sigue que  $\widehat{W}$  es un proceso Gaussiano.

Ahora, observemos que la función media de  $\widehat{W}$ , es  $\mu_{\widehat{W}}(t) = E(\sqrt{c} W_{t/c}) = \sqrt{c} \mu_W(t/c) = 0$ , para toda  $t \geq 0$ .

Por otro lado, para  $s, t \geq 0$  la función de covarianza de  $\widehat{W}$  es  $\kappa_{\widehat{W}}(s, t) = c \kappa_W(s/c, t/c) = \min\{s, t\}$ . Además, las propiedades (c) y (d) de la Definición 70 se siguen inmediatamente, lo que concluye la prueba.  $\square$

**Teorema 73 (Propiedad de inversión de tiempo)** Sea  $W$  un movimiento Browniano. Definimos para cada  $t \geq 0$

$$W_t^* = \begin{cases} t W_{\frac{1}{t}} & \text{si } t > 0, \\ 0 & \text{si } t = 0. \end{cases}$$

Entonces  $W^* = \{W_t^* \mid t \geq 0\}$  es un movimiento Browniano.

DEMOSTRACIÓN. Sean  $0 \leq t_1 < t_2 < \dots < t_n$  y  $\alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{R}$ , para  $n \in \mathbb{N}$ . La combinación lineal

$$\sum_{i=1}^n \alpha_i W_{t_i}^* = \sum_{i=1}^n \alpha_i t_i W_{\frac{1}{t_i}}, \quad (3.2)$$

puede expresarse como  $\sum_{i=1}^n \beta_i W_{\frac{1}{t_i}}$ , definiendo  $\beta_i = t_i \alpha_i, i = 1, \dots, n$ . Así, por la Definición 70, y el Teorema 65, la combinación lineal en (3.2) tiene distribución normal. En consecuencia,  $W^*$  es un proceso Gaussiano.

Calculemos la función media de  $W^*$ . Si  $t = 0$ , trivialmente  $\mu_{W^*}(0) = 0$ ; así, supongamos que  $t > 0$  entonces  $\mu_{W^*}(t) = t \mu_W(\frac{1}{t}) = 0$ , para toda  $t \geq 0$ .

Ahora calculemos la función de covarianza de  $W^*$ . Si alguno de  $s$  o  $t$  es cero, es claro que  $\kappa_{W^*}(s, t) = 0$ , pues supongamos que  $s = 0$ , entonces  $\kappa_{W^*}(s, t) = E(0) = 0 = \min\{0, t\}$ .

Supongamos ahora que  $0 < s < t$  entonces  $\kappa_{W^*}(s, t) = st\kappa_W(\frac{1}{s}, \frac{1}{t}) = st \min\{\frac{1}{s}, \frac{1}{t}\}$  y ya que  $0 < s < t$  sucede si, y sólo si,  $\frac{1}{t} < \frac{1}{s}$ , entonces  $\kappa(s, t) = s = \min\{s, t\}$ .

Mostraremos que  $W^*$  es un proceso con incrementos independientes. Sean  $0 \leq t_1 < t_2 < \dots < t_n$  tiempos arbitrarios y consideremos el vector aleatorio

$$X = (t_2W_{1/t_2} - t_1W_{1/t_1}, \dots, t_nW_{1/t_n} - t_{n-1}W_{1/t_{n-1}}).$$

Notemos que si  $t_i < t_j$  entonces  $t_i \leq t_{j-1}$ , por lo que

$$\begin{aligned} \text{Cov}(W_{t_i}^* - W_{t_{i-1}}^*, W_{t_j}^* - W_{t_{j-1}}^*) &= \text{Cov}(t_iW_{1/t_i}, t_jW_{1/t_j}) - \text{Cov}(t_iW_{1/t_i}, t_{j-1}W_{1/t_{j-1}}) \\ &\quad - \text{Cov}(t_{i-1}W_{1/t_{i-1}}, t_jW_{1/t_j}) \\ &\quad + \text{Cov}(t_{i-1}W_{1/t_{i-1}}, t_{j-1}W_{1/t_{j-1}}) \\ &= \min\{t_i, t_j\} - \min\{t_i, t_{j-1}\} - \min\{t_{i-1}, t_j\} \\ &\quad + \min\{t_{i-1}, t_{j-1}\} \\ &= 0. \end{aligned}$$

Lo anterior implica que la matriz de covarianza del vector  $X$  es una matriz diagonal y, a su vez, esto implica la independencia de los incrementos.

Por otra parte, las trayectorias de  $W^*$ , al ser producto de funciones continuas resultan ser continuas en  $(0, \infty)$  con probabilidad uno y, más fuerte aún, son continuas en  $[0, \infty)$  con probabilidad uno, ya que en todo elemento de  $\{[a, b] \cap \mathbb{Q} \mid 0 \leq a < b, a, b \in \mathbb{Q}\}$  se cumple que

$$(W_{t_1}^*, \dots, W_{t_n}^*) \stackrel{D}{=} (W_{t_1}, \dots, W_{t_n}),$$

para cualesquiera tiempos  $a \leq t_1 < \dots < t_n \leq b$  y cualquier  $n \in \mathbb{N}$ .

De esta manera,  $\{W_t^* \mid t \in \mathbb{Q} \cap [0, 1]\}$  y  $\{W_t \mid t \in \mathbb{Q} \cap [0, 1]\}$  tienen el mismo comportamiento estocástico, por lo que

$$0 \stackrel{\text{c.s.}}{=} \lim_{\substack{t \rightarrow 0^+ \\ t \in \mathbb{Q}}} W_t \stackrel{\text{c.s.}}{=} \lim_{\substack{t \rightarrow 0^+ \\ t \in \mathbb{Q}}} W_t^*.$$

Luego, como el conjunto de los racionales es denso en  $\mathbb{R}$  y, además,  $W^*$  tiene trayectorias continuas en  $(0, \infty)$ , entonces se concluye que

$$\lim_{t \rightarrow 0^+} W_t^* = 0 \text{ casi seguramente,}$$

mostrando que las trayectorias de  $W^*$  son continuas casi seguramente en  $[0, \infty)$ . Por último, notemos que la propiedad (d) de la Definición 70 se verifica directamente, por lo que  $W^*$  es un movimiento Browniano.  $\square$

## 3.2. Puente Browniano

El puente Browniano, también conocido como “movimiento Browniano atado”, es un proceso estocástico útil para modelar fenómenos aleatorios que parten de un estado inicial y al final de la observación se espera que regresen a tal estado [32] y es por esto que, coloquialmente, está “atado”. Ahora, procedamos a introducirlo formalmente. Para cada  $0 \leq t \leq 1$ , definimos

$$B_t^0 = \sum_{j=1}^{\infty} \sum_{k=0}^{2^j-1} Z_{j,k} \int_0^t H_{j,k}(\tau) d\tau, \quad (3.3)$$

donde  $\{Z_{j,k} \mid j, k \in \mathbb{N} \cup \{0\}, \text{ con } 0 \leq k < 2^j - 1\}$  es una colección de v.a. i.i.d. con distribución  $Z_{j,k} \sim N(0, 1)$  y las funciones  $H_{j,k}$  son como en la ecuación (A.2).

Entonces  $B^0 = \{B_t^0 \mid 0 \leq t \leq 1\}$  es un puente Browniano. Observe que las ecuaciones (A.6) y (A.8) implican que  $B_0^0 = B_1^0 = 0$ .

**Nota 74** (I) Observemos que al omitir el primer sumando de la ecuación (A.9) se obtiene la ecuación (3.3); de esto, es intuitivamente claro que podemos entender al puente Browniano como un movimiento Browniano definido en el intervalo  $[0, 1]$  condicionado en los extremos.

(II) Directamente de la ecuación (3.3) sabemos que el puente Browniano está bien definido y que sus trayectorias son continuas con probabilidad uno, ya que los resultados de la Sección 3.1 de este capítulo se mantienen válidos.

Del inciso (I) de la nota anterior se desprende el siguiente resultado.

**Teorema 75** Sea  $W$  un movimiento Browniano definido en el intervalo  $[0, 1]$ . Si  $B^0$  es un puente Browniano, entonces

$$B_t^0 \stackrel{D}{=} W_t - tW_1,$$

para toda  $0 \leq t \leq 1$ .

DEMOSTRACIÓN. De la ecuación (A.9) sabemos que

$$\begin{aligned} W_t &= \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} Z_{j,k} \int_0^t H_{j,k}(\tau) d\tau \\ &= Z_{0,0} \int_0^t H_{0,0}(\tau) d\tau + \sum_{j=1}^{\infty} \sum_{k=0}^{2^j-1} Z_{j,k} \int_0^t H_{j,k}(\tau) d\tau. \end{aligned}$$

Luego, la ecuación (A.5) implica que

$$\begin{aligned} Z_{0,0} \int_0^t H_{0,0}(\tau) d\tau + \sum_{j=1}^{\infty} \sum_{k=0}^{2^j-1} Z_{j,k} \int_0^t H_{j,k}(\tau) d\tau &= tZ_{0,0} \int_0^1 H_{0,0}(\tau) d\tau + B_t^0 \\ &= tW_1 + B_t^0, \end{aligned}$$

en consecuencia,  $B_t^0 \stackrel{D}{=} W_t - tW_1$ .  $\square$

Para ilustrar la conexión entre el puente Browniano y el movimiento Browniano, en la figura siguiente se observa una trayectoria del puente, la cual al compararla con la Figura 3.1 puede apreciarse una cierta similitud en su comportamiento.

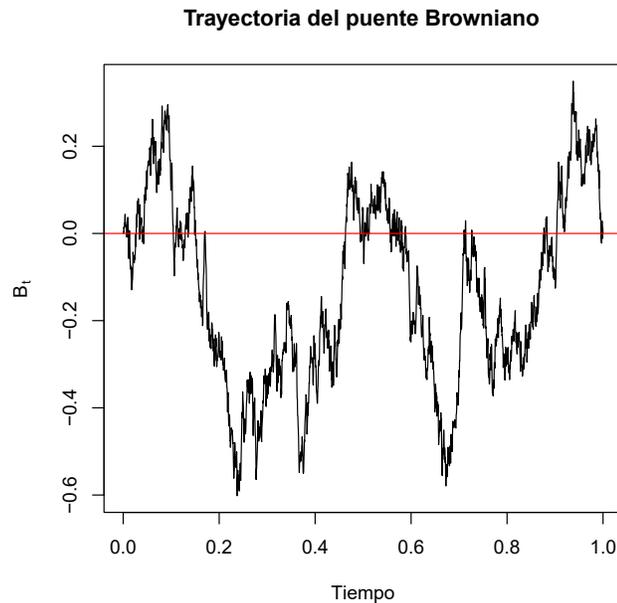


Figura 3.2: Una posible trayectoria del puente Browniano.

Motivados por el resultado anterior, la cual nos permite ver la conexión entre el puente Browniano y el movimiento Browniano, introduciremos la definición del puente Browniano como un proceso Gaussiano, para ello determinaremos su función media y su función de covarianza.

**Teorema 76** Sea  $B^0$  un puente Browniano. Entonces  $B^0$  es un proceso Gaussiano con función media

$$\mu_{B^0}(t) = 0, \text{ para cada } 0 \leq t \leq 1,$$

y con función de covarianza

$$\kappa_{B^0}(s, t) = \min\{s, t\} - st, \text{ para cada } s, t \in [0, 1].$$

DEMOSTRACIÓN. El Teorema 98 implica que

$$\mu_{B^0}(t) = \sum_{j=1}^{\infty} \sum_{k=0}^{2^j-1} \lambda_{j,k} \Delta_{j,k}(t) E(Z_{j,k}) = 0.$$

Por otro lado, al aplicar el Teorema 75, se tiene que

$$\begin{aligned} \kappa_{B^0}(s, t) &= \text{Cov}(B_s^0, B_t^0) = \text{Cov}(W_s, W_t) - s\text{Cov}(W_1, W_t) - t\text{Cov}(W_s, W_1) \\ &\quad + st\text{Cov}(W_1, W_1) \\ &= \min\{s, t\} - s \min\{1, t\} - t \min\{s, 1\} + st \min\{1, 1\} \\ &= \min\{s, t\} - st. \end{aligned}$$

Notemos lo siguiente

$$\sum_{i=1}^n \lambda_i B_{t_i}^0 = \sum_{i=1}^n \lambda_i [W_{t_i} - t_i W_1] = \sum_{i=1}^n \lambda_i W_{t_i} - \sum_{i=1}^n \lambda_i t_i W_1, \quad (3.4)$$

donde la primer suma de la ecuación (3.4) tiene una distribución normal, pues es una combinación lineal de las componentes de  $W$ . Luego, por definir  $\widehat{\lambda}_1 := -\sum_{i=1}^n \lambda_i t_i$  se tiene  $\widehat{\lambda}_1 W_1$  tiene distribución normal; de esta manera la combinación lineal de las componentes de  $B_t^0$  tiene distribución normal, pues la suma de v.a. con distribución normal tiene distribución normal. En consecuencia,  $B^0$  es un proceso Gaussiano.  $\square$

A continuación, veamos un corolario del teorema anterior.

**Proposición 77** Sea  $B^0$  un puente Browniano. Entonces  $B_t^0 \sim N(0, t(1-t))$ , para cada  $0 \leq t \leq 1$ .

Al inicio de esta sección, se mencionó que podemos entender a un puente Browniano como un movimiento Browniano en el intervalo  $[0, 1]$  condicionado en uno de sus extremos. En lo siguiente, exploraremos con mayor detalle tal afirmación.

**Teorema 78** Sea  $B^0$  un puente Browniano. Si  $W$  es un movimiento Browniano en  $[0, 1]$ , entonces

$$B_t^0 \stackrel{D}{=} (W_t \mid W_1 = 0),$$

para toda  $0 \leq t \leq 1$ .

DEMOSTRACIÓN. Sea  $f_{W_t|W_1}$  la función de distribución de la v.a.  $W_t$  condicionada por  $W_1$ . Sabemos que

$$f_{W_t|W_1}(x | 0) = \frac{f_{W_t, W_1}(x, 0)}{f_{W_1}(0)}, \quad (3.5)$$

donde  $f_{W_t, W_1}$  es la función de densidad conjunta de  $(W_t, W_1)$  y  $f_{W_1}$  es la función de densidad de  $W_1$ .

Puesto que los eventos

$$\{\omega \in \Omega \mid W_t(\omega) = x, W_1(\omega) = 0\} \quad \text{y} \quad \{\omega \in \Omega \mid W_t(\omega) = x, W_1(\omega) - W_t = -x\},$$

son equivalentes, entonces aplicando la propiedad de incrementos independientes, podemos escribir la ecuación (3.5) de la manera siguiente

$$f_{W_t|W_1}(x | 0) = \frac{f_{W_t}(x)f_{W_1-W_t}(-x)}{f_{W_1}(0)}.$$

Luego, la propiedad de incrementos estacionarios implica que  $W_1 - W_t \stackrel{D}{=} W_{1-t}$ , por lo tanto

$$\begin{aligned} f_{W_t|W_1}(x | 0) &= \frac{1}{\sqrt{2\pi t(1-t)}} e^{-\frac{1}{2}(x^2/t)} e^{-\frac{1}{2}(x^2/(1-t))} \\ &= \frac{1}{\sqrt{2\pi t(1-t)}} e^{-\frac{1}{2} \frac{x^2}{t(1-t)}}. \end{aligned}$$

En consecuencia,  $(W_t \mid W_1 = 0) \sim N(0, t(1-t))$ ; así, la Proposición 77 implica que  $W_t - tW_1 \stackrel{D}{=} (W_t \mid W_1 = 0)$ .  $\square$

### 3.3. Probabilidades de cruce

A lo largo de esta sección presentaremos cálculos explícitos para las probabilidades de cruce del movimiento y puente Browniano. Con probabilidad de cruce nos referimos a la probabilidad de que, dada una función lineal  $f(t)$ , el movimiento o el puente Browniano la cruce en algún momento. Asimismo, realizaremos una revisión de resultados previos. La metodología para resolver el resultado principal de esta sección será similar a la usada en [42], [50], [59].

El siguiente teorema, tomado de [16], constituye la primera herramienta en que se cimienta este trabajo de investigación.

**Teorema 79** Sea  $W$  un movimiento Browniano y  $m, b \in \mathbb{R}$ . Entonces

$$P\left(\sup_{t \geq 0} \{W_t \geq mt + b\}\right) = P\left(\bigcup_{t \geq 0} \{W_t \geq mt + b\}\right) = \begin{cases} e^{-2mb}, & \text{si } m, b > 0 \\ 1, & \text{en otro caso.} \end{cases}$$

Para apreciar las implicaciones del resultado anterior enunciaremos el siguiente teorema y procederemos a su prueba.

**Teorema 80** Sea  $W$  un movimiento Browniano y sean  $b, k, m, t, x \in \mathbb{R}$  tales que  $m, k > 0$  y  $mt + b > x$ . Entonces

$$P\left(\bigcup_{t \geq k} \{W_t \geq mt + b\}\right) = 1 - \Phi\left(m\sqrt{k} + \frac{b}{\sqrt{k}}\right) + e^{-2mb}\Phi\left(-m\sqrt{k} + \frac{b}{\sqrt{k}}\right)$$

y 1, en otro caso. Aquí,  $\Phi$  denota la f.d. de una v.a. normal estándar.

