116

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

LA LEY ARCOSENO EN CAMINATAS ALEATORIAS, MOVIMIENTO BROWNIANO Y EN EL PROCESO POISSON COMPUESTO

T E S I S

QUE PARA OBTENER EL TÍTULO DE
ACTUARIO

PRESENTA:

GERÓNIMO FRANCISCO URIBE BRAVO

DIRECTORA DE TESIS:

DRA. MARÍA EMILIA CABALLERO ACOSTA

2002

M. EN C. ELENA DE OTEYZA DE OTEYZA
Jefa de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunico a usted que hemos revisado el trabajo escrito: La ley arcoseno en caminatas aleatorias movimiento browniano y en el proceso Poisson compuesto

realizado por Gerónimo Francisco Uribe Bravo

con número de cuenta 9757687-5, quién cubrió los créditos de la carrera de Actuaría

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis
Propietario Dra. María Emilia Caballero Acosta

Propietario Dr. Guillermo Grabinsky Steider

Propietario Dra. Ana Meda Guardiola

Suplente Dr. Alberto Contreras Cristán

Suplente Dra. Eliane Regina Rodrigues Caloni Eliane Regina Rodrigues

Consejo Departamental de Matemáticas

M. en C. José Antonio Flores Díaz

CONSEJO DEPARTAMENTAL DE MATEMATICAS

La ley arcoseno en caminatas aleatorias, movimiento browniano y en el proceso Poisson compuesto

Gerónimo F. Uribe

A mis padres, Adriana y Paco, y a mi hermano Gabriel, con mucho cariño.

A Abi, con todo mi amor.

Reconocimientos

En este espacio, no podré nombrar a todas las personas que me apoyaron para la realización de la tesis, tanto en aspectos académicos como también moralmente.

Quiero agradecer especialmente el apoyo que me brindó y el tiempo que me dedicó durante la carrera y en la elaboración de la tesis a la Dra. María Emilia Caballero.

Al Dr. Guillermo Grabinsky, le reconozco haber sido uno de los mejores profesores que he tenido y le agradezco el tiempo que dedicó a la revisión de este trabajo, así como les doy las gracias a los Doctores Alberto Contreras, Ana Meda y Eliane Rodrigues por los comentarios que hicieron en la revisión final.

A mis amigos, por los momentos alegres que vivimos y a mi numerosa familia de la que siempre recibí cariño y ayuda cuando lo necesite: muchas gracias.

Índice general

Abreviaturas y símbolos III

Introducción V
Planteamiento del problema V
Estructura de la tesis VII

Capítulo 1. Las caminatas aleatorias y la ley arcoseno 1
1.1. Introducción 1
1.2. La caminata aleatoria simple y la ley arcoseno 2
1.3. Cantidad de elementos positivos en caminatas aleatorias 16
Notas 30

Capítulo 2. El Movimiento Browniano y la ley arcoseno 33
2.1. Introducción 33
2.2. El principio de invariancia de Erdös-Kac 33
2.3. El principio de invariancia de Donsker 38
2.4. La ley arcoseno 74
Notas 77

Capítulo 3. El proceso Poisson compuesto y la ley arcoseno 79
3.1. Convergencia débil en el espacio de Skorohod 79
3.2. Procesos de Lévy 81
3.3. Un teorema límite para ciertos procesos de Lévy 86
3.4. La ley arcoseno 91
Notas 94

Conclusión 95

Apéndice A. Algunas propiedades de los espacios métricos 97

Bibliografía 101

Abreviaturas y símbolos

càdlàg	Continua por la derecha y con límites por la izquierda.
(c.s. rel. $\mathbb{P}$)	Casi seguramente respecto a $\mathbb{P}$.
$\mathbf{1}_C$	La función indicadora del conjunto C.
$\mathbb{N}$	El conjunto de los números naturales.
$\mathbb{Q}$	El conjunto de los números racionales.
$\mathbb{R}$	El conjunto de los números reales.
$\overline{\mathbb{R}}$	$\mathbb{R} \cup -\infty, \infty$.
$[t]$	Mayor entero menor o igual a t.
$\left(\begin{array}{c} \alpha \\ n \end{array} \right)$	Para $\alpha \in \mathbb{R}$ y $n \in \mathbb{N}$, se define como $\frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}$.
$\wedge$	$a \wedge b$ es el mínimo de $\{a, b\}$.
Γ	La función Gamma.
$\sim$	$f \sim g$ si $\lim_{x\to\infty} (f(x)/g(x)) = 1$
F_X	Distribución de la variable aleatoria real X.
$\mathbb{P}$	Medida de probabilidad
$\mathbb{E}$	Esperanza respecto a $\mathbb{P}$.
$\mathrm{Var}(X)$	Varianza de la variable aleatoria X.
$\mathscr{C}_{[0,\infty)}$	El espacio de funciones continuas de $[0, \infty)$ en $\mathbb{R}$.
$\|f\|$	La norma uniforme de $f$sobre $[0,1]$:$\sup_{t\in[0,1]} \|f(t)\|$.
$\omega^T(f, \delta)$	Módulo de continuidad de f sobre $[0, T]$ valuado en δ.
$\mathscr{B}_M$	La σ-álgebra de Borel asociada al espacio topológico M.
$\sigma(\mathscr{C})$	σ-álgebra generada por $\mathscr{C}$.
$\mathscr{F}_1 \otimes \mathscr{F}_2$	σ-álgebra generada por $\{C_1 \times C_2 : C_i \in \mathscr{F}_i, i = 1, 2\}$.
$\mathbb{P}_1 \otimes \mathbb{P}_2$	Producto de las medidas de probabilidad $\mathbb{P}_1$ y $\mathbb{P}_2$.
$\mathbb{P}X^{-1}(A)$	Igual a $\mathbb{P}(X \in A)$.
$\Rightarrow$	Convergencia débil.
$\mathscr{P}(C)$	Conjunto potencia de C.
$B_\varepsilon(x)$	Bola de radio ε con centro en x.
$\partial(C)$	Frontera de C.
$\overline{C}$	Cerradura de C.
$\mathrm{Int}\, C$	Interior de C.

Introducción

En este trabajo se estudia un problema clásico dentro de la probabilidad: la fracción de tiempo que un proceso estocástico es positivo. Asociado a este problema están los nombres de grandes probabilistas de la primera mitad del siglo pasado, como P. Lévy, P. Erdös, M. Kac, W. Feller y M. Donsker, en el orden en el que constribuyeron a la solución del problema para cierta clase de procesos estocásticos: las caminatas aleatorias y su generalización a tiempo continuo, los procesos de Lévy. Sus nombres quedan unidos a través del desarrollo de un concepto fundamental dentro de la probabilidad, la convergencia débil, en la cual, el objetivo fundamental no es el calcular de manera exacta alguna probabilidad que deseamos conocer, sino aproximarla. Los inicios de esta teoría se remontan al descubrimiento del teorema de De-Moivre y Laplace, una versión del teorema límite central para la caminata aleatoria simple. Desde ahí hasta el desarrollo de la teoría de la convergencia en distribución para variables aleatorias con valores en un espacio métrico y su aplicación al teorema límite central funcional (también conocido como el principio de invariancia de Donsker) se ha recorrido un largo camino, que se ha servido del desarrollo de teorías muy importantes dentro de la matemática, principalmente del análisis matemático. En este sentido, este trabajo permite echar un vistazo a las técnicas que se utilizaban para el cálculo aproximado de probabilidades a principios del siglo pasado, así como a aquellas que se empezaron a utilizar a mediados del mismo. Sin embargo, cabe la siguiente aclaración: el problema que se propuso al inicio de este proyecto no fue motivado por la necesidad de hacer un recuento histórico del desarrollo de la convergencia débil, sino de dar explicación a un resultado que se obtiene al simular un proceso estocástico ligado a la teoría actuarial, situación que se expone a continuación.

Planteamiento del problema

Consideremos la siguiente situación, presentada en una compañía de seguros: Las reclamaciones, a partir del instante inicial 0, llegan a ella en los instantes $T_1 \leq T_2 \leq \cdots$, cada una por un monto de X_i. Si se define para $t \geq 0$

$$N_t = \begin{cases} 0 & \text{si } t \in [0, T_1) \\ k & \text{si } t \in [T_k, T_{k+1}) \end{cases},$$

entonces N_t denota la cantidad de reclamaciones que han llegado a la compañía hasta el instante t. Por otro lado, si $S_0 = 0$ y

$$S_n = \sum_{i=1}^{n} X_i$$

para $n \geq 1$, entonces S_n es la reclamación total cuando han ocurrido n siniestros, y por lo tanto

$$Y_t = S_{N_t}, \quad t \geq 0$$

representa la reclamación por concepto de siniestros que ha recibido la aseguradora en el instante t.

Los siguientes supuestos son usuales para algunas coberturas:

1. Los montos de los siniestros son independientes entre si y tienen la misma distribución. Esto es, $(X_i)_{i=1}^{\infty}$ son variables aleatorias independientes e idénticamente distribuidas.
2. $(N_t)_{t\geq 0}$ es un proceso Poisson de parámetro $\lambda > 0$.
3. Las familias de variables aleatorias

   $$\{X_i : i \geq 1\} \text{ y } \{N_t : t \geq 0\}$$

   son independientes.

Así, el proceso $(Y_t)_{t\geq 0}$ es un proceso Poisson compuesto. Si la vigencia de la póliza es de T unidades de tiempo y la prima se cobra al inicio del periodo, siendo igual a $\mathbb{E}(Y_T) = \lambda\mu T$, donde $\mu = \mathbb{E}(X_1)$, entonces al instante $t \in [0, T]$, la aseguradora cuenta con un capital de $\lambda\mu T - Y_t$ para hacer frente a las reclamaciones. A la esperanza de dicha cantidad se le conoce como la prima no devengada, que tiene el valor de $\lambda\mu(T - t)$, y es la cantidad que se tendría que devolver al asegurado en caso de una cancelación prematura de la póliza en el instante t. Es por esto que las compañías de seguros deben de tener esa cantidad constituida en una reserva llamada reserva de riesgos en curso, de acuerdo a la ley mexicana. Sin embargo, para una cierta realización del proceso Y, la cantidad $\lambda\mu T - Y_t$ podría ser menor a su media para $t \in [0, T]$, con lo cual la compañía de seguros estaría obligada a aportar capital adicional para constituir la reserva de riesgos en curso. En este trabajo nos avocamos a aproximar la distribución de la fracción del tiempo en el cual esto ocurre, es decir, a aproximar la distribución de la variable aleatoria

$$\frac{1}{T}\int_{[0,T]} \mathbf{1}_{(\lambda\mu T - Y_t < \lambda\mu(T-t))}\, dt = \frac{1}{T}\int_{[0,T]} \mathbf{1}_{(Y_t > \lambda\mu t)}\, dt,$$

bajo las anteriores hipótesis, además de una adicional: usualmente las aseguradoras ponen un límite a su responsabilidad, por lo que podemos suponer sin ningún problema que las variables X_i son acotadas.

Estructura de la tesis

En el primer capítulo, se estudia la versión discreta del problema presentado en la sección anterior, primero para el caso de la caminata aleatoria simple y después para caminatas aleatorias en general. En el segundo capítulo, se estudia el mismo problema para el movimiento browniano, mediante el uso de la teoría de la convergencia débil y de los resultados obtenidos en el primer capítulo. Finalmente, en el capítulo 3, se resuelve el problema planteado en la sección anterior, otra vez mediante argumentos de convergencia débil.

CAPÍTULO 1

Las caminatas aleatorias y la ley arcoseno

1.1. Introducción

En este capítulo estudiaremos un problema relacionado con el propuesto en la introducción. Primero en el caso más sencillo que se puede presentar, a saber, el de la caminata aleatoria simple. El estudio de este ejemplo, aparte de proporcionarnos una distribución importante, nos ayudará a encontrar la solución para el problema original. Seguido de esto, se presenta un estudio de un problema análogo pero para caminatas aleatorias más generales, para finalmente obtener algunos teoremas sobre límites débiles haciendo algunos supuestos sobre la distribución que gobierna los saltos de las caminatas aleatorias. En ambos casos, el estudio de las variables aleatorias involucradas se hace mediante análisis combinatorio y a través de las funciones generadoras de ciertas sucesiones de probabilidades, ya que estas funciones nos sirven para representar de manera más compacta a las sucesiones y nos ayudan a obtener ciertas relaciones que éstas satisfacen.

Una caminata aleatoria es uns sucesión de variables aleatorias reales $(S_n)_{n=0}^{\infty}$ tal que $S_0 = 0$ y las variables aleatorias $(S_i - S_{i-1})_{i=1}^{\infty}$ son independientes e idénticamente distribuidas. Si $X_i = S_i - S_{i-1}$, entonces a la distribución de X_i, que es independiente de i, y a la que daremos el nombre de distribución del salto, gobierna a la caminata aleatoria y las características de esta última se obtienen a través del estudio de dicha distribución. Si X_i tiene una distribución Bernoulli de parámetro 1/2, es decir

$$\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2},$$

a la caminata aleatoria se le conoce como caminata aleatoria simple y ciertas distribuciones relacionadas con esta sucesión se encuentran mediante métodos combinatorios; sin embargo, para el caso general, los métodos a utilizar generalmente son más complicados y el objetivo de su estudio es expresar la solución al problema planteado en términos de las caminatas aleatorias a través de la función de distribución. A veces tal solución es muy complicada y no es de utilidad práctica, por lo que además de la solución exacta, generalmente buscamos expresiones aproximadas más sencillas. Este enfoque se puede utilizar aún cuando la solución exacta a nuestro problema nos sea desconocida.

El problema que trataremos es el siguiente: dado un recorrido de longitud fija n de una caminata aleatoria $(S_i)_{i=0}^{\infty}$, calcular la distribución de la fracción de tiempo en el cual la caminata es positiva. Debemos precisar el significado de la frase anterior, puesto que no se trata de un proceso a tiempo continuo como el proceso Poisson. Una interpretación sencilla es la siguiente: si $\mathbf{1}_A$ denota a la función indicadora del conjunto A,[1] encontrar la distribución de

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{(S_n>0)},$$

que representa una medida discreta de la cantidad que nos interesa. Este problema se trata para caminatas aleatorias arbitrarias. Sin embargo, ya que la caminata aleatoria simple no cruza el eje del tiempo sin pasar por el cero, esto nos da la pauta para estudiar otra variable aleatoria que también se puede interpretar como la referida en el problema planteado al principio del párrafo: si definimos

$$Y_t = (1-(t-[t]))\,S_{[t]} + (t-[t])\,S_{[t]+1},$$

que se construye a partir de una caminata aleatoria $(S_n)_{n=0}^{\infty}$ interpolando linealmente entre los puntos (n, S_n), encontrar la distribución de

$$\frac{1}{n}\int_0^n \mathbf{1}_{(Y_t>0)}\,dt,$$

lo cual se hará en el caso de la caminata aleatoria simple.

1.2. La caminata aleatoria simple y la ley arcoseno

Como se mencionó anteriormente, lo primero que haremos será encontrar la distribución de la variable aleatoria

$$F_n = \frac{1}{n}\int_0^n \mathbf{1}_{(Y_t>0)}\,dt,$$

donde la caminata aleatoria es simple, para lo cual utilizaremos el que esta caminata aleatoria no cambia de signo sin tomar el valor 0, y que al tomar el valor cero, la caminata vuelve a empezar. Lo que haremos será análogo al análisis del primer paso: nos fijaremos en el primer regreso a 0 dentro de los primeros n pasos (o en n, si la caminata aleatoria no se hace cero en los primeros n pasos.) Es por esto que el primer paso hacia la solución del problema será encontrar la distribución del primer regreso a cero para la caminata aleatoria simple.

Se decidió presentar el método clásico, que consiste en hacer análisis combinatorio, así como el método de la función generadora. Este último nos permitirá conocer la distribución de la variable aleatoria F_n, puesto que un análisis combinatorio se vuelve más complicado. Para la caminata aleatoria simple, nos aprovechamos de la solución exacta como función de n para encontrar una distribución límite que nos permite aproximar las probabilidades asociadas a la distribución de F_n mediante el cálculo de una integral. Es

[1] No se utiliza el nombre de función característica, pues éste tiene otra acepción dentro de la teoría de la probabilidad.

importante que para la caminata aleatoria simple, podemos calcular explícitamente la distribución de F_n.

1.2.1. El primer regreso a cero.

1.2.1.1. Método Combinatorio. Notemos que la evolución de la caminata aleatoria hasta el tiempo n queda determinada por el camino que sigue. Definimos a un camino como un conjunto de puntos

$$\{(i, j_i) : i = 0 \dots n\} \subset \mathbb{N} \times \mathbb{Z},$$

para el cual

$$j_i = j_{i-1} \pm 1,$$

con $i = 2 \dots n$.[2] Así, para encontrar la probabilidad de que la evolución de la caminata aleatoria simple hasta el tiempo n tenga cierta propiedad, una posibilidad es encontrar el número de caminos para los cuales se cumple dicha propiedad y multiplicar por $1/2^n$, que es la probabilidad de que la caminata siga un camino específico durante n pasos.[3] Éste es un método combinatorio.

Denotaremos por T el tiempo del primer regreso a cero y por $C_n(p)$ el número de caminos de n pasos con la propiedad p. Es claro que el primer regreso a cero ocurre necesariamente en un número par de realizaciones y que es mayor o igual a 2. Por otro lado,

$$\mathbb{P}(T = n) = C_n(S_1 \neq 0, \dots S_{n-1} \neq 0, S_n = 0)\,/2^n, \quad n \geq 2.$$

Para calcular la probabilidad del evento $\{T = n\}$, sólo debemos de contar el número de caminos que no tocan cero más que al tiempo cero y al tiempo n. Como cualquiera de estos caminos yace íntegramente por arriba del eje del tiempo (eje de la abscisa) o por debajo del mismo, y más aún, como el número de caminos que yace por debajo del eje del tiempo es igual al número de caminos por encima del mismo, surge la igualdad

$$C_n(S_i \neq 0, i = 1 \dots n-1, S_n = 0) = 2C_n(S_i > 0, i = 1 \dots n-1, S_n = 0)\,.$$

Notemos que a cada camino de n pasos

$$\{(i, j_i) : i = 0 \dots n\}$$

para el cual

1. $j_i > 0$, $i = 1 \dots n-1$ y
2. $j_n = 0$

[2] En la pareja ordenada (i, j_i) perteneciente a un camino, i es el número de pasos y j_i es la posición de la caminata aleatoria.

[3] Pues es igual al inverso de la cantidad de caminos que la caminata puede seguir y bajo las condiciones impuestas, los caminos a seguir son equiprobables.

le podemos asociar un camino de $n-2$ pasos

$$\{(i, j'_i) : i = 0 \dots n-2\}$$

tal que

1. $j'_0 = 1$,
2. $j'_i > 0$ para $i = 1 \dots n-3$ y
3. $j'_{n-2} = 1$

(haciendo $j'_i = j_{i+1}$) y que esta asociación es biunívoca, de donde

$$C_n(S_i > 0, i = 1 \dots n, S_n = 0) = C_{n-2}(S_0 = 1, S_i > 0, i = 1 \dots n-3, S_{n-2} = 1).$$

Consideremos ahora el total de caminos que empiezan y terminan en 1 y constan de $n-2$ pasos, éste será la suma del número de caminos positivos que empiezan y terminan en 1 y constan de $n-2$ pasos y el número de caminos que empiezan y terminan en 1, que constan de $n-2$ pasos y que en algún momento tocan el eje del tiempo, esto es

$$\begin{aligned} C_{n-2}(S_0 = 1, S_{n-2} = 1) &= C_{n-2}(S_0 = 1, S_i > 0, i = 1 \dots n-3, S_{n-2} = 1) \\ &+ C_{n-2}(S_0 = 1, S_{n-2} = 1, \text{ existe } i \in \{1, \dots, n-3\} \text{ con } S_i = 0). \end{aligned} \tag{1}$$

Podemos calcular el lado izquierdo de la ecuación anterior y mediante un argumento combinatorio, llamado Principio de Reflexión, podemos calcular el segundo sumando del lado derecho y obtener mediante su diferencia, el número de caminos que nos interesa. Vamos por pasos:

1. En general, para calcular $C_{n-2}(S_0 = k, S_{n-2} = l)$, consideremos cualquier camino que satisfaga la propiedad anterior, digamos $\{(i, j_i) : i = 0, \dots, n-2\}$. Al considerar los incrementos $j_i - j_{i-1} \in \{-1, 1\}$ para $i = 1, \dots, n-2$, vemos que la cantidad de incrementos iguales a 1 debe ser igual a $((l-k) + (n-2))/2$,[4] mientras que la cantidad de incrementos iguales a -1 debe ser igual a $n-2$ menos la anterior cantidad. Esto es porque la diferencia entre la cantidad de incrementos iguales a 1 y la cantidad de incrementos iguales a -1 debe ser igual a $k-l$ y la suma de las dos cantidades anteriores debe ser igual a $n-2$. Por otro lado, podemos recuperar al camino a partir de los incrementos, de manera única. Esto quiere decir que $C_{n-2}(S_0 = k, S_{n-2} = l)$ es igual a la cantidad de vectores $(x_1, \dots, x_{n-2}) \in \{-1, 1\}^{n-2}$ (en los que pensaremos como los incrementos) para las cuales la cantidad de entradas iguales a 1 sea igual $((l-k) + (n-2))/2$ y la cantidad de entradas iguales a -1 sea igual a $n-2$ menos la cantidad anterior, que es igual a la cantidad de formas en las que podemos dividir a un conjunto de $n-2$ elementos en dos subconjuntos complementarios, uno de ellos con cardinalidad $((l-k) + (n-2))/2$. Esto implica que $((l-k) + (n-2))/2$ debe ser

[4] Si $((l-k) + (n-2))/2$ no es un entero no negativo, entonces $C_{n-2}(S_0 = k, S_{n-2} = l) = 0$.

un entero positivo para que $C_{n-2}(S_0 = k, S_{n-2} = l)$ sea distinto de cero. En este caso, obtenemos la igualdad

(2)
$$C_{n-2}(S_0 = k, S_{n-2} = l) = \binom{n-2}{\frac{(l-k)+(n-2)}{2}}.$$

Por medio de (2), se llega a la igualdad

(3)
$$\mathbb{P}(S_n = l) = C_n(S_0 = 0, S_n = l)\,/2^n = \binom{n}{\frac{l+n}{2}}/2^n$$

cuando $(l+n)/2$ es un entero no negativo y $\mathbb{P}(S_n = l)$ es cero en otro caso.

2. Por otro lado, para calcular

$$C_{n-2}(S_0 = 1, S_{n-2} = 1, \text{ existe } i \in \{1, \ldots, n-3\} \text{ con } S_i = 0)\,,$$

sea $\{(i, j_i) : i = 0 \ldots n-2\}$ un camino de 1 a 1 que toca el eje del tiempo, esto es,

a) $j_0 = 1 = j_{n-2}$ y
b) existe $i \in \{1, \ldots, n-3\}$ tal que $j_i = 0$.