DEMOSTRACIÓN. Observemos que

$$P\left(\bigcup_{t \geq k} \{W_t \geq mt + b\}\right) = \int_{\mathbb{R}} P\left(\bigcup_{t \geq k} \{W_t \geq mt + b\} \mid W_k = x\right) f_{W_k}(x) dx, \quad (3.6)$$

donde  $f_{W_k}$  es la f.d. de  $W_k$ .

Notemos que

$$P\left(\bigcup_{t \geq k} \{W_t \geq mt + b\} \mid W_k = x\right) = P\left(\bigcup_{t \geq k} \{W_t - W_k \geq mt + b - x\} \mid W_k = x\right).$$

Ahora, ya que  $W$  es un movimiento Browniano entonces tiene incrementos independientes; así, dado los tiempos  $k \leq t$  se tiene que las variables aleatorias  $W_k, W_t - W_k$  son independientes, por lo que

$$\begin{aligned} & P\left(\bigcup_{t \geq k} \{W_t - W_k \geq mt + b - x\} \mid W_k = x\right) \\ &= P\left(\bigcup_{t \geq k} \{W_t - W_k \geq mt + b - x\}\right). \end{aligned}$$

Luego, como  $W$  es un proceso con incrementos estacionarios se sigue que

$$P\left(\bigcup_{t \geq k} \{W_t - W_k \geq mt + b - x\}\right) = P\left(\bigcup_{t \geq k} \{W_{t-k} \geq mt + b - x\}\right).$$

Así, tomando  $\tau = t - k$  se puede reindexar los conjuntos, lo que implica

$$\begin{aligned} P\left(\bigcup_{t \geq 0} \{W_t \geq mt + b\}\right) &= P\left(\bigcup_{\tau \geq 0} \{W_\tau \geq m(\tau + k) + b - x\}\right) \\ &= P\left(\bigcup_{\tau \geq 0} \{W_\tau \geq m\tau + (mk + b - x)\}\right). \end{aligned}$$

Tomando como  $m_0 = m$  y  $b_0 = mk + b - x$  se sigue del Teorema 79 que

$$P\left(\bigcup_{t \geq k} \{W_t \geq mt + b\} \mid W_k = x\right) = \begin{cases} e^{-2m(mk+b-x)}, & \text{si } m > 0, mk + b > x \\ 1, & \text{en otro caso.} \end{cases}$$

Usando lo anterior, en la ecuación (3.6), se tiene que si  $m \leq 0$  o  $mk + b \leq x$  es directo notar que

$$P\left(\bigcup_{t \geq k} \{W_t \geq mt + b\}\right) = 1.$$

Entonces

$$\begin{aligned} \int_{\mathbb{R}} P\left(\bigcup_{t \geq k} \{W_t \geq mt + b\} \mid W_k = x\right) f_{W_k}(x) dx &= \int_{-\infty}^{mk+b} e^{-2m(mk+b-x)} \left(\frac{1}{\sqrt{2\pi k}} e^{-\frac{x^2}{2}}\right) dx \\ &\quad + \int_{mk+b}^{\infty} \frac{1}{\sqrt{2\pi k}} e^{-\frac{x^2}{2}} dx, \end{aligned}$$

cuando  $m > 0$  y  $mk + b > x$ .

Por un lado, el primer sumando

$$\begin{aligned} &\int_{-\infty}^{mk+b} e^{-2m(mk+b-x)} \left(\frac{1}{\sqrt{2\pi k}} e^{-\frac{x^2}{2}}\right) dx \\ &= \frac{1}{\sqrt{2\pi k}} \int_{-\infty}^{mk+b} \exp(-2(m^2k + mb)) \exp\left(\left(4mx - \frac{x^2}{k}\right)/2k\right) dx \\ &= \frac{1}{\sqrt{2\pi k}} \exp(-2(m^2k + mb)) \int_{-\infty}^{mk+b} \exp\left(\left(4mx - \frac{x^2}{k} + 4m^2k - 4m^2k\right)/2\right) dx \\ &= \frac{1}{\sqrt{2\pi k}} \exp(-2mb) \int_{-\infty}^{mk+b} \exp\left(-\frac{1}{2} \left(\frac{x - 2mk}{\sqrt{k}}\right)^2\right) dx \\ &= \exp(-2mb) P(Y \leq mk + b). \end{aligned}$$

Donde  $Y \sim N(2mk, k)$ . De esta manera, si  $Z$  es una v.a. tal que  $Z \sim N(0, 1)$ , se cumple que  $\frac{Y - 2mk}{\sqrt{k}} \stackrel{D}{=} Z$ , permitiéndonos reescribir lo anterior como

$$\begin{aligned} \exp(-2mb)P(Y \leq mk + b) &= \exp(-2mb)P\left(Z \leq -m\sqrt{k} + \frac{b}{\sqrt{k}}\right) \\ &= \exp(-2mb)\Phi\left(-m\sqrt{k} + \frac{b}{\sqrt{k}}\right). \end{aligned} \quad (3.7)$$

Por otro lado,

$$\int_{mk+b}^{\infty} \frac{1}{\sqrt{2\pi k}} e^{-\frac{x^2}{2}} dx = P(\tilde{Y} \geq mk + b).$$

Donde  $\tilde{Y} \sim N(0, k)$ , en consecuencia  $\frac{\tilde{Y}}{\sqrt{k}} \stackrel{D}{=} Z$ , por lo que

$$\begin{aligned} P(Y \geq mk + b) &= P\left(Z \geq m\sqrt{k} + \frac{b}{\sqrt{k}}\right) \\ &= 1 - P\left(Z \leq m\sqrt{k} + \frac{b}{\sqrt{k}}\right) \\ &= 1 - \Phi\left(m\sqrt{k} + \frac{b}{\sqrt{k}}\right). \end{aligned} \quad (3.8)$$

De las ecuaciones (3.7) y (3.8) se sigue que

$$\begin{aligned} P\left(\bigcup_{\tau \geq 0} \{W_t \geq mt + b\}\right) &= \\ &= 1 - \Phi\left(m\sqrt{k} + \frac{b}{\sqrt{k}}\right) + \exp(-2mb)\Phi\left(-m\sqrt{k} + \frac{b}{\sqrt{k}}\right). \end{aligned}$$

Y por tanto, se obtiene el resultado.  $\square$

La versatilidad del Teorema 79 se refleja también en probabilidades de cruce que involucran al puente Browniano pero, para proseguir, realizaremos un pequeño recordatorio respecto al puente Browniano.

**Nota 81** Previamente probamos que si  $T = 1$  entonces el puente Browniano  $B^0 = \{B_t^0 \mid 0 \leq t \leq T\}$  satisface  $B_t \stackrel{D}{=} W_t - \frac{t}{T}W_T$ , para  $0 \leq t \leq T$ , donde  $W = \{W_t \mid t \geq 0\}$  es un movimiento Browniano.

Y más aún, probamos que  $W_t - \frac{t}{T}W_T \stackrel{D}{=} (W_t \mid W_T = 0)$ , lo cual nos permite trabajar con el puente Browniano a través de considerar al movimiento Browniano condicionado,

lo cual jugará un rol crucial pues esto nos permitirá emplear los resultados ya conocidos de probabilidades de cruce del movimiento Browniano para abordar el problema central de este trabajo.

**Teorema 82** Sea  $W$  un movimiento Browniano y  $b, k, m \in \mathbb{R}$ . Si  $b > 0$  y  $mk + b > x$ , entonces

$$P\left(\bigcup_{0 \leq t \leq k} \{W_t \geq mt + b\} \mid W_k = x\right) = e^{-2\left(mb + \frac{b(b-x)}{k}\right)}$$

y 1, en otro caso.

DEMOSTRACIÓN. Usando que  $W$  es un proceso estocástico continuo a tiempo continuo y la propiedad de inversión de tiempo (ver Teorema 73) se tiene que

$$\begin{aligned} P\left(\bigcup_{0 < t \leq k} \{W_t \geq mt + b\} \mid W_k = x\right) &= P\left(\bigcup_{0 < t \leq k} \{tW_{\frac{1}{t}} \geq mt + b\} \mid kW_{\frac{1}{k}} = x\right) \\ &= P\left(\bigcup_{0 < t \leq k} \left\{W_{\frac{1}{t}} \geq m + \frac{b}{t}\right\} \mid W_{\frac{1}{k}} = \frac{x}{k}\right). \end{aligned}$$

Haciendo  $\tau = \frac{1}{t}$

$$\begin{aligned} P\left(\bigcup_{0 < t \leq k} \left\{W_{\frac{1}{t}} \geq m + \frac{b}{t}\right\} \mid W_{\frac{1}{k}} = \frac{x}{k}\right) &= P\left(\bigcup_{\tau \geq \frac{1}{k}} \{W_\tau \geq m + b\tau\} \mid W_{\frac{1}{k}} = \frac{x}{k}\right) \\ &= P\left(\bigcup_{\tau \geq \frac{1}{k}} \left\{W_\tau - W_{\frac{1}{k}} \geq m + b\tau - \frac{x}{k}\right\} \mid W_{\frac{1}{k}} = \frac{x}{k}\right). \end{aligned}$$

Tras aplicar la propiedad de incrementos independientes y estacionarios obtenemos que

$$\begin{aligned} P\left(\bigcup_{\tau \geq \frac{1}{k}} \left\{W_\tau - W_{\frac{1}{k}} \geq m + b\tau - \frac{x}{k}\right\} \mid W_{\frac{1}{k}} = \frac{x}{k}\right) &= P\left(\bigcup_{\tau \geq \frac{1}{k}} \left\{W_\tau - W_{\frac{1}{k}} \geq m + b\tau - \frac{x}{k}\right\}\right) \\ &= P\left(\bigcup_{\tau \geq \frac{1}{k}} \left\{W_{\tau - \frac{1}{k}} \geq m + b\tau - \frac{x}{k}\right\}\right). \end{aligned}$$

Así, introduciendo el cambio de variable  $t = \tau - \frac{1}{k}$  podemos reindexar los conjuntos de la siguiente manera

$$\begin{aligned} P\left(\bigcup_{\tau \geq \frac{1}{k}} \left\{W_{\tau - \frac{1}{k}} \geq m + b\tau - \frac{x}{k}\right\}\right) &= P\left(\bigcup_{t \geq 0} \left\{W_t \geq m + b\left(t + \frac{1}{k}\right) - \frac{x}{k}\right\}\right) \\ &= P\left(\bigcup_{t \geq 0} \left\{W_t \geq bt + \left(m + \frac{b-x}{k}\right)\right\}\right). \end{aligned}$$

Aplicando el Teorema 79 a la expresión anterior y tomando como  $m_0 = b$  y  $b_0 = m + \frac{b-x}{k}$  se tiene que si  $m + \frac{b-x}{k} > 0$ , o equivalentemente  $mk + b > x$ , y  $b > 0$  entonces

$$\begin{aligned} P\left(\bigcup_{0 \leq t \leq k} \{W_t \geq mt + b\} \mid W_k = x\right) &= \exp\left(-2b\left(m + \frac{b-x}{k}\right)\right) \\ &= \exp\left(-2\left(mb + b\left(\frac{b-x}{k}\right)\right)\right) \end{aligned}$$

y 1, en otro caso.  $\square$

De la siguiente proposición (ver ecuación (4.3) de [16]), se desprende un resultado crucial de este trabajo.

**Proposición 83** Sea  $W$  un movimiento Browniano y  $\alpha, \beta, y, z \in \mathbb{R}$ . Si  $\alpha, z \geq 0$  y  $\beta, y > 0$  entonces se satisface

$$P(-(\alpha t + \beta) \leq W_t \leq zt + y, \forall t \geq 0) = \begin{cases} 1 - \sum_{i=1}^{\infty} [e^{-2A_i} + e^{-2B_i} - e^{-2C_i} - e^{-2D_i}] & \\ 0, & \text{otro caso.} \end{cases} \quad (3.9)$$

Donde

$$\begin{aligned} A_i &= i^2 zy + (i-1)^2 \alpha \beta + i(i-1)(z\beta + y\alpha), \\ B_i &= (i-1)^2 zy + i^2 \alpha \beta + i(i-1)(z\beta + y\alpha), \\ C_i &= i^2(zy + \alpha\beta) + i(i-1)z\beta + i(i+1)y\alpha, \quad y \\ D_i &= i^2(zy + \alpha\beta) + i(i+1)z\beta + i(i-1)y\alpha. \end{aligned} \quad (3.10)$$

En particular, si  $\alpha = z \geq 0$  y  $\beta = y > 0$ , entonces

$$P(|W_t| \leq zt + y, \forall t \geq 0) = \begin{cases} 1 - 2 \sum_{i=1}^{\infty} (-1)^{i+1} e^{-2i^2 zy} & \\ 0, & \text{otro caso.} \end{cases}$$

**Nota 84** Consideremos el caso  $\alpha = z \geq 0$  y  $\beta = y > 0$  entonces de la ecuación (3.9) se tiene que

$$P(|W_t| \leq zt + y, \forall t \geq 0) = 1 - \sum_{i=1}^{\infty} [e^{-2A_i} + e^{-2B_i} - e^{-2C_i} - e^{-2D_i}]. \quad (3.11)$$

Con

$$\begin{aligned} A_i &= zy(i^2 + (i-1)^2 + 2i(i-1)) = zy(i + (i-1))^2 = zy(2i-1)^2 \\ &= B_i \end{aligned}$$

y

$$\begin{aligned} C_i &= zy(2i^2 + i(i-1) + i(i+1)) = 4i^2zy \\ &= D_i. \end{aligned}$$

Por lo tanto

$$\begin{aligned} 2 \sum_{i=1}^{\infty} (-1)^{i+1} e^{-2i^2zy} &= \sum_{i=1}^{\infty} [e^{-2A_i} + e^{-2B_i} - e^{-2C_i} - e^{-2D_i}] \\ &= 2 \sum_{i=1}^{\infty} e^{-2(2i-1)^2zy} - e^{-8i^2zy}. \end{aligned}$$

Es decir, la probabilidad de que el movimiento Browniano quede dentro de la región determinada por las rectas  $zt + y$  y su negativo, en todo momento, es

$$1 - 2 \sum_{i=1}^{\infty} e^{-2(2i-1)^2zy} - e^{-8i^2zy}.$$

Presentamos ahora el resultado principal de este trabajo, el cual nos permitirá estudiar la distribución límite del estadístico de prueba Smirnov-Kolmogorov.

**Teorema 85** Sea  $W$  un movimiento Browniano y  $b, k, m, x \in \mathbb{R}$ . Si  $b \geq 0$  y  $mk + b > x$ , entonces

$$\begin{aligned} &P\left(\bigcup_{0 \leq t \leq k} \{|W_t| \geq mt + b\} \mid W_k = x\right) \\ &= 2 \sum_{i=1}^{\infty} \left[ \exp\left(-2(2i-1)^2 \left(mb + \frac{b^2}{k}\right)\right) \cosh\left(-2(-2i+1)^2 \left(\frac{bx}{k}\right)\right) \right. \\ &\quad \left. - \exp\left(-8i^2 \left(mb + \frac{b^2}{k}\right)\right) \cosh\left(-4i \left(\frac{bx}{k}\right)\right) \right] \end{aligned}$$

y 1, en otro caso.

DEMOSTRACIÓN. Sabemos que

$$P\left(\bigcup_{0 \leq t \leq k} \{|W_t| \geq mt + b\} \mid W_k = x\right) = 1 - P\left(\bigcap_{0 \leq t \leq k} \{|W_t| < mt + b\} \mid W_k = x\right). \quad (3.12)$$

Para calcular el segundo sumando de la ecuación (3.12) usamos que  $W$  es un proceso estocástico continuo a tiempo continuo y la propiedad de inversión de tiempo (ver Teorema 73), por tanto

$$\begin{aligned} P\left(\bigcap_{0 \leq t \leq k} \{|W_t| < mt + b\} \mid W_k = x\right) &= P\left(\bigcap_{0 < t \leq k} \{|W_t| < mt + b\} \mid W_k = x\right) \\ &= P\left(\bigcap_{0 < t \leq k} \{t|W_{\frac{1}{t}}| < mt + b\} \mid kW_{\frac{1}{k}} = x\right). \end{aligned}$$

Introduciendo el cambio de variable  $\tau = \frac{1}{t}$  se obtiene

$$\begin{aligned} P\left(\bigcap_{0 < t \leq k} \{t|W_{\frac{1}{t}}| < mt + b\} \mid kW_{\frac{1}{k}} = x\right) \\ &= P\left(\bigcap_{\tau \geq \frac{1}{k}} \{|W_\tau| < b\tau + m\} \mid W_{\frac{1}{k}} = \frac{x}{k}\right) \\ &= P\left(\bigcap_{\tau \geq \frac{1}{k}} \{-(b\tau + m) < W_\tau < b\tau + m\} \mid W_{\frac{1}{k}} = \frac{x}{k}\right) \\ &= P\left(\bigcap_{\tau \geq \frac{1}{k}} \left\{-\left(b\tau + m + \frac{x}{k}\right) < W_\tau - W_{\frac{1}{k}} < b\tau + m - \frac{x}{k}\right\} \mid W_{\frac{1}{k}} = \frac{x}{k}\right). \end{aligned} \quad (3.13)$$

Aplicando la propiedad (a) de la Definición 70 y la Nota 69 en la ecuación (3.13) se obtiene

$$\begin{aligned} P\left(\bigcap_{\tau \geq \frac{1}{k}} \left\{-\left(b\tau + m + \frac{x}{k}\right) < W_\tau - W_{\frac{1}{k}} < b\tau + m - \frac{x}{k}\right\} \mid W_{\frac{1}{k}} = \frac{x}{k}\right) \\ &= P\left(\bigcap_{\tau \geq \frac{1}{k}} \left\{-\left(b\tau + m + \frac{x}{k}\right) < W_\tau - W_{\frac{1}{k}} < b\tau + m - \frac{x}{k}\right\}\right) \\ &= P\left(\bigcap_{\tau \geq \frac{1}{k}} \left\{-\left(b\tau + m + \frac{x}{k}\right) < W_{\tau - \frac{1}{k}} < b\tau + m - \frac{x}{k}\right\}\right). \end{aligned}$$

Tomando  $t = \tau - \frac{1}{k}$

$$\begin{aligned} P \left( \bigcap_{\tau \geq \frac{1}{k}} \left\{ - \left( b\tau + m + \frac{x}{k} \right) < W_{\tau - \frac{1}{k}} < b\tau + m - \frac{x}{k} \right\} \right) \\ = P \left( \bigcap_{t \geq 0} \left\{ - \left( bt + m + \frac{b+x}{k} \right) \leq W_t \leq bt + m + \frac{b-x}{k} \right\} \right). \end{aligned}$$

Note que en la igualdad anterior hemos usado que las trayectorias de  $W$  son continuas con probabilidad uno.

Ahora, considerando  $\alpha = b$ ,  $\beta = m + \frac{b+x}{k}$ ,  $z = b$  y  $y = m + \frac{b-x}{k}$  en la Proposición 3.9 se tiene que

$$\begin{aligned} \alpha\beta = mb + \frac{b^2 + bx}{k}, \quad z\beta = mb + \frac{b^2 + bx}{k}, \quad y\alpha = mb + \frac{b^2 - bx}{k}, \\ zy = mb + \frac{b^2 - bx}{k}, \quad z\beta + y\alpha = 2 \left( mb + \frac{b^2}{k} \right), \quad \text{y} \quad zy + \alpha\beta = 2 \left( mb + \frac{b^2}{k} \right). \end{aligned}$$

De esta manera, sustituyendo los coeficientes en (3.10) de la Proposición 83, se obtiene

$$\begin{aligned} A_i &= i^2 zy + (i-1)^2 \alpha\beta + i(i-1)(z\beta + y\alpha) \\ &= i^2 \left[ \left( mb + \frac{b^2 - bx}{k} \right) + \left( mb + \frac{b^2 + bx}{k} \right) + 2 \left( mb + \frac{b^2}{k} \right) \right] \\ &\quad + i \left[ -2 \left( mb + \frac{b^2 + bx}{k} \right) - 2 \left( mb + \frac{b^2}{k} \right) \right] + \left( mb + \frac{b^2 + bx}{k} \right) \\ &= i^2 \left[ 4mb + \frac{4b^2}{k} \right] - 2i \left[ 2mb + \frac{2b^2 + bx}{k} \right] + \left( mb + \frac{b^2 + bx}{k} \right) \\ &= 4i^2 \left[ mb + \frac{b^2}{k} \right] - 2i \left[ 2mb + \frac{2b^2 + bx}{k} \right] + \left( mb + \frac{b^2 + bx}{k} \right) \\ &= \left( mb + \frac{b^2}{k} \right) (2i-1)^2 + \frac{bx}{k} (-2i+1). \end{aligned}$$

$$\begin{aligned}
B_i &= (i-1)^2 zy + i^2 \alpha \beta + i(i-1)(z\beta + y\alpha) \\
&= i^2 \left[ \left( mb + \frac{b^2 - bx}{k} \right) + \left( mb + \frac{b^2 + bx}{k} \right) + 2 \left( mb + \frac{b^2}{k} \right) \right] \\
&\quad + i \left[ -2 \left( mb + \frac{b^2 - bx}{k} \right) - 2 \left( mb + \frac{b^2}{k} \right) \right] + \left( mb + \frac{b^2 - bx}{k} \right) \\
&= i^2 \left[ 4mb + \frac{4b^2}{k} \right] - 2i \left[ 2mb + \frac{2b^2 - bx}{k} \right] + \left( mb + \frac{b^2 - bx}{k} \right) \\
&= 4i^2 \left[ mb + \frac{b^2}{k} \right] - 2i \left[ 2mb + \frac{2b^2 - bx}{k} \right] + \left( mb + \frac{b^2 - bx}{k} \right) \\
&= \left( mb + \frac{b^2}{k} \right) (2i-1)^2 - \frac{bx}{k} (-2i+1).
\end{aligned}$$

$$\begin{aligned}
C_i &= i^2(zy + \alpha\beta) + i(i-1)z\beta + i(i+1)y\alpha \\
&= i^2 \left[ 2 \left( mb + \frac{b^2}{k} \right) + \left( mb + \frac{b^2 + bx}{k} \right) + \left( mb + \frac{b^2 - bx}{k} \right) \right] \\
&\quad + i \left[ - \left( mb + \frac{b^2 + bx}{k} \right) + \left( mb + \frac{b^2 - bx}{k} \right) \right] \\
&= 4i^2 \left[ mb + \frac{b^2}{k} \right] - 2i \left[ \frac{bx}{k} \right].
\end{aligned}$$

$$\begin{aligned}
D_i &= i^2(zy + \alpha\beta) + i(i+1)z\beta + i(i-1)y\alpha \\
&= i^2 \left[ 2 \left( mb + \frac{b^2}{k} \right) + \left( mb + \frac{b^2 + bx}{k} \right) + \left( mb + \frac{b^2 - bx}{k} \right) \right] \\
&\quad + i \left[ \left( mb + \frac{b^2 + bx}{k} \right) - \left( mb + \frac{b^2 - bx}{k} \right) \right] \\
&= 4i^2 \left[ mb + \frac{b^2}{k} \right] + 2i \left[ \frac{bx}{k} \right].
\end{aligned}$$

Por lo tanto, reemplazando en la expresión (3.9) de la Proposición 83 obtenemos

$$\begin{aligned}
& e^{-2A_i} + e^{-2B_i} - e^{-2C_i} - e^{-2D_i} \\
&= \exp\left(-2\left(mb + \frac{b^2}{k}\right)(2i-1)^2\right) \left[ \exp\left(-2\left(\frac{bx}{k}\right)(-2i+1)^2\right) \right. \\
&\quad \left. + \exp\left(2\left(\frac{bx}{k}\right)(-2i+1)^2\right) \right] \\
&\quad - \exp\left(-8i^2\left(mb + \frac{b^2}{k}\right)\right) \left[ \exp\left(4i\left(\frac{bx}{k}\right)\right) \right. \\
&\quad \left. + \exp\left(-4i\left(\frac{bx}{k}\right)\right) \right] \\
&= \exp\left(-2(2i-1)^2\left(mb + \frac{b^2}{k}\right)\right) 2 \cosh\left(-2(-2i+1)^2\left(\frac{bx}{k}\right)\right) \\
&\quad - \exp\left(-8i^2\left(mb + \frac{b^2}{k}\right)\right) 2 \cosh\left(-4i\left(\frac{bx}{k}\right)\right) \\
&= 2 \left[ \exp\left(-2(2i-1)^2\left(mb + \frac{b^2}{k}\right)\right) \cosh\left(-2(-2i+1)^2\left(\frac{bx}{k}\right)\right) \right. \\
&\quad \left. - \exp\left(-8i^2\left(mb + \frac{b^2}{k}\right)\right) \cosh\left(-4i\left(\frac{bx}{k}\right)\right) \right].
\end{aligned}$$

Por tanto, usando que

$$\begin{aligned}
& P\left(\bigcap_{t \geq 0} \left\{ -\left(bt + m + \frac{b+x}{k}\right) \leq W_t \leq bt + m + \frac{b-x}{k} \right\}\right) \\
&= P\left(-\left(bt + m + \frac{b+x}{k}\right) \leq W_t \leq bt + m + \frac{b-x}{k}, \forall t \geq 0\right)
\end{aligned}$$

se verifica que, por la Proposición 83, si  $\alpha = b \geq 0$ ,  $\beta = m + \frac{b+x}{k} > 0$ ,  $x = b \geq 0$  y  $y = m + \frac{b-x}{k} > 0$

$$\begin{aligned}
& P\left(\bigcup_{0 \leq t \leq k} \{|W_t| \geq mt + b\} \mid W_k = x\right) \\
&= 1 - P\left(-\left(bt + m + \frac{b+x}{k}\right) \leq W_t \leq bt + m + \frac{b-x}{k}, \forall t \geq 0\right) \\
&= 2 \sum_{i=1}^{\infty} \left[ \exp\left(-2(2i-1)^2\left(mb + \frac{b^2}{k}\right)\right) \cosh\left(-2(-2i+1)^2\left(\frac{bx}{k}\right)\right) \right. \\
&\quad \left. - \exp\left(-8i^2\left(mb + \frac{b^2}{k}\right)\right) \cosh\left(-4i\left(\frac{bx}{k}\right)\right) \right]
\end{aligned}$$

y 1, en otro caso. □

Del resultado anterior se desprende el siguiente corolario que será de gran utilidad para la construcción del estadístico de prueba de Smirnov-Kolmogorov.

**Corolario 86** Sea  $W$  un movimiento Browniano y  $b \in \mathbb{R}$ . Si  $b > 0$  entonces

$$P\left(\max_{0 \leq t \leq 1} |B_t^0| \geq b\right) = 2 \sum_{i=1}^{\infty} (-1)^{i+1} e^{-2i^2 b^2}$$

y 1, en otro caso.

DEMOSTRACIÓN.

Por el Teorema 78

$$P\left(\max_{0 \leq t \leq 1} |B_t^0| \geq b\right) = P\left(\max_{0 \leq t \leq 1} \{|W_t| \geq b\} \mid W_1 = 0\right) = P\left(\bigcup_{0 \leq t \leq 1} \{|W_t| \geq b\} \mid W_1 = 0\right).$$

Luego, tomando  $k = 1$ ,  $m = 0$  y  $x = 0$  en el Teorema 85

$$\begin{aligned} & P\left(\bigcup_{0 \leq t \leq 1} \{|W_t| \geq b\} \mid W_1 = 0\right) \\ &= 2 \sum_{i=1}^{\infty} \left[ \exp\left(-2(2i-1)^2 \left(mb + \frac{b^2}{k}\right)\right) \cosh\left(-2(-2i+1)^2 \left(\frac{bx}{k}\right)\right) \right. \\ &\quad \left. - \exp\left(-8i^2 \left(mb + \frac{b^2}{k}\right)\right) \cosh\left(-4i \left(\frac{bx}{k}\right)\right) \right] \\ &= 2 \sum_{i=1}^{\infty} \left[ \exp(-2(2i-1)^2 b^2) - \exp(-8i^2 b^2) \right]. \end{aligned}$$

Luego, por la Nota 84 se concluye que si  $b > 0$

$$\sum_{i=1}^{\infty} \left[ \exp(-2(2i-1)^2 b^2) - \exp(-8i^2 b^2) \right] = \sum_{i=1}^{\infty} (-1)^{i+1} e^{-2i^2 b^2}.$$

En consecuencia, para  $b > 0$

$$P\left(\bigcup_{0 \leq t \leq 1} \{|W_t| \geq b\} \mid W_1 = 0\right) = 2 \sum_{i=1}^{\infty} (-1)^{i+1} e^{-2i^2 b^2}$$

y 1, en otro caso. □

# Estadístico Smirnov-Kolmogorov

Este capítulo es resultado de todo lo discutido previamente, es por esto que da nombre a este trabajo de tesis. Recordemos que en el Capítulo 2 se definió la f.d.e.  $F_n$  (ver Definición 48), la cual resulta ser un estimador natural de la f.d.  $F$  asociada a una muestra aleatoria dada. Esto es de suma utilidad en contextos donde desconocemos a  $F$ .

La ley fuerte de los grandes números (ver Teorema 28) nos permite asegurar que  $F_n \xrightarrow{c.s.} F$ . Además, el teorema del límite central (ver Teorema 29) implica que

$$\sqrt{n}(F_n(x) - F(x)) \xrightarrow{D} N(0, F(x)(1 - F(x))).$$

Por otro lado, el teorema de Glivenko-Cantelli (ver Sección 2.1) extiende el alcance de la ley fuerte de los grandes números, proporcionándonos convergencia uniforme. Finalmente, a la distancia  $\sqrt{n}D_n$  (ver Teorema 61) se le conoce como el estadístico Smirnov-Kolmogorov.

Es importante notar que la f.d.e.  $F_n$  es una función aleatoria. De esta manera, la extensión del teorema del límite central a un límite central “funcional” recurre al uso de temas más sofisticados. Por tal motivo, introduciremos el siguiente resultado sin demostración, cuyo análisis y estudio queda fuera de los alcances de este trabajo. Las personas interesadas pueden consultar [15], [18], [19].

**Teorema 87 (T. Donsker)** Sea  $\{X_n\}_{n \in \mathbb{N}}$  una sucesión de v.a. i.i.d., con distribución común  $F$ , tales que  $E(X_n) = 0$  y  $\text{Var}(X_n) = 1$ . Además, sea  $W = \{W_t \mid 0 \leq t \leq 1\}$  un movimiento Browniano. Si  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ , entonces

$$\sqrt{n} \bar{X}_{\lfloor nt \rfloor} \xrightarrow{D} W_t,$$

donde  $\lfloor \cdot \rfloor$  denota la función menor entero.

Ahora, estamos listos para estudiar la distribución del estadístico Smirnov-Kolmogorov. Para poder utilizar el Teorema 87 comenzaremos por considerar la siguiente f.d.e.

$$G_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{Y_i \leq x\}}, \quad 0 \leq x \leq 1, \quad (4.1)$$

donde  $Y_1, \dots, Y_n$  son v.a. i.i.d. tales que  $Y_i \sim Unif(0, 1)$ , para  $i = 1, \dots, n$ .

Si consideramos  $U_1^n < \dots < U_n^n$  los  $n$  estadísticos de orden asociados a las v.a.  $Y_1, \dots, Y_n$ , se satisface que

$$G_n(y) - y = \frac{1}{n} \sum_{m=1}^n \mathbf{1}_{\{U_m^n \leq y\}} - y.$$

Ahora supongamos que  $U_1^n < \dots < U_{m-1}^n < U_m^n \leq y$ , para algún  $y \in (0, 1)$ , entonces

$$G_n(y) - y = G_n(U_m^n) - y = \frac{m}{n} - y \leq \frac{m}{n} - U_m^n,$$

de esta manera, tomando los supremos correspondientes se sigue que

$$\sup_{0 \leq y \leq 1} G_n(y) - y \leq \sup_{m=1, \dots, n} \frac{m}{n} - U_m^n. \quad (4.2)$$

Por otro lado, como  $\left\{ \frac{m}{n} - U_m^n \mid m = 1, \dots, n \right\} \subseteq \{G_n(y) - y \mid 0 \leq y \leq 1\}$ , también se satisface que

$$\sup_{1 \leq m \leq n} \frac{m}{n} - U_m^n \leq \sup_{0 \leq y \leq 1} G_n(y) - y. \quad (4.3)$$

En consecuencia, las ecuaciones (4.2) y (4.3) implican que

$$\sup_{0 \leq y \leq 1} G_n(y) - y = \sup_{1 \leq m \leq n} \frac{m}{n} - U_m^n. \quad (4.4)$$

El siguiente resultado, cuya demostración puede ser consultada en [13, Capítulo 5], nos facilitará el trabajo algebraico.

**Lema 88** Sean  $\{V_n\}_{n \in \mathbb{N}}$  una sucesión de v.a. i.i.d. tales que  $V_i \sim exp(1)$ . Definamos

$$Z_n = \sum_{i=1}^n V_i. \quad (4.5)$$

Entonces

$$U_m^n \stackrel{D}{=} \frac{Z_m}{Z_{n+1}},$$

para toda  $m = 1, \dots, n$ .

Los siguientes teoremas representan los resultados más importantes de este capítulo. El Teorema 89 nos permite aproximar el estadístico Smirnov-Kolmogorov a través del puente Browniano.

**Teorema 89** Sea  $W$  un movimiento Browniano. Entonces  $D_n$  definido en la ecuación (2.2) satisface que

$$\sqrt{n}D_n \xrightarrow{D} \max_{0 \leq t \leq 1} |B_t^0|,$$

donde  $B_t^0$  es el puente Browniano definido en (3.3).

DEMOSTRACIÓN. Sea  $F$  una f.d. continua y  $F^{-1}$  su transformación inversa (ver Definición 54). Dada  $x \in \mathbb{R}$  existe  $y \in (0, 1)$  tal que  $x = F^{-1}(y)$ , entonces el inciso (a) del Teorema 57 implica que

$$F(x) = F(F^{-1}(y)) \geq y. \quad (4.6)$$

Por otro lado, el inciso (b) del Teorema 54 sabemos que

$$F(x^-) = F(F^{-1}(y)^-) \leq y, \quad (4.7)$$

donde  $F(F^{-1}(y)^-)$  denota el límite de  $F$  por la izquierda de  $F^{-1}(y)$ .