Consideremos

$$k = \min\{m \in \{1, \ldots, n-3\} : j_m = 0\}\,,$$

este último conjunto es no vacío ya que i pertenece a él. Si ahora reflejamos el camino con respecto al eje del tiempo antes del momento k, obtenemos un camino que va de -1 a 1 en $n-2$ pasos, como se puede apreciar en la Figura 1. Hemos considerado la transformación

$$\{(i, j_i) : i = 0 \ldots n-2\} \mapsto \{(i, j_i') : i = 0 \ldots n-2\}\,,$$

donde

a) $j_i' = -j_i$ si $i < k$ y
b) $j_i' = j_i$ si $i \geq k$,

por lo que

a) $j_0' = -j_0 = -1$ pues $k > 0$ y
b) $j_{n-2}' = j_{n-2}$ ya que $k < n-2$

y para verificar que es un camino, notemos que

$$j_{k+1}' = j_{k+1} = j_k \pm 1 = \pm 1 = j_k' \pm 1$$

y como

$$j_k = j_k' = 0 \text{ y } j_{k-1} = 1,$$

entonces

$$j_{k-1}' = -1 \text{ y } j_k = j_{k-1}' + 1.$$

Esta asociación es biunívoca, pues dado un camino que va de -1 a 1, podemos considerar el primer instante en el que toca cero y aplicar el mismo procedimiento de reflexión para obtener un camino de 1 a 1 que toca cero, notando que al aplicar

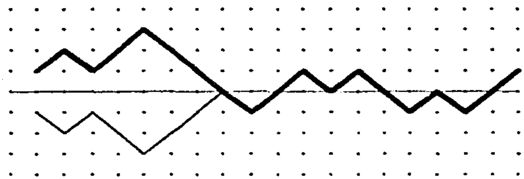

FIGURA 1. Ilustración del Principio de Reflexión

ambas asociaciones, la primera seguida de la segunda o la segunda seguida de la primera, obtenemos el camino con el que empezamos.

Así, hemos visto que

$$C_{n-2}(S_0 = 1, S_{n-2} = 1, \text{ existe } i \in \{1, \ldots, n-3\} \text{ con } S_i = 0) = C_{n-2}(S_0 = -1, S_{n-2} = 1).$$

Para que $C_{n-2}(S_0 = -1, S_{n-2} = 1)$ sea mayor a cero es necesario que $n/2$ sea un entero mayor o igual a 2, de acuerdo al inciso anterior. En ese caso, de acuerdo a (2), se obtiene

$$C_{n-2}(S_0 = 1, S_{n-2} = 1, \text{ existe } i \in \{1, \ldots, n-3\} \text{ con } S_i = 0) = \binom{n-2}{\frac{n}{2}},$$

de lo cual, al considerar (1) y simplificar la expresión resultante:

$$\begin{aligned} C_{n-2}(S_0 = 1, S_i > 0, i = 1 \ldots n-3, S_{n-2} = 1) &= \binom{n-2}{\frac{n-2}{2}} - \binom{n-2}{\frac{n}{2}} \\ &= \frac{2}{n}\binom{n-2}{\frac{n-2}{2}} \end{aligned}$$

y así, finalmente obtenemos, para n para,

$$\mathbb{P}(T = n) = \frac{2}{2^n}\frac{2}{n}\binom{n-2}{\frac{n-2}{2}}.$$

Para concluir con el método combinatorio, se pide al lector comprobar la igualdad

$$\frac{2}{2^n}\frac{2}{n}\binom{n-2}{\frac{n-2}{2}} = \frac{1}{n}\mathbb{P}(S_{n-2} = 0),$$

válida para $n \geq 2$ par como una consecuencia de (3). Esta igualdad nos relaciona la densidad discreta de T con la de S_n y se obtendrá de nueva cuenta a continuación.

1.2.1.2. Método de la Función Generadora. Sean

$$u_n = \mathbb{P}(S_n = 0)$$

y

$$f_n = \mathbb{P}(T = n)$$

para $n \in \mathbb{N}$. Notemos que $u_0 = 1$, que $u_n = f_n = 0$ si n es impar y que para $n \geq 1$,

$$u_{2n} = \sum_{i=1}^{n} f_{2i} u_{2(n-i)},$$

puesto que $S_{2n} = 0$ ($n \geq 1$) si y sólo si existe $i \in \{1, \ldots, n\}$ tal que el primer regreso a cero ocurre en el instante $2i$ y de ahí, independientemente de lo ocurrido anteriormente, debemos regresar a cero en $2(n-i)$ pasos. Si

$$F(s) = \sum_{n=1}^{\infty} f_{2n} s^{2n}$$

es la función generadora de $\{f_k\}_{k\in\mathbb{N}}$ y

$$P(s) = \sum_{n=0}^{\infty} u_{2n} s^{2n}$$

es la de $\{u_k\}_{k\in\mathbb{N}}$, entonces

$$\begin{aligned} P(s) &= 1 + \sum_{n=1}^{\infty} u_{2n} s^{2n} \\ &= 1 + \sum_{n=1}^{\infty} s^{2n} \sum_{i=1}^{n} f_{2i} u_{2(n-i)} \end{aligned}$$

Notemos que la última serie en la ecuación anterior es el producto de $P(s)$ y $F(s)$, de donde $P(s) = 1 + P(s)F(s)$.

Para calcular $P(s)$, utilizamos (3) y la definición de u_n para obtener:

$$\begin{aligned} u_{2n} &= \frac{1}{2^{2n}} \binom{2n}{n} = \frac{1}{2^n n!} \frac{\prod_{i=0}^{n-1}(2i+1) \prod_{i=1}^{n}(2i)}{2^n \prod_{i=1}^{n} i} \\ &= \frac{1}{2^n n!} \prod_{i=0}^{n-1}(2i+1) = \frac{(-1)^n}{n!} \prod_{i=0}^{n-1}(-1/2 - i). \end{aligned}$$

Si se define para a un número real arbitrario,

$$\binom{a}{n} = \frac{a(a-1)\cdots(a-n+1)}{n!},$$

entonces

$$u_{2n} = (-1)^n \binom{-1/2}{n}.$$

Podemos utilizar la serie binóminca, que se puede consultar en [1], para obtener

$$P(s) = \sum_{n=0}^{\infty} u_{2n}s^{2n} = \sum_{n=0}^{\infty} s^{2n}(-1)^n \binom{-1/2}{n} = (1-s^2)^{-1/2}$$

de donde,

$$F(s) = 1 - (1-s^2)^{1/2}.$$

Si utilizamos de nueva cuenta la serie binómica, notemos que

$$\begin{aligned} F(s) &= 1 - \sum_{n=0}^{\infty}(-1)^n s^{2n} \binom{1/2}{n} \\ &= \sum_{n=1}^{\infty}(-1)^{n+1} s^{2n} \binom{1/2}{n}, \end{aligned}$$

pero al analizar con más detalle, se observa que

$$\begin{aligned} (-1)^{n+1}\binom{1/2}{n} &= \frac{(-1)^{n+1}}{n!}\prod_{i=0}^{n-1}(1/2-i) \\ &= \frac{-1}{n!2^n}\prod_{i=0}^{n-1}(2i-1) \\ &= \frac{1}{n!2^n}\prod_{i=1}^{n-1}(2i-1) \\ &= \frac{1}{n!2^n}\prod_{i=1}^{n-1}(2i-1)\frac{\prod_{i=1}^{n-1}2i}{(n-1)!2^{n-1}} \\ &= \frac{1}{2^{2n-1}n}\frac{(2(n-1))!}{(n-1)!^2} \\ &= \frac{1}{2n}u_{2(n-1)}, \end{aligned}$$

donde la última igualdad es consecuencia de la definición de u_n y de la expresión que se obtiene para u_n al utilizar (3). Esto significa que

$$\sum_{n=1}^{\infty} s^{2n} f_{2n} = F(s) = \sum_{n=1}^{\infty} s^{2n}\frac{1}{2n}u_{2(n-1)},$$

de lo cual se infiere la igualdad $f_{2n} = \frac{1}{2n}u_{2(n-1)}$, que coincide con la expresión que encontramos previamente.

1.2.2. La ley Arcoseno. Ahora calcularemos la distribución de F_n, la variable aleatoria definida en la Sección 1.2, aunque sólo sea para n par, puesto que utilizaremos al primer regreso a cero para la caminata aleatoria simple y esta variable siempre es par. Además, como veremos, la distribución de F_n se vuelve muy complicada cuando n es grande, por lo que el hecho de que las distribuciones de F_n converjan a una distribución cuando n tiende a infinito se vuelve muy conveniente. Estos son los dos aspectos de la caminata aleatoria simple que exploraremos a continuación, el primero mediante el uso de funciones generadoras y el segundo mediante la aproximación de Stirling, que si recordamos, es la que nos proporciona una aproximación para el factorial, de donde se obtiene la aproximación normal para la distribución binomial de parámetros n y p cuando n es grande, distribución relacionada con la de S_n, puesto que $(S_n + n)/2$ tiene distribución binomial de parámetros n y $p = 1/2$. Esto se puede verificar ya que

$$(S_n + n)/2 = \sum_{i=1}^{n} (S_i - S_{i-1} + 1)/2,$$

donde $\{(S_i - S_{i-1} + 1)/2\}_{i=1}^{n}$ son independientes y su distribución común es Bernoulli de parámetro $1/2$.

1.2.2.1. La distribución Arcoseno discreta. Calcularemos ahora la probabilidad de que la caminata aleatoria se encuentre $2k$ unidades de tiempo en el lado positivo en un recorrido de $2n$ pasos. Diremos que la caminata pasa una unidad de tiempo del lado positivo en el intevalo $[m, m+1]$ si S_m es positivo o S_{m+1} es positivo. Con esta definición, la probabilidad que deseamos calcular es $\mathbb{P}(F_{2n} = k/n)$. Además, una caminata de $2n$ pasos se encuentra durante $2k$ unidades de tiempo en el lado positivo si y sólo si su reflexión a lo largo del eje del tiempo pasa $2(n-k)$ unidades de tiempo en el lado positivo (o la caminata pasa $2(n-k)$ unidades de tiempo en el lado negativo.) Si denotamos por $u_{2k,2n}$ a la probabilidad que nos interesa, la anterior afirmación se reduce a la igualdad $u_{2k,2n} = u_{2(n-k),2n}$, por lo que la densidad discreta de la cantidad de tiempo en el que la caminata aleatoria simple es positiva es simétrica alrederor del punto n.

Si una caminata de $2n$ pasos se encuentra durante $2k$ unidades de tiempo en el lado positivo y $0 < k < n$, entonces necesariamente pasa por el cero, digamos en el momento $2i$. Si S_1 es positivo, se sigue que la caminata debe pasar $2(k-i)$ unidades de tiempo en el lado positivo entre el momento $2i$ y el $2n$, lo cual sucede con probabilidad

$$\frac{1}{2} f_{2i} u_{2(k-i),2(n-i)},$$

o si S_1 es negativo, debe pasar $2k$ unidades de tiempo en el lado positivo entre los instantes $2i$ y $2n$, que sucede con probabilidad

$$\frac{1}{2} f_{2i} u_{2k,2(n-i)}.$$

Aquí ocupamos el que la caminata aleatoria simple vuelva a empezar la primera vez que llega a cero. Utilizando la aditividad finita de la probabilidad, ésto da lugar a la relación

$$u_{2k,2n} = \frac{1}{2}\sum_{i=1}^{n} f_{2i}u_{2(k-i),2(n-i)} + \frac{1}{2}\sum_{i=1}^{n} f_{2i}u_{2k,2(n-i)} \tag{4}$$

válida para $0 < k < n$ e incorrecta para el caso $k = 0$ ó $k = n$, puesto que no contempla la posibilidad de que la caminata regrese a cero por primera vez en un instante posterior al $2n$. Esto se remedia facilmente, pues una caminata de $2n$ pasos se encuentra durante cero unidades de tiempo en el lado positivo sin regresar a cero en ese lapso con probabilidad $\frac{1}{2}\sum_{k=1}^{\infty} f_{2(n+k)}$. Análogamente, se tiene una expresión idéntica para el caso en el que pasa $2n$ unidades de tiempo en el lado positivo y así, se obtiene:

$$u_{0,2n} = \frac{1}{2}\sum_{i=1}^{n} f_{2i}u_{-2i,2(n-i)} + \frac{1}{2}\sum_{i=1}^{n} f_{2i}u_{0,2(n-i)} + \frac{1}{2}\sum_{k=1}^{\infty} f_{2(n+k)}$$

$$u_{2n,2n} = \frac{1}{2}\sum_{i=1}^{n} f_{2i}u_{2(n-i),2(n-i)} + \frac{1}{2}\sum_{i=1}^{n} f_{2i}u_{2n,2(n-i)} + \frac{1}{2}\sum_{k=1}^{\infty} f_{2(n+k)}.$$

Notemos que el primer sumando del lado derecho de la primera igualdad es cero, así como el segundo sumando del lado derecho de la segunda igualdad, pero se incluyen para preservar cierta analogía con el caso $0 < k < n$.

Si ahora consideramos

$$U_{2n}(s) = \sum_{k=0}^{n} s^{2k}u_{2k,2n},$$

notemos que de la relación (4) se tiene que

$$\begin{aligned} U_{2n}(s) &= \frac{1}{2}\sum_{k=0}^{n}\sum_{i=1}^{n} s^{2k} f_{2i}u_{2(k-i),2(n-i)} + \frac{1}{2}\sum_{k=0}^{n}\sum_{i=1}^{n} s^{2k} f_{2i}u_{2k,2(n-i)} \\ &\quad + \frac{1}{2}(1+s^{2n})\sum_{k=1}^{\infty} f_{2(n+k)} \\ &= \frac{1}{2}\sum_{i=1}^{n} s^{2i} f_{2i}U_{2(n-i)}(s) + \frac{1}{2}\sum_{i=1}^{n} f_{2i}U_{2(n-i)}(s) \\ &\quad + \frac{1}{2}(1+s^{2n})\sum_{k=1}^{\infty} f_{2(n+k)}, \end{aligned} \tag{5}$$

donde la última igualdad se obtiene al cambiar el orden de las dos sumas.

Calculemos ahora la función generadora conjunta $G(s,t)$ de $\{u_{2k,2n}\}$ valuada en (s,t), igual a

$$G(s,t) = \sum_{n=1}^{\infty}\sum_{k=0}^{n} u_{2k,2n} s^{2k} t^{2n}.$$

La función G valuada en (s,t) no es más que la función generadora de $\{U_{2n}(s)\}_{n\in\mathbf{N}}$ valuada en t (definiendo $U_0 = 0$), que podemos encontrar al multiplicar (5) por t^{2n} y sumar n desde 1 hasta ∞:

$$\begin{aligned}
(6) \qquad 2G(s,t) &= 2\sum_{n=1}^{\infty} t^{2n} U_{2n}(s) \\
&= \sum_{n=1}^{\infty} t^{2n} \sum_{i} s^{2i} f_{2i} U_{2(n-i)}(s) + \sum_{n=1}^{\infty} t^{2n} \sum_{i=1}^{n} f_{2i} U_{2(n-i)}(s) \\
&+ \sum_{n=1}^{\infty} t^{2n} \sum_{k=1}^{\infty} f_{2(n+k)} + \sum_{n=1}^{\infty} t^{2n} s^{2n} \sum_{k=1}^{\infty} f_{2(n+k)}.
\end{aligned}$$

La primera suma que aparece en el lado derecho de la última igualdad en (6) es igual al producto de la función generadora de $\{f_{2n}\}$ valuada en st (que denotaremos por $F(st)$) con la función generadora de $\{U_{2n}(s)\}$ valuada en t (que es igual a $G(s,t)$). El segundo sumando del lado derecho de la última igualdad en (6) corresponde al producto de la función generadora de $\{f_n\}$ valuada en t y la generadora de $\{U_n(s)\}$ valuada en t (igual a $G(s,t)$). El tercer sumando se calcula a continuación:

$$\begin{aligned}
\sum_{n} t^{2n}(f_{2(n+1)} + f_{2(n+2)} + \cdots) &= \sum_{n\in\mathbf{N}} \sum_{k=n+1}^{\infty} t^{2n} f_{2k} \\
&= \sum_{k\in\mathbf{N}} \sum_{n=0}^{k-1} t^{2n} f_{2k} \\
&= \sum_{k\in\mathbf{N}} \frac{1-t^{2k}}{1-t^2} f_{2k} \\
&= \frac{1-F(t)}{1-t^2}.
\end{aligned}$$

Se ha utilizado el hecho de que $\sum_{n\in\mathbf{N}} f_{2n} = 1$, igualdad que se justifica pues la anterior suma es igual a $F(1)$, donde F es la función generadora obtenida en la subsección anterior. La cuarta suma es parecida a la tercera, pero debemos substituir t por st. Así, hemos obtenido la relación:

$$2G(s,t) = F(st)G(s,t) + F(t)G(s,t) + \frac{1-F(t)}{1-t^2} + \frac{1-F(st)}{1-s^2t^2},$$

de donde

$$G(s,t) = (1-t^2)^{-1/2}(1-s^2t^2)^{-1/2}.$$

Finalmente, la probabilidad $u_{2k,2n}$ es el coeficiente de $s^{2k}t^{2n}$ en la expansión en serie de potencias de la función G, que se puede obtener al utilizar la serie binómica:

$$\begin{aligned} G(s,t) &= \left(\sum_{n=0}^{\infty} t^{2n} \binom{-1/2}{n} (-1)^n\right) \left(\sum_{n=0}^{\infty} s^{2n} t^{2n} \binom{-1/2}{n} (-1)^n\right) \\ &= \sum_{n=0}^{\infty} t^{2n} \sum_{k=0}^{n} s^{2k} \binom{-1/2}{k} \binom{-1/2}{n-k}. \end{aligned}$$

Este último resultado nos dice que la probabilidad buscada $u_{2k,2n}$ es igual a

$$(-1)^n \binom{-1/2}{k} \binom{-1/2}{n-k}$$

y como comprobamos en la sección anterior la igualdad

$$(-1)^k \binom{-1/2}{k} = \mathbb{P}(S_{2k} = 0),$$

podemos finalmente encontrar una expresión para la probabilidad de que la caminata aleatoria pase $2k$ unidades de tiempo en el lado positivo en un trayecto de $2n$ pasos:

(7) $$u_{2k,2n} = \mathbb{P}(S_{2k} = 0)\, \mathbb{P}(S_{2(n-k)} = 0).$$

De aquí que la fracción de tiempo que nuestra caminata aleatoria de longitud $2n$ pasa en el lado positivo (la variable aleatoria F_{2n}) tiene la distribución arcoseno discreta,[5] dada por

$$u_{2k,2n} = \mathbb{P}\left(F_{2n} = \frac{k}{n}\right) = \mathbb{P}(S_{2k} = 0)\, \mathbb{P}(S_{2(n-k)} = 0).$$

1.2.2.2. Convergencia a la distribución arcoseno. Así como la fórmula de Stirling nos provee de una aproximación para probabilidades relacionadas con la distribución binomial, al aproximar la función factorial, que toman forma en el teorema límite central para la caminata aleatoria simple, también la podemos utilizar para aproximar a la familia de distribuciones que encontramos en el inciso anterior. La fórmula de Stirling establece que

$$n! \sim \sqrt{2\pi}\, n^{n+1/2} e^{-n},$$

cuyo significado es que el cociente de ambas cantidades converge a 1 cuando $n \to \infty$, de lo cual se sigue:

$$\mathbb{P}(S_{2k} = 0) = \binom{2k}{k} 2^{-2k} \sim \frac{1}{\sqrt{\pi n}\sqrt{k/n}}.$$

[5] La razón por la cual se le da este nombre quedará más clara en la siguiente subsección.

por lo cual para $\varepsilon > 0$ existe $N_1 \in \mathbb{N}$ tal que $k, n-k \geq N_1$ implica que

$$(1-\varepsilon)\frac{1}{\pi n\sqrt{k/n}\sqrt{1-k/n}} \leq \mathbb{P}\left(F_{2n} = \frac{k}{n}\right) \leq (1+\varepsilon)\frac{1}{\pi n\sqrt{k/n}\sqrt{1-k/n}}.$$

Si $x, y \in (0,1)$ son tales que $x < y$, entonces

$$\mathbb{P}(F_{2n} \in (x,y)) = \sum_{\{k: k/n \in (x,y)\}} \mathbb{P}\left(F_{2n} = \frac{k}{n}\right)$$

y si $N \geq \max(N_1 x, N_1(1-y))$ y $n \geq N$, entonces $k/n \in (x,y)$ implica que $k \geq N_1$ y $n-k \geq N_1$, por lo que

$$\begin{aligned}(1-\varepsilon) \sum_{\{k: k/n \in (x,y)\}} \frac{1}{n}\frac{1}{\pi\sqrt{(k/n)(1-k/n)}} &\leq \mathbb{P}(F_{2n} \in (x,y)) \\ &\leq (1+\varepsilon) \sum_{\{k: k/n \in (x,y)\}} \frac{1}{n}\frac{1}{\pi\sqrt{(k/n)(1-k/n)}}.\end{aligned}$$

Esto implica que

$$\mathbb{P}(F_{2n} \in (x,y))$$

y

$$\sum_{\{k: k/n \in (x,y)\}} \frac{1}{n}\frac{1}{\pi\sqrt{(k/n)(1-k/n)}}$$

tienen el mismo límite cuando $n \to \infty$, que es $\int_{[x,y]} 1/\left(\pi\sqrt{t(1-t)}\right) dt$, puesto que la última expresión es casi una suma de Riemman de la función $1/\left(\pi\sqrt{t(1-t)}\right)$ para una partición de $[x,y]$ de norma menor o igual a $1/n$, donde la palabra casi se explica pues pueden faltar el primer y el último sumando (que son irrelevantes para el límite.) Hemos demostrado entonces que si $x, y \in (0,1)$ con $x < y$, entonces tiene lugar la relación

$$\lim_{n\to\infty} \mathbb{P}(F_{2n} \in (x,y)) = \int_{[x,y]} \frac{1}{\pi\sqrt{t(1-t)}}\, dt.$$

La validez de esta ecuación para $x = 0$ y $y \in (0,1)$ se demuestra mediante la simetría de la densidad discreta de F_{2n} y de la función $1/\sqrt{t(1-t)}$ respecto a $1/2$, así como la igualdad (que se demuestra en el siguiente párrafo)

$$\int_{[0,1]} \frac{1}{\pi\sqrt{t(1-t)}}\, dt = 1$$

pues entonces para $y \in (0, 1/2)$:

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(F_{2n} \in (0, y)) &= \lim_{n\to\infty} 1/2 - 1/2\mathbb{P}(F_{2n} \in (y, y + 1/2)) \\ &= 1/2 - 1/2 \int_{[y,y+1/2]} \frac{1}{\pi\sqrt{t(1-t)}}\, dt \\ &= \int_{[0,y]} \frac{1}{\pi\sqrt{t(1-t)}}\, dt.\end{aligned}$$

Para $y \in [1/2, 1]$, se aplica el resultado anterior a $y - 1/2$ y se utilizan las simetrías mencionadas anteriormente. En resumen, se tiene el siguiente teorema límite para la fracción de tiempo en el lado positivo, F_{2n}:

$$\lim_{n\to\infty} \mathbb{P}(F_{2n} \leq x) = \int_{[0,x]} \frac{1}{\pi\sqrt{t(1-t)}}\, dt, \quad x \in [0, 1].$$

Si A es una variable aleatoria continua con densidad $1_{[0,1]}(t)/\left(\pi\sqrt{t(1-t)}\right)$, el resultado anterior nos dice que F_{2n} converge a A en distribución. A la ley de la variable A se le llama distribución arcoseno, lo que explica el nombre de distribución arcoseno discreta para la ley de F_{2n}.