Luego, como  $F$  es continua entonces  $F(x^-) = F(x)$ . En consecuencia, de las ecuaciones (4.6) y (4.7) se obtiene que

$$F(x) = F(F^{-1}(y)) = y.$$

Así, podemos concluir lo siguiente

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = \sup_{0 \leq y \leq 1} |G_n(y) - y|.$$

Además, esto nos permite observar que si  $F$  es continua,  $\sqrt{n}D_n$  no depende de la distribución  $F$  (ver ecuaciones (2.2) y (2.2') de [16]).

Lo anterior implica que  $\sqrt{n}D_n = \sqrt{n} \sup_{0 \leq y \leq 1} |G_n(y) - y|$ ; así, por la ecuación (4.4) y el Lema 88 se tiene la siguiente igualdad

$$\begin{aligned} \sqrt{n}D_n &\stackrel{D}{=} \sqrt{n} \sup_{1 \leq m \leq n} \left| \frac{m}{n} - U_m^n \right| \\ &\stackrel{D}{=} \sqrt{n} \sup_{1 \leq m \leq n} \left| \frac{m}{n} - \frac{Z_m}{Z_{n+1}} \right|, \end{aligned}$$

donde  $Z_m$  y  $Z_{n+1}$  satisfacen la ecuación (4.5).

Reescribiendo

$$\begin{aligned} \sqrt{n}D_n &\stackrel{D}{=} \frac{n}{Z_{n+1}} \sup_{1 \leq m \leq n} \left| \left( \frac{mZ_{n+1}}{n} - Z_m \right) \frac{1}{\sqrt{n}} \right| \\ &\stackrel{D}{=} \frac{n}{Z_{n+1}} \sup_{1 \leq m \leq n} \left| \left( \frac{Z_m}{\sqrt{n}} - \frac{mZ_{n+1}}{n^{3/2}} \right) \right|. \end{aligned}$$

Al trabajar en un conjunto finito, el máximo se alcanza y ya que éste existe, coincide con el supremo. De esta manera

$$\begin{aligned}\sqrt{n}D_n &\stackrel{D}{=} \frac{n}{Z_{n+1}} \max_{1 \leq m \leq n} \left| \left( \frac{Z_m}{\sqrt{n}} - \frac{mZ_{n+1}}{n^{3/2}} \right) \right| \\ &\stackrel{D}{=} \frac{n}{Z_{n+1}} \max_{1 \leq m \leq n} \left| \left( \frac{Z_m - m}{\sqrt{n}} - \frac{m(Z_{n+1} - n)}{n^{3/2}} \right) \right|.\end{aligned}$$

Ahora, definamos

$$B_n(t) = \begin{cases} \frac{Z_m - m}{\sqrt{n}}, & \text{si } t = m/n \text{ con } m \in \{0, 1, \dots, n\}, \\ \text{lineal,} & \text{sobre } \left[ \frac{m-1}{n}, \frac{m}{n} \right]. \end{cases} \quad (4.8)$$

Si  $t = \frac{m}{n}$ , para algún  $m \in \{0, 1, \dots, n\}$ , entonces

$$\begin{aligned}\frac{Z_m - m}{\sqrt{n}} - \frac{m(Z_{n+1} - n)}{n^{3/2}} &= B_n(t) - t \left( \frac{Z_{n+1} - n}{\sqrt{n}} \right) \\ &= B_n(t) - t \left( \frac{Z_{n+1} - n}{\sqrt{n}} \right) + tB_n(1) - tB_n(1).\end{aligned}$$

Como  $B_n(1) = \frac{Z_n - n}{\sqrt{n}}$ , la ecuación anterior se reescribe de la siguiente manera

$$\begin{aligned}B_n(t) - t \left( \frac{Z_{n+1} - n}{\sqrt{n}} \right) + tB_n(1) - tB_n(1) &= B_n(t) - t \left( \frac{Z_{n+1} - n}{\sqrt{n}} + B_n(1) - \frac{Z_n - n}{\sqrt{n}} \right) \\ &= B_n(t) - t \left( \frac{Z_{n+1} - Z_n}{\sqrt{n}} + B_n(1) \right).\end{aligned} \quad (4.9)$$

Notemos que por la ley fuerte de los grandes números (ver Teorema 28)

$$\frac{Z_n}{n} = \frac{1}{n} \sum_{i=1}^n V_i \xrightarrow{\text{c.s.}} 1 \quad \text{y} \quad \frac{Z_{n+1}}{n} = \frac{1}{n} \sum_{i=1}^{n+1} V_i \xrightarrow{\text{c.s.}} 1,$$

pues  $E(V_i) = 1$ , ya que  $V_i \sim \exp(1)$ .

Lo anterior implica que

$$\frac{Z_{n+1} - Z_n}{\sqrt{n}} \xrightarrow{\text{c.s.}} 0 \quad (4.10)$$

y por lo tanto, en el Teorema de Donsker al tomar como  $X_m = V_m - 1$ , se cumple que  $E(X_m) = 0$  y  $\text{Var}(X_m) = 1$ , pues  $V_m \sim \exp(1)$ ; así

$$B_n(\cdot) \xrightarrow{D} W(\cdot),$$

donde  $W$  es un movimiento Browniano.

Como consecuencia de aplicar la ecuación (4.10) en la ecuación (4.9) se obtiene que

$$\sqrt{n}D_n \xrightarrow{D} \max_{1 \leq m \leq n} |B_n(t) - tB_n(1)|,$$

donde, por el Teorema 75, podemos reconocer al lado derecho de la ecuación anterior como el puente Browniano  $B^0$ . En consecuencia,

$$\sqrt{n}D_n \xrightarrow{D} \max_{0 \leq t \leq 1} |B_t^0|.$$

□

En contextos de aplicación, trabajar con  $D_n$  resulta computacionalmente muy costoso (ver [17] y [41]). Así, presentamos el resultado que se utiliza más frecuentemente en la práctica; para tener un trabajo más completo hemos incluido un ejemplo de un problema de bondad de ajuste el cual abordamos empleando la prueba Smirnov-Kolmogorov en la Sección 4.1.

**Teorema 90 (Distribución límite del estadístico Smirnov-Kolmogorov)** Consideremos  $B^0$  un puente Browniano y sea  $\alpha \in \mathbb{R}$ . Si  $\alpha > 0$ , entonces

$$P \left( \max_{0 \leq t \leq 1} |B_t^0| \leq \alpha \right) = 1 - 2 \sum_{i=1}^{\infty} (-1)^{i+1} e^{-2i^2\alpha^2} \quad (4.11)$$

y 0, en otro caso.

DEMOSTRACIÓN. Por aplicar el Teorema 78, se tiene que

$$\max_{0 \leq t \leq 1} |B_t^0| \stackrel{D}{=} \max_{0 \leq t \leq 1} |(W_t | W_1 = 0)|.$$

Luego, por el Corolario 86 obtenemos que para  $\alpha > 0$

$$P \left( \max_{0 \leq t \leq 1} |B_t^0| \leq \alpha \right) = 1 - 2 \sum_{i=1}^{\infty} (-1)^{i+1} e^{-2i^2\alpha^2}$$

y 0 en otro caso. □

Para aplicar estos resultados en la práctica es necesario una muestra de tamaño  $n$  *suficientemente* grande, a saber  $n \geq 80$  de acuerdo con [44]. Sin embargo, en ocasiones es necesario conocer la distribución de  $D_n$  para una  $n$  dada, este trabajo fue originalmente desarrollado por [2] de donde se desprende el siguiente resultado.

**Proposición 91 (Distribución para tamaño de muestra finito)** Si  $n \in \mathbb{N}$  y  $D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|$  es el estadístico Smirnov-Kolmogorov entonces

$$P\left(D_n \leq \frac{1}{2n} + x\right) = \begin{cases} 0, & \text{si } x \leq 0, \\ \int_{\frac{1}{2n}-x}^{\frac{1}{2n}+x} \int_{\frac{1}{3n}-x}^{\frac{1}{3n}+x} \cdots \int_{\frac{2n-1}{2n}-x}^{\frac{2n-1}{2n}+x} f(u_1, u_2, \dots, u_n) du_n \dots du_1, & \text{si } 0 < x \leq \frac{2n-1}{2n}, \\ 1, & \text{si } x \geq \frac{2n-1}{2n}, \end{cases}$$

$$\text{donde } f(u_1, \dots, u_n) = \begin{cases} n!, & \text{si } 0 < u_1 < \dots < u_n \\ 1, & \text{en otro caso.} \end{cases}$$

Para profundizar en el tema recomendamos consultar [2] y [22]; además, para perspectiva computacional de este resultado recomendamos consultar [17] y [41].

## 4.1. Ejemplo de aplicación de la prueba Smirnov-Kolmogorov

La prueba Smirnov-Kolmogorov es la prueba de bondad de ajuste más ampliamente utilizada en contextos de aplicación. A continuación veremos un ejemplo, ejecutado en el software de cómputo científico R [43], en el cual se ilustra la utilidad de la prueba.

En el archivo `housing.txt`, obtenido del repositorio de datos del libro de Zelterman [61], se reporta el costo de renta promedio para departamentos y casas en cada uno de los 50 estados y el Distrito de Columbia de Estados Unidos. De esta manera, tenemos un par de v.a.  $X_d$  y  $X_c$  que representan el costo de renta promedio para departamentos y casas, respectivamente; además, tenemos una muestras aleatorias de tamaño  $n = 51$  para cada variable.

Importemos la base de datos, es decir, las realizaciones de las muestras aleatorias.

```
housing <- read.table("housing.txt", header = TRUE, row.names = 1)
```

Visualicemos los primeros diez datos, para verificar que se ha importado correctamente vea la Tabla 4.1.

```
head(housing, 10)
```

```

##      Apartment House
## AK          949 237.8
## AL          631 121.5
## AR          606 105.7
## AZ          866 229.2
## CA         1135 467.0
## CO          848 242.2
## CT          970 306.0
## DC         1011 474.1
## DE          917 250.9
## FL          947 218.7

```

Figura 4.1: Output de la instrucción `head(housing,10)`.

Hemos anexado la Tabla 4.1, en donde pueden verse datos contenidos en el archivo `housing.txt`.

	Apartment	House		Apartment	House		Apartment	House
AK	949	237.8	KY	578	118.4	NY	953	318.9
AL	631	121.5	LA	698	132.4	OH	667	140.2
AR	606	105.7	MA	991	353.6	OK	614	105.5
AZ	866	229.2	MD	1074	341.2	OR	780	273.3
CA	1135	467	ME	702	180.2	PA	726	164.7
CO	848	242.2	MI	706	151.3	RI	850	286
CT	970	306	MN	734	213.8	SC	675	138.7
DC	1011	474.1	MO	657	141.5	SD	569	126.2
DE	917	250.9	MS	638	99.7	TN	660	138.6
FL	947	218.7	MT	631	180.3	TX	768	126.8
GA	787	169.1	NC	694	154.5	UT	784	236
HI	1298	560.2	ND	534	112.5	VA	934	269.6
IA	607	120.7	NE	626	126.5	VT	797	214.7
ID	690	183.7	NH	914	264.7	WA	874	308.1
IL	811	214.9	NJ	1068	364.1	WI	704	173.3
IN	670	125.2	NM	668	165.1	WV	528	95.9
KS	654	125.7	NV	1011	271.5	WY	636	188.2

Tabla 4.1: Datos del archivo `housing.txt`.

La columna etiquetada por “Apartment” contiene las realizaciones de  $X_d$ ; mientras que, la columna etiquetada por “House” contiene las realizaciones de  $X_c$ . En el libro de

Zel'terman se menciona que al aplicar una estandarización a  $X_d$ , es decir

$$Y_d := \frac{X_d - E(X_d)}{\sqrt{\text{Var}(X_d)}},$$

se satisface que  $Y_d \sim N(0, 1)$ .

Esta afirmación proviene de observar un comportamiento similar a la campana de Gauss en la Figura 4.2.

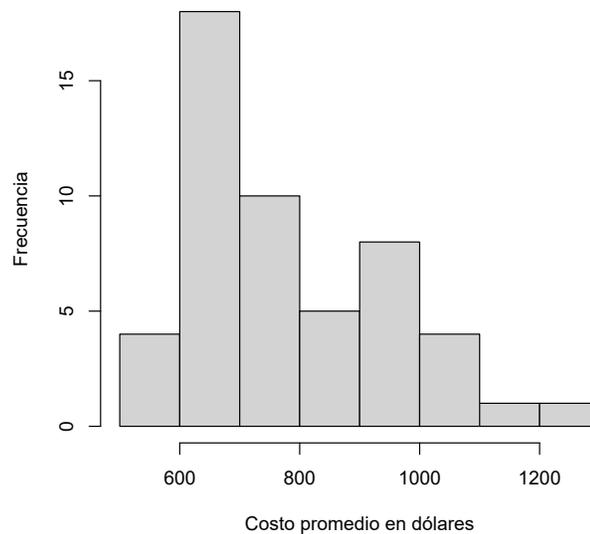


Figura 4.2: Histograma de los costos de renta promedio para departamentos.

La Figura 4.2 fue obtenida empleando las siguientes sentencias de código.

```
hist(housing$Apartment, main = "", ylab = "Frecuencia",
     xlab = "Costo promedio en dólares")
```

Al no conocer los parámetros reales de la f.d. se prefiere la estandarización, por simplicidad para el análisis. Ahora, para verificar la afirmación sobre la distribución recurriremos a la prueba Smirnov-Kolmogorov. Comenzaremos por generar un vector que almacene las realizaciones de  $X_d$ .

```
y<-housing$Apartment
```

Ahora procederemos a aplicar la estandarización.

```
y <-(y - mean(y)) / sd(y)
```

Al observar la Figura 4.3 podemos realizar un análisis visual del comportamiento entre la f.d.e.  $F_{51}$ , construida con las 51 realizaciones de la estandarización de  $Y_d$ , y la f.d.  $F$  de una variable aleatoria normal estándar.

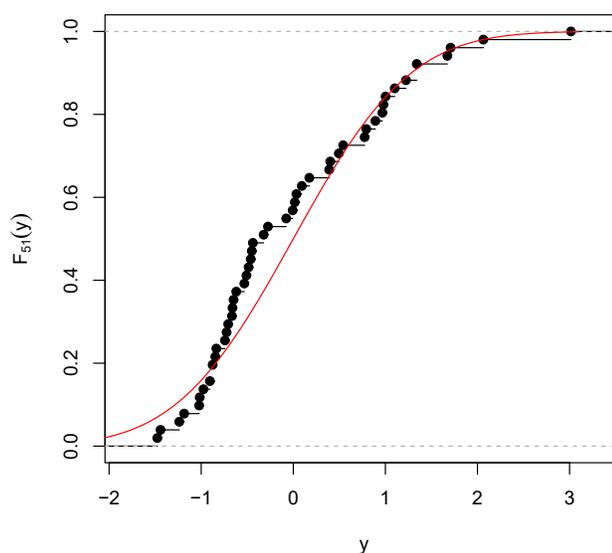


Figura 4.3: Relación entre la f.d.e.  $F_{51}$  y la f.d.  $F$ .

En la figura anterior tenemos en color rojo la gráfica de  $F$  y en color negro la gráfica de  $F_{51}$ . Esta figura fue construida a través de las siguientes sentencias de código.

```
plot(ecdf(y),main=" ",xlab="y", ylab =expression('F' [51 ](y)))
intervalo<-seq(-3.1,3.1,0.1)
lines(intervalo,pnorm(intervalo,0,1), col="red")
```

Notemos que dada la cercanía entre ambas gráficas, podemos inferir que la hipótesis

$$H_0: Y_d \sim N(0, 1) \quad (4.12)$$

es cierta. Llamaremos a la función `ks.test` para llevar a cabo la prueba de bondad de ajuste Smirnov-Kolmogorov y formalizar la aceptación o rechazo de  $H_0$ .

```
ks.test(y,"pnorm",0,1)
```

En la sentencia anterior, el fragmento `"pnorm",0,1` hace referencia a la f.d. de una variable aleatoria normal estándar, por lo que con esta línea de código verificaremos la validez de la hipótesis descrita en (4.12) empleando la prueba Smirnov-Kolmogorov.

La prueba concluye (ver Figura 4.4) que la distancia, con la métrica del supremo, entre la f.d.e. asociada a las 51 realizaciones de la estandarización de  $X_d$  y la f.d. de una variable aleatoria normal estándar es

$$D_{51} = 0.15968.$$

Así, si la experta o el experto que analiza estos datos considera que dicha distancia es *suficientemente pequeña*, aceptaríamos que  $Y_d \sim N(0, 1)$ . Además, para tomar una decisión más objetiva tenemos que el  $p$ -valor es 0.1484; por tanto, si la experta o el experto han fijado un nivel de significancia

$$\alpha < 0.1484,$$

se concluiría que hay evidencia suficiente para aceptar la hipótesis  $H_0$  con un nivel de significancia  $\alpha$ .

```
##
## One-sample Kolmogorov-Smirnov test
##
## data: y
## D = 0.15968, p-value = 0.1484
## alternative hypothesis: two-sided
```

Figura 4.4: Output de la instrucción `ks.test(y, "pnorm", 0, 1)`.

Para hacer un análisis usando la distribución límite emplearemos la función `cont_ks_c_cdf` contenida en la paquetería `KSgeneral` que se desprende del trabajo de Dimitrova, Kaishev y Tan [14]. Para cargar la paquetería escribimos la siguiente sentencia de código.

```
library(KSgeneral)
```

Ahora, la función `cont_ks_c_cdf` nos permitirá calcular la probabilidad  $P(D_{51} \geq x) = \alpha$  para  $\alpha, x \in \mathbb{R}$ . En la práctica, usualmente  $\alpha$  toma el valor de 0.05; así, partiendo de lo anterior, realizaremos la prueba de manera directa: calculando los estadísticos de prueba y aceptando  $H_0$  si

$$D_{51} = 0.15968 \leq q,$$

donde  $q$  es el 0.05 cuantil superior de la v.a.  $D_{51}$ , es decir,  $q$  satisface la ecuación

$$P(D_{51} \geq q) = 0.05.$$

Al implementar la sentencia de código:

```
cont_ks_c_cdf(0.186588, 51)
```

se tiene que  $q = 0.186588$ , pues a  $P(D_{51} \geq 0.186588) \approx 0.05$ .

```
[1] 0.05000191
```

Figura 4.5: Output de la instrucción `cont_ks_c_cdf(0.186588, 51)`.

Recordemos que el estadístico evaluado en las realizaciones es 0.15968; además, se cumple que

$$0.15968 \leq q = 0.186588.$$

En consecuencia, hay suficiente evidencia para aceptar la hipótesis  $H_0 : Y_d \sim N(0, 1)$  con un nivel de significancia  $\alpha = 0.05$ .

Notemos que la razón de elegir a  $q$  como el 0.05 cuantil superior es que la probabilidad  $P(D_{51} \geq q)$  es *pequeña*, lo cual puede observarse en la Figura 4.6.

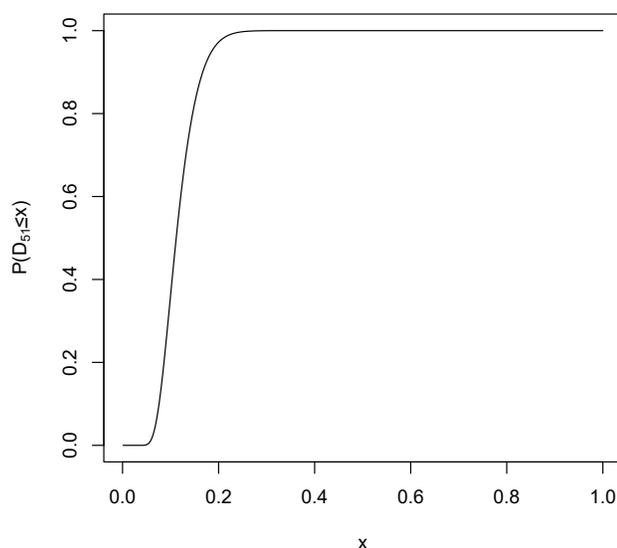


Figura 4.6: Gráfica de la función de distribución de la v.a.  $D_{51}$ .

La Figura 4.6 fue construida usando las siguientes sentencias de código.

```
intervalo<-seq(0.001,1,0.001)
l<-length(intervalo)
x<-rep(0.001,l)
j=0.001
for (i in 0:l){
x[i]<-cont_ks_cdf(j, 51)
j=j+0.001
}
```

```
plot(intervalo,x,type = "l", lty = 1,xlab="x",
ylab=expression('P(D' [51]*'<x)'))
```

La función `ks.test` viene integrada por default en las funciones básicas del software R y, para complementar este ejemplo, se ha decidido incluir el código de esta función.

```
getAnywhere(ks.test)
```

Figura 4.7: Instrucción para obtener el código de la función `ks.test`.

```
A single object matching 'ks.test' was found
It was found in the following places
package:stats
namespace:stats
with value

function (x, y, ..., alternative = c("two.sided", "less",
"greater"), exact = NULL)
{
  alternative <- match.arg(alternative)
  DNAME <- deparse1(substitute(x))
  x <- x[!is.na(x)]
  n <- length(x)
  if (n < 1L)
    stop("not enough 'x' data")
  PVAL <- NULL
  if (is.numeric(y)) {
    DNAME <- paste(DNAME, "and", deparse1(substitute(y)))
    y <- y[!is.na(y)]
    n.x <- as.double(n)
    n.y <- length(y)
    if (n.y < 1L)
      stop("not enough 'y' data")
    if (is.null(exact))
      exact <- (n.x * n.y < 10000)
    METHOD <- "Two-sample Kolmogorov-Smirnov test"
    TIES <- FALSE
    n <- n.x * n.y / (n.x + n.y)
    w <- c(x, y)
    z <- cumsum(ifelse(order(w) <= n.x, 1/n.x, -1/n.y))
```

```
if (length(unique(w)) < (n.x + n.y)) {
  if (exact) {
    warning("cannot compute exact p-value with ties")
    exact <- FALSE
  }
  else warning("p-value will be approximate in the presence of ties")
  z <- z[c(which(diff(sort(w)) != 0), n.x + n.y)]
  TIES <- TRUE
}
STATISTIC <- switch(alternative, two.sided = max(abs(z)),
  greater = max(z), less = -min(z))
nm_alternative <- switch(alternative, two.sided = "two-sided",
  less = "the CDF of x lies below that of y",
  greater = "the CDF of x lies above that of y")
if (exact && (alternative == "two.sided") && !TIES)
  PVAL <- 1 - .Call(C_pSmirnov2x, STATISTIC, n.x, n.y)
}
else {
  if (is.character(y))
    y <- get(y, mode = "function", envir = parent.frame())
  if (!is.function(y))
    stop(print(
      "'y' must be numeric or a function or a string naming a valid function"))
  METHOD <- "One-sample Kolmogorov-Smirnov test"
  TIES <- FALSE
  if (length(unique(x)) < n) {
    warning("ties should not be present for the Kolmogorov-Smirnov test")
    TIES <- TRUE
  }
  if (is.null(exact))
    exact <- (n < 100) && !TIES
  x <- y(sort(x), ...) - (0:(n - 1))/n
  STATISTIC <- switch(alternative, two.sided = max(c(x,
    1/n - x)), greater = max(1/n - x), less = max(x))
  if (exact) {
    PVAL <- 1 - if (alternative == "two.sided")
      .Call(C_pKolmogorov2x, STATISTIC, n)
  }
  else {
    pkolmogorov1x <- function(x, n) {
      if (x <= 0)
```

```

return(0)
if (x >= 1)
return(1)
j <- seq.int(from = 0, to = floor(n * (1 -
x)))
1 - x * sum(exp(lchoose(n, j) + (n - j) * log(1 -
x - j/n) + (j - 1) * log(x + j/n)))
}
pkolmogorov1x(STATISTIC, n)
}
}
nm_alternative <- switch(alternative, two.sided = "two-sided",
less = "the CDF of x lies below the null hypothesis",
greater = "the CDF of x lies above the null hypothesis")
}
names(STATISTIC) <- switch(alternative, two.sided = "D",
greater = "D^+", less = "D^-")
if (is.null(PVAL)) {
pkstwo <- function(x, tol = 1e-06) {
if (is.numeric(x))
x <- as.double(x)
else stop("argument 'x' must be numeric")
p <- rep(0, length(x))
p[is.na(x)] <- NA
IND <- which(!is.na(x) & (x > 0))
if (length(IND))
p[IND] <- .Call(C_pKS2, p = x[IND], tol)
p
}
PVAL <- if (alternative == "two.sided")
1 - pkstwo(sqrt(n) * STATISTIC)
else exp(-2 * n * STATISTIC^2)
}
PVAL <- min(1, max(0, PVAL))
RVAL <- list(statistic = STATISTIC, p.value = PVAL,
alternative = nm_alternative,
method = METHOD, data.name = DNAME)
class(RVAL) <- "htest"
return(RVAL)
}

```

<bytecode: 0x00000000049631f0>

<environment: namespace:stats>



# Conclusiones

Históricamente, la distribución límite del estadístico de prueba Smirnov-Kolmogorov

$$\sqrt{n}D_n = \sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|,$$

con  $F_n$  la f.d.e., fue hallada a partir de considerar un problema de probabilidad de cruce para el movimiento Browniano (ver [16]). En este trabajo se presentó una alternativa para hallar la distribución límite de tal estadístico de prueba que resulta más natural, pues al considerar un movimiento Browniano en  $[0, k]$ ,  $W = \{W_t \mid 0 \leq t \leq k\}$ , y  $b, m, n, x \in \mathbb{R}$ , el problema de probabilidad de cruce

$$P \left( \bigcup_{0 \leq t \leq k} \{|W_t| \geq mt + b\} \mid W_k = x \right),$$

puede reescribirse en términos del puente Browniano  $B^0$ , al tomar  $k = 1$ ,  $m = 0$  y  $x = 0$ , de la manera siguiente

$$P \left( \max_{0 \leq t \leq 1} |B_t^0| \geq b \right),$$

usando las propiedades del puente Browniano.

Gracias al Teorema de Donsker sabemos que

$$\sqrt{n}D_n \xrightarrow{D} \max_{0 \leq t \leq 1} |B_t^0|,$$

de esta manera, en este trabajo hemos hallado la distribución límite del estadístico de prueba Smirnov-Kolmogorov tomando como único objeto de estudio al puente Browniano y la probabilidad de que este proceso estocástico cruce la función constante  $b$ .

En el contexto de estadística, a la constante  $b$  se denomina nivel de significancia, y usualmente se denota por  $\alpha$ . En una prueba de hipótesis

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_1$$

el nivel de significancia se define como la probabilidad de cometer el error del tipo I, es decir rechazar la hipótesis  $H_0$  cuando en realidad era correcta.

En consecuencia, a través de los resultados de este trabajo hemos proporcionado una alternativa para el cálculo del nivel significancia en la prueba Smirnov-Kolmogorov cuando se tiene una observación de una muestra aleatoria de tamaño  $n$ , con  $n$  suficientemente grande. Esto resulta importante pues la prueba Smirnov-Kolmogorov es una de las pruebas de bondad de ajuste más usadas en la práctica.

# Bibliografía

- [1] Aslam, M., “Introducing Kolmogorov–Smirnov Tests under Uncertainty: An Application to Radioactive Data”, *ACS Omega*, vol. 5, n.º 1, págs. 914-917, 2020.
- [2] Birnbaum, Z. W., “Numerical tabulation of the distribution of Kolmogorov’s statistic for finite sample size”, *J. Amer. Statist. Assoc.*, vol. 47, págs. 425-441, 1952.
- [3] Borel, É., “Les probabilités dénombrables et leurs applications arithmétiques”, *Rendiconti del Circolo Matematico di Palermo*, vol. 27, n.º 1, págs. 247-271, 1909.
- [4] Cantelli, F., “Cantelli. Sulla probabilità come limite della frequenza”, *Rom. Acc. L. Rend.*, vol. 26, n.º 1, págs. 39-45, 1917.
- [5] Cantelli, F., “Sulla determinazione empirica delle leggi di probabilità”, *Gion. Ist. Ital. Attauri.*, vol. 4, págs. 421-424, 1933.
- [6] Casella, G. y Berger, R. L., *Statistical Inference*. California, E.U.A.: Duxbury, 2002.
- [7] Chebyshev, P. L., “Des valeurs moyennes”, *Journal de Mathématiques Pures et Appliquées*, vol. 12, n.º 2, págs. 177-184, 1867.
- [8] Chen, G. y N., B., “A General Purpose Approximate Goodness-of-Fit Test”, *Journal of Quality Technology*, vol. 27, n.º 2, págs. 154-161, 1995.
- [9] Chui, C. K., *An Introduction to Wavelets*, ép. Wavelet Analysis and Its Applications 1. Elsevier Academic Press, 1992.
- [10] Chung, K. L., *A course in probability theory*, Third. Academic Press, Inc., San Diego, CA, 2001.
- [11] D’Agostino, R. B. y Stephens, M. A., *Goodness-of-Fit Techniques*. Nueva York, E.U.A.: Marcel Dekker, Inc., 1986.
- [12] Darling, D. A., “The Kolmogorov-Smirnov, Cramér-von Mises Tests”, *The Annals of Mathematical Statistics*, vol. 28, n.º 4, págs. 823-838, 1957.
- [13] Devroye, L., *Non-Uniform Random Variate Generation*, 1st. Springer-Verlag, New York, 1986.

- [14] Dimitrova, D. S., Kaishev, V. K. y Tan, S., “Computing the Kolmogorov-Smirnov Distribution When the Underlying CDF is Purely Discrete, Mixed, or Continuous”, *Journal of Statistical Software, Articles*, vol. 95, n.º 10, págs. 1-42, 2020.
- [15] Donsker, M. D., “An invariance principle for certain probability limit theorems”, *Memoirs of the American Mathematical Society*, vol. 6, pág. 12, 1951.
- [16] Doob, J. L., “Heuristic approach to the Kolmogorov-Smirnov theorems”, *Ann. Math. Statistics*, vol. 20, págs. 393-403, 1949.
- [17] Drew, J. H., Glen, A. G. y Leemis, L. M., “Computing the cumulative distribution function of the Kolmogorov–Smirnov statistic”, *Computational Statistics & Data Analysis*, vol. 34, n.º 1, págs. 1-15, 2000.
- [18] Dudley, R. M., *Uniform Central Limit Theorems*, 2nd, ép. Studies in Advanced Mathematics. Cambridge University Press, 2014.
- [19] Durrett, R., *Probability: Theory and Examples*. Cambridge, Reino Unido: Cambridge University Press, 2019.
- [20] Faber, G., “Über die Orthogonalfunktionen des Herrn Haar”, *Jahresbericht der Deutschen Mathematiker-Vereinigung*, vol. 19, págs. 104-112, 1910.
- [21] Feller, W., “On the Kolmogorov-Smirnov limit theorems for empirical distributions”, *Annals of Mathematical Statistics*, vol. 19, págs. 177-189, 1948.
- [22] Gibbons, J. D. y Chakraborti, S., *Nonparametric Statistical Inference*, 4th ed., rev. and expanded, ép. Statistics, textbooks and monographs 168. Marcel Dekker, 2003.
- [23] Gilchrist, W., *Statistical Modelling with Quantile Functions*. Florida, E.U.A.: Chapman & Hall/CRC, 2000.
- [24] Gilchrist, W., “Galton misrepresented”, *Significance*, vol. 2, n.º 3, págs. 136-137, 2005.
- [25] Glivenko, V., “Sulla determinazione empirica delle leggi di probabilità”, *Gion. Ist. Ital. Attauri.*, vol. 4, págs. 92-99, 1933.
- [26] Haar, A., “Zur Theorie der orthogonalen Funktionensysteme”, *Math. Ann.*, vol. 69, n.º 3, págs. 331-371, 1910.
- [27] Haar, A., “Zur Theorie der orthogonalen Funktionensysteme”, *Math. Ann.*, vol. 71, n.º 1, págs. 38-53, 1911.
- [28] Hirsch, F., Lacombe, G. y Levy, S., *Elements of Functional Analysis*, 1.<sup>a</sup> ed., ép. Graduate Texts in Mathematics. Springer, 1999.

- 
- [29] Hoel, P. G., *Introduction to Mathematical Statistics*. Nueva York, E.U.A.: John Wiley & Sons, Inc., 1966.
- [30] Hogg, R. V., McKean, J. W. y Craig, A. T., *Introduction to Mathematical Statistics*. Boston, E.U.A.: Pearson Education, Inc., 2019.
- [31] Hollander, M. y Wolfe, D. A., *Nonparametric Statistics*. Nueva York, E.U.A.: John Wiley & Sons, Inc., 1973.
- [32] Ibe, O. C., *Markov Processes for Stochastic Modeling*, Second Edition. Oxford, U.K.: Elsevier, 2013.
- [33] Jacod, J. y Protter, P., *Probability Essentials*, Softcover reprint of the original 1st, ép. Universitext. Springer Berlin Heidelberg, 2000.
- [34] Janson, S., *Gaussian Hilbert spaces*, ép. Cambridge Tracts in Mathematics. Cambridge University Press, 1997.
- [35] Klafter, J., Shlesinger, M. y Zumofen, G., “Beyond Brownian motion”, *Physics today*, vol. 49, n.º 2, págs. 33-39, 1996.
- [36] Knuth, D. E., *The Art of Computer Programming*, 2nd. Addison-Wesley Publishing, 1981.
- [37] Kolmogorov, A. N., “Sulla Determinazione Empirica di una Legge di Distribuzione”, *Giornale dell’Istituto Italiano degli Attuari*, vol. 4, págs. 83-91, 1933.
- [38] Kolmogorov, A. N., *Foundations of the Theory of Probability: Second English Edition*. Courier Dover Publications, 2018.
- [39] Kuiper, N. H., “Tests Concerning Random Points on a Circle”, *Indagationes Mathematicae (Proceedings)*, vol. 63, n.º 1, págs. 38-47, 1960.
- [40] Lilliefors, H. W., “On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown”, *Journal of the American Statistical Association*, vol. 62, n.º 318, págs. 399-402, 1967.
- [41] Marsaglia, G., Tsang, W. W. y Wang, J., “Evaluating Kolmogorov’s Distribution”, *Journal of Statistical Software, Articles*, vol. 8, n.º 18, págs. 1-4, 2003.
- [42] Peralta, L., “Distribution Function of the Blow up Time of the Solution of an Anticipating Random Fatigue Equation”, *Acta Math. Sin. (Engl. Ser.)*, vol. 37, n.º 4, págs. 551-564, 2021.
- [43] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2020.