Para calcular de manera explícita la distribución arcoseno, utilizamos la substitución $s = t - 1/2$, que para $x \in [0, 1]$ nos dice que:

$$\int_{[0,x]} \frac{1}{\pi\sqrt{t(1-t)}}\, dt = \int_{[-1/2,x-1/2]} \frac{1}{\pi\sqrt{1/4 - s^2}}\, ds$$

que es igual a $1/2 + \arcsin(2x - 1)/\pi$ (esta expresión se encuentra al utilizar la substitución $s = \sin(r)/2$.) Note que esto en particular nos dice que

$$\int_{[0,1]} \frac{1}{\pi\sqrt{t(1-t)}}\, dt = \frac{\arcsin(1) - \arcsin(-1)}{\pi} = 1.$$

Podemos encontrar una expresión más sencilla para la distribución arcoseno al utilizar la formula de adición para la función seno, que establece la igualdad $\sin(a + b) = \sin(a)\cos(b) + \sin(b)\cos(a)$, así como el caso particular $\sin(2x) = 2\cos(x)\sin(x)$, y que $\cos\arcsin(x) = \sqrt{1 - x^2}$ para $x \in [0, 1]$, pues entonces para x en el mismo rango:

$$\sin(\arcsin(2x - 1) + \pi/2) = \cos\arcsin(2x - 1) = 2\sqrt{x}\sqrt{(1 - x)} = \sin 2\arcsin(\sqrt{x}),$$

de donde

$$\frac{\arcsin(2x - 1)}{\pi} + \frac{1}{2} = \frac{2}{\pi}\arcsin(\sqrt{x}).$$

La última expresión es importante no sólo por ser más sencilla sino por razones históricas, es la forma en la que Paul Lévy presentó a esta distribución límite en relación a un problema parecido al que nosotros consideramos.

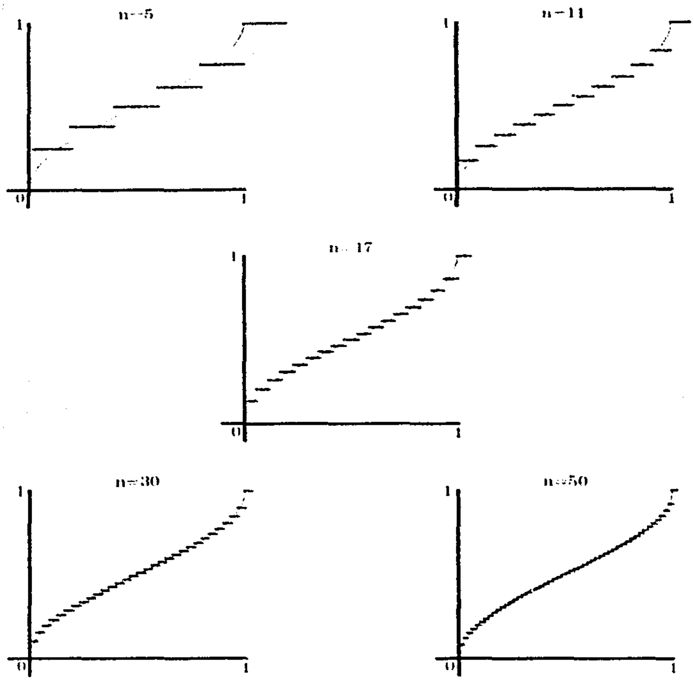

FIGURA 2. Ilustración del teorema límite para la caminata aleatoria simple

En la Figura 2 se puede observar la convergencia de la distribución arcoseno discreta hacia la distribución arcoseno. Se ha gráficado la distribución arcoseno discreta para los valores de $n = 5, 11, 17, 30$ y 50 junto con la distribución arcoseno. Como la densidad arcoseno se concentra en los puntos 0 y 1, lo mismo sucede con la densidad discreta de la variable aleatoria F_{2n} para n grande.

1.3. Cantidad de elementos positivos en caminatas aleatorias

El objetivo de esta sección es investigar la distribución de la cantidad de elementos positivos en los primeros n pasos de una caminata aleatoria $(S_n)_{n\in\mathbb{N}}$ y como en la sección anterior, investigar si existen condiciones suficientes en términos de la distribución del salto de la caminata aleatoria para la existencia del límite débil para la fracción de elementos positivos en los primeros n pasos de una caminata aleatoria conforme n se hace grande. Esto se hace mediante la identidad de Sparre Andersen, la cual estipula que la primera variable aleatoria mencionada tiene la misma distribución que el índice en el cual ocurre por primera vez el máximo de las sumas parciales de las n variables aleatorias. Las técnicas que se utilizan para estudiar esta última variable aleatoria nos proveen, en consecuencia, de resultados acerca de la primera. El material de esta sección proviene de [7].

1.3.1. Identidad de Eric Sparre Andersen. Sean $(X_i)_{i=1}^n$ variables aleatorias independientes con distribución común F. Definimos $S_0 = 0$ y para $j \geq 1$, $S_j = \sum_{i=1}^j X_i$, que conforman la sucesión de sumas parciales asociada (la caminata aleatoria que consideraremos.) Si $\pi_n = \sum_{i=1}^n \mathbf{1}_{(S_i>0)}$, esta variable representa la cantidad de sumas parciales positivas asociadas a $(X_i)_{i=1}^n$. Para proceder con su estudio demostraremos el siguiente resultado, para el cual consideraremos las permutaciones de un vector $x \in \mathbb{R}^n$. Recordemos que un vector en $\mathbb{R}^n$ es una n-ada ordenada $x = (x_1, \ldots, x_n)$, por lo que si $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ es biyectiva, podemos considerar el vector $x^\sigma = (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$, al cual nos referimos como una permutación de x. Recordemos que hay $n!$ funciones biyectivas de $\{1, \ldots, n\}$ en si mismo.

LEMA 1.1. *Sean $(x_1, \ldots, x_n)$ números reales arbitrarios. Si a cada permutación x^ν*

$$(x_{\nu(1)}, \ldots, x_{\nu(n)})$$

de $(x_1, \ldots, x_n)$ le asociamos las $n+1$ sumas parciales

$$(0, x_{\nu(1)}, x_{\nu(1)} + x_{\nu(2)}, \ldots, x_{\nu(1)} + \cdots + x_{\nu(n)}),$$

entonces el número de permutaciones para las cuales la sucesión de sumas parciales tiene r términos estrictamente positivos es igual al número de permutaciones para las cuales el índice de la primera ocurrencia del máximo en la sucesión de sumas parciales es r.[6]

EJEMPLO. Consideremos $(x_1, \ldots, x_5) = (1, 1, 1, -1, -1)$. De las $5! = 120$ permutaciones, solamente las siguientes $5!/3!2! = 10$ son distinguibles:

(1,1,1,-1,-1) (1,1,-1,1,-1) (1,-1,1,1,-1)
(-1,1,1,1,-1) (1,1,-1,-1,1)
(1,-1,1,-1,1) (-1,1,1,-1,1)
(1,-1,-1,1,1) (-1,1,-1,1,1) (-1,-1,1,1,1)

[6] Se hace la siguiente convención: En el vector de sumas parciales $(0, s_1, \ldots, s_n)$ con $n+1$ entradas, los índices son $0, 1, \ldots, n$ y no $1, 2, \ldots, n+1$.

que corresponden a $3!2! = 12$ permutaciones cada una. Las sumas parciales asociadas a estas permutaciones son:

$$(0,1,2,3,2,1) \qquad (0,1,2,1,2,1) \qquad (0,1,0,1,2,1)$$

$$(0,-1,0,1,2,1) \qquad (0,1,2,1,0,1)$$

$$(0,1,0,1,0,1) \qquad (0,-1,0,1,0,1)$$

$$(0,1,0,-1,0,1) \qquad (0,-1,0,-1,0,1) \qquad (0,-1,-2,-1,0,1)$$

que se pueden visualizar en la Figura 3. Notemos que las cantidades de sumas parciales positivas para las permutaciones son

mientras que los son

por lo cual hay $2 \cdot 2!3! = 24$ permutaciones de $(x_1, \ldots, x_5)$ para las cuales la sucesión de sumas parciales asociada tiene r elementos positivos y hay la misma cantidad de permutaciones para las cuales el primer máximo de la sucesión de sumas parciales tiene índice r, con $r \in \{1, \ldots, 5\}$. Para $r \notin \{1, \ldots, 5\}$, no hay permutaciones con r sumas parciales positivas o con el índice del primer máximo igual a r, por lo que en cualquier caso, se cumple la afirmación del Lema 1.1.

Demostración del Lema 1.1. Sean

$$A_r$$

la cantidad de permutaciones para las cuales la sucesión de sumas parciales asociada tiene r sumas estrictamente positivas y

$$B_r$$

la cantidad de permutaciones para las cuales la sucesión de sumas parciales alcanza por primera vez el máximo en el lugar r, por lo que el problema que nos ocupa es verificar la igualdad

$$A_r = B_r$$

(dichas cantidades pueden ser distintas a cero sólo para $r = 0, \ldots, n$.) Notemos que las $n!$ permutaciones de $(x_1, \ldots, x_n)$ se pueden obtener fijando un elemento, digamos x_l, en el último o primer lugar, permutando los restantes $n-1$ elementos y variando $l \in \{1, \ldots, n\}$.

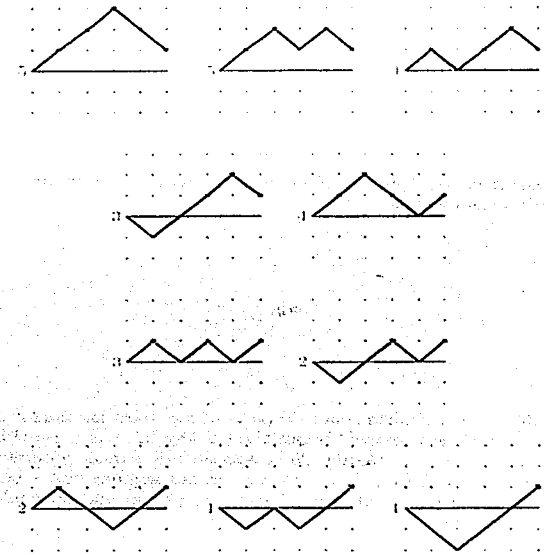

FIGURA 3. Sumas parciales positivas e para las permutaciones distinguibles de $(1,1,1,-1,-1)$.

Sea $(x_1, \ldots, \hat{x}_l, \ldots, x_n)$, el vector cuyas entradas se obtienen de $(x_1, \ldots, x_n)$ quitando la entrada x_l, esto es:

$$\begin{aligned}(x_1, \ldots, \hat{x}_l, \ldots, x_n) &= (x_1, \ldots, x_{l-1}, x_{l+1}, \ldots, x_n) \quad \text{si } l \in \{2, \ldots, n-1\}\\ (\hat{x}_1, x_2, \ldots, x_n) &= (x_2, \ldots, x_n)\\ (x_1, \ldots, x_{n-1}, \hat{x}_n) &= (x_1, \ldots, x_{n-1}).\end{aligned}$$

Así, toda permutación de $(x_1, \ldots, x_n)$ es de la forma (y, x_l) (o (x_l, y)) con y una permutación del vector que acabamos de definir.

Probaremos el resultado por inducción sobre n, la dimensión del vector. Es válido para $n = 1$ pues sólo hay una permutación, la identidad, si $x_1 > 0$, entonces $A_0 = B_0 = 0, A_1 = B_1 = 1$ y si $x_1 \leq 0$ entonces $A_0 = B_0 = 1, A_1 = B_1 = 0$. Supongamos ahora que el resultado es cierto para $n = k$ y sea $(x_1, \ldots, x_{k+1}) \in \mathbb{R}^{k+1}$, por lo que $x_1 + \cdots + x_{k+1} > 0$ o $x_1 + \cdots + x_{k+1} \leq 0$. Veamos que en ambos casos, vale la conclusión del lema:

a) Supongamos que $x_1 + \cdots + x_{k+1} \leq 0$. Sea A_r^l el número de permutaciones de $(x_1, \ldots, x_{k+1})$ que tienen a x_l como última coordenada y cuya sucesión de sumas parciales tiene r elementos estrictamente positivos, por lo que

$$A_r = \sum_{l=1}^{k+1} A_r^l.$$

De manera análoga, definimos B_r^l el número de permutaciones de

$$(x_1, \ldots, x_{k+1})$$

que tienen a x_l como última coordenada y cuyas sumas parciales alcanzan el primer máximo en el lugar r, por lo que se tiene la igualdad

$$B_r = \sum_{l=1}^{k+1} B_r^l.$$

Como $x_1 + \cdots + x_{k+1} \leq 0$, el número de sumas parciales positivas asociadas a una permutación $(x_{\nu(1)}, \ldots, x_{\nu(k+1)})$ es igual al número de sumas parciales positivas asociadas a la permutación $(x_{\nu(1)}, \ldots, x_{\nu(k)})$, y la misma hipótesis implica que el índice del primer máximo de las sumas parciales asociadas a una permutación $(x_{\nu(1)}, \ldots, x_{\nu(k+1)})$ es igual al índice del primer máximo asociado a la permutación $(x_{\nu(1)}, \ldots, x_{\nu(k)})$. Por lo anterior, A_r^l es el número de permutaciones de $(x_1, \ldots, \hat{x}_l, \ldots, x_{k+1})$ que tienen asociadas r sumas parciales positivas y B_r^l es el número de permutaciones del vector anterior que tienen asociadas a r como índice del primer máximo. Por hipótesis de inducción, se cumple la igualdad

$$A_r^l = B_r^l,$$

por lo que en este caso, $A_r = B_r$.

b) Si $x_1 + \cdots + x_{k+1} > 0$, el razonamiento es similar al caso anterior, pues una permutación (y, x_l) tiene asociadas r sumas parciales positivas si y sólo si y tiene asociadas $r - 1$ sumas positivas, que da lugar a la igualdad

$$A_r = \sum_{l=1}^{k+1} A_{r-1}^l,$$

donde A_r^l es el número de permutaciones que tienen a x_l como última coordenada y que tienen asociadas r sumas parciales positivas. Este número coincide con el número de permutaciones del vector $(x_1, \ldots, \hat{x}_l, \ldots, x_{k+1})$ que tienen asociadas

r sumas parciales positivas. Asimismo, una permutación (x_l, y) de $(x_1, \ldots, x_{k+1})$ tiene como índice del primer máximo (de las sumas parciales) a r si y sólo si y tiene como índice a $r-1$, puesto que el índice no puede ser cero al existir una suma parcial $(x_1 + \cdots + x_{k+1})$ mayor a cero.[7] Como consecuencia, tenemos la igualdad

$$B_r = \sum_{l=1}^{k+1} B_{r-1}^l,$$

con B_r^l igual a la cantidad de permutaciones de la forma (x_l, y) que tienen como índice del primer máximo a r, que es igual al número de permutaciones de $(x_1, \ldots, \hat{x}_l, \ldots, x_{k+1})$ que tienen como índice del primer máximo a r. Por hipótesis de inducción, se cumple la igualdad la igualdad $A_r^l = B_r^l$, de la cual se deduce que $A_r = B_r$.

□

TEOREMA 1.1 (Identidad de Sparre Andersen). *Si M_n es el índice del primer máximo de $(S_0, S_1, \ldots, S_n)$, entonces M_n y π_n tienen la misma distribución.*

DEMOSTRACIÓN. Para cada permutación $\nu : \{1, \ldots, n\} \to \{1, \ldots, n\}$, consideremos las variables

$$\begin{aligned} S_0^\nu &= 0 \\ S_i^\nu &= \sum_{j=1}^{i} X_{\nu(i)}, \quad , i = 1, \ldots, n \\ Y^\nu &= \mathbf{1}_{\{\text{Hay r elementos positivos en } (S_0^\nu, \ldots, S_n^\nu)\}} \\ Z^\nu &= \mathbf{1}_{\{\text{El índice del primer término máximo de } (S_0^\nu, \ldots, S_n^\nu) \text{ es r}\}}. \end{aligned}$$

Como $(X_i)_{i=1}^n$ son variables aleatorias independientes e idénticamente distribuidas, se sigue que son intercambiables, por lo que las variables Y^ν (indicadas por las permutaciones de $\{1, \ldots, n\}$) son idénticamente distribuidas, así como las variables Z^ν, de donde, al denotar por Id a la permutación identidad,

$$\begin{aligned} \mathbb{E}(Y^\nu) &= \mathbb{E}(Y^{\mathrm{Id}}) \\ \mathbb{E}(Z^\nu) &= \mathbb{E}(Z^{\mathrm{Id}}). \end{aligned}$$

Finalmente notemos que por el Lema 1.1,

$$\sum_\nu Y^\nu = \sum_\nu Z^\nu.$$

---

[7]Note que aquí hemos considerado a las permutaciones que tienen a x_l como primera coordenada y no como la última.

por lo que

$$\mathbb{P}(\pi_n = r) = \mathbb{E}(Y^{\mathrm{Id}}) = \frac{1}{n!}\sum_{\nu}\mathbb{E}(Y^{\nu}) = \frac{1}{n!}\sum_{\nu}\mathbb{E}(Z^{\nu}) = \mathbb{E}(Z^{\mathrm{Id}}) = \mathbb{P}(M_n = r)\,.$$

□

1.3.2. El Índice del primer máximo. Afortunadamente, la distribución del índice de la primera vez que ocurre el máximo de las n primeras sumas parciales es una variable más fácil de estudiar, pues podemos utilizar la hipótesis de independencia más directamente. Más especificamente, si $(X_i)_{i=1}^n$ son variables aleatorias independientes con distribución común, definimos $(S_i)_{i=0}^n$ como en la subsección anterior, a la variable

$$M = \max_{0\leq i\leq n} S_i$$

y a

$$M_n = \min\{i \in \{0,\ldots,n\} : S_i = M\}\,.$$

El Teorema de Sparre Andersen establece que M_n y π_n tienen la misma distribución. Notemos que el evento $\{M_n = k\}$, para $k \in \{1,\ldots,n-1\}$ es igual a

$$\{S_i < S_k, i \in \mathbb{Z}\cap[0,k-1], S_i \leq S_k, i \in \mathbb{Z}\cap[k,n]\}\,,$$

de lo cual se sigue la igualdad

$$\{M_n = k\} = \{S_k - S_i > 0, i \in \mathbb{Z}\cap[0,k]\} \cap \{S_i - S_k \leq 0, i \in \mathbb{Z}\cap[k,n]\}\,.$$

Los dos eventos que ocurren en el lado derecho de la anterior igualdad son independientes, puesto que el primero depende de las variables $X_1,\ldots X_k$ y el segundo de las variables $X_{k+1},\ldots X_n$, ya que $S_i - S_k = X_{k+1} + \cdots + X_i$ para $i \in \{k+1,\ldots,n\}$. De aquí se sigue también que

$$(S_{k+1} - S_k,\ldots,S_n - S_k) \text{ y } (S_1,\ldots,S_{n-k})$$

tienen la misma distribución. Además, como $S_k - S_i = X_{i+1} + \cdots + X_k$ para $i \in \{0,\ldots,k-1\}$, tenemos que

$$(S_k - S_{k-1},\ldots,S_k - S_0) \text{ y } (S_1,\ldots,S_k)$$

tienen la misma distribución. Así, se sigue que si $k \in \{1,\ldots,n-1\}$, entonces

$$\mathbb{P}(M_n = k) = \mathbb{P}(S_1 > 0,\ldots,S_k > 0)\,\mathbb{P}(S_1 \leq 0,\ldots,S_{n-k} \leq 0) = p_k q_{n-k}, \tag{8}$$

donde definimos para $k \geq 1$,

$$p_k = \mathbb{P}(S_1 > 0,\ldots,S_k > 0)$$
$$q_k = \mathbb{P}(S_1 \leq 0,\ldots,S_k \leq 0)\,.$$

Por otro lado el evento $\{M_n = 0\}$ es igual a $\{S_i \leq 0, i \in \{1,\ldots n\}\}$ cuya probabilidad es lo que hemos definido como q_n y el evento $\{M_n = n\}$ es igual a $\{S_n - S_i > 0, i = 0,\ldots,n-1\}$ y dado que $(S_n - S_{n-1},\ldots S_n - S_0)$ tiene la misma distribución que $(S_1,\ldots,S_n)$, se ve que la probabilidad del evento $\{M_n = n\}$ es p_n. Así, si definimos $p_0 = q_0 = 1$, la ecuación

(8) es válida para $k = 0, \dots, n$. Sean P y Q las funciones generadoras de las sucesiones $\{p_n : n = 0, 1, \dots\}$ y $\{q_n : n = 0, 1, \dots\}$ respectivamente. Se introducen para estudiar a las sucesiones mediante representaciones más compactas, así como operar con ellas. Trabajaremos ahora con la función P, utilizando el siguiente razonamiento: Como $(X_i)_{i=1}^{n}$ son independientes e idénticamente distribuidas, se sigue que

$$\begin{aligned} p_n &= \mathbb{P}(S_1 > 0, \dots, S_n > 0) \\ &= \mathbb{P}(X_1 > 0, X_1 + X_2 > 0, \dots, X_1 + \cdots + X_n > 0) \\ &= \mathbb{P}(X_n > 0, X_n + X_{n-1} > 0, \dots, X_n + \cdots + X_1 > 0) \\ &= \mathbb{P}(S_n - S_{n-1} > 0, \dots, S_n - S_1 > 0) \end{aligned}$$

Decimos que $n > 0$ es un índice de crecimiento si $S_n > S_0 = 0, \dots, S_n > S_{n-1}$,[8] por lo que p_n es igual a la probabilidad de que n sea índice de crecimiento, para $n \geq 1$. El primer índice de crecimiento coincide con el tiempo de entrada de la caminata $(S_n)_{n\in\mathbb{N}}$ a $(0, \infty)$. En otras palabras, el evento $\{n$ es el primer índice de crecimiento$\}$ coincide con el evento $\{S_n > 0, S_k \leq 0, k \leq n-1\}$. Sea τ_n la probabilidad de que n sea el primer índice de crecimiento y notemos que si n es índice de crecimiento, entonces el primer índice de crecimiento, k, satisface $1 \leq k \leq n$, por lo cual, para $n \geq 1$,

$$\begin{aligned} \{n \text{ es índice de crecimiento}\} &= \bigcup_{k=1}^{n} \left(\begin{array}{l} \{n \text{ es índice de crecimiento}\} \\ \cap \{\ k \text{ tiempo de entrada a } (0,\infty)\} \end{array} \right) \\ &= \bigcup_{k=1}^{n} \left(\begin{array}{l} \{S_k > 0, S_i \leq 0, i \leq k-1\} \\ \cap \{S_j < S_n, j \leq n-1\} \end{array} \right) \\ &= \bigcup_{k=1}^{n-1} \left(\begin{array}{l} \{S_k > 0, S_i \leq 0, i \leq k-1\} \\ \cap \{S_n - S_j > 0, k \leq j \leq n-1\} \end{array} \right) \\ &\quad \bigcup \{S_n > 0, S_i \leq 0, i \leq n-1\}. \end{aligned}$$

Como los eventos $\{S_k > 0, S_i \leq 0, i \leq k-1\}$ y $\{S_n - S_j > 0, k \leq j \leq n-1\}$ son independientes y

$$\mathbb{P}(S_n - S_j > 0, k \leq j \leq n-1) = \mathbb{P}(S_j > 0, 1 \leq j \leq n-k) = p_{n-k},$$

se sigue que

$$p_n = \mathbb{P}(n \text{ es índice de crecimiento}) = \sum_{k=1}^{n} \tau_k p_{n-k}, \ n \geq 1$$

[8] En general puede no existir algún índice de crecimiento y si existe, no tiene porque ser único.

donde $\tau_k = \mathbb{P}(k$ es tiempo de entrada a $(0,\infty))$. Si denotamos por τ la generadora de la sucesión $\{\tau_n : n \in \mathbb{N}\}$ (con $\tau_0 = 0$), de la ecuación anterior obtenemos

$$P(s) = \sum_{n=0}^{\infty} p_n s^n = 1 + \sum_{n=1}^{\infty} s^n \sum_{k=1}^{n} \tau_k p_{n-k} = 1 + P(s)\tau(s),$$

por lo que finalmente

$$P(s) = \frac{1}{1-\tau(s)}.$$

Para proceder con el estudio de τ, probaremos el siguiente resultado, en el cual se mencionan las permutaciones cíclicas de un vector $x \in \mathbb{R}^n$. Si $x = (x_1, \ldots, x_n)$, un reordenamiento cíclico es simplemente un corrimiento de los índices, por ejemplo,

$$(x_3, \ldots, x_n, x_1, x_2),$$

de los cuales existen n. Si $\nu \in \{1, \ldots, n-1\}$ el reordenamiento ν del vector x, al cual denotaremos por x^ν es

$$x^\nu = (x_{\nu+1}, \ldots, x_n, x_1, \ldots, x_\nu),$$

y el reordenamiento x^n, que es el único que falta, es simplemente x. Así, el índice de cualquier reordenamiento cíclico es el índice de la entrada que ocupa la última coordenada.

LEMA 1.2. *Sean $(x_1, \ldots, x_n)$ reales fijos, tales que $x_1 + \cdots + x_n > 0$. Si a cada reordenamiento cíclico x^ν de $(x_1, \ldots, x_n)$ le asociamos la sucesión de sumas parciales $(s_0^\nu, \ldots, s_n^\nu)$ y denotamos por r el número de reordenamientos para los cuales n es un índice de crecimiento para $(s_0^\nu, \ldots, s_n^\nu)$, entonces r es mayor o igual a 1. Si ν es uno de estos reordenamientos, r es la cantidad de índices de crecimiento asociados a s^ν.*

EJEMPLO. Consideremos al vector $x = (-1, 1, 1, -1, 1, 1) \in \mathbb{R}^6$. Las sumas parciales asociadas a los seis reordenamientos de este vector son

$$s^1 = (1,2,1,2,3,2) \qquad s^2 = (1,0,1,2,1,2) \qquad s^3 = (-1,0,1,0,1,2)$$
$$s^4 = (1,2,1,2,3,2) \qquad s^5 = (1,0,1,2,1,2) \qquad s^6 = (-1,0,1,0,1,2),$$

ordenadas como se indica anteriormente. Éstas se pueden visualizar en la Figura 4, en la que además podemos verificar que $n = 6$ es índice de crecimiento para los reordenamientos cíclicos 3 y 6 y que los índices de crecimiento de las sumas parciales asociadas a x^3 o a x^6 son 4 y 6. Así, hay más de un reordenamiento para el cual n es índice de crecimiento, de hecho hay 2 de estos reordenamientos, y en ellos hay 2 índices de crecimiento. Esto coincide con las dos afirmaciones del lema anterior.

DEMOSTRACIÓN DEL LEMA 1.2. Sean $s^\nu = (s_0^\nu, \ldots, s_n^\nu)$ el vector de sumas parciales asociado al reordenamiento cíclico ν y $s = (s_0, \ldots, s_n)$ el vector de sumas parciales asociado al vector dado inicialmente. Notemos que entonces, si $\nu \in \{1, \ldots, n-1\}$, tiene lugar

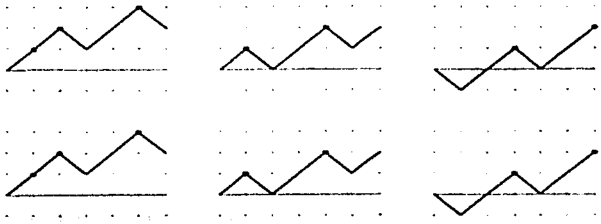

FIGURA 4. Sumas parciales e índices de crecimiento de las permutaciones cíclicas de (-1,1,1,-1,1,1).

la igualdad

$$s_k^\nu = \begin{cases} s_{k+\nu} - s_\nu & \text{si } k \in \{0,\dots,n-\nu\} \\ s_n - s_\nu + s_{k-(n-\nu)} & \text{si } k \in \{n-\nu,\dots,n\}. \end{cases} \tag{9}$$

Como $x_1 + \cdots + x_n > 0$, se sigue que el máximo de las sumas parciales es mayor a cero. Si ν denota el índice del primer máximo de s, se sigue que $s_i < s_\nu$ si $i < \nu$ y $s_i \leq s_\nu$ si $i \geq \nu$, por lo que al fijarnos en el reordenamiento cíclico ν, haciendo uso de la igualdad (9), obtenemos que si $k \in \{0,\dots,n-\nu\}$, entonces

$$s_k^\nu = s_{\nu+k} - s_\nu \leq 0$$

y como $s_n > 0$, obtenemos

$$s_n^\nu = s_n > s_k^\nu;$$

pero si $n - \nu \leq k < n$, entonces

$$s_n^\nu - s_k^\nu = s_n - s_k^\nu = s_\nu - s_{k-(n-\nu)}$$

y como en este caso $0 \leq n - (k - \nu) < \nu$, se sigue que

$$s_\nu - s_{k-(n-\nu)} > 0.$$

Así, podemos conlcuir que $s_k^\nu < s_n = s_n^\nu$ para $k < n$, esto es, que n es índice de crecimiento para el reordenamiento ν y por lo tanto la primera afirmación del lema es verdadera.

Para verificar la segunda afirmación, supongamos sin pérdida de generalidad que n es índice de crecimiento para s, por lo que si hay r índices de crecimiento en s, digamos $\nu_1 < \cdots < \nu_r = n$, notemos que n es índice de crecimiento para s^m si y sólo si m es índice de crecimiento para la sucesión original, o de otra manera, si $m \in \{\nu_1,\dots,\nu_r\}$. Esto sucede pues si m es índice de crecimiento para s, tenemos dos opciones, $m = n$ y $m < n$. En el primer caso, $s^m = s$, por lo que n es índice de crecimiento para s^m. En el segundo, notemos que como $s_m > 0$ y n es índice de crecimiento para s, si $k \in \{0, \ldots, n-m\}$, entonces

$$s_k^m = s_{k+m} - s_m < s_{k+m} < s_n.$$

Si $k \in \{n-m, \ldots, n-1\}$, entonces $k-(n-m) \in \{0, \ldots, m-1\}$, por lo que

$$s_{k-(n-m)} - s_m < 0,$$

por ser m un índice de crecimiento, de donde para $k \in \{n-m, \ldots, n-1\}$, se tiene

$$s_k^m = s_n - s_m + s_{k-(n-m)} < s_n = s_n^m.$$

Esto quiere decir que n es índice de crecimiento para la permutación m. Por otro lado, si n es índice de crecimiento para la permutación m, entonces $s_n^m > s_k^m$ para $k \in \{0, \ldots, m\}$. Si observamos que

$$s_k = s_{k+n-m}^m - s_{n-m}^m$$

para $k \in \{0, \ldots, m\}$, entonces

$$s_m = s_n^m - s_{n-m}^m > s_{k+n-m}^m - s_{n-m}^m = s_k$$

para $k \in \{0, \ldots, m-1\}$, o sea, m es índice de crecimiento para s. Así, la cantidad de permutaciones para las cuales n es un índice de crecimiento es igual a la cantidad de índices de crecimiento que hay en aquellas permutaciones en las cuales n es un índice de crecimiento. □

NOTA. Si definimos a los índices de crecimiento débiles como aquellos k para los cuales $s_i \leq s_k$ para $i \leq k$, el lema anterior permanece válido si substituimos la hipótesis $s_n > 0$ por $s_n \geq 0$ y nos referimos a los índices de crecimiento débiles en vez de a los índices de crecimiento. Esta aclaración es importante para extender el resultado que obtendremos para P a la función Q.

El siguiente resultado utiliza el lema anterior para obtener una expresión para τ, de donde obtendremos una para P:

PROPOSICIÓN 1.1. *Tiene lugar la igualdad*

$$\log \frac{1}{1-\tau(s)} = \sum_{n=0}^{\infty} \frac{s^n}{n} \mathbb{P}(S_n > 0).$$

DEMOSTRACIÓN. Sea

$$Y_r^\nu = \mathbf{1}_{(\text{El } r\text{-ésimo índice de crecimiento para } (S_0^\nu, \ldots, S_n^\nu) \text{ es } n)}.$$

Como las variables aleatorias $X_1, \ldots, X_n$ son independientes e idénticamente distribuidas, entonces son intercambiables, por lo que las variables Y_r^ν son idénticamente distribuidas. Además por el lema anterior, $\sum_\nu Y_r^\nu$ toma el valor 0 o r, ya que si algún sumando es mayor que cero, entonces hay r índices de crecimiento en $(S_0, \ldots, S_n)$ y n es el r-ésimo índice de crecimiento en los reordenamientos cíclicos asociados a estos índices de crecimiento,[9] por lo que en este caso la suma $\sum_\nu Y_r^\nu$ tiene el valor r. Por otro lado, si todos los sumandos son cero, entonces $\sum_\nu Y_r^\nu = 0$.

Así, al denotar por τ_n^r a la probabilidad de que el r-ésimo índice de crecimiento de la sucesión de sumas parciales asociada a $(X_i)_{i=1}^\infty$ sea n, obtenemos

$$\tau_n^r = \mathbb{E}(Y^n) = \frac{1}{n}\sum_\nu \mathbb{E}(Y^\nu) = \frac{r}{n}\mathbb{P}\left(\sum_\nu Y^\nu = r\right).$$

Otra consecuencia del lema anterior es que los eventos

$$\left\{\sum_\nu Y_r^\nu = r\right\}, \qquad r = 1, \ldots, n$$

son ajenos, puesto que $\sum_\nu Y_r^\nu = r$ si y sólo si r es la cantidad de reordenamientos para los cuales n es índice de crecimiento, así, sólo una de las sumas $\sum_\nu Y_r^\nu$ puede ser igual a r. Pero además si $S_n > 0$ entonces el lema anterior nos permite afirmar que existe r entre 1 y n tal que $\sum_\nu Y_r^\nu = r$ y si alguna de las sumas $\sum_\nu Y_r^\nu$ es igual a r, entonces n es índice de crecimiento para algún reordenamiento cíclico y por ende, S_n es mayor a cero. En resumen, hemos verificado que

$$\mathbb{P}(S_n > 0) = \mathbb{P}\left(\text{Existe } r \in \{1, \ldots, n\} \text{ tal que } \sum_\nu Y_r^\nu = r\right) = \sum_{r=1}^n \mathbb{P}\left(\sum_\nu Y_r^\nu = r\right).$$

Por lo anterior

$$\sum_r \frac{1}{r}\tau_n^r = \frac{1}{n}\mathbb{P}(S_n > 0).$$

Si denotamos por $\tau_r(s)$ a $\sum_{n\geq 1} \tau_n^r s^n$, entonces $\tau_r(s) = (\tau(s))^r$, pues se tiene la igualdad

$$\tau_n^r = \sum_{k=1}^{n-1} \tau_k^{r-1}\tau_{n-k}^1, \tag{10}$$

ya que n es el r-ésimo índice de crecimiento para $(S_0, \ldots, S_n)$ si y sólo si existe algún k entre 1 y $n-1$ que sea el índice de crecimiento $r-1$ para $(S_0, \ldots, S_n)$ e independientemente, $n-k$ debe ser el primer índice de crecimiento para $(0, S_{k+1} - S_k, \ldots, S_n - S_k)$. La probabilidad del último evento considerado es igual a la probabilidad de que $n - k$ sea índice de crecimiento para $(S_0, \ldots, S_{n-k})$.

De acuerdo a la igualdad (10):

$$\sum_{n\geq 1} \frac{1}{n}s^n\mathbb{P}(S_n > 0) = \sum_{n\geq 1} s^n \sum_{r\geq 1} \frac{1}{r}\tau_n^r = \sum_{r\geq 1}\frac{1}{r}\tau(s)^r = \log\frac{1}{1-\tau(s)},$$

[9]Ver la prueba de la segunda afirmación del Lema 1.2.

puesto que para cualquier $s \in (0,1)$,

$$\log \frac{1}{(1-s)} = \sum_{n \geq 1} \frac{s^n}{n}$$

y $\tau(s) \in (0,1)$ para $s \in (0,1)$. $\square$

NOTA. Mediante la proposición anterior, podemos afirmar que

$$\log P(s) = \sum_{n=1}^{\infty} \frac{s^n}{n} \mathbb{P}(S_n > 0)$$

y mediante la aclaración que sigue al Lema 1.2, aplicando una técnica análoga a la utilizada en la demostración de la Proposición 1.1, podemos concluir que

$$\log Q(s) = \sum_{n=1}^{\infty} \frac{s^n}{n} \mathbb{P}(S_n \leq 0)\,.$$

Ahora, podemos utilizar estos resultados para obtener explícitamente la solución al problema que originalmente nos habíamos planteado, encontrar la distribución del número de sumas parciales positivas, como consecuencia de encontrar la del índice del primer máximo.

1.3.3. Resultados exactos y asintóticos para el número de sumas positivas.

Como consecuencia del análisis de la subsección anterior, podemos calcular explícitamente la densidad discreta del número de sumas parciales positivas para algunos casos particulares. Esto es, si $(X_i)_{i=1}^{\infty}$ son variables aleatorias independientes e idénticamente distribuidas con distribución común F, $(S_n)_{n\in\mathbb{N}}$ es la sucesión de sumas parciales asociada a $(X_i)_{i\in\mathbb{N}}$ y $\pi_n = \sum_{i=1}^{n} 1_{(S_i>0)}$ es la cantidad de sumas parciales positivas (asociadas a $(S_i)_{i=1}^{n}$), para ciertas funciones F es fácil obtener la distribución de π_n. El primer caso es cuando F es continua y simétrica, puesto que entonces $\mathbb{P}(S_n > 0) = \mathbb{P}(S_n \leq 0) = 1/2$ para cualquier $n \geq 1$, por lo que

$$\log P(s) = \frac{1}{2}\sum_{n=1}^{\infty} \frac{s^n}{n} = \frac{1}{2}\log\frac{1}{1-s} = \log(1-s)^{-1/2}$$

$$\log Q(s) = \frac{1}{2}\sum_{n=1}^{\infty} \frac{s^n}{n} = \frac{1}{2}\log\frac{1}{1-s} = \log(1-s)^{-1/2}$$

y así, $P(s) = Q(s) = (1-s)^{-1/2}$, por lo que $p_k = q_k = (-1)^k \binom{-1/2}{k}$, de donde podemos obtener la distribución de π_n:

$$\mathbb{P}(\pi_n = k) = p_k q_{n-k} = \binom{-1/2}{k}\binom{-1/2}{n-k}\,.$$

Esta es la misma expresión que encontramos en el caso de la caminata aleatoria simple[10] (aunque en ese caso, para una variable aleatoria distinta,) por lo que tenemos el mismo teorema límite para π_n/n, a saber, que su distribución converge a la distribución arcoseno cuando $n \longrightarrow \infty$.

Por otro lado, como una generalización del caso en el que la distribución sea continua y simétrica, podemos considerar una distribución F para la cual $\mathbb{P}(S_n > 0) = \delta$ independientemente de n. En este caso,

$$\log P(s) = \delta \sum_{n=1}^{\infty} \frac{s^n}{n} = \delta \log \frac{1}{1-s} = \log(1-s)^{-\delta}$$

$$\log Q(s) = (1-\delta) \sum_{n=1}^{\infty} \frac{s^n}{n} = (1-\delta) \log \frac{1}{1-s} = \log(1-s)^{1-\delta}$$

y así, $P(s) = (1-s)^{-\delta}$ y $Q(s) = (1-s)^{-(1-\delta)}$, por lo que

$$p_k = (-1)^k \binom{-\delta}{k} \quad \text{y}$$

$$q_k = (-1)^k \binom{-(1-\delta)}{k}$$

de donde obtenemos

$$p_k = (-1)^k \binom{-\delta}{k} = (-1)^k \frac{\prod_{i=0}^{k-1}(-\delta - i)}{k!} = \frac{\prod_{i=0}^{k-1}(\delta + i)}{k!} = \frac{\Gamma(k+\delta)}{\Gamma(k+1)\,\Gamma(\delta)}$$

y de manera análoga,

$$q_k = \frac{\Gamma(k+1-\delta)}{\Gamma(k+1)\,\Gamma(1-\delta)},$$

por lo que

$$\begin{aligned} p_k q_{n-k} &= \frac{\Gamma(k+\delta)}{\Gamma(k+1)} \frac{\Gamma(n-k+1-\delta)}{\Gamma(n-k+1)} \frac{1}{\Gamma(\delta)\,\Gamma(1-\delta)} \\ &= \frac{\sin \pi\delta}{\pi} \frac{\Gamma(k+\delta)}{\Gamma(k+1)} \frac{\Gamma(n-k+1-\delta)}{\Gamma(n-k+1)}, \end{aligned}$$

al utilizar la relación $\Gamma(p)\,\Gamma(1-p) = \pi / \sin \pi p$. Si utilizamos la aproximación de Stirling para la función Gamma, nos damos cuenta de que si $p \in (0,1)$, entonces

$$\Gamma(k+p)/\Gamma(k+1) \sim 1/k^{(1-p)},$$

[10]Ver la Sección 1.2.

puesto que

$$\frac{\Gamma(k+p)}{\Gamma(k+1)} \sim \frac{e^{-k-p}}{e^{-k-1}} \frac{(k+p)^{k+p+1/2}}{(k+1)^{k+1+1/2}} = e^{1-p} \frac{1}{(k+1)^{1-p}} \left(\frac{k+p}{k+1}\right)^{k+p+1/2}$$

y como el último factor se puede escribir como

$$\frac{(1+p/k)^{k+p+1/2}}{(1+1/k)^{k+p+1/2}},$$

que converge a e^{p-1} conforme $k \to \infty$, se sigue que

$$\Gamma(k+p)/\Gamma(k+1) \sim 1/(k+1)^{1-p} \sim 1/k^{1-p}.$$

Así, como en el caso de la caminata aleatoria simple, podemos conlcuir que para $x, y \in (0,1)$ con $x < y$,

$$\lim_{n\to\infty} \mathbb{P}((\pi_n/n) \in [x,y]) = \frac{\sin \pi\delta}{\pi} \int_{[x,y]} \frac{1}{t^{1-\delta}(1-t)^{\delta}}\, dt.$$

Sean $\varepsilon > 0$, $x \in (0,1)$ y A una variable con densidad

$$f(x) = \mathbf{1}_{(0,1)}(x) \frac{\sin \pi\delta}{\pi x^{\delta}(1-x)^{1-\delta}}.$$

Como

$$\lim_{n\to\infty} \mathbb{P}(A \in [0,1/n) \cup (1-1/n,1]) = \mathbb{P}(A \in \{0,1\}) = 0,$$

se sigue que existe $x_1 \in (0,(1/2) \wedge x)$ tal que

$$\mathbb{P}(A \in [0,x_1) \cup (1-x_1,1]) < \varepsilon/4$$

y dado que

$$\lim_{n\to\infty} \mathbb{P}((\pi_n/n) \in [x_1,x]) = \mathbb{P}(A \in [x_1,x]) > 1-\varepsilon/4,$$

se sigue que existe $N_1 \in \mathbb{N}$ tal que si $n \geq N_1$, entonces

$$\mathbb{P}((\pi_n/n) \in [0,x_1)) \leq \mathbb{P}((\pi_n/n) \in [0,x_1) \cup (1-x_1,1]) < \varepsilon/4.$$

Sea $N_2 \in \mathbb{N}$ tal que si $n \geq N_2$ entonces $\mathbb{P}((\pi_n/n) \in (x_1,x])$ se encuentre en la vecindad de radio $\varepsilon/2$ con centro en $\mathbb{P}(A \in (x_1,x])$. Entonces si $n \geq N = \max(N_1,N_2)$:

$$\begin{aligned} |F_{\pi_n/n}(x) - F_A(x)| \leq & |\mathbb{P}((\pi_n/n) \in [0,x_1]) - \mathbb{P}(A \in [0,x_1])| \\ & + |\mathbb{P}((\pi_n/n) \in (x_1,x]) - \mathbb{P}(A \in (x_1,x])| \\ < & \varepsilon/2 + \varepsilon/2 = \varepsilon. \end{aligned}$$

Del anterior razonamiento se desprenden dos resultados: que $F_{\pi_n/n}(0) \to 0 = F_A(0)$, ya que para toda $\varepsilon > 0$ podemos encontrar $x_1 = x_1(\varepsilon)$ y $N \in \mathbb{N}$ tal que

$$F_{\pi_n/n}(0) \leq F_{\pi_n/n}(x_1) < \varepsilon$$

y además, que $\lim_{n\to\infty} F_{\pi_n/n}(x) = F_A(x)$ si $x \in (0,1)$. Como $F_{\pi/n}(x) = F_A(x)$ si $x \in (-\infty, 0) \cup [1, \infty)$, se sigue que π/n converge en distribución a A, a cuya distribución le llamaremos ley arcoseno generalizada de parámetro δ.