- 
- [44] Resnick, S., *Adventures on Stochastic Processes*. Nueva York, E.U.A.: Birkhäuser Boston, 1992.
- [45] Resnick, S., *A Probability Path*. Nueva York, E.U.A.: Birkhäuser Boston, 1999.
- [46] Rincón, L., *Estadística Descriptiva*. Ciudad de México, México: Prensas de Ciencias, 2017.
- [47] Royden, H. L., *Real analysis*, Third. Macmillan Publishing Company, New York, 1988.
- [48] Rudin, W., *Real and complex analysis*, Third. McGraw-Hill Book Co., New York, 1987.
- [49] Schauder, J., “Eine Eigenschaft des Haarschen Orthogonalsystems”, *Math. Z.*, vol. 28, n.º 1, págs. 317-320, 1928.
- [50] Scheike, T. H., “A boundary-crossing result for Brownian motion”, *J. Appl. Probab.*, vol. 29, n.º 2, págs. 448-453, 1992.
- [51] Schilling, R., Partzsch, L. y Bottcher, B., *Brownian Motion: An Introduction to Stochastic Process*. Berlin, Alemania: Walter de Gruyter & Co., 2012.
- [52] Shorak, G. R. y Wellner, J. A., *Empirical Processes with Application to Statistics*. Pensilvania, E.U.A.: Society for Industrial y Applied Mathematics, 2009.
- [53] Smirnov, N. V., “Estimate of Deviation between Empirical Distribution Functions in Two Independent Samples”, *Bulletin Moscow University*, vol. 2, n.º 2, págs. 3-16, 1939.
- [54] Spivak, M., *Calculus*, Fourth. Publish or Perish, 2008.
- [55] Steele, J. M., *Stochastic Calculus and Financial Applications*. Nueva York, E.U.A.: Springer-Verlag, 2001.
- [56] Taylor, J. C., *An Introduction to Measure and Probability*. Springer-Verlag, New York, 1997.
- [57] Tucker, H. G., *A Graduate Course in Probability*, ép. Probability and Mathematical Statistics, Vol. 2. Academic Press, Inc., New York-London, 1967.
- [58] Vaart, A. W. y Wellner, J. A., *Weak Convergence and Empirical Processes*, ép. Springer Series in Statistics. Springer-Verlag, New York, 1996.
- [59] Velázquez, R., “Probabilidades de cruce para el movimiento Browniano y puente Browniano”, Tesis de Licenciatura, Universidad Autónoma del Estado de Hidalgo, 2020.

- 
- [60] Wasserman, L., *All of Nonparametric Statistics*. Nueva York, E.U.A.: Springer, 2006.
- [61] Zelterman, D., *Applied multivariate statistics with R*, ép. Statistics for Biology and Health. Springer, Cham, 2015.
- [62] Zhao, D., Bu, L., Alippi, C. y Wei, Q., “A Kolmogorov-Smirnov Test to Detect Changes in Stationarity in Big Data”, *IFAC-PapersOnLine*, vol. 50, n.º 1, págs. 14 260-14 265, 2017.



# A. Construcción del movimiento Browniano

Habiendo definido al movimiento Browniano surge una pregunta fundamental sobre si existe algún objeto tal que satisfaga alguna de las definiciones previas; para esto, realizaremos la construcción de Lévy-Ciesielski para el movimiento Browniano en el intervalo  $[0, 1]$ , basándonos en [55].

Consideraremos el espacio de funciones  $L^2[0, 1]$  definido por

$$L^2[0, 1] := \left\{ f : [0, 1] \rightarrow \mathbb{R} \mid f \text{ es medible y } \int_0^1 |f(x)|^2 dx < \infty \right\},$$

y lo dotaremos con el producto interno

$$\langle f, g \rangle = \int_0^1 f(x)g(x)dx,$$

para toda  $f, g \in L^2[0, 1]$ .

Luego, como todo producto interno induce una métrica, al considerar al espacio  $L^2[0, 1]$  como un espacio métrico, se tiene un espacio de Hilbert, es decir, un espacio métrico completo (ver [34], [28]).

Ahora, podemos proceder con el siguiente lema.

**Lema 92** Sea  $H : [0, 1] \rightarrow \mathbb{R}$  un función definida por

$$H(t) = \begin{cases} 1, & \text{si } 0 \leq t < \frac{1}{2}, \\ -1, & \text{si } \frac{1}{2} \leq t \leq 1, \\ 0, & \text{en otro caso.} \end{cases} \quad (\text{A.1})$$

Para  $j \geq 1$ ,  $0 \leq k < 2^{j-1}$  y  $t \in [0, 1]$  definimos las funciones

$$H_{0,0}(t) = 1, \quad (\text{A.2})$$

$$H_{j,k}(t) = 2^{(j-1)/2} H(2^{j-1}t - k).$$

Entonces el conjunto de funciones  $\{H_{j,k}\}_{j,k \in \mathbb{N} \cup \{0\}}$  es ortonormal en  $L^2[0, 1]$ .

**Nota 93** (I) Las funciones  $H$  son conocidas como ondula madre. Estas funciones han sido ampliamente usadas a partir del trabajo de Alfred Haar a principios del siglo XX (ver [26], [27], [9]). A las funciones definidas en (A.2) se les conoce como funciones de Haar.

(II) El conjunto de las funciones de Haar  $\{H_{j,k}\}_{j,k \in \mathbb{N} \cup \{0\}}$  es un conjunto numerable, pues todo número  $n \in \mathbb{N}$  puede escribirse como  $n = 2^{j-1} + k$ , con  $j, k \in \mathbb{N} \cup \{0\}$  tales que  $j \geq 1$  y  $0 \leq k < 2^j$ , entonces podemos reindexar de la manera siguiente

$$\{H_{j,k}\}_{j,k \in \mathbb{N} \cup \{0\}} = \{H_n\}_{n \in \mathbb{N} \cup \{0\}}.$$

(III) El conjunto  $\{H_n\}_{n \in \mathbb{N} \cup \{0\}}$  además de ser un conjunto ortonormal, es una base ortonormal para  $L^2[0, 1]$ , es decir, satisface que:

(a) Para toda  $f \in L^2[0, 1]$

$$\lim_{n \rightarrow \infty} \left\| f - \sum_{i=0}^n \langle f, H_i \rangle H_i \right\| = 0.$$

(b) Para toda  $f \in L^2[0, 1]$

$$\sum_{i=1}^{\infty} \langle f, H_i \rangle^2 = \|f\|^2.$$

DEMOSTRACIÓN. Notemos que para  $j \geq 1$  y  $0 \leq k < 2^{j-1}$  se tiene que

$$H_{j,k}(t) = \begin{cases} 2^{(j-1)/2}, & \text{si } \frac{k}{2^{j-1}} \leq t < \frac{1+2k}{2^j}, \\ -2^{(j-1)/2}, & \text{si } \frac{1+2k}{2^j} \leq t \leq \frac{1+k}{2^{j-1}}, \\ 0, & \text{en otro caso.} \end{cases}$$

En consecuencia,

$$\begin{aligned}
\langle H_{j,k}, H_{j,k} \rangle &= \int_0^1 (H_{j,k}(t))^2 dt \\
&= \int_0^{\frac{k}{2^{j-1}}} 0 dt + \int_{\frac{k}{2^{j-1}}}^{\frac{1+2k}{2^j}} (2^{(j-1)/2})^2 dt \\
&\quad + \int_{\frac{1+2k}{2^j}}^{\frac{1+k}{2^{j-1}}} (-2^{(j-1)/2})^2 dt + \int_{\frac{1+k}{2^{j-1}}}^1 0 dt \\
&= \int_{\frac{k}{2^{j-1}}}^{\frac{1+k}{2^{j-1}}} 2^{(j-1)} dt \\
&= 2^{j-1} \left( \frac{1+k}{2^{j-1}} - \frac{k}{2^{j-1}} \right) \\
&= 1.
\end{aligned}$$

Luego, es directo notar que  $\langle H_0, H_0 \rangle = 1$ ; de esta manera, hemos mostrado que todo elemento de  $\{H_{j,k}\}_{j,k \in \mathbb{N} \cup \{0\}}$  tiene norma igual a uno. Ahora, basta mostrar la ortogonalidad del conjunto.

Consideremos dos parejas  $j, k$  y  $j^*, k^*$ , tales que se cumpla al menos una de las siguientes condiciones,  $j \neq j^*$  ó  $k \neq k^*$ . Notemos que para  $j \geq 0$  fija el soporte de  $H_{j,k}$ ,  $\text{supp}(H_{j,k})$ , es el intervalo  $[k/2^{j-1}, 1+k/2^{j-1}]$ ; por lo que, las funciones  $H_{j,0}, \dots, H_{j,2^{j-1}-1}$  son traslaciones de  $H_{j,0}$ , de donde se tiene que la intersección de  $\text{supp}(H_{j,k})$  y  $\text{supp}(H_{j,k^*})$  es a lo más un punto (ver Figura A.1), es decir,

$$\text{card}(\text{supp}(H_{j,k}) \cap \text{supp}(H_{j,k^*})) \leq 1,$$

donde  $\text{card}()$  denota la cardinalidad del conjunto.

Lo anterior, implica que  $H_{j,k}(t)H_{j,k^*}(t) \neq 0$  para, a lo más, un punto. En consecuencia,

$$\int_0^1 H_{j,k}(t)H_{j,k^*}(t)dt = 0.$$

En general, para  $H_{j,k}$  y  $H_{j^*,k^*}$ , con  $j \neq j^*$  y  $k \neq k^*$ , se cumple que sus soportes están contenidos uno dentro de otro o son disjuntos (ver Figura A.2). En el primer caso, como  $H_{j^*,k^*}$  es simétrica respecto al punto  $\frac{1+2k^*}{2^{j^*}}$ , es decir,

$$H_{j^*,k^*} \left( -t + \frac{1+2k^*}{2^{j^*}} \right) = -H_{j^*,k^*} \left( t + \frac{1+2k^*}{2^{j^*}} \right), \quad 0 \leq t \leq 1.$$

De esta manera, se cumple que la longitud del intervalo en donde  $H_{j^*,k^*}$  es positiva,  $\left[ \frac{k^*}{2^{j^*-1}}, \frac{1+2k^*}{2^{j^*}} \right)$ , intersecado con el intervalo en donde  $H_{j,k}$  es positiva tiene la misma

longitud que el intervalo en donde  $H_{j^*,k^*}$  es negativa,  $\left[\frac{1+2k^*}{2^{j^*}}, \frac{1+k^*}{2^{j^*-1}}\right]$ , intersecado con el intervalo en donde  $H_{j,k}$  es positiva, y similarmente cuando se intersecamos con en el intervalo donde  $H_{j,k}$  es negativa, esto es,

$$\ell\left(\left[\frac{k}{2^{j-1}}, \frac{1+2k}{2^j}\right] \cap \left[\frac{k^*}{2^{j^*-1}}, \frac{1+2k^*}{2^{j^*}}\right]\right) = \ell\left(\left[\frac{k}{2^{j-1}}, \frac{1+2k}{2^j}\right] \cap \left[\frac{1+2k^*}{2^{j^*}}, \frac{1+k^*}{2^{j^*-1}}\right]\right)$$

y

$$\ell\left(\left[\frac{1+2k}{2^j}, \frac{1+k}{2^{j-1}}\right] \cap \left[\frac{k^*}{2^{j^*-1}}, \frac{1+2k^*}{2^{j^*}}\right]\right) = \ell\left(\left[\frac{1+2k}{2^j}, \frac{1+k}{2^{j-1}}\right] \cap \left[\frac{1+2k^*}{2^{j^*}}, \frac{1+k^*}{2^{j^*-1}}\right]\right),$$

donde  $\ell$  es la longitud de los intervalos.

Por lo anterior, definiendo

$$A^+ = \left[\frac{k}{2^{j-1}}, \frac{1+2k}{2^j}\right] \cap \left[\frac{k^*}{2^{j^*-1}}, \frac{1+2k^*}{2^{j^*}}\right], \quad A^- = \left[\frac{k}{2^{j-1}}, \frac{1+2k}{2^j}\right] \cap \left[\frac{1+2k^*}{2^{j^*}}, \frac{1+k^*}{2^{j^*-1}}\right]$$

$$B^+ = \left[\frac{1+2k}{2^j}, \frac{1+k}{2^{j-1}}\right] \cap \left[\frac{k^*}{2^{j^*-1}}, \frac{1+2k^*}{2^{j^*}}\right], \quad B^- = \left[\frac{1+2k}{2^j}, \frac{1+k}{2^{j-1}}\right] \cap \left[\frac{1+2k^*}{2^{j^*}}, \frac{1+k^*}{2^{j^*-1}}\right]$$

se obtiene que

$$\ell(A^+) - \ell(A^-) = 0 \quad \text{y} \quad \ell(B^+) - \ell(B^-) = 0,$$

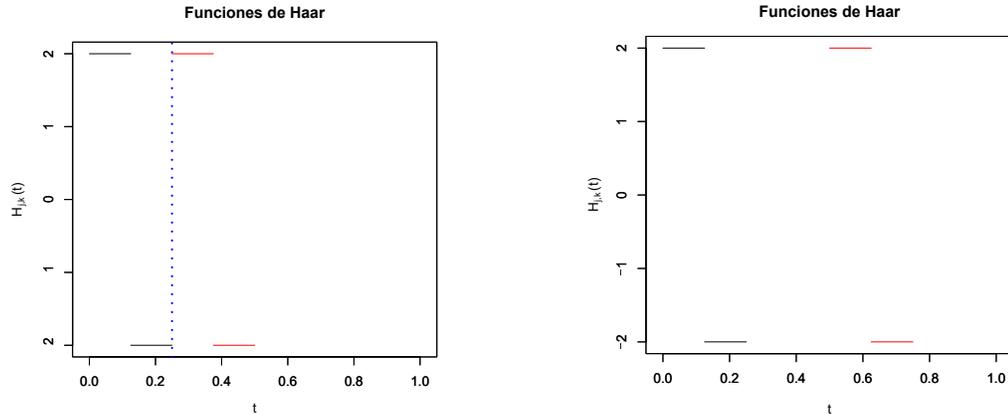
lo que implica que  $\int_0^1 H_{j,k}(t)H_{j^*,k^*}(t)dt = 0$ , pues

$$\begin{aligned} \int_0^1 H_{j,k}(t)H_{j^*,k^*}(t)dt &= 2^{(j-1)/2}2^{(j^*-1)/2} \left( \int_{A^+} dt - \int_{A^-} dt - \int_{B^+} dt + \int_{B^-} dt \right) \\ &= 2^{(j-1)/2}2^{(j^*-1)/2} (\ell(A^+) - \ell(A^-) - \ell(B^+) + \ell(B^-)) \\ &= 0. \end{aligned}$$

Por otra parte, cuando los soportes son disjuntos trivialmente se tiene que  $H_{j,k}(t)H_{j^*,k^*}(t)$  para toda  $0 \leq t \leq 1$ , por lo que  $\int_0^1 H_{j,k}(t)H_{j^*,k^*}(t)dt = 0$ . Así, habiendo mostrado la ortogonalidad del conjunto, se concluye la prueba.  $\square$

**Ejemplo 94** A continuación, ilustraremos el comportamiento de las funciones de Haar. En la Figura A.1 hemos fijado a  $j = 3$ , por lo que  $k$  puede tomar los valores 0, 1, 2 y 3. En la Figura A.1a se observa en color negro la gráfica de  $H_{3,0}$  y en color rojo la gráfica de  $H_{3,1}$ ; además, hemos resaltado el punto en donde  $\text{supp}(H_{3,0})$  y  $\text{supp}(H_{3,1})$  se intersecan trazando la recta vertical  $y = 1/4$  como línea punteada de color azul.

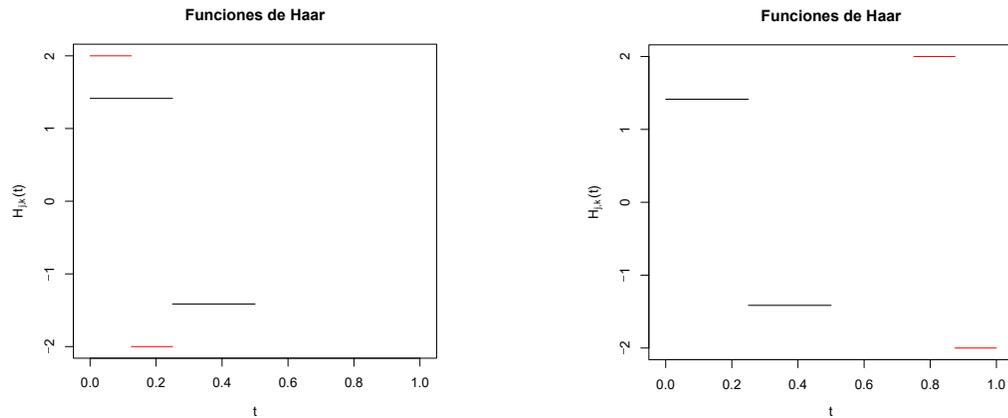
Luego, en la Figura A.1b, nuevamente tenemos en color negro a la gráfica de  $H_{3,0}$ , pero en color rojo tenemos la gráfica de la función  $H_{3,2}$ . De esta manera, observamos que sus soportes tienen intersección vacía y, más aún, al considerar las Figuras A.1a y A.1b se verifica visualmente que para  $j$  fija  $H_{j,k}$  es una traslación de  $H_{j,0}$ ,  $k = 0, 1, \dots, 2^j - 1$ .