Como ejemplo de la situación descrita en el párrafo anterior, consideremos $(X_i)_{i\in\mathbb{N}}$ una sucesión de variables aleatorias independientes con densidad común

$$f(x) = \frac{1}{\pi\left(1 + (x-a)^2\right)},$$

esto es, X_i tiene distribución Cauchy de parámetro a. La función característica de X_i cuando $a = 0$ es $\Phi_{X_i}(\omega) = e^{-|\omega|}$ (se puede probar mediante cálculo de residuos,) por lo que la función característica de X_i para cualquier a es $e^{i\omega a - |\omega|}$, de donde la función carácteristica de S_n/n es igual a la de X_1 y por lo tanto $\mathbb{P}(S_n > 0) = \mathbb{P}(X_1 > 0) = \arctan(-a)/\pi + 1/2$. Si $a = -1$, entonces $\mathbb{P}(S_n > 0) = 1/4$, mientras que si $a = 1$, entonces $\mathbb{P}(S_n > 0) = 3/4$. En la Figura 5 podemos apreciar la densidad y la distribución arcoseno generalizada de parámetros $1/4, 1/2$ y $3/4$, que corresponde, en términos del anterior ejemplo, a los valores de $a = -1, 0$ y 1.

Notas

En este capítulo se calculó la distribución de la fracción de tiempo que la caminata aleatoria es positiva en una trayectoria de $2n$ pasos, F_{2n}, por medio del uso de la técnica de la función generadora. Cuando se discute a la variable F_{2n}, en la mayoría de los textos prueban que la distribución arcoseno discreta es su distribución por inducción. Ésto sucede, por dar dos ejemplos, en las referencias [8] y [6]. En esta última, una nota al pie de la página 82 hace alusión a *métodos complicados en la primera edición* para el cálculo de dicha distribución, en constraste con la prueba por inducción que presentan. Si se consulta la referencia [5], se encuentra el método de la función generadora que no utiliza de antemano el conocimiento de la distribución de F_{2n}, como lo hace la otra prueba. Los métodos sencillos a los que se refiere Feller, son los que utilizan como base la identidad de Andersen, misma que fué descubierta utilizando técnicas muy elaboradas pero que ha encontrado una demostración un tanto más sencilla, que sin embargo utiliza funciones generadoras.

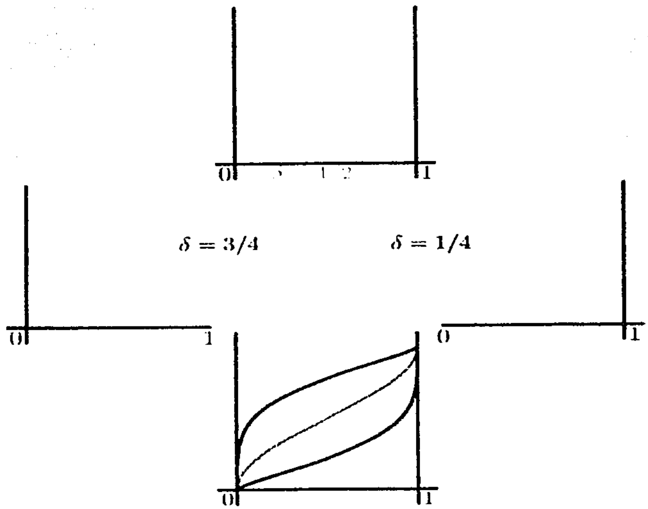

FIGURA 5. Densidad y distribución arcoseno generalizada, $\delta = 1/4, 1/2$ y $3/4$

32

CAPÍTULO 2

El Movimiento Browniano y la ley arcoseno

2.1. Introducción

En el capítulo anterior, hemos calculado la distribución límite de la fracción de elementos positivos de una caminata aleatoria, al suponer la relación

$$\mathbb{P}(S_n > 0) = \delta \in (0,1).$$

En este capítulo, investigaremos la convergencia débil de la misma sucesión de variables aleatorias, haciendo el supuesto de que la varianza de S_1 es finita y estrictamente mayor a cero (esto es, que no es la variable aleatoria constante) y que su esperanza es cero. En particular, esto implica que podemos utilizar la aproximación contenida en el Teorema Central del Límite, primero de manera directa, para obtener un resultado bautizado como principio de invarianza, pues estipula que la distribución límite no depende de la distribución del salto. Después, explicaremos de otra manera el principio de invarianza, al probar un resultado sobre convergencia débil para variables aleatorias con valores en un espacio de funciones, para lo cual se vuelve necesario introducir al Movimiento Browniano real. El resultado nos revela una relación entre caminatas aleatorias con varianza finita y el Movimiento Browniano, así como una infinidad de principios de invarianza como el que nosotros consideramos. Finalmente, explotamos la relación obtenida entre las caminatas aleatorias y el Movimiento Browniano para calcular la distribución de la fracción de tiempo en el que el Movimiento Browniano es positivo, que resulta ser la ley arcoseno y esto generalizará los resultados obtenidos en el capítulo anterior.

2.2. El principio de invariancia de Erdös-Kac

Utilizaremos ahora el Teorema Central del Límite y el estudio sobre la fracción de sumas positivas en caminatas aleatorias cuya distribución del salto es absolutamente continua y simétrica respecto a cero para demostrar un enunciado similar para caminatas aleatorias un poco más generales. Consideraremos una sucesión de variables aleatorias independientes e idénticamente distribuidas $(X_n)_{n\in\mathbb{N}}$ tal que $\mathbb{E}(X_i) = 0, \mathrm{Var}(X_i) = 1$ y si $(S_n)_{n\in\mathbb{N}}$ es la sucesión de sumas parciales asociada a $(X_n)_{n\in\mathbb{N}}$, $S_n = \sum_{i=1}^n X_i$, definimos $N_n = \sum_{i=1}^n \psi(S_i)$, donde $\psi(x) = \mathbf{1}_{(0,\infty)}(x)$. El resultado a verificar es que N_n/n converge en distribución a la ley arcoseno encontrada en el capítulo anterior independientemente de la distribución de X_i. Dado que la distribución arcoseno es continua, probar la convergencia en distribución se reduce a verificar que $\lim_{n\to\infty} \mathbb{P}(N_n/n < \alpha) = p(\alpha)$ para todo $\alpha \in \mathbb{R}$, donde $p(\alpha)$ es la distribución arcoseno.

Sean $k \in \mathbb{N}$, $n_j = [jn/k]$ para $j = 0 \dots k$ y

$$D_n = \frac{1}{n} \sum_{j=1}^{k} \sum_{r=n_{j-1}+1}^{n_j} \left(\psi(S_r) - \psi(S_{n_j})\right).$$

Notemos que, al separar a D_n en dos sumas, obtenemos las igualdades

$$D_n = \frac{1}{n} \sum_{j=1}^{k} \sum_{r=n_{j-1}+1}^{n_j} \left(\psi(S_r) - \psi(S_{n_j})\right) = N_n/n - \frac{1}{n} \sum_{j=1}^{k} \psi(S_{n_j}) \left(n_j - n_{j-1}\right), \tag{11}$$

puesto que todo entero entre 1 y n pertenece a uno (y solamente uno) de los conjuntos $[n_{j-1} + 1, n_j] \cap \mathbb{N}$ para j entre 1 y k, lo cual explica el primer sumando del tercer miembro de (11).

La idea detrás de la definición de la variable aleatoria D_n es dividir el camino de n pasos en k partes cuya longitud sea muy similar, casi igual a n/k, para comparar la cantidad de elementos positivos que hay en todo el bloque con un múltiplo de la indicadora de que el primer elemento del bloque sea positivo, verificando que podemos acotar la diferencia entre estas dos variables aleatorias.

Primero estimaremos la probabilidad de que $|D_n|$ sea mayor a δ por medio de la desigualdad de Markov, para lo cual, notamos que

$$\mathbb{P}(|D_n| > \delta) \leq \frac{1}{\delta}\mathbb{E}(|D_n|) \leq \frac{1}{n\delta} \sum_{i=1}^{k} \sum_{r=n_{i-1}+1}^{n_i} \mathbb{E}(|\psi(S_r) - \psi(S_{n_i})|).$$

Como el valor absoluto de la diferencia de las dos funciones indicadoras es igual a la indicadora de la diferencia simétrica de los eventos que representan, entonces

$$\mathbb{E}(|\psi(S_r) - \psi(S_{n_i})|) = \mathbb{P}(S_r > 0, S_{n_i} \leq 0) + \mathbb{P}(S_r \leq 0, S_{n_i} > 0).$$

Podemos mayorar las dos probabilidades anteriores utilizando la desigualdad de Tchebyshev:

$$\begin{aligned}
\mathbb{P}(S_{n_i} > 0, S_r \leq 0) &= \mathbb{P}(S_{n_i} \geq \varepsilon\sqrt{n_i}, S_r \leq 0) + \mathbb{P}(0 < S_{n_i} < \varepsilon\sqrt{n_i}, S_r \leq 0) \\
&\leq \mathbb{P}(S_{n_i} - S_r \geq \varepsilon\sqrt{n_i}) + \mathbb{P}(0 < S_{n_i} < \varepsilon\sqrt{n_i}) \\
&\leq \frac{n_i - r}{\varepsilon^2 n_i} + \mathbb{P}(0 < S_{n_i} < \varepsilon\sqrt{n_i}), \\
\mathbb{P}(S_{n_i} \leq 0, S_r > 0) &= \mathbb{P}(S_{n_i} \leq -\varepsilon\sqrt{n_i}, S_r > 0) + \mathbb{P}(-\varepsilon\sqrt{n_i} < S_{n_i} \leq 0, S_r > 0) \\
&\leq \mathbb{P}(S_{n_i} - S_r \leq -\varepsilon\sqrt{n_i}) + \mathbb{P}(-\varepsilon\sqrt{n_i} < S_{n_i} \leq 0) \\
&\leq \frac{n_i - r}{\varepsilon^2 n_i} + \mathbb{P}(-\varepsilon\sqrt{n_i} < S_{n_i} \leq 0).
\end{aligned}$$

De esta manera, obtenemos la desigualdad

$$\mathbb{E}(|\psi(S_r) - \psi(S_{n_i})|) \leq \frac{2(n_i - r)}{\varepsilon^2 n_i} + \mathbb{P}(|S_{n_i}| < \varepsilon\sqrt{n_i}),$$

que nos permite obtener una mayoración para la cola de la distribución de $|D_n|$, en términos de su esperanza:

$$\mathbb{E}(|D_n|) \leq \frac{2}{n}\sum_{i=1}^{k}\sum_{r=n_{i-1}+1}^{n_i}\frac{(n_i - r)}{\varepsilon^2 n_i} + \sum_{i=1}^{k}\frac{n_i - n_{i-1}}{n}\mathbb{P}(|S_{n_i}| < \varepsilon\sqrt{n_i})$$

$$(12) \qquad = \sum_{i=1}^{k}\frac{1}{\varepsilon^2}\frac{(n_i - n_{i-1} - 1)(n_i - n_{i-1})}{2nn_i} + \sum_{i=1}^{k}\frac{n_i - n_{i-1}}{n}\mathbb{P}(|S_{n_i}| < \varepsilon\sqrt{n_i}),$$

donde se ha utilizado la igualdad

$$\sum_{r=n_{i-1}+1}^{n_i}\frac{n_i - r}{n_i} = \sum_{r=1}^{n_i - n_{i-1} - 1}\frac{r}{n_i} = \frac{(n_i - n_{i-1} - 1)(n_i - n_{i-1})}{2n_i}.$$

A la cota para $\mathbb{E}(|D_n|)$ dada en (12) la llamaremos $R(n,\varepsilon,k)$ y podemos encontrar su límite cuando $n \longrightarrow \infty$ al utilizar el Teorema Central del Límite y el que n_i/n tienda a i/k conforme $n \longrightarrow \infty$, puesto que entonces:

$$\lim_{n\to\infty}\frac{n_i - n_{i-1} - 1}{n} = \frac{i - (i-1)}{k} = \frac{1}{k}$$

y

$$\lim_{n\to\infty}\frac{n_i - n_{i-1}}{n_i} = \frac{1}{i}$$

por lo que

$$\lim_{n\to\infty} R(n,\varepsilon,k) = \frac{1}{k\varepsilon^2}\sum_{i=1}^{k}\frac{1}{i} + \frac{\sqrt{2}}{\sqrt{\pi}}\int_0^{\varepsilon} e^{-x^2/2}dx$$

y como el doble la densidad normal estándar es menor o igual a 1 (de lo cual se sigue que su integral de 0 a ε se encuentre acotada por ε) y la suma de los primeros k enteros positivos está acotada por $1 + \log k$, pues

$$\sum_{i=1}^{k} i = 1 + \sum_{i=2}^{k} i \leq 1 + \sum_{i=2}^{k}\int_{[i-1,i]}\frac{1}{x}\,dx = 1 + \log k,$$

entonces

$$\lim_{n\to\infty} R(n,\varepsilon,k) \leq \frac{1+\log k}{k\varepsilon^2} + \varepsilon.$$

Así, finalmente se obtiene una cota superior para $\mathbb{P}(|D_n| \geq \delta)$, a saber,

$$R(n,\varepsilon,k)/\delta.$$

Para dos variables aleatorias X y Y se tienen las cotas

$$\begin{aligned}\mathbb{P}(Y < \alpha - \delta) - \mathbb{P}(|X - Y| \geq \delta) &\leq \mathbb{P}(X < \alpha) \\ &\leq \mathbb{P}(Y < \alpha + \delta) + \mathbb{P}(|X - Y| \geq \delta),\end{aligned}$$

puesto que

$$\begin{aligned}\mathbb{P}(X < \alpha) &= \mathbb{P}(X < \alpha, |X - Y| < \delta) + \mathbb{P}(X < \alpha, |X - Y| \geq \delta) \\ &\leq \mathbb{P}(Y < \alpha + \delta) + \mathbb{P}(|X - Y| \geq \delta)\end{aligned}$$

y

$$\begin{aligned}\mathbb{P}(X < \alpha) &\geq \mathbb{P}(X < \alpha, Y < \alpha - \delta) \\ &\geq \mathbb{P}(Y < \alpha - \delta, |X - Y| < \delta) \\ &\geq \mathbb{P}(Y < \alpha - \delta) - \mathbb{P}(Y < \alpha - \delta, |X - Y| \geq \delta) \\ &\geq \mathbb{P}(Y < \alpha - \delta) - \mathbb{P}(|X - Y| \geq \delta).\end{aligned}$$

Al aplicarlas con $X_n = N_n/n$ y $Y_{n,k} = \frac{1}{n}\sum_{i=1}^{k}(n_i - n_{i-1})\psi(S_{n_i})$, al recordar que de acuerdo a (11), $X_n - Y_{n,k} = D_n$ y también la mayoración para $\mathbb{P}(|D_n| \geq \delta)$, obtenemos

$$\begin{aligned}\mathbb{P}(Y_{n,k} < \alpha - \delta) - \frac{R(n,\varepsilon,k)}{\delta} &\leq \mathbb{P}(N_n < \alpha) \\ &\leq \mathbb{P}(Y_{n,k} < \alpha + \delta) + \frac{R(n,\varepsilon,k)}{\delta}.\end{aligned} \tag{13}$$

Ahora notemos que $(Y_{n,k})_{n\in\mathbb{N}}$ converge en distribución para cada $k \in \mathbb{N}$ fija, a un límite continuo independiente de la distribución de X_i, esto es, $\mathbb{P}(Y_{k,n} < \alpha) \to p_k(\alpha)$ conforme $n \to \infty$, donde p_k no depende de la distribución de X_i. Esto sucede, pues al utilizar el Teorema Central del Límite, sabemos que

$$\frac{1}{\sqrt{n}}(S_{n_1}, \ldots, S_{n_k})$$

converge en débilmente a un vector gaussiano Z con media cero y matriz de varianza-covarianza

$$V = \left(\frac{i \wedge j}{k}\right)_{1\leq i,j\leq k}.$$

Como la asignación

$$(s_1, \ldots, s_k) \mapsto (\psi(s_1), \ldots, \psi(s_k))$$

es continua salvo en un conjunto al que la medida de probabilidad inducida por el vector gaussiano es cero (por ejemplo, es continua en $\{s_1 \neq 0, \ldots, s_n \neq 0\}$), se sigue que

$$\left(\frac{1}{k}\psi(S_{n_1}), \ldots, \frac{1}{k}\psi(S_{n_k})\right) = \left(\frac{1}{k}\psi(S_{n_1}/\sqrt{n}), \ldots, \frac{1}{k}\psi(S_{n_k}/\sqrt{n})\right)$$

converge en distribución a $(\psi(Z_1)/k, \ldots, \psi(Z_k)/k)$, donde Z es el anterior vector gaussiano. Finalmente, como $(n_i - n_{i-1})/n$ converge a $1/k$, vemos que

$$\left(\frac{n_1 - n_0}{n}\psi(S_{n_1}), \ldots, \frac{n_k - n_{k-1}}{n}\psi(S_{n_k})\right)$$

converge en distribución a $(\psi(Z_1)/k, \ldots, \psi(Z_k)/k)$, por lo que finalmente

$$Y_{n,k} = \sum_{i=1}^{k} \frac{n_i - n_{i-1}}{n}\psi(S_{n_i})$$

converge en distribución a

$$\frac{1}{k}\sum_{i=1}^{k}\psi(Z_i),$$

por lo que, como se afirmó, $Y_{n,k}$ converge débilmente cuando $n \to \infty$ a un límite cuya distribución no depende de la distribución de X_i.

Para culminar, se tiene que verificar que N_n/n converge débilmente a un límite cuya distribución no depende de la distribución de X_i. Para esto, tomemos límites inferiores y superiores cuando $n \to \infty$ en (13), para obtener

$$\begin{aligned} p_k(\alpha - \delta) - \frac{1 + \log k}{k\delta\varepsilon^2} - \frac{\varepsilon}{\delta} &\leq \liminf_n \mathbb{P}\left(\frac{N_n}{n} < \alpha\right) \\ &\leq \limsup_n \mathbb{P}\left(\frac{N_n}{n} < \alpha\right) \\ &\leq p_k(\alpha + \delta) + \frac{1 + \log k}{k\delta\varepsilon^2} + \frac{\varepsilon}{\delta}, \end{aligned} \tag{14}$$

donde $p_k(\alpha)$ es igual al límite conforme $n \to \infty$ de $\mathbb{P}(Y_{n,k} < \alpha)$, que es independiente de la distribución de X_i. Como en el caso en el que X_i se distribuye normal estandar se ha probado en el capítulo anterior que

$$\lim_{n\to\infty} \mathbb{P}(N_n/n < \alpha) = p(\alpha),$$

con p igual a la distribución arcoseno, se sigue que, reemplazando primero α por $\alpha + \delta$ en la primera desigualdad de 14, α por $\alpha - \delta$ en la segunda y tomando el límite cuando $n \to \infty$, para esta distribución particular se tiene que:

$$\begin{aligned} p(\alpha - \delta) - \frac{1 + \log k}{k\varepsilon^2\delta} - \frac{\varepsilon}{\delta} &\leq p_k(\alpha) \\ &\leq p(\alpha + \delta) + \frac{1 + \log k}{k\varepsilon^2\delta} + \frac{\varepsilon}{\delta}, \end{aligned}$$

de lo cual obtenemos una cota para p_k que no depende de la distribución de X_i, por lo cual utilizando de nueva cuenta las desigualdades (13), obtenemos

$$\begin{aligned} p(\alpha - 2\delta) - 2\frac{1+\log k}{k\varepsilon^2\delta} - 2\frac{\varepsilon}{\delta} &\leq \liminf_n \mathbb{P}\left(\frac{N_n}{n} < \alpha\right) \\ &\leq \limsup_n \mathbb{P}\left(\frac{N_n}{n} < \alpha\right) \\ &\leq p(\alpha + 2\delta) + 2\frac{1+\log k}{k\varepsilon^2\delta} + 2\frac{\varepsilon}{\delta}. \end{aligned} \tag{15}$$

Finalmente, si hacemos $\varepsilon = k^{-1/4}$, $\delta = k^{-1/8}$ en (15) y tomamos el límite cuando $k \to \infty$, utilizando la continuidad de p, se concluye la ley arcoseno para el número de sumas positivas:

$$\lim_{n\to\infty} \mathbb{P}(N_n/n < \alpha) = p(\alpha),$$

lo cual sucede independientemente de la distribución de X, siempre y cuando tenga esperanza cero y varianza uno.[1]

2.3. El principio de invariancia de Donsker

En la sección anterior, consideramos una caminata aleatoria $S_n = \sum_{n=1}^{\infty} X_i$, donde $0 < \mathrm{Var}(X_i) < \infty$ y $\mathbb{E}(X_i) = 0$ y probamos un resultado sobre la convergencia en distribución de la fracción de elementos positivos. La técnica fué encontrar un principio de invarianza para la distribución límite y verificar la veracidad del resultado para una caminata particular. Ahondaremos un poco en la generalización de este procedimiento, pues permite entender la relación existente entre las caminatas aleatorias y otro proceso estocástico llamado Movimiento Browniano, que tiene como consecuencia una infinidad de principios de invariancia como el obtenido anteriormente.

El Movimiento Browniano es una colección de variables aleatorias $(B_t)_{t\geq 0}$ definidas en un espacio de probabilidad $(\Omega, \mathscr{F}, \mathbb{P})$ con las siguientes características:

i) $\mathbb{P}(B_0 = 0) = 1$.
ii) $B_t - B_s$ se distribuye $N(0, t-s)$ y es independiente de $\sigma(B_u : 0 \leq u \leq s)$.
iii) La función $t \mapsto B_t(\omega)$ es continua casi seguramente con respecto a $\mathbb{P}$.

Una de las primeras cuestiones a establecer es la existencia de un proceso estocástico que cumpla la descripción anterior. Este punto se esclarecerá al mismo tiempo que se encuentre la relacion entre las caminatas aleatorias y el Movimiento Browniano para lo cual se vuelve necesario introducir ciertos conceptos, como el de convergencia en distribución en espacios más generales que $\mathbb{R}$ o $\mathbb{R}^d$. En concreto, trabajaremos con un espacio de funciones, el espacio de funciones continuas de $[0,\infty)$ en $\mathbb{R}$, al que podemos dotar de una métrica adecuada para que resulte un espacio métrico completo y separable ó espacio

[1]Esto último no representa una restricción ya que si $\mathrm{Var}(X_i) > 0$, podemos considerar a la variable $X_i/\sqrt{\mathrm{Var}(X_i)}$, para la cual el resultado es válido, sin modificar a la fracción de sumas positivas.

polaco, y tal que la convergencia en esa métrica sea la misma que la convergencia uniforme sobre compactos. Sobre este espacio construiremos una medida de probabilidad, la medida de Wiener, bajo la cual las proyecciones $(f \mapsto f(t))_{t\geq 0}$ constituyen un Movimiento Browniano, y la medida será dada como un límite, llamado límite debil, de medidas de probabilidad asociadas a caminatas aleatorias. A mi parecer, la relación que se obtiene entre las caminatas aleatorias y el Movimiento Browniano contenida en el principio de invariancia de Donsker es importante pues permite entender mejor a ambos procesos al poder traducir ciertos resultados sobre caminatas aleatorias al movimiento browniano y viceversa. Un ejemplo de esto es que podemos formalizar el argumento heurístico detrás del principio de reflexión para el Movimiento Browniano a través de aproximaciones por caminatas aleatorias.

El plan de acción es el siguiente:

1. Consideraremos el espacio de funciones continuas de $[0,\infty)$ en $\mathbb{R}$, al cual dotaremos de una métrica que lo convierta en un espacio métrico completo y separable.
2. Daremos algunas propiedades de este espacio que resultarán de utilidad.
3. Se presentará el concepto de convergencia débil de medidas de probabilidad en espacios más generales que $\mathbb{R}$.
4. Se especializará lo anterior al espacio métrico que nos concierne.
5. Se introducirá el concepto de distribuciones finito dimensionales, que son distribuciones en $\mathbb{R}^n$, y que caracterizan a la ley de un proceso estocástico continuo, situación similar a la del teorema de Kolmogorov[2], pero distinta en escencia. Estas distribuciones son útiles para estudiar la convergencia débil.
6. Finalmente, se dará el resultado principal, el principio de invariancia de Donsker, que permite entender una relación entre el movimiento browniano y las caminatas aleatorias.

2.3.1. El espacio de funciones continuas de $[0,\infty)$ en $\mathbb{R}$. Consideremos a

$$S = \mathscr{C}([0,\infty)),$$

el espacio de funciones continuas de $[0,\infty)$ en $\mathbb{R}$, y a $d : S\times S \to \mathbb{R}$ dada por

$$d(f_1,f_2) = \sum_{n=1}^{\infty} \frac{1}{2^n} \max_{0\leq t\leq n} \left(|f_1(t) - f_2(t)| \wedge 1\right).$$

Entonces d es una métrica en S, pues:

1. $d(f_1,f_2) \geq 0$ por ser cada uno de los sumandos que componen a la serie que define a esta cantidad no-negativos.
2. $d(f_1,f_2) = 0$ si y sólo si $f_1 = f_2$, ya que si $d(f_1,f_2) = 0$, entonces

$$\max_{0\leq t\leq n} \left(|f_1(t) - f_2(t)| \wedge 1\right) = 0 \quad \forall n \in \mathbb{N},$$

[2] Ver [12].

por lo que $f_1(t) = f_2(t)$ para todo $t \in [0, n]$ para toda $n \in \mathbb{N}$, de donde $f_1(t) = f_2(t)$ para todo $t \in [0, \infty)$. Por otro lado, $d(f, f) = 0$.

3. $d(f_1, f_2) = d(f_2, f_1)$ ya que $|f_1(t) - f_2(t)| = |f_2(t) - f_1(t)|$.
4. $d(f_1, f_3) \leq d(f_1, f_2) + d(f_2, f_3)$, puesto que
$$a \leq b \text{ implica } a \wedge 1 \leq b \wedge 1$$
y
$$0 \leq a \leq b \text{ implica } (a + b) \wedge 1 \leq (a \wedge 1) + (b \wedge 1),$$
que aplicado a
$$0 \leq |f_3(t) - f_1(t)| \leq |f_2(t) - f_1(t)| + |f_3(t) - f_2(t)|$$
junto con la desigualdad
$$\max_{x \in [a,b]} (f(x) + g(x)) \leq \max_{x \in [a,b]} f(x) + \max_{x \in [a,b]} g(x)$$
válida para funciones f, g continuas en $[a, b]$ nos da la desigualdad
$$\max_{0 \leq t \leq n} (|f_3(t) - f_1(t)| \wedge 1) \leq \left(\max_{0 \leq t \leq n} |f_2(t) - f_1(t)| \wedge 1\right) + \left(\max_{0 \leq t \leq n} |f_3(t) - f_2(t)| \wedge 1\right),$$
de la cual sigue la desigualdad del triángulo.

Aún más, se tiene lo siguiente:

TEOREMA 2.1. *(S, d) es un espacio métrico completo y separable.*

DEMOSTRACIÓN. Utilizaremos la notación
$$a_n(f, g) = \max_{0 \leq t \leq n} (|f(t) - g(t)| \wedge 1).$$

Sea $(f_k)_{k \in \mathbb{N}}$ una sucesión de Cauchy en (S, d). Notemos que si $\varepsilon = 1/2^n$, entonces existe $N \in \mathbb{N}$ tal que si $l, m \geq N$ se cumple la desigualdad $d(f_l, f_m) < \varepsilon/2^n$ y que esta desigualdad se puede cumplir sólo si $a_k(f_l, f_m) < \varepsilon$ para $k = 1, \ldots, n$, por lo que para cada $\varepsilon > 0$ existe $N \in \mathbb{N}$ tal que si $l, m \geq N$, la norma uniforme de $f_l - f_m$ sobre $[0, n]$,
$$\max_{t \in [0,n]} |f_l(t) - f_m(t)|,$$
es menor a ε. Esto significa que la sucesión $(f_n)_{n \in \mathbb{N}}$ es de Cauchy uniforme sobre todo intervalo $[0, n]$, por lo que la sucesión numérica $(f_k(t))_{k \in \mathbb{N}}$ converge para todo $t \in [0, \infty)$. De hecho, por ser de Cauchy uniforme en $[0, n]$, la sucesión $(f_k)_{k \in \mathbb{N}}$ converge uniformemente en $[0, n]$ a una función continua f^n y por la unicidad del límite, la igualdad $f^n = f^m$ se cumple sobre $[0, n \wedge m]$. Así, podemos definir $f : [0, \infty) \to \mathbb{R}$ por $f(t) = f^n(t)$ para $t \in [0, n]$ y sólo falta demostrar que $d(f_n, f) \to 0$ conforme $n \to \infty$. Sea $\varepsilon > 0$, $N \in \mathbb{N}$ tal

que $(1/2^{N-1}) < \varepsilon$ y $K \in \mathbb{N}$ tal que si $k \geq K$ entonces $a_n(f_k, f) < \varepsilon/2$ para $n \leq N$ (se puede lograr gracias a la convergencia uniforme de $(f_k)_{k \in \mathbb{N}}$ a f en $[0, N]$). Entonces para $k \geq K$, como $a_i(f_k, f) \leq 1$:

$$\begin{aligned}
d(f_k, f) &= \sum_{i=1}^{N} \frac{1}{2^i} a_i(f_k, f) + \sum_{i=N+1}^{\infty} \frac{1}{2^i} a_i(f_k, f) \\
&< \frac{\varepsilon}{2} \sum_{i=1}^{N} \frac{1}{2^i} + \sum_{i=N+1}^{\infty} \frac{1}{2^n} \\
&= \frac{\varepsilon}{2}(1 - 1/2^N) + \frac{1}{2^N} \\
&< \frac{\varepsilon}{2}(1 - 1/2^n) + \varepsilon/2 \\
&< \varepsilon,
\end{aligned}$$

por lo que $d(f_n, f) \longrightarrow 0$ y (S, d) es un espacio métrico completo.

Para ver que (S, d) es separable, sea $f \in S$, de donde f es continua en $[0, n]$ para cualquier $n \in \mathbb{N}$, por lo que para cualquier $\varepsilon > 0$ existe un polinomio p_ε con coeficientes racionales tal que $\|f - p_\varepsilon\|_{[0,n]} < \varepsilon/2$. Si $n \in \mathbb{N}$ es tal que $1/2^{n-1} < \varepsilon$, entonces, como en el párrafo anterior, $d(f, p_\varepsilon) < \varepsilon$ y así, el conjunto de polinomios con coeficientes racionales es denso en (S, d). Como el anterior conjunto es numerable, se sigue que (S, d) es separable. $\square$

En el espacio (S, d) utilizaremos la σ-álgebra de Borel $\mathscr{B}_S$. Esta se define como la σ-álgebra más pequeña que contiene a la topología inducida en S por la métrica d. Si consideramos a las proyecciones $\pi_t : S \longrightarrow \mathbb{R}$ dadas por $\pi_t(f) = f(t)$, entonces estas funciones son continuas, puesto que si $(f_n)_{n \in \mathbb{N}}$ converge a f en (S, d), en particular converge puntualmente, de donde $\pi_t(f_n) = f_n(t) \longrightarrow f(t) = \pi_t(x)$ conforme $n \longrightarrow \infty$. La continuidad de estas funciones hace que sean $(\mathscr{B}_S, \mathscr{B}_{\mathbb{R}})$-medibles, esto es, que para cualquier Boreliano de $\mathbb{R}$, B, $\pi_t^{-1}(B) \in \mathscr{B}_S$. Por medio de las proyecciones, podemos definir a los cilindros finito dimensionales, que son conjuntos de la forma $\{(f(t_1), \ldots, f(t_n)) \in B\}$, para B un elemento de la σ-álgebra generada por la topología usual de $\mathbb{R}^n$, $\mathscr{B}_{\mathbb{R}^n}$, $n \geq 1$ y $t_1, \ldots, t_n \in [0, \infty)$. Dado que $f \longmapsto (f(t_1), \ldots, f(t_n))$ es continua, entonces es $(\mathscr{B}_S, \mathscr{B}_{\mathbb{R}^n})$-medible, por lo que cualquier cilindro pertenece a $\mathscr{B}_S$. Si denotamos por $\mathscr{C}$ a la familia de cilindros finito dimensionales, la anterior observación se reduce a $\mathscr{C} \subset \mathscr{B}_S$, y lo que ahora verificaremos es que de hecho, $\sigma(\mathscr{C}) = \mathscr{B}_S$. Denotaremos por $\mathscr{G}$ a $\sigma(\mathscr{C})$.

La contención $\mathscr{G} \subset \mathscr{B}_S$ se sigue de $\mathscr{C} \subset \mathscr{B}_S$ y la definición de σ-álgebra generada por una familia de subconjuntos, por lo que sólo hace falta demostrar que $\mathscr{B}_S \subset \mathscr{G}$. Para esto, sea

$$\mathscr{C}_1 = \left\{\pi_t^{-1}(A) : A \in \mathscr{B}_{\mathbb{R}}\right\},$$

la familia de los cilindros 1-dimensionales (contenida en $\mathscr{C}$ y por lo tanto en $\mathscr{B}_S$) y $\mathscr{G}_1 = \sigma(\mathscr{C}_1)$. Entonces

$$\mathscr{G}_1 \subset \mathscr{G} \subset \mathscr{B}_S$$

y para demostrar que de hecho todas las σ-álgebras son iguales, bastaría verificar que todo abierto en S pertenece a $\mathscr{G}_1$, cosa que haremos a continuación: recordemos que todo espacio métrico separable (X, ρ) admite una base numerable de la topología inducida en X por la métrica ρ. Explícitamente, la demostración de este resultado nos dice que si $U \subset X$ es abierto, entonces U se puede escribir como

$$U = \bigcup_{i=1}^{\infty} B_{\epsilon_i}(f_i) .$$

Así, para verificar que todo abierto es elemento de $\mathscr{G}_1$, es suficiente verificar que una bola $B_\epsilon(f_0)$ pertenece a $\mathscr{G}_1$, puesto que las σ-álgebras son cerradas bajo uniones numerables. Como la función π_t es $(\mathscr{G}_1, \mathscr{B}_{\mathbf{R}})$-medible, se sigue que la función $f \mapsto |f(t) - f_0(t)| \wedge 1$ es $(\mathscr{G}_1, \mathscr{B}_{\mathbf{R}})$-medible. Por la continuidad de f y f_0, se cumple:

$$\sup_{0 \leq t \leq n} (|f(t) - f_0(t)| \wedge 1) = \sup_{\substack{0 \leq t \leq n \\ t \in \mathbf{Q}}} (|f(t) - f_0(t)| \wedge 1) ,$$

entonces $f \mapsto \sup_{0 \leq t \leq n} (|f(t) - f_0(t)| \wedge 1)$ es $(\mathscr{G}_1, \mathscr{B}_{\mathbf{R}})$-medible como supremo numerable de funciones medibles respecto a esas σ-álgebras. Se sigue que cada uno de los sumandos que define a $d(f, f_0)$ son $(\mathscr{G}_1, \mathscr{B}_{\mathbf{R}})$-medibles como función de f para f_0 fija, por lo que la función $f \mapsto d(f, f_0)$ es $(\mathscr{G}_1, \mathscr{B}_{\mathbf{R}})$-medible como límite de funciones medibles respecto a esas σ-álgebras y por lo tanto $B_\epsilon(f_0) = \{f : d(f, f_0) < \varepsilon\} \in \mathscr{G}_1$ como imagen inversa de un boreliano bajo una función medible. Así, hemos verificado el siguiente teorema:

TEOREMA 2.2. *Sean $\mathscr{C} \subset \mathscr{P}(S)$ la familia que consta de los cilindros finito dimensionales de S, $\mathscr{C}_1$ la familia de los cilindros 1-dimensionales y $\mathscr{B}_S$ la σ-álgebra generada por la topología inducida en S por la métrica d. Entonces $\sigma(\mathscr{C}_1) = \sigma(\mathscr{C}) = \mathscr{B}_S$.*

El Teorema 2.2 nos permite afirmar que si dos medidas de probabilidad, $\mathbb{P}$ y $\mathbb{P}'$, sobre $(S, \mathscr{B}_S)$ coinciden en $\mathscr{C}$, entonces también coindicen en $\mathscr{B}_S$. Esto sucede pues $\mathscr{C}$ es un π-sistema, ya que $S = \{f \in S : f(0) \in \mathbb{R}\}$ y si $C_1, C_2 \in \mathscr{C}$, donde

$$C_1 = \{(f_{t_1}, \ldots, f_{t_n}) \in B_1\} ,$$
$$C_2 = \{(f_{s_1}, \ldots, f_{s_m}) \in B_2\} ,$$

con $B_1 \in \mathscr{B}_{\mathbf{R}^n}$ y $B_2 \in \mathscr{B}_{\mathbf{R}^m}$, entonces

$$C_1 = \{(f(t_1), \ldots, f(t_n), f(s_1), \ldots, f(s_m)) \in B_1 \times \mathbb{R}^m\} ,$$
$$C_2 = \{(f(t_1), \ldots, f(t_n), f(s_1), \ldots, f(s_m)) \in \mathbb{R}^n \times B_2\} \text{ y}$$
$$C_1 \cap C_2 = \{(f(t_1), \ldots, f(t_n), f(s_1), \ldots, f(s_m)) \in B_1 \times B_2\} ,$$

por lo que $C_1 \cap C_2 \in \mathscr{C}$ (pues $B_1 \times B_2 \in \mathscr{B}_{\mathbb{R}^{n+m}}$). Además, la familia de elementos de $\mathscr{B}_S$ en el cual las dos medidas de probabilidad, $\mathbb{P}$ y $\mathbb{P}'$ coinciden, que denotaremos por Λ, es un λ-sistema, puesto que S pertenece a él y si $A, B \in \Lambda$ y $A \subset B$, entonces $B \setminus A \in \Lambda$, pues

$$\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A) = \mathbb{P}'(B) - \mathbb{P}'(A) = \mathbb{P}'(B \setminus A).$$

Además, si $A_n \in \Lambda$ y $A_n \subset A_{n+1}$ para $n \in \mathbb{N}$, entonces $\cup_{n\in\mathbb{N}} A_n \in \Lambda$, puesto que

$$\mathbb{P}(\cup_{n\in\mathbb{N}} A_n) = \lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} \mathbb{P}'(A_n) = \mathbb{P}'(\cup_{n\in\mathbb{N}} A_n).$$

Como $\Lambda \subset \mathscr{B}_S$ por definición, si $\mathbb{P}$ y $\mathbb{P}'$ coinciden en $\mathscr{C}$, o dicho de otra manera, si $\mathscr{C} \subset \Lambda$, el lema de clases de Sierpinski nos permite afirmar que $\mathscr{B}_S = \sigma(\mathscr{C}) \subset \Lambda$, por lo que $\Lambda = \mathscr{B}_S$ y por lo tanto $\mathbb{P}$ y $\mathbb{P}'$ coinciden en $\mathscr{B}_S$.

Otra consecuencia del Teorema 2.2 es que si $(\Omega, \mathscr{F})$ es un espacio medible y tenemos una función $X : \Omega \to S$, entonces X es $(\mathscr{F}, \mathscr{B}_S)$-medible si y sólo si $\pi_t \circ X$ es $(\mathscr{F}, \mathscr{B}_{\mathbb{R}})$-medible para cada $t \in [0, \infty)$. Esto sucede pues para verificar que X es $(\mathscr{F}, \mathscr{B}_S)$-medible, basta probar que la imagen inversa de un cilindro de dimensión 1 bajo X pertenece a $\mathscr{F}$, lo cual claramente sucede si $\pi_t \circ X$ es $(\mathscr{F}, \mathscr{B}_{\mathbb{R}})$-medible para cada $t \in [0, \infty)$. La otra implicación es consecuencia de la continuidad de π_t. Si suponemos por un momento que tenemos una colección de variables aleatorias $X_t : \Omega \to \mathbb{R}$ para cada $t \geq 0$ tal que para cada $\omega \in \Omega$ la aplicación $\omega \mapsto X_t(\omega)$ sea continua, podemos definir $X : \Omega \to S$ mediante $\pi_t \circ X = X_t$, y el argumento anterior nos muestra que X es $(\mathscr{F}, \mathscr{B}_S)$-medible.

Otra aplicación de la igualdad $\mathscr{C} = \mathscr{B}_S$ será demostrar que la ley de dos procesos continuos está caracterizada por las distribuciones finito dimensionales, cosa que haremos más adelante.

2.3.2. Convergencia en Distribución. Un ejemplo de convergencia en distribución lo encontramos en el Teorema Central del Límite. Este resultado afirma que si $(X_i)_{i=1}^{\infty}$ son variables aleatorias independientes e idénticamente distribuidas, con varianza finita σ^2 y media cero, y $S_n = \sum_{i=1}^{n} X_i$, entonces

$$\mathbb{P}\Big(S_n/\sqrt{n\sigma^2} \leq x\Big) \to \int_{-\infty}^{x} \Big(e^{-t^2/2}/\sqrt{2\pi}\Big)\, dt.$$

En otras palabras, la distribución de $S_n/\sqrt{n\sigma^2}$ converge a la distribución normal estándar puntualmente cuando $n \to \infty$. Si X y $X_i, i = 1, 2, \ldots$ son variables aleatorias con valores en $\mathbb{R}$, decimos que X_i converge en distribución a X si la sucesión de funciones de distribución asociadas a las variables $(X_i)_{i\geq 1}$, $(F_{X_i}(x))_{i=1}^{\infty}$, converge a la distribución de X, $F_X(x)$, conforme $i \to \infty$ cuando x es punto de continuidad de la función F_X. Recordemos que la función de distribución se define para $\mathbb{R}$, y que se puede extender el concepto a $\mathbb{R}^n$; sin embargo, no tiene sentido en espacios más generales, en los que se substituye la función de distribución por la ley que podemos inducir a partir de la variable aleatoria. Es por esto que la extensión del concepto de convergencia en distribución a espacios más generales no es directa, de hecho, esta extensión se basa en una equivalencia para la convergencia en distribución en $\mathbb{R}$, a saber, que $\mathbb{E}(f(X_n)) \longrightarrow \mathbb{E}(f(X))$ conforme $n \longrightarrow \infty$ para cualquier función $f : \mathbb{R} \longrightarrow \mathbb{R}$ que sea continua y acotada.

DEFINICIÓN. Consideremos (M, ρ) un espacio métrico. Sean $(\Omega, \mathscr{F}, \mathbb{P})$ y $(\Omega_k, \mathscr{F}_k, \mathbb{P}_k)$, espacios de probabilidad para cada k de 1 en adelante. Si $X : \Omega \longrightarrow M$ y $X_k : \Omega_n \longrightarrow M$ son funciónes medibles, con respecto a la σ-álgebra de Borel en M, decimos que X_k converge en distribución a X, denotado por $X_k \xrightarrow{\mathscr{L}} X$ cuando $k \longrightarrow \infty$, si $\int f(X_k)\, d\mathbb{P}_k \longrightarrow \int f(X)\, d\mathbb{P}$ cuando $k \longrightarrow \infty$ para cualquier función $f : M \longrightarrow \mathbb{R}$ continua y acotada.

Si consideramos una varable aleatoria Y definida en $(\Omega, \mathscr{F}, \mathbb{P})$, un espacio de probabilidad, y con valores en un espacio métrico (M, ρ), podemos inducir una medida $\mathbb{P}_Y$ sobre $(M, \mathscr{B}_M)$ mediante la asignación $\mathbb{P}_Y(A) = \mathbb{P}(Y \in A)$ para $A \in \mathscr{B}_M$. Si $f : M \longrightarrow \mathbb{R}$ es una función medible tal que la integral de $f(Y)$ esté definida, entonces f es integrable respecto a $\mathbb{P}_Y$ y se tiene la fórmula de cambio de variable

$$\int_\Omega f \circ Y \, d\mathbb{P} = \int_M f \, d\mathbb{P}_Y.$$

Si consideramos el escenario presentado en la definición anterior, entonces las integrales $f \circ X_n$ y $f \circ X$ existen (de hecho son finitas) y podemos escribir la condición para la convergencia en distribución como

$$\int f \, d\mathbb{P}_{X_n} \longrightarrow \int f \, d\mathbb{P}_X \text{ conforme } k \longrightarrow \infty.$$

La discusión anterior se hace con el objeto de presentar la definición de convergencia débil, que tomaremos como punto de partida para las investigaciones subsecuentes.

DEFINICIÓN. Sea (M, ρ) un espacio métrico y $(\mathbb{P}_k)_{k \in \mathbb{N}}$ medidas de probabilidad sobre $\mathscr{B}_M$, la σ-álgebra generada por los abiertos de M. Si $\mathbb{P}$ es otra medida de probabilidad definida sobre $\mathscr{B}_M$, decimos que la sucesión $(\mathbb{P}_k)_{k \in \mathbb{N}}$ converge débilmente a $\mathbb{P}$ cuando $k \longrightarrow \infty$ si para cada función $f : M \longrightarrow \mathbb{R}$ que sea continua y acotada, la sucesión $(\int f \, d\mathbb{P}_k)_{k \in \mathbb{N}}$ converge a $\int f \, d\mathbb{P}$ cuando $k \longrightarrow \infty$. Utilizaremos la notación $\Rightarrow$ para la convergencia débil.

El siguiente resultado nos dá algunas equivalencias para la convergencia débil, que resultan convenientes al trabajar con este concepto, se le conoce como el Teorema Portmanteau. Una consecuencia importante de este teorema es que nos proporciona condiciones suficientes sobre un conjunto $A \in \mathscr{B}_M$ para que se cumpla la relación

$$\lim_{n \to \infty} \mathbb{P}_k(A) = \mathbb{P}(A),$$

que en el caso en que M sea $\mathbb{R}$ se reducen a $\mathbb{P}(\{x\}) = 0$ para conjuntos A de la forma $(-\infty, x)$.