(a) La intersección de los soportes es un punto. (b) La intersección de los soportes es el conjunto vacío.

Figura A.1: Gráficas de las funciones Haar para  $j$  fija.

En la Figura A.2 presentamos el caso cuando  $j \neq j^*$  y/o  $k \neq k^*$ , pues en negro tenemos la gráfica de la función  $H_{2,0}$  y para la Figura A.2a, en rojo tenemos la gráfica de  $H_{3,0}$ , en este caso podemos ver que  $\text{supp}(H_{3,0}) \subset \text{supp}(H_{2,0})$ . Por otra parte, en la Figura A.2b, se muestra en rojo la gráfica de la función  $H_{3,3}$  y podemos ver que los soportes son disjuntos.



(a) Un soporte contiene a otro. (b) Los soportes son disjuntos.

Figura A.2: Gráficas de las funciones Haar con  $j \neq j^*$  y/o  $k \neq k^*$ .

**Lema 95** El conjunto de funciones de Haar,  $\{H_{j,k}\}_{j,k \in \mathbb{N} \cup \{0\}}$ , definido como en el lema anterior, es una base ortonormal para  $L^2[0, 1]$ .

DEMOSTRACIÓN. Mostraremos que el conjunto de las funciones de Haar son densas en el conjunto de las funciones indicadoras de los intervalos diádicos contenidos en el intervalo  $[0, 1]$ .

Sea

$$D = \left\{ \left[ \frac{x}{2^z}, \frac{y}{2^z} \right] \subseteq [0, 1] \mid x, y, z \in \mathbb{N} \cup \{0\} \text{ tales que } x < y \right\}$$

el conjunto de intervalos diádicos contenidos en el intervalo  $[0, 1]$  y consideremos un intervalo arbitrario y fijo de la forma  $\left[ \frac{r}{2^s}, \frac{r+1}{2^s} \right] = I_{r,s}$ , con  $r, s \in \mathbb{N} \cup \{0\}$  y  $0 \leq r < 2^s$ .

Comenzaremos por mostrar que las funciones de Haar, satisfacen las condiciones de base ortonormal (ver inciso (III) de la Nota 93) cuando se consideran las funciones indicadoras de los intervalos de la forma  $I_{r,s}$ .

Notemos que, por su definición,  $I_{r,s}$  está contenido en algún intervalo de la forma  $\left[ \frac{m}{2^i}, \frac{m+1/2}{2^i} \right]$  ó  $\left[ \frac{m+1/2}{2^i}, \frac{m+1}{2^i} \right]$ , para algún  $m \in \mathbb{N} \cup \{0\}$  si  $i < s$ , entonces para  $1 \leq j \leq s$  se satisface que

$$\sum_{k=0}^{2^{j-1}-1} \left( \int_{[0,1]} \mathbf{1}_{I_{r,s}}(t) H_{j,k}(t) dt \right)^2 = \left( \frac{\pm 2^{(j-1)/2}}{2^s} \right)^2 = \frac{2^{j-1}}{2^{2s}}.$$

Por otra parte, para cualquier  $j \geq s+1$  es directo observar que

$$\left( \int_{[0,1]} \mathbf{1}_{I_{r,s}}(t) H_{j,k}(t) dt \right)^2 = 0.$$

Y por último, respecto a  $H_{0,0}$  se tiene que

$$\left( \int_{[0,1]} \mathbf{1}_{I_{r,s}}(t) H_{0,0}(t) dt \right)^2 = \frac{1}{2^{2s}}.$$

Por lo que

$$\begin{aligned} \sum_{j,k} \langle \mathbf{1}_{I_{r,s}}, H_{j,k} \rangle^2 &= \sum_{j,k} \left( \int_{[0,1]} \mathbf{1}_{I_{r,s}}(t) H_{j,k}(t) dt \right)^2 = \frac{1}{2^{2s}} + \sum_{j=1}^s \left( \frac{2^{j-1}}{2^{2s}} \right) \\ &= \frac{1}{2^{2s}} + \frac{2^s - 1}{2^{2s}} = \frac{1}{2^s} = \|\mathbf{1}_{I_{r,s}}\|^2. \end{aligned}$$

Usando lo anterior, aunado a la ortonormalidad del conjunto  $\{H_{j,k}\}_{j,k \in \mathbb{N} \cup \{0\}}$ , se tiene que

$$\begin{aligned}
\left\| \sum_{j,k} \langle \mathbf{1}_{I_{r,s}}, H_{j,k} \rangle H_{j,k} \right\|^2 &= \left\langle \sum_{j,k} \langle \mathbf{1}_{I_{r,s}}, H_{j,k} \rangle H_{j,k}, \sum_{j,k} \langle \mathbf{1}_{I_{r,s}}, H_{j,k} \rangle H_{j,k} \right\rangle \\
&= \int_{[0,1]} \left( \sum_{j,k} \langle \mathbf{1}_{I_{r,s}}, H_{j,k} \rangle H_{j,k}(t) \right)^2 dt \\
&= \int_{[0,1]} \sum_{j,k} \langle \mathbf{1}_{I_{r,s}}, H_{j,k} \rangle^2 H_{j,k}^2(t) dt \\
&\quad + \sum_{\substack{j \neq j^* \\ k \neq k^*}} \langle \mathbf{1}_{I_{r,s}}, H_{j,k} \rangle \langle \mathbf{1}_{I_{r,s}}, H_{j^*,k^*} \rangle H_{j,k}(t) H_{j^*,k^*}(t) dt \quad (\text{A.3}) \\
&= \int_{[0,1]} \sum_{j,k} \langle \mathbf{1}_{I_{r,s}}, H_{j,k} \rangle^2 H_{j,k}^2(t) dt \\
&= \int_{[0,1]} \sum_{j,k} \langle \mathbf{1}_{I_{r,s}}, H_{j,k} \rangle^2 dt \\
&= \|\mathbf{1}_{I_{r,s}}\|^2.
\end{aligned}$$

Ahora, notemos lo siguiente

$$\begin{aligned}
\left\| \mathbf{1}_{I_{r,s}} - \sum_{j,k} \langle \mathbf{1}_{I_{r,s}}, H_{j,k} \rangle H_{j,k} \right\|^2 &= \left\langle \mathbf{1}_{I_{r,s}} - \sum_{j,k} \langle \mathbf{1}_{I_{r,s}}, H_{j,k} \rangle H_{j,k}, \mathbf{1}_{I_{r,s}} - \sum_{j,k} \langle \mathbf{1}_{I_{r,s}}, H_{j,k} \rangle H_{j,k} \right\rangle \\
&= \int_{[0,1]} \mathbf{1}_{I_{r,s}}^2(t) - 2f(t) \sum_{j,k} \langle \mathbf{1}_{I_{r,s}}, H_{j,k} \rangle H_{j,k}(t) dt + \\
&\quad \int_{[0,1]} \left( \sum_{j,k} \langle \mathbf{1}_{I_{r,s}}, H_{j,k} \rangle H_{j,k}(t) \right)^2 dt \\
&= \|\mathbf{1}_{I_{r,s}}\|^2 - 2 \int_{[0,1]} \sum_{j,k} \langle \mathbf{1}_{I_{r,s}}, H_{j,k} \rangle \mathbf{1}_{I_{r,s}}(t) H_{j,k}(t) dt \\
&\quad + \left\| \sum_{j,k} \langle \mathbf{1}_{I_{r,s}}, H_{j,k} \rangle H_{j,k} \right\|^2 \\
&= \|\mathbf{1}_{I_{r,s}}\|^2 - 2 \int_{[0,1]} \sum_{j,k} \langle \mathbf{1}_{I_{r,s}}, H_{j,k} \rangle^2 dt + \left\| \sum_{j,k} \langle \mathbf{1}_{I_{r,s}}, H_{j,k} \rangle H_{j,k} \right\|^2 \\
&= \|\mathbf{1}_{I_{r,s}}\|^2 - \left\| \sum_{j,k} \langle \mathbf{1}_{I_{r,s}}, H_{j,k} \rangle H_{j,k} \right\|^2.
\end{aligned}$$

De lo anterior, usando la ecuación (A.3), se tiene que

$$\left\| \mathbf{1}_{I_{r,s}} - \sum_{j,k} \langle \mathbf{1}_{I_{r,s}}, H_{j,k} \rangle H_{j,k} \right\| = 0.$$

Notemos que podemos extender lo anterior para intervalos de longitud  $\frac{1}{2^s}$  abiertos de uno o ambos extremos, pues la argumentación sería completamente análoga. En consecuencia, como cualquier elemento de  $D$  puede escribirse como una unión disjunta de intervalos de longitud  $\frac{1}{2^s}$ , las funciones de Haar  $\{H_{j,k}\}_{j,k \in \mathbb{N} \cup \{0\}}$  son densas en el conjunto de funciones indicadoras de los intervalos diádicos, puesto que podemos escribir la función indicadora de cualquier elemento de  $D$  como una suma de funciones indicadoras de intervalos de longitud  $\frac{1}{2^s}$ .

Finalmente, dado que las funciones Riemann integrables pueden ser aproximadas por funciones indicadoras, considerando combinaciones lineales de ellas, es decir, considerar funciones simples [48], se obtiene que, en efecto, para cualquier función  $f \in L^2[0, 1]$  las funciones de Haar satisfacen la definición de base ortonormal.  $\square$

**Nota 96** Como el conjunto  $\{H_{j,k}\}_{j,k \in \mathbb{N} \cup \{0\}}$  es una base ortonormal para  $L^2[0, 1]$  se cumple la identidad Parseval, para espacios de Hilbert; esto es, para toda  $f, g \in L^2[0, 1]$

$$\langle f, g \rangle = \sum_{j,k} \langle f, H_{j,k} \rangle \langle g, H_{j,k} \rangle. \quad (\text{A.4})$$

Para la construcción de un objeto estocástico como lo es el movimiento Browniano, es claro que no es suficiente con considerar únicamente funciones de  $L^2[0, 1]$ ; sin embargo, lo anterior, nos será de utilidad para estudiar las ondículas triangulares, las cuales juegan un rol crucial en la construcción. Ahora, comenzaremos a trabajar con v.a. en el siguiente lema, para agregar la parte aleatoria a nuestra construcción.

Habiendo recordado lo anterior, procedamos con el siguiente resultado para la construcción del movimiento Browniano.

**Lema 97** Sea  $\{Z_n\}_{n \in \mathbb{N} \cup \{0\}}$  una sucesión de v.a. i.i.d. tal que  $Z_n \sim N(0, 1)$ . Si  $n \geq 2$ , entonces existe una variable aleatoria  $C$  tal que

$$|Z_n| \leq C \sqrt{\log(n)} \quad \text{y} \quad P(C < \infty) = 1.$$

**DEMOSTRACIÓN.** Usando que  $Z_n$  tiene distribución normal estándar y que para toda  $x \geq 1$  se satisface que

$$e^{-x^2/2} \leq x e^{-x^2/2}.$$

Obtenemos

$$\begin{aligned} P(|Z_n| \geq x) &= \frac{2}{\sqrt{2\pi}} \int_x^\infty e^{-u^2/2} du \\ &\leq \sqrt{\frac{2}{\pi}} \int_x^\infty u e^{-u^2/2} du \\ &= \sqrt{\frac{2}{\pi}} e^{-x^2/2}. \end{aligned}$$

Notemos que para  $\alpha > 1$  y  $n \geq 2$  se satisface que

$$2\alpha \log(n) > 1,$$

de esta manera, por lo anterior, se tiene que

$$\begin{aligned} \sum_{n=2}^{\infty} P\left(|Z_n| \geq \sqrt{2\alpha \log(n)}\right) &\leq \sum_{n=2}^{\infty} \sqrt{\frac{2}{\pi}} \exp(-\alpha \log(n)) \\ &= \sqrt{\frac{2}{\pi}} \sum_{n=2}^{\infty} \frac{1}{n^\alpha} < \infty, \end{aligned}$$

donde la serie  $\sum_{n=1}^{\infty} \frac{1}{n^\alpha} < \infty$  por el criterio de las  $p$ -series, y por tanto

$$\sum_{n=2}^{\infty} P\left(|Z_n| \geq \sqrt{2\alpha \log(n)}\right) < \infty.$$

Sea  $A_n = \{|Z_n| \geq \sqrt{2\alpha \log(n)}\}$ . Por el lema de Borel-Cantelli (ver [3] y [4]), se tiene que

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) = 0,$$

esto significa que con probabilidad uno, solo una cantidad finita de eventos  $A_n$  son simultáneos.

Entonces si definimos

$$C = \sup_{\substack{n \in \mathbb{N} \\ n \geq 2}} \frac{|Z_n|}{\sqrt{\log(n)}},$$

se tiene que  $C$  es finita c.s., puesto que al ocurrir de manera simultánea solo una colección finita de los  $A_n$ , esto implica que la cantidad de veces que  $|Z_n|$  es mayor que  $\sqrt{2\alpha \log(n)}$

es, a lo más, finita; así  $\frac{|Z_n|}{\sqrt{\log(n)}} < \sqrt{2\alpha} < \infty$  c.s. y tomando el supremo sobre los  $n \in \mathbb{N}$

tales que  $n \geq 2$ , se tiene lo deseado.  $\square$

En el siguiente resultado estableceremos la relación entre las ondículas de Haar y la sucesión de variables aleatorias con distribución normal estándar, para ello, es oportuno introducir a las funciones de Faber–Schauder (ver [20], [49]), también conocidas como ondículas triangulares.

Las ondículas triangulares se generan a partir de la función  $\Delta : [0, 1] \rightarrow \mathbb{R}$ , definida por

$$\Delta(t) = \begin{cases} 2t, & \text{si } 0 \leq t < \frac{1}{2}. \\ 2(1-t), & \text{si } \frac{1}{2} \leq t \leq 1. \end{cases}$$

Es directo verificar que  $\Delta$  es dos veces la integral de la función  $H$ , definida previamente en la ecuación (A.1), en efecto, para  $0 \leq t < \frac{1}{2}$

$$\int_0^t H(\tau) d\tau = \int_0^t 1 d\tau = t.$$

Por otra parte, para  $\frac{1}{2} \leq t \leq 1$  se tiene que

$$\int_0^t H(\tau) d\tau = \int_0^{1/2} 1 d\tau + \int_{1/2}^t -1 d\tau = \frac{1}{2} + \frac{1}{2} - t = 1 - t.$$

En general, cada ondícula triangular, multiplicada por una constante de proporcionalidad, resulta ser la integral de una función de Haar, es decir, para cada  $t \in [0, 1]$

$$\lambda_{0,0} \Delta_{0,0}(t) = \int_0^t H_{0,0}(\tau) d\tau = t, \quad (\text{A.5})$$

donde  $\lambda_{0,0} = 1$ .

Para  $j, k \in \mathbb{N} \cup \{0\}$  tales que  $j \geq 1$  y  $0 \leq k < 2^j$ , se tiene que para  $0 \leq t < \frac{k}{2^{j-1}}$

$$\int_0^t H_{j,k}(\tau) d\tau = \int_0^t 0 d\tau = 0.$$

Para  $\frac{k}{2^{j-1}} \leq t < \frac{1+2k}{2^j}$ ,

$$\begin{aligned} \int_0^t H_{j,k}(\tau) d\tau &= \int_{\frac{k}{2^{j-1}}}^t 2^{(j-1)/2} d\tau = 2^{(j-1)/2} \left( t - \frac{k}{2^{j-1}} \right) = 2^{(j-1)/2} \left( \frac{2^{j-1}t - k}{2^{j-1}} \right) \\ &= \left( \frac{2^{j-1}t - k}{2^{(j-1)/2}} \right). \end{aligned} \quad (\text{A.6})$$

Por otro lado, para  $\frac{1+2k}{2^j} \leq t < \frac{1+k}{2^{j-1}}$ , se tiene que

$$\begin{aligned} \int_0^t H_{j,k}(\tau) d\tau &= \int_{\frac{k}{2^{j-1}}}^{\frac{1+2k}{2^j}} 2^{(j-1)/2} d\tau + \int_{\frac{1+2k}{2^j}}^t -2^{(j-1)/2} d\tau = \frac{1}{2^{(j+1)/2}} + 2^{(j-1)/2} \left( \frac{1+2k}{2^j} - t \right) \\ &= \frac{1}{2^{(j+1)/2}} + 2^{(j-1)/2} \left( \frac{1/2 + k - 2^{j-1}t}{2^{j-1}} \right) = \frac{1}{2^{(j+1)/2}} + \left( \frac{1/2 + k - 2^{j-1}t}{2^{(j-1)/2}} \right) \\ &= \frac{1}{2^{(j+1)/2}} + \left( \frac{1/2 + k - 2^{j-1}t}{2^{(j-1)/2}} \right) = \frac{1+k-2^{j-1}t}{2^{(j-1)/2}}. \end{aligned} \tag{A.7}$$

Finalmente, para  $\frac{1+k}{2^{j-1}} < t \leq 1$

$$\begin{aligned} \int_0^t H_{j,k}(\tau) d\tau &= \int_{\frac{k}{2^{j-1}}}^{\frac{1+2k}{2^j}} 2^{(j-1)/2} d\tau + \int_{\frac{1+2k}{2^j}}^{\frac{1+k}{2^{j-1}}} -2^{(j-1)/2} d\tau + \int_{\frac{1+k}{2^{j-1}}}^t 0 d\tau \\ &= \frac{1}{2^{(j+1)/2}} + 2^{(j-1)/2} \left( \frac{1+2k}{2^j} - \frac{1+k}{2^{j-1}} \right) = \frac{1}{2^{(j+1)/2}} - \frac{1}{2^{(j+1)/2}} \\ &= 0. \end{aligned} \tag{A.8}$$

De esta manera,

$$\lambda_{j,k} \Delta_{j,k}(t) = \lambda_{j,k} \Delta(2^{j-1}t - k) = \int_0^t H_{j,k}(\tau) d\tau,$$

con  $\lambda_{j,k} = \frac{2^{-(j-1)/2}}{2}$ .