TEOREMA 2.3 (Teorema Portmanteau). *Sean (M, ρ) un espacio métrico, $(\mathbb{P}_k)_{k \in \mathbb{N}}$ medidas de probabilidad sobre $\mathscr{B}_M$ y $\mathbb{P}$ otra medida de probabilidad sobre $\mathscr{B}_M$. Son equivalentes:*

a) $\mathbb{P}_k \Rightarrow \mathbb{P}$.
b) $\int f\, d\mathbb{P}_k \longrightarrow \int f\, d\mathbb{P}$ *conforme* $k \to \infty$ *para cualquier función* $f : M \longrightarrow \mathbb{R}$ *que sea uniformemente continua y acotada.*
c) $\limsup_{k\to\infty} \mathbb{P}_k(F) \leq \mathbb{P}(F)$ *para cualquier* F *cerrado.*
d) $\liminf_{k\to\infty} \mathbb{P}_k(G) \leq \mathbb{P}(G)$ *para cualquier* G *abierto.*
e) $\mathbb{P}_k(A) \longrightarrow \mathbb{P}(A)$ *cuando* $k \to \infty$ *para cualquier* $A \in \mathscr{B}_M$ *que cumpla* $\mathbb{P}(\operatorname{Int} A) = \mathbb{P}(\overline{A})$.

NOTA. Si $A \in \mathscr{B}_M$ es tal que $\mathbb{P}(\operatorname{Int} A) = \mathbb{P}(\overline{A})$, decimos que A es un conjunto de continuidad de $\mathbb{P}$.

DEMOSTRACIÓN.

a)⇒b): Esta implicación es inmediata pues cualquier función uniformemente continua y acotada es continua y acotada.

b)⇒c): Sea $F \in \mathscr{B}_M$ cerrado. Para verificar *c)*, aproximaremos la función $\mathbf{1}_F$ por una función que sea uniformemente continua y acotada. Para este efecto, si definimos $\rho(x, A)$ por

$$\rho(x, A) = \inf\{\rho(x,y) : y \in A\},$$

que para cualquier $A \subset M$ distinto del vacío está bien definida pues el conjunto $\{\rho(x,y) : y \in A\}$ es entonces no vacío y acotado inferiormente por 0, entonces la asignación $x \longmapsto \rho(x, A)$ es continua, propiedad que verificamos a continuación: Dado que

$$\rho(x, A) \leq \rho(x, z) \text{ para toda } z \in A,$$

entonces la desigualdad del triángulo nos indica que para cualquier $y \in M$

$$\rho(x, A) \leq \rho(x,y) + \rho(y,z) \text{ para toda } z \in A. \tag{16}$$

Al considerar el ínfimo de las cantidades que aparecen del lado derecho de la desigualdad (16) al variar a $z \in A$, obtenemos la desigualdad $\rho(x, A) \leq \rho(x,y) + \rho(y, A)$ y como x e y son arbitrarios, encontramos que

$$|\rho(x, A) - \rho(y, A)| \leq \rho(x,y),$$

por lo que de hecho, la asignación $x \longmapsto \rho(x, A)$ es uniformemente continua. Sea $f_\varepsilon : M \longrightarrow \mathbb{R}$ dada por

$$f_\varepsilon(x) = (1 - \rho(x, F)/\varepsilon)^+,$$

por lo que

$$|f_\varepsilon(x) - f_\varepsilon(y)| \leq \frac{|\rho(x,F) - \rho(y,F)|}{\varepsilon} \leq \frac{\rho(x,y)}{\varepsilon},$$

de donde concluimos que f_ε es uniformemente continua. Además, $\mathbf{1}_F \leq f_\varepsilon \leq 1$, por lo que f_ε es acotada y de la desigualdad anterior obtenemos

$$\limsup_{k\to\infty} \mathbb{P}_k(F) \leq \limsup_{k\to\infty} \int f_\varepsilon \, d\mathbb{P}_k = \int f_\varepsilon \, d\mathbb{P}.$$

Finalmente, notemos que $\lim_{\varepsilon\downarrow 0} f_\varepsilon = \mathbf{1}_F$, puesto que si $\lim_{\varepsilon\downarrow 0} f_\varepsilon(x) = 1$, entonces $\rho(x, F) = 0$, por lo que para cada $n \in \mathbb{N}$ existe $x_n \in F$ tal que $\rho(x, x_n) < 1/n$, de donde $x \in \overline{F}$, y como F es cerrado, entonces $x \in F$. Como $x \in F$ implica que $\rho(x, F) = 0$ y $f_\varepsilon(x) = 1$, se sigue que $\lim_{\varepsilon\downarrow 0} f_\varepsilon(x) = 1$ si y sólo si $x \in F$. Por otro lado, $f_\varepsilon(x) < 1$ si y sólo si $\rho(x, F) > 0$, y si esto último ocurre, entonces para cada $\varepsilon < \rho(x, F)$, $f_\varepsilon(x) = 0$, de donde $\lim_{\varepsilon\downarrow 0} f_\varepsilon(x) = 0$. Así, como $f_\varepsilon \downarrow \mathbf{1}_F$, y las funciones $(f_\varepsilon)_{\varepsilon>0}$ son uniformemente acotadas, el teorema de convergencia acotada (aplicado a cualquier sucesión $(\varepsilon_n)_{n\in\mathbb{N}}$ tal que $\varepsilon_n \downarrow 0$ conforme $n \to \infty$) nos dice que

$$\int f_{\varepsilon_n} \, d\mathbb{P} \to \mathbb{P}(F)$$

cuando $n \to \infty$, de donde

$$\int f_\varepsilon \, d\mathbb{P} \to \mathbb{P}(F)$$

cuando $\varepsilon \downarrow 0$ y por lo tanto:

$$\limsup_{k\to\infty} \mathbb{P}_k(F) \leq \mathbb{P}(F).$$

c)⇒d): Si $G \in \mathscr{B}_M$ es abierto, entonces G^c es cerrado, por lo que

$$1 - \liminf_{k\to\infty} \mathbb{P}_k(G) = \limsup_{k\to\infty} \mathbb{P}_k(G^c) \leq \mathbb{P}(G^c) = 1 - \mathbb{P}(G)$$

de donde obtenemos

$$\liminf_{k\to\infty} \mathbb{P}_k(G) \geq \mathbb{P}(G).$$

d)⇒e): Si $A \in \mathscr{B}_M$ es tal que

$$\mathbb{P}(\operatorname{Int} A) = \mathbb{P}(\overline{A}),$$

entonces por monotonía de $\mathbb{P}$,

$$\mathbb{P}(A) = \mathbb{P}(\overline{A}).$$

Además, por hipótesis,

$$\liminf_{k\to\infty} \mathbb{P}_k(\overline{A}^c) \geq \mathbb{P}(\overline{A}^c) \quad \text{y}$$
$$\liminf_{k\to\infty} \mathbb{P}_k(\operatorname{Int} A) \geq \mathbb{P}(\operatorname{Int} A),$$

por lo que

$$\begin{aligned}\mathbb{P}(\operatorname{Int} A) &\leq \liminf_{k\to\infty} \mathbb{P}_k(\operatorname{Int} A) \leq \liminf_{k\to\infty} \mathbb{P}_k(A) \\ &\leq \limsup_{k\to\infty} \mathbb{P}_k(A) \leq \limsup_{k\to\infty} \mathbb{P}_k(\overline{A}) \\ &\leq \mathbb{P}(\overline{A}),\end{aligned}$$

de donde obtenemos la igualdad $\lim_{k\to\infty} \mathbb{P}_k(A) = \mathbb{P}(A)$.

c)⇒a): Sea $f : M \longrightarrow \mathbb{R}$ una función continua y acotada. Suponiendo que $\mathbb{P}_k(A) \to \mathbb{P}(A)$ para cualquier $A \in \mathscr{B}_M$ que cumpla $\mathbb{P}(\operatorname{Int} A) = \mathbb{P}(\overline{A})$, debemos demostrar que $\int f\, d\mathbb{P}_k \longrightarrow \int f\, d\mathbb{P}$. Si C es una cota para f, entonces $f + C$ es continua, acotada y no-negativa y $\int f\, d\mathbb{P}_k = \int (f + C)\, d\mathbb{P}_k - C$, que converge a $\int f\, d\mathbb{P}$ cuando $k \longrightarrow \infty$ si y sólo si $\int (f + C)\, d\mathbb{P}_k$ converge a $\int (f + C)\, d\mathbb{P}$ cuando $k \longrightarrow \infty$ y así, basta probar el resultado deseado para funciones continuas, acotadas y no-negativas. Necesitamos un lema previo a la demostración.

LEMA 2.1. *Consideremos (X, S, μ) un espacio de medida arbitrario y $f : X \longrightarrow \mathbb{R}$ una función $(S, \mathscr{B}_{\mathbb{R}})$-medible y no-negativa. Entonces la función de $[0, \infty)$ en $\overline{\mathbb{R}}$ cuya regla de correspondencia es*

$$t \longmapsto \mu\{x \in X : f(x) > t\}$$

es $(\mathscr{B}_{[0,\infty)}, \mathscr{B}_{\overline{\mathbb{R}}})$-medible y

$$\int f\, d\mu = \int_{[0,\infty)} \mu(\{x \in X : f(x) > t\})\, dt.$$

La demostración es como sigue: sea $g : [0, \infty) \longrightarrow \mathbb{R}$ dada por

$$g(t) = \mu(\{x \in X : f(x) > t\}).$$

Como

$$\{x \in X : f(x) > t\} \subset \{x \in X : f(x) > s\}$$

si $t \geq s$, se sigue que g es decreciente. Dado que si $(t_n)_{n\geq 1}$ es una sucesión de reales no-negativos tal que $t_n \uparrow t$ cuando $n \longrightarrow \infty$, entonces

$$\bigcup_{n\geq 1} \{x \in X : f(x) > t_n\} = \{x \in X : f(x) > t\},$$

podemos concluir que g es continua por la izquierda, pues la familia de conjuntos cuya unión estamos considerando es creciente. Para verificar la medibilidad de g, basta verificar que $g^{-1}([c, \infty]) \in \mathscr{B}_{[0,\infty)}$ para cualquier $c \in \mathbb{R}$. Si $c \in \mathbb{R}$ y el conjunto $\{t \in [0, \infty) : g(t) \geq c\}$ es vacío, entonces la imagen inversa de $[c, \infty]$ bajo g es el conjunto vacío, que pertenece a cualquier σ-álgebra y si el conjunto $\{t \in [0, \infty) : g(t) \geq c\}$ es distinto del vacío, sea c^* su supremo (el supremo de un conjunto no acotado superiormente será por convención infinito), se afirma que en este caso, $g^{-1}([c,\infty]) = [0,c^*]$. La razón es que si $t_n \uparrow c^*$ cuando $n \to \infty$, con $g(t_n) \geq c$, entonces $g(t_n) \downarrow g(c^*)$ cuando $n \to \infty$, por lo que $g(c^*) \geq c$. Si $0 \leq t \leq c^*$, entonces $g(t) \geq g(c^*) \geq c$ y si $t > c^*$, entonces $g(t) < c$, por lo que t no pertenece a $g^{-1}([c,\infty])$. Así, g es $(\mathscr{B}_{[0,\infty)}, \mathscr{B}_{\overline{\mathbb{R}}})$-medible.

La segunda afirmación del lema se probará mediante aproximación: Si f es una función simple y no-negativa, con descomposición canónica $\sum_{i=1}^n a_i \mathbf{1}_{A_i}$, donde $a_0 = 0 \leq a_1 < \cdots < a_n$, $A_i \in S$, $i = 1\ldots n$, $A_i \cap A_j$ si $i \neq j$ y $\cup_i A_i = X$, entonces por un lado $\int f\,d\mu = \sum_{i=1}^n a_i \mu(A_i)$ y por otro lado si g está definida como en el párrafo anterior y λ denota a la medida de Lebesgue restringida a $\mathscr{B}_{[0,\infty)}$, entonces

$$\begin{aligned}
\int g\,d\lambda &= \sum_{i=1}^n \int_{[a_{i-1},a_i)} g\,d\lambda \\
&= \sum_{i=1}^n \sum_{j=i}^n (a_i - a_{i-1})\mu(A_j) \\
&= \sum_{j=1}^n a_j \mu(A_j),
\end{aligned}$$

por lo que tiene lugar la igualdad

$$\int f\,d\mu = \int g\,d\lambda = \int_{[0,\infty)} \mu(\{x \in X : f(x) > t\})\,dt,$$

y la conclusión del lema es correcta en el caso en el que f sea simple. En general, si f es no negativa, sabemos que existe una sucesión de funciones simples y no-negativas $(f_n)_{n\geq 1}$ que converge de manera creciente a f cuando $n \to \infty$. Pero esto último implica la igualdad $\cup_{n\geq 1} \{x \in X : f_n(x) > t\} = \{x \in X : f(x) > t\}$ y dado que la sucesión de conjuntos cuya unión estamos considerando es creciente, se sigue que la sucesión de funciones no-negativas $(\mu(\{x \in X : f_n(x) > t\}))_{n\geq 1}$ converge de manera creciente a la función $\mu(\{x \in X : f_n(x) > t\})$. El teorema de convergencia monótona y el resultado ya verificado para funciones simples implican

$$\begin{aligned}
\int f\,d\mu &= \lim_{n\to\infty} \int f_n\,d\mu \\
&= \lim_{n\to\infty} \int_{[0,\infty)} \mu(\{x \in X : f_n(x) > t\})\,d\lambda \\
&= \int_{[0,\infty)} \mu(\{x \in X : f(x) > t\})\,d\lambda.
\end{aligned}$$

Vamos ahora con la demostración de la implicación que nos concierne, si $f : M \to \mathbb{R}$ es una función continua, acotada (digamos por C) y no-negativa, sean $g_k, g : [0,\infty) \to \mathbb{R}$ dadas por $g_k(t) = \mathbb{P}_k(\{x \in M : f(x) > t\})$ (análogamente se define g utilizando a $\mathbb{P}$). Entonces g es decreciente, por lo que tiene una cantidad a lo más numerable de discontinuidades y como g es continua en t si y sólo si $\mathbb{P}(\{x \in M : f(x) = t\}) = 0$, la contención $\partial(\{x \in M : f(x) > t\}) \subset \{x \in M : f(x) = t\}$ nos permite afirmar que, excepto en un conjunto a lo más numerable,

$$\mathbb{P}(\mathrm{Int}\,\{x \in M : f(x) > t\}) = \mathbb{P}\left(\overline{\{x \in M : f(x) > t\}}\right),$$

por lo que $g_k(t) \to g(t)$ cuando $k \to \infty$ con excepción de un conjunto a lo más numerable. Finalmente, como $\int f\, d\mathbb{P}_k = \int_{[0,C]} g_n\, d\lambda$ y $\int f\, d\mathbb{P} = \int_{[0,C]} g\, d\lambda$ (donde λ es la medida de Lebesgue restringida a $[0, C]$, ya que $|g_k| \leq 1 \in \mathscr{L}_1(\lambda)$ para toda k, el teorema de convergencia acotada nos permite afirmar que

$$\int f\, d\mathbb{P}_k \to \int f\, d\mathbb{P}.$$

□

Si nos imaginamos por un momento que el conjunto de medidas de probabilidad sobre la σ-álgebra de Borel de un espacio métrico fuese metrizable de tal manera que la convergencia utilizando esa métrica fuera igual a la convergencia en distribución, entonces la compacidad de una familia de esas medidas de probabilidad sería equivalente a la compacidad por sucesiones de la misma. Las propiedades de los conjuntos compactos, tales como la existencia de puntos de acumulación en sus subconjuntos infinitos, podrían llevarnos a verificar la existencia de límites débiles para medidas de probabilidad, cosa que puede resultar difícil.[3] Es en este sentido que se introduce el siguiente concepto.

DEFINICIÓN. Sea (M, ρ) un espacio métrico y Π una familia de medidas de probabilidad sobre $\mathscr{B}_M$. Decimos que Π es relativamente compacta si cada sucesión de elementos de Π admite una subsucesión que converge débilmente.

Por ejemplo, si consideramos a $\Pi = \{\mathbb{P}_k : k \in \mathbb{N}\}$ una sucesión de medidas de probabilidad sobre la σ-álgebra de Borel de un espacio métrico que converge débilmente a otra medida de probabilidad $\mathbb{P}$ (sobre la misma σ-álgebra), entonces Π es relativamente compacta. Pero hemos dado de antemano la existencia de un límite débil, y nuestro objetivo, como ya se mencionó es verificar la existencia de límites débiles, por lo que requerimos otra manera de verificar que una familia de medidas de probabilidad sea secuencialmente compacta. Para eso es que se introduce el concepto de tensión de una familia de medidas de probabilidad y se enuncia el Teorema de Prohorov, que relaciona ambos conceptos.

[3]De hecho, si el espacio métrico es separable, sí se puede dotar al espacio de medidas de probabilidad definidas sobre la σ-álgebra de Borel del espacio de una métrica que sea equivalente a la convergencia en distribución. A esta métrica se le conoce como la métrica de Prohorov. Consultar, por ejemplo, la referencia [3].

DEFINICIÓN. Sea (M,ρ) un espacio métrico y Π una familia de medidas de probabilidad sobre $\mathscr{B}_M$. Decimos que Π es tensa si para cada $\varepsilon > 0$ existe $K \subset M$ compacto (que por supuesto pertenece a $\mathscr{B}_M$ por ser cerrado) tal que $\mathbb{P}(K) > 1-\varepsilon$ para cada $\mathbb{P} \in \Pi$.

TEOREMA 2.4 (Teorema de Prohorov). *Sea (M,ρ) un espacio métrico y Π una familia de medidas de probabilidad sobre la σ-álgebra de Borel de M. Entonces:*

a) Si M es completo y separable y Π es relativamente compacta, entonces Π es tensa.

b) Si Π es tensa entonces es relativamente compacta.

DEMOSTRACIÓN.

a) Se afirma lo siguiente: Si $(G_k)_{k\in\mathbb{N}}$ es una sucesión de conjuntos abiertos en M tal que $G_k \subset G_{k+1}$ para $k \in \mathbb{N}$ y $\cup_{k\in\mathbb{N}} G_k = M$, entonces para cada $\varepsilon > 0$ existe $k \in \mathbb{N}$ tal que $\mathbb{P}(G_k) > 1-\varepsilon$ para cualquier $\mathbb{P} \in \Pi$. La razón: Si lo anterior no fuese cierto, entonces existiría $\varepsilon > 0$ tal que para cualquier $k \in \mathbb{N}$, existe $\mathbb{P}_k \in \Pi$ tal que $\mathbb{P}_k(G_k) \le 1-\varepsilon$. Si escogemos una subsucesión $(\mathbb{P}_{k_n})_{n\in\mathbb{N}}$ de $(\mathbb{P}_k)_{k\in\mathbb{N}}$ que converja débilmente (por la compacidad secuencial) a $\mathbb{P}$, entonces el teorema Portmanteau nos permite concluir que

$$\mathbb{P}(G_k) \le \liminf_{n\to\infty} \mathbb{P}_{k_n}(G_k) \le \liminf_{n\to\infty} \mathbb{P}_{k_n}(G_{k_n}) \le 1-\varepsilon,$$

siendo la anterior desigualdad imposible pues el límite débil de medidas de probabilidad es una medida de probabilidad (ya que la función constante igual a 1 es continua y acotada) y entonces $1 = \mathbb{P}(M) = \lim_{k\to\infty} \mathbb{P}(G_k)$.

Ahora, la separabilidad implica que existe un conjunto numerable y denso $D \subset M$ y como para cada $k \in \mathbb{N}$, $M = \cup_{d\in D} B_{1/k}(d)$, si ε es mayor a cero, existe $n_k \in \mathbb{N}$ tal que $\mathbb{P}\big(\cup_{i=1}^{n_k} B_{1/k}(d_i)\big) > 1-\varepsilon/2^k$ para toda $\mathbb{P} \in \Pi$, donde $(d_i)_{i\in\mathbb{N}}$ es una enumeración de los elementos de D. Como $\cap_{k\ge 1} \cup_{i=1}^{n_k} B_{1/k}(d_i)$ es un conjunto totalmente acotado, si K es su cerradura, entonces es completa, pues M lo es, por lo que K es compacto (consultar el Teorema A.3) Finalmente, notemos que

$$\mathbb{P}(K^c) \le \sum_{k\ge 1} 1 - \mathbb{P}\big(\cup_{i=1}^{n_k} B_{1/k}(d_i)\big) < \sum_{k\ge 1} \varepsilon/2^k = \varepsilon, \quad \forall \mathbb{P} \in \Pi$$

por lo que $\mathbb{P}(K) > 1-\varepsilon$ para toda $\mathbb{P} \in \Pi$ y por lo tanto, Π es tensa.

b) Sea $(\mathbb{P}_k)_{k\ge 1}$ una sucesión de elementos de Π, por lo que debemos verificar que alguna subsucesión converge débilmente. Como Π es tensa, para cada $u \ge 1$, existe un compacto K'_u tal que $\mathbb{P}_k(K'_u) \ge 1 - 1/u$ para toda $k \in \mathbb{N}$. Si $K_u = \cup_{i=1}^{u} K'_u$, entonces K_u es compacto, $K_u \subset K_{u+1}$ y para toda $k \in \mathbb{N}$

$$\mathbb{P}_k(K_u) \ge \mathbb{P}_k(K'_u) \ge 1 - 1/u.$$

Como K_u es métrico compacto, entonces es separable, pues para cada $k \ge 1$, existen $n_k \ge 1$ y $x_1^k \dots x_{n_k}^k$ tal que $K_u \subset \cup_{i=1}^{n_k} B_{1/k}(x_i^k)$, de donde el conjunto a lo más numerable $D = \{x_i^k : k \ge 1, 1 \le i \le n_k\}$ es denso. Así, se sigue que el conjunto $\cup_{u\geq 1} K_u$ es separable, pues basta tomar la unión (numerable) de los subconjuntos densos y a lo más numerables de cada K_u como un subconjunto denso y a lo más numerable de la unión. Mediante el uso del Teorema A.2, se concluye que existe una familia a lo más numerable $\mathscr{A}$ que consta de conjuntos abiertos en M tal que si G es abierto en M y $x \in G \cap \cup_{u\geq 1} K_u$, entonces existe $A \in \mathscr{A}$ tal que $x \in A \subset \overline{A} \subset G$. Consideremos

$$\mathscr{H} = \{\emptyset\} \cup \left\{ \left(\overline{A_1} \cup \cdots \cup \overline{A_n}\right) \cap K_u : u \geq 1, n \geq 1, A_1, \ldots, A_n \in \mathscr{A} \right\}$$

teniendo en cuenta que la familia es a lo más numerable y que cada uno de sus elementos es un compacto. Ahora utilizaremos el método diagonal de Cantor para construir una primera aproximación a la medida de probabilidad que estamos buscando: Sea $\{H_i : i \geq 1\}$ una enumeración de $\mathscr{H}$. Como para cada $H \in \mathscr{H}$, la sucesión numérica $(\mathbb{P}_k(H))_{k\geq 1}$ es acotada, el Teorema de Bolzano-Weierstrass[4] nos permite afirmar que existe una subsucesión $(\mathbb{P}_k^1)_{k\geq 1}$ de $(\mathbb{P}_k)_{k\geq 1}$ que es convergente en H_1. Como para cada $H \in \mathscr{H}$, la sucesión $(\mathbb{P}_k^1(H))_{k\geq 1}$ es acotada, entonces existe una subsucesión $(\mathbb{P}_k^2)_{k\geq 1}$ de $(\mathbb{P}_k^1)_{k\geq 1}$ que es convergente en H_2 (y entonces también en H_1). Continuando recursivamente, para cada $n \in \mathbb{N}$, podemos construir una subsusucesión $(\mathbb{P}_k^{n+1})_{k\geq 1}$ de $(\mathbb{P}_k^n)_{k\geq 1}$ que sea convergente en $H_1 \cdots H_{n+1}$, por lo que la subsucesión diagonal $(\mathbb{P}_k^k)_{k\geq 1}$ de $(\mathbb{P}_k)_{k\geq 1}$ es convergente en todos los elementos de $\mathscr{H}$. Si denotamos por $(\mathbb{P}_{k_i})_{i\geq 1}$ a esta subsucesión, consideremos, para $H \in \mathscr{H}$,

$$\alpha(H) = \lim_{i\to\infty} \mathbb{P}_{k_i}(H).$$

Lo que haremos a continuación será construir una medida de probabilidad $\mathbb{P}$ tal que si G es abierto en M entonces $\mathbb{P}(G) = \sup_{H\subset G} \alpha(H)$, puesto que para esa medida de probabilidad se cumple

$$\liminf_{i\to\infty} \mathbb{P}_{n_i}(G) \geq \liminf_{i\to\infty} \mathbb{P}_{n_i}(H) = \alpha(H)$$

si $H \in \mathscr{H}$ y $H \subset G$. Pero entonces, al considerar el supremo de $\alpha(H)$ para $H \subset G$, obtenemos

$$\liminf_{i\to\infty} \mathbb{P}_{n_i}(G) \geq \mathbb{P}(G),$$

de donde el teorema Portmanteau nos permite afirmar que la sucesión $(\mathbb{P}_{n_i})_{i\geq 1}$ converge débilmente a $\mathbb{P}$. Para construir dicha medida de probabilidad, hagamos notar las siguientes propiedades de α, que hereda de las medidas $\mathbb{P}_{k_i}$:

1. $\alpha(\emptyset) = 0$.
2. $\alpha(H_1 \cup H_2) = \alpha(H_1) + \alpha(H_2)$ si $H_1, H_2 \in \mathscr{H}$ y $H_1 \cap H_2 = \emptyset$.
3. $\alpha(H_1 \cup H_2) \leq \alpha(H_1) + \alpha(H_2)$ si $H_1, H_2 \in \mathscr{H}$.
4. $\alpha(H_1) \leq \alpha(H_2)$ si $H_1 \subset H_2$ y ambos conjuntos pretenecen a $\mathscr{H}$.