Habiendo definido las ondículas triangulares, podemos introducir el siguiente resultado.

**Lema 98** Sea  $\{Z_{j,k} \mid j, k \in \mathbb{N} \cup \{0\}, \text{ con } 0 \leq k < 2^j\}$  una colección de v.a. i.i.d. tal que  $Z_{j,k} \sim N(0, 1)$ ; además, consideremos los  $\lambda_{j,k}$  y  $\Delta_{j,k}$  definimos previamente. Si

$$W_t = \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} \lambda_{j,k} Z_{j,k} \Delta_{j,k}(t), \quad 0 \leq t \leq 1, \tag{A.9}$$

entonces  $W_t$  converge uniforme y absolutamente en el intervalo  $[0, 1]$  con probabilidad uno.

**DEMOSTRACIÓN.** Sea  $N \in \mathbb{N}$  un número arbitrario y fijo. Recordemos que el inciso (II) de la Nota 93 implica que la colección de v.a. puede reindexarse de la manera siguiente

$$\{Z_{j,k} \mid j, k \in \mathbb{N} \cup \{0\}, \text{ con } 0 \leq k < 2^j\} = \{Z_n\}_{n \in \mathbb{N} \cup \{0\}},$$

tomando  $n = 2^{j-1} + k$ . De esta manera, podemos entender a la colección de v.a. previa como una sucesión de v.a.

Así, en consecuencia del Lema 97, se tiene que

$$\begin{aligned} \sum_{j=N}^{\infty} \sum_{k=0}^{2^{j-1}-1} |\lambda_{j,k} Z_{j,k} \Delta_{j,k}(t)| &= \sum_{j=N}^{\infty} \sum_{k=0}^{2^{j-1}-1} \lambda_{j,k} |Z_{j,k}| \Delta_{j,k}(t) \\ &\leq \sum_{j=N}^{\infty} \sum_{k=0}^{2^{j-1}-1} C(\omega) \lambda_{j,k} \sqrt{\log(2^{j-1} + k)} \Delta_{j,k}(t). \end{aligned}$$

Como  $2^{j-1} + k < 2^{j-1} + 2^{j-1} = 2^j$ , entonces  $\log(2^{j-1} + k) < \log(2^j) = j \log(2) \leq j$ , implicando que

$$\sum_{j=N}^{\infty} \sum_{k=0}^{2^{j-1}-1} C(\omega) \lambda_{j,k} \sqrt{\log(2^{j-1} + k)} \Delta_{j,k}(t) \leq \frac{C(\omega)}{2} \sum_{j=N}^{\infty} \sum_{k=0}^{2^{j-1}-1} 2^{-(j-1)/2} \sqrt{k} \Delta_{j,k}(t).$$

Como  $\Delta_{j,k}(t) = 0$  en  $[0, 1] \setminus \left[ \frac{k}{2^{j-1}}, \frac{1+k}{2^{j-1}} \right]$  y  $|\Delta_{j,r}| \leq 1$ , se tiene que

$$\frac{C(\omega)}{2} \sum_{j=N}^{\infty} \sum_{k=0}^{2^{j-1}-1} 2^{-(j-1)/2} \sqrt{k} \Delta_{j,k}(t) \leq \frac{C(\omega)}{2} \sum_{j=N}^{\infty} 2^{(j-1)/2} \sqrt{k}.$$

Donde

$$\frac{C(\omega)}{2} \sum_{j=N}^{\infty} 2^{(j-1)/2} \sqrt{k} \xrightarrow{c.s.} 0,$$

puesto que  $C$  es finita c.s., esto implica que la cantidad de eventos en los que la serie anterior diverge es, a lo más, finita; en consecuencia,  $W_t$  converge absolutamente con probabilidad uno en  $[0, 1]$ . Luego, como la convergencia absoluta implica la convergencia uniforme, entonces se ha probado lo deseado.  $\square$

**Nota 99** Observemos que en el resultado anterior, las sumas parciales de  $W_t$  son continuas, ya que cada sumando es producto de funciones continuas; por lo tanto el límite es continuo.

Ahora, teniendo los lemas previos, podemos construir al movimiento Browniano en el  $[0, 1]$ .

**Teorema 100** Sean  $\lambda_{j,k}$ ,  $Z_{j,k}$  y  $\Delta_{j,k}$ ,  $j, k \in \mathbb{N} \cup \{0\}$  con  $0 \leq k < 2^j$ , como en el lema anterior. Si para cada  $0 \leq t \leq 1$  definamos  $W_t$  como en la ecuación (A.9), entonces  $W = \{W_t \mid 0 \leq t \leq 1\}$  es un movimiento Browniano en el intervalo  $[0, 1]$ .

DEMOSTRACIÓN. Recordemos que por el inciso (II) de la Nota 93, podemos escribir cualquier  $n \in \mathbb{N} \cup \{0\}$  como  $n = 2^j + k$  para algunas  $j, k \in \mathbb{N} \cup \{0\}$  tales que  $0 \leq k < 2^j$ . De esta manera  $Z_{j,k}$  es equivalente a escribir  $Z_n$  y es completamente análogo para  $\lambda_{j,k}$  y  $\Delta_{j,k}$ .

Ahora, sean  $s, t \in [0, 1]$ ,  $0 \leq s < t$ . Usando lo anterior y la definición de  $W_t$  (ver ecuación (A.9)) se tiene que

$$\begin{aligned} W_t - W_s &= \lim_{N \rightarrow \infty} \sum_{n=0}^N Z_n \int_0^t H_n(\tau) d\tau - \sum_{n=0}^N Z_n \int_0^s H_n(\tau) d\tau \\ &= \lim_{N \rightarrow \infty} \sum_{n=0}^N Z_n \int_s^t H_n(\tau) d\tau. \end{aligned}$$

Entonces, si  $0 \leq s \leq t \leq u \leq v$  y consideramos el vector aleatorio  $X^T = (W_t - W_s, W_v - W_u)$  se tiene que la función característica de  $X$ , para todo  $a^T = (a_1, a_2) \in \mathbb{R}^2$  es

$$\begin{aligned} E(e^{ia^T X}) &= E(\exp[ia_1(W_t - W_s) + ia_2(W_v - W_u)]) \\ &= E\left(\exp\left[ia_1 \lim_{N \rightarrow \infty} \sum_{n=0}^N Z_n \int_s^t H_n(\tau) d\tau + ia_2 \lim_{N \rightarrow \infty} \sum_{n=0}^N Z_n \int_u^v H_n(\tau) d\tau\right]\right) \\ &= E\left(\exp\left[i \lim_{N \rightarrow \infty} \sum_{n=0}^N Z_n \left(a_1 \int_s^t H_n(\tau) d\tau + a_2 \int_u^v H_n(\tau) d\tau\right)\right]\right) \\ &= E\left(\lim_{N \rightarrow \infty} \exp\left[i \sum_{n=0}^N Z_n \left(a_1 \int_s^t H_n(\tau) d\tau + a_2 \int_u^v H_n(\tau) d\tau\right)\right]\right) \\ &= \lim_{N \rightarrow \infty} \prod_{n=0}^N E\left(\exp\left[i Z_n \left(a_1 \int_s^t H_n(\tau) d\tau + a_2 \int_u^v H_n(\tau) d\tau\right)\right]\right). \end{aligned}$$

Notemos que  $a_1 \int_s^t H_n(\tau) d\tau + a_2 \int_u^v H_n(\tau) d\tau \in \mathbb{R}$ , por lo que cada factor es la función característica de una variable normal estándar multiplicada por una constante, por lo que

$$\begin{aligned} &\lim_{N \rightarrow \infty} \prod_{n=0}^N E\left(\exp\left[i Z_n \left(a_1 \int_s^t H_n(\tau) d\tau + a_2 \int_u^v H_n(\tau) d\tau\right)\right]\right) \\ &= \lim_{N \rightarrow \infty} \prod_{n=0}^N \exp\left(it0 - \frac{1}{2} \left[a_1 \int_s^t H_n(\tau) d\tau + a_2 \int_u^v H_n(\tau) d\tau \in \mathbb{R}\right]^2\right) \\ &= \lim_{N \rightarrow \infty} \exp\left(-\frac{1}{2} \sum_{n=0}^N \left[a_1 \int_s^t H_n(\tau) d\tau + a_2 \int_u^v H_n(\tau) d\tau \in \mathbb{R}\right]^2\right). \end{aligned}$$

Expandiendo el binomio al cuadrado se tiene que

$$\begin{aligned} \lim_{N \rightarrow \infty} \exp \left( -\frac{1}{2} \sum_{n=0}^N \left[ a_1 \int_s^t H_n(\tau) d\tau + a_2 \int_u^v H_n(\tau) d\tau \in \mathbb{R} \right]^2 \right) = \\ \lim_{N \rightarrow \infty} \exp \left( -\frac{1}{2} \sum_{n=0}^N a_1^2 \left( \int_s^t H_n(\tau) dt \right)^2 + a_1 a_2 \int_s^t H_n(\tau) dt \int_u^v H_n(\tau) dt + a_2^2 \left( \int_u^v H_n(\tau) dt \right)^2 \right). \end{aligned} \quad (\text{A.10})$$

Ahora, como  $[s, t)$  y  $[u, v)$  son intervalos disjuntos se tiene que

$$\langle \mathbf{1}_{[s,t)}, \mathbf{1}_{[u,v)} \rangle = 0.$$

Por otro lado, usando la identidad de Parseval (ver ecuación (A.4)) se tiene que

$$\begin{aligned} \langle \mathbf{1}_{[s,t)}, \mathbf{1}_{[u,v)} \rangle &= \sum_{n=0}^{\infty} \langle \mathbf{1}_{[s,t)}, H_n \rangle \langle \mathbf{1}_{[u,v)}, H_n \rangle \\ &= \sum_{n=0}^{\infty} \int_s^t H_n(\tau) dt \int_u^v H_n(\tau) dt. \end{aligned}$$

Por lo que el segundo sumando de la ecuación (A.10) es cero; así, usando lo antes mencionado y la definición de base ortonormal, tiene que

$$\begin{aligned} E \left( e^{ia^T X} \right) &= \lim_{N \rightarrow \infty} \exp \left( -\frac{1}{2} \sum_{n=0}^N a_1^2 \left( \int_s^t H_n(\tau) dt \right)^2 + a_2^2 \left( \int_u^v H_n(\tau) dt \right)^2 \right) \\ &= \lim_{N \rightarrow \infty} \exp \left( -\frac{1}{2} a_1^2 \sum_{n=0}^N \left( \int_s^t H_n(\tau) dt \right)^2 \right) \exp \left( -\frac{1}{2} a_2^2 \sum_{n=0}^N \left( \int_u^v H_n(\tau) dt \right)^2 \right) \\ &= \exp \left( -\frac{1}{2} a_1^2 \sum_{n=0}^{\infty} \langle \mathbf{1}_{[s,t)}, H_n \rangle^2 \right) \exp \left( -\frac{1}{2} a_2^2 \sum_{n=0}^{\infty} \langle \mathbf{1}_{[u,v)}, H_n \rangle^2 \right) \\ &= \exp \left( -\frac{1}{2} a_1^2 \|\mathbf{1}_{[s,t)}\|^2 \right) \exp \left( -\frac{1}{2} a_2^2 \|\mathbf{1}_{[u,v)}\|^2 \right) \\ &= \exp \left( -\frac{1}{2} a_1^2 (t - s) \right) \exp \left( -\frac{1}{2} a_2^2 (v - u) \right). \end{aligned}$$

Con lo anterior, hemos mostrado que dados dos incrementos, éstos son independientes; sin embargo, podemos extender este argumento para cualquier cantidad finita de incrementos lo que implica que  $W$  es un proceso con incrementos independientes.

Por otra parte, como  $W_t$  converge absolutamente en  $[0, 1]$  (ver Lema 98) se tiene que

$$E(W_t) = E\left(\sum_{n=0}^{\infty} \lambda_n Z_n \Delta_n(t)\right) = \sum_{n=0}^{\infty} \lambda_n \Delta_n(t) E(Z_n) = 0,$$

para toda  $t \in [0, 1]$ . En consecuencia, la función media de  $W$  es la función idénticamente cero, es decir,  $\mu_W(t) = 0$ , para toda  $0 \leq t \leq 1$ .

Respecto a la función de covarianza de  $W$ , se tiene que si  $s, t \in [0, 1]$ , entonces

$$\begin{aligned} \kappa_W(s, t) &= E([W_s - E(W_s)][W_t - E(W_t)]) = E(W_s W_t) \\ &= E\left(\left[\sum_{n=0}^{\infty} \lambda_n Z_n \Delta_n(t)\right] \left[\sum_{m=0}^{\infty} \lambda_m Z_m \Delta_m(s)\right]\right) \\ &= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} E\left(\left[Z_n \int_0^t H_n(\tau) d\tau\right] \left[Z_m \int_0^t H_n(\tau) d\tau\right]\right) \\ &= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} E(Z_n Z_m) \int_0^t H_n(\tau) d\tau \int_0^t H_n(\tau) d\tau \tag{A.11} \\ &= \sum_{n=0}^{\infty} E(Z_n^2) \int_0^t H_n(\tau) d\tau \int_0^t H_n(\tau) d\tau \\ &\quad + \sum_{\substack{n=0 \\ m=0 \\ n \neq m}}^{\infty} E(Z_n Z_m) \int_0^t H_n(\tau) d\tau \int_0^t H_n(\tau) d\tau. \end{aligned}$$

Luego, ya que  $\{Z_n\}_{n \in \mathbb{N}}$  es una sucesión de v.a. i.i.d. tales que  $Z_n \sim N(0, 1)$  tenemos que

$$E(Z_n Z_m) = \begin{cases} 0, & \text{si } m \neq n. \\ 1, & \text{si } m = n. \end{cases}$$

Así, la identidad de Parseval (ver ecuación (A.4)), implica que la ecuación (A.11) es igual a

$$\begin{aligned} \kappa_W(s, t) &= \sum_{n=0}^{\infty} \int_0^t H_n(\tau) d\tau \int_0^t H_n(\tau) d\tau = \sum_{n=0}^{\infty} \langle \mathbf{1}_{[0,t]}, H_n \rangle \langle \mathbf{1}_{[0,s]}, H_n \rangle \\ &= \langle \mathbf{1}_{[0,t]}, \mathbf{1}_{[0,s]} \rangle = \min\{s, t\}. \end{aligned}$$

Por último, de la Nota 99, sabemos que las trayectorias son continuas con probabilidad uno mientras que de la definición de  $W_0$  se tiene que es igual a cero; por lo que, en particular es igual a cero con probabilidad uno. De esta manera,  $W$  satisface la Definición 70 del movimiento Browniano como un proceso Gaussiano, lo que concluye la prueba.  $\square$

**Nota 101** A partir de la construcción del teorema anterior, podemos construir al movimiento Browniano en intervalo  $[0, \infty)$  [51], [55]. Para esto, es necesario considerar procesos independientes, esto es, diremos que los procesos  $X$  y  $Y$  son independientes si para cualesquiera tiempos  $t_1, \dots, t_n \in T_1$  y  $\tau_1, \dots, \tau_n \in T_2$ , los vectores aleatorios

$$(X_{t_1}, \dots, X_{t_n}) \quad \text{y} \quad (Y_{\tau_1}, \dots, Y_{\tau_n})$$

son independientes.

Así, considerando  $\{W^k\}_{k \in \mathbb{N} \cup \{0\}}$ , con  $W^k = \{W_t^k \mid 0 \leq t \leq 1\}$ , una sucesión de movimiento Brownianos definidos en el mismo espacio de probabilidad si para cada  $t \geq 0$  definimos

$$W_t = \begin{cases} W_t^0, & \text{si } t \in [0, 1). \\ W_1^0 + W_{t-1}^1, & \text{si } t \in [1, 2). \\ \sum_{k=0}^{j-1} W_1^k + W_{t-j}^j, & \text{si } t \in [j, j+1), j \geq 2. \end{cases}$$

Entonces  $\mathcal{W} = \{W_t \mid t \geq 0\}$  es un movimiento Browniano en  $[0, \infty)$ .