[4] Ver [1].

Sean $\beta(G) = \sup_{H\subset G} \alpha(H)$, para G abierto y $\gamma(A) = \inf_{A\subset G} \beta(G)$. Lo que haremos es demostrar que γ es una medida exterior y que la σ-álgebra de los conjuntos γ-medibles contiene a los conjuntos cerrados, pues esto implicará que contiene a $\mathscr{B}_M$. Notemos que para cada conjunto abierto G, se cumple la igualdad $\beta(G) = \alpha(G)$, por lo que sólo restaría verificar que al definir como $\mathbb{P}$ a la restricción de γ a $\mathscr{B}_M$, $\mathbb{P}$ es una medida de probabilidad (que evidentemente satisface $\mathbb{P}(G) = \sup_{H\subset G} \alpha(H)$.) Esto se verifica pues cada K_u es elemento de $\mathscr{H}$, por lo que

$$1 \geq \mathbb{P}(M) \geq \sup_{u\geq 1} \alpha(K_u) \geq \sup_{u\geq 1} 1 - 1/u = 1.$$

Así, para terminar la demostración de esta implicación, sólo hace falta verificar que γ es una medida exterior y que la σ-álgebra de los conjuntos γ-medibles contiene a los conjuntos cerrados. Esto se hará por etapas.

a) Si F es un conjunto cerrado contenido en el conjunto abierto G y en H, elemento de $\mathscr{H}$, entonces existe $H' \in \mathscr{H}$ tal que $F \subset H' \subset G$. Esto sucede pues F es compacto, al estar contenido en H, que lo es, y por ser cerrado. Además, sabemos que para cada $x \in G$, existe $A_x \in \mathscr{A}$ tal que $x \in A_x \subset \overline{A_x} \subset G$. Si consideramos la familia $\{A_x : x \in F\}$, ésta forma una cubierta abierta de F y por lo tanto existen $x_1, \cdots, x_n \in F$ tal que $F \subset A_{x_1} \cup \cdots \cup A_{x_n}$. Pero, como H está contenido, por definición de $\mathscr{H}$, en algún K_u, tenemos que $F \subset K_u \cap (\overline{A_{x_1}} \cup \cdots \cup \overline{A_{x_n}}) \subset G$, por lo que podemos tomar como H' al conjunto $K_u \cap (\overline{A_{x_1}} \cup \cdots \cup \overline{A_{x_n}})$.

b) β es finitamente subaditiva en los conjuntos abiertos. Razón: supongamos que G_1 y G_2 son dos conjuntos abiertos y que $H \subset G_1 \cup G_2$, con $H \in \mathscr{H}$. Sean

$$F_1 = \{x \in H : \rho(x, G_1^c) \geq \rho(x, G_2^c)\}$$
$$F_2 = \{x \in H : \rho(x, G_2^c) \geq \rho(x, G_1^c)\},$$

por lo que $F_1 \cup F_2 = H$. Además, como las asignaciones $x \mapsto \rho(x, G_1^c)$ y $x \mapsto \rho(x, G_2^c)$ son continuas, se sigue que F_1 y F_2 son cerrados, pues son la imagen inversa de $[0, \infty)$ y $(-\infty, 0]$ bajo la funcion $x \mapsto \rho(x, G_1^c) - \rho(x, G_2^c)$ respectivamente. Finalmente, como G_1^c y G_2^c son cerrados, se sigue que si $x \in F_1$ y suponemos que $x \notin G_1$, entonces $x \in G_2$, por lo que $0 = \rho(x, G_1^c) \geq \rho(x, G_2^c) > 0$, que es una contradicción. Así, concluimos que $F_1 \subset G_1$ y de manera análoga $F_2 \subset G_2$. Dado que $F_1, F_2 \subset H$, el inciso anterior nos permite concluir que existen $H_1, H_2 \in \mathscr{H}$ tal que $F_i \subset H_i \subset G_i$, $i = 1, 2$, por lo que

$$\alpha(H) \leq \alpha(H_1 \cup H_2) \leq \alpha(H_1) + \alpha(H_2) \leq \beta(G_1) + \beta(G_2)$$

y como la anterior serie de desigualdades es válida para cualquier $H \subset G_1 \cup G_2$, se sigue la desigualdad

$$\beta(G_1 \cup G_2) \leq \beta(G_1) + \beta(G_2).$$

Finalmente, la subaditividad finita de β se comprueba por inducción.

c) β es σ-subaditiva en los abiertos, pues si G_i es abierto para $i \geq 1$ y $H \subset \bigcup_{i\geq 1} G_i$, con H un elemento de $\mathscr{H}$, como H es compacto, existe $n \geq 1$ tal que $H \subset G_1 \cup \cdots \cup G_n$. Entonces

$$\alpha(H) \leq \beta(G_1 \cup \cdots \cup G_n) \leq \sum_{i=1}^{n} \beta(G_i) \leq \sum_{i\geq 1} \beta(G_i)$$

y como la anterior desigualdad es válida para cualquier H contenido en la unión de los G_i, obtenemos

$$\beta\left(\bigcup_{i\geq 1} G_i\right) \leq \sum_{i\geq 1} \beta(G_i).$$

d) γ es una medida exterior: como $\gamma(\emptyset) = 0$ y γ hereda la monotonía y la no-negatividad de β (que a su vez las herada de α), entonces sólo hace falta verificar que γ es σ-subaditiva. Sean $A_i, i \geq 1$ subconjuntos arbitrarios de M y $G_i, i \geq 1$ abiertos tal que $A_i \subset G_i$ y $\gamma(A_i) > \beta(G_i) - \varepsilon/2^i$. Como

$$\bigcup_{i\geq 1} A_i \subset \bigcup_{i\geq 1} G_i,$$

entonces

$$\gamma\left(\bigcup_{i\geq 1} A_i\right) \leq \beta\left(\bigcup_{i\geq 1} G_i\right) \leq \sum_{i\geq 1} \beta(G_i) \leq \varepsilon + \sum_{i\geq 1} \gamma(A_i)$$

y como la anterior serie de desigualdades es válida para cualquier $\varepsilon > 0$, concluimos que

$$\gamma\left(\bigcup_{i\geq 1} A_i\right) \leq \sum_{i\geq 1} \gamma(A_i).$$

e) Si F es cerrado y G es abierto, entonces $\beta(G) \geq \gamma(F \cap G) + \gamma(F^c \cap G)$. El inciso anterior nos servirá para verificar que los conjuntos cerrados son γ-medibles, recordemos que hasta ahora sólo tenemos que γ es una medida exterior. Para esto, dada $\varepsilon > 0$, ya que $G \cap F^c$ es abierto existe $H_1 \in \mathscr{H}$ tal que $H_1 \subset G \cap F^c$ y $\beta(G \cap F^c) < \alpha(H_1) + \varepsilon$. Como H_1 es cerrado, por ser elemento de $\mathscr{H}$, entonces $G \cap H_1^c$ es abierto y además, contiene a $G \cap F$. Consideremos a $H_2 \in \mathscr{H}$ tal que $H_2 \subset G \cap H_1^c$ y $\gamma(G \cap H_1^c) < \alpha(H_2) + \varepsilon$. Como $H_1 \cup H_2 \in \mathscr{H}$, $H_1 \cap H_2 = \emptyset$ y $H_1 \cup H_2 \subset G$, se sigue que

$$\beta(G) \geq \alpha(H_1 \cup H_2) = \alpha(H_1) + \alpha(H_2) > \gamma(G \cap F^c) + \gamma(G \cap F) - 2\varepsilon$$

y dado que la anterior serie de desigualdades es válida para cualquier $\varepsilon > 0$, hemos verificado la desigualdad deseada.

f) Los conjuntos cerrados son γ-medibles. Esto sucede ya que si A es cualquier subconjunto de M, F es cerrada y consideramos a G un conjunto abierto tal que $A \subset G$, entonces el inciso anterior y la monotonía de γ nos permiten afirmar que

$$\beta(G) \geq \gamma(G \cap F^c) + \gamma(G \cap F) \geq \gamma(A \cap F^c) + \gamma(A \cap F),$$

por lo que el ínfimo de las cantidades del lado derecho al considerar a $A \subset G$ y G abierto, que es igual a $\gamma(A)$, obtenemos $\gamma(A) \geq \gamma(A \cap F^c) + \gamma(A \cap F)$.

□

El siguiente teorema nos dice que la convergencia débil se preserva ante transformaciones continuas entre dos espacios métricos. Para contextualizarlo, consideremos dos espacios métricos M y M' y una función continua f de M en M'. Dada cualquier medida de probabilidad $\mathbb{P}$ sobre $\mathscr{B}_M$, podemos definir una en $\mathscr{B}_{M'}$ mediante la asignación $\mathbb{P}f^{-1}(A) = \mathbb{P}(f^{-1}(A))$ para $A \in \mathscr{B}_{M'}$. La relación entre estas dos medidas está dada mediante la fórmula de cambio de variable, dada $g: M' \to \mathbb{R}$ una función $(\mathscr{B}_{M'}, \mathscr{B}_{\mathbb{R}})$-medible, g es integrable respecto a $\mathbb{P}f^{-1}$ si y sólo si $g \circ f$ lo es respecto a $\mathbb{P}$ y en ese caso,

$$\int_{M'} g \, d\mathbb{P}f^{-1} = \int_M g \circ f \, d\mathbb{P}.$$

TEOREMA 2.5. *Sean (M, ρ), (M', ρ'), $f: M \to M'$ una función continua y $(\mathbb{P}_k)_{k\in\mathbb{N}}$ es una sucesión de medidas de probabilidad que convergen débilmente a $\mathbb{P}$. Entonces la sucesión de medidas $(\mathbb{P}_k f^{-1})_{k\in\mathbb{N}}$ converge débilmente a $\mathbb{P}f^{-1}$.*

DEMOSTRACIÓN. Es consecuencia inmediata del teorema de cambio de variable. Si $g: M' \to \mathbb{R}$ es una función continua y acotada, entonces es $(\mathscr{B}_{M'}, \mathscr{B}_{\mathbb{R}})$-medible e integrable respecto a $\mathbb{P}_k f^{-1}$ (y a $\mathbb{P}f^{-1}$) y $g \circ f$ es continua y acotada, por lo que

$$\lim_{k\to\infty} \int g \, d\mathbb{P}_k f^{-1} = \lim_{k\to\infty} \int g \circ f \, d\mathbb{P}_k = \int g \circ f \, d\mathbb{P} = \int g \, d\mathbb{P}f^{-1}.$$

□

2.3.2.1. Espacios producto, convergencia débil e independencia. Consideremos a dos espacios métricos (M_1, ρ_1), (M_2, ρ_2) y a $(\Omega, \mathscr{F}, \mathbb{P})$ un espacio de probabilidad. Al espacio producto $M_1 \times M_2$ le podemos asignar la métrica $((x_1, x_2), (y_1, y_2)) \mapsto \rho_1(x_1, y_1) \vee \rho_2(x_2, y_2)$, que denotaremos por ρ y tiene la ventaja de metrizar la topología producto de los espacios topológicos, que es la topología que tiene a los conjuntos de la forma $U_1 \times U_2$, donde U_i es abierto en M_i, como subbase. Si definimos las proyecciones $\pi_i: M_1 \times M_2 \to M_i$ por $\pi_i(x_1, x_2) = x_i$, $i = 1, 2$, una consecuencia de la definición de esta topología es que una función $f: X \to M_1 \times M_2$ definida en un espacio topológico es continua si y sólo si $\pi_i \circ f$ es continua para $i = 1, 2$. Como ahora contamos con una métrica en el espacio producto, lo podemos dotar de la σ-álgebra de Borel $\mathscr{B}_{M_1 \times M_2}$. Dado que las proyecciones son continuas, π_i es $(\mathscr{B}_{M_1 \times M_2}, \mathscr{B}_{M_i})$-medible y la pregunta natural es si se tiene una caracterización de las funciones $(\mathscr{S}, \mathscr{B}_{M_1 \times M_2})$-medibles a través de su composición con las proyecciones, a saber que $f : X \to M_1 \times M_2$ definida en un espacio medible $(X, \mathscr{S})$ sea $(\mathscr{S}, \mathscr{B}_{M_1 \times M_2})$-medible si y sólo si $\pi_i \circ X$ sea $(\mathscr{S}, \mathscr{B}_{M_1 \times M_2})$-medible para $i = 1, 2$. En general, esta caracterización es incorrecta, pero en el caso de que los dos espacios sean separables sí es correcta.

TEOREMA 2.6. *Si (M_1, ρ_1) y (M_2, ρ_2) son dos espacios métricos separables, entonces*

$$\mathscr{B}_{M_1 \times M_2} = \mathscr{B}_{M_1} \otimes \mathscr{B}_{M_2}$$

y si $(X, \mathscr{S})$ es cualquier espacio médible, $f : X \to M_1 \times M_2$ es $(\mathscr{S}, \mathscr{B}_{M_1 \times M_2})$-medible si y sólo si $\pi_i \circ f$ es $(\mathscr{S}, \mathscr{B}_{M_i})$-medible, $i = 1, 2$.

DEMOSTRACIÓN. Si D_i es un subconjunto denso y numerable para M_i, $i = 1, 2$, entonces $D_1 \times D_2$ es un subconjunto denso y numerable para $M_1 \times M_2$ y por lo tanto,

$$\mathscr{U} = \{B_q(d) : d \in D_1 \times D_2, q \in \mathbb{Q} \cap [0, \infty)\}$$

es una base numerable para la topología de $M_1 \times M_2$. Como para $r > 0$ y $(x_1, x_2) \in M_1 \times M_2$,

$$\begin{aligned} B_r(x_1, x_2) &= \{(y_1, y_2) \in M_1 \times M_2 : \rho_1(x_1, y_1) \wedge \rho_2(x_2, y_2) < r\} \\ &= \{(y_1, y_2) \in M_1 \times M_2 : \rho_1(x_1, y_1) \leq r \text{ y } \rho_2(x_2, y_2) < r\} \\ &= B_r(x_1) \times B_r(x_2), \end{aligned}$$

entonces cada elemento de $\mathscr{U}$ es elemento de $\mathscr{B}_{M_1} \times \mathscr{B}_{M_1}$ que se define como

$$\{A \times B : A \in \mathscr{B}_{M_1}, B \in \mathscr{B}_{M_2}\}.$$

Si $U \subset M_1 \times M_2$ es abierto, entonces se puede representar como la unión numerable de elementos de $\mathscr{U}$, por lo que pertenece a $\sigma(\mathscr{B}_{M_1} \times \mathscr{B}_{M_2})$, que es por definición $\mathscr{B}_{M_1} \otimes \mathscr{B}_{M_2}$. Esto demuestra que la topología de $M_1 \times M_2$ está contenida en $\mathscr{B}_{M_1} \otimes \mathscr{B}_{M_2}$, por lo que $\mathscr{B}_{M_1 \times M_2} \subset \mathscr{B}_{M_1} \otimes \mathscr{B}_{M_2}$. Por otro lado, puesto que las proyecciones son continuas, entonces son medibles, por lo que si $A \times B \in \mathscr{B}_{M_1} \times \mathscr{B}_{M_2}$, como $A \times B = \pi_1^{-1}(A) \cap \pi_2^{-1}(B)$, se sigue que $\mathscr{B}_{M_1} \times \mathscr{B}_{M_2} \subset \mathscr{B}_{M_1 \times M_2}$ y por lo tanto $\mathscr{B}_{M_1 \times M_2} = \mathscr{B}_{M_1} \otimes \mathscr{B}_{M_2}$. La segunda afirmación del teorema se sigue pues si U_i es abierto en M_i, $i = 1, 2$, entonces

$$f^{-1}(U_1 \times U_2) = f^{-1}((U_1 \times M_2) \times (M_1 \times U_2)) = (\pi_1 \circ f)^{-1}(U_1) \cap (\pi_2 \circ f)^{-1}(U_2) \in \mathscr{S}.$$

□

Si $\mathbb{P}$ es cualquier medida de probabilidad sobre $\mathscr{B}_{M_1 \times M_2}$, podemos definir dos medidas de probabilidad asociadas a ésta:

$$\mathbb{P}_i = \mathbb{P}\pi_i^{-1}, \quad i = 1, 2$$

y como las proyecciones son continuas, si $(\mathbb{P}^k)_{k \in \mathbb{N}}$ es una sucesión de medidas de probabilidad sobre $\mathscr{B}_{M_1 \times M_2}$ que converge débilmente a $\mathbb{P}$ cuando $k \to \infty$, entonces $(\mathbb{P}_i^k)_{k \in \mathbb{N}}$ converge débilmente a $\mathbb{P}_i$ cuando $k \to \infty$, $i = 1, 2$. Si $\mathbb{P}$ es una medida producto, digamos

$\mathbb{P} = \mathbb{P}' \otimes \mathbb{P}''$, entonces $\mathbb{P}' = \mathbb{P}_1$ y $\mathbb{P}'' = \mathbb{P}_2$. A las medidas $\mathbb{P}_i$, $i = 1, 2$ se les llama marginales, y en general, no podemos garantizar que si convergen débilmente las dos sucesiones de marginales, converja débilmente la sucesión original. Un resultado positivo lo obtenemos en el caso de medidas producto en el producto de espacios separables, como se ilustra a continuación.

TEOREMA 2.7. *Sean (M_1, ρ_1), (M_2, ρ_2) dos espacios métricos separables, $(\mathbb{P}^k)_{k \in \mathbb{N}}$ y $\mathbb{P}$ medidas de probabilidad sobre $\mathscr{B}_{M_1 \times M_2}$. Entonces*

a) $(\mathbb{P}^k)_{k \in \mathbb{N}}$ converge débilmente a $\mathbb{P}$ cuando $k \to \infty$ si $\mathbb{P}^k(A_1 \times A_2) \to \mathbb{P}(A_1 \times A_2)$ siempre que $\mathbb{P}_i(\partial(A_i)) = 0$, $i = 1, 2$.

b) $(\mathbb{P}_1^k \otimes \mathbb{P}_2^k)_{k \in \mathbb{N}}$ converge a $\mathbb{P}_1 \otimes \mathbb{P}_2$ débilmente cuando $k \to \infty$ si y sólo si $(\mathbb{P}_i^k)_{k \in \mathbb{N}}$ converge a $\mathbb{P}_i$ débilmente, $i = 1, 2$.

DEMOSTRACIÓN.

a) Consideremos a la familia

$$\mathscr{A} = \{A_1 \times A_2 : \mathbb{P}_i(\partial(A_i)) = 0, i = 1, 2\},$$

que es un π-sistema, pues $\emptyset$ pertenece a ella y si $A_1 \times A_2$ y $B_1 \times B_2$ pertenecen a ella, dado que

$$(A_1 \times A_2) \cap (B_1 \times B_2) = (A_1 \cap B_1) \times (A_2 \times B_2)$$

y como $\partial(A \cap B) \subset \partial(A) \cup \partial(B)$,[5] entonces

$$\mathbb{P}(\partial(A_1 \cap B_1) \times M_2) \leq \mathbb{P}(\partial(A_1) \times M_2) + \mathbb{P}(\partial(B_1) \times M_2) = 0$$

con una desigualdad análoga para el otro índice, por lo que

$$\mathbb{P}_i(\partial(A_i \cap B_i)) = 0, \quad i = 1, 2.$$

Además, $\mathbb{P}_k(A) \to \mathbb{P}(A)$ cuando $k \to \infty$ para cada $A \in \mathscr{A}$, por lo que si $A_1, \ldots, A_n \in \mathscr{A}$, la fórmula de inclusión-exclusión implica que

$$\mathbb{P}^k\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n (-1)^{i-1} \sum_{1 \leq j_1 < \ldots < j_i \leq n} \mathbb{P}^k(A_{j_1} \cap \ldots \cap A_{j_i})$$

$$\longrightarrow \sum_{i=1}^n (-1)^{i-1} \sum_{1 \leq j_1 < \ldots < j_i \leq n} \mathbb{P}(A_{j_1} \cap \ldots \cap A_{j_i}) = \mathbb{P}\left(\bigcup_{i=1}^n A_i\right).$$

Si logramos probar que cada conjunto abierto $U \subset M_1 \times M_2$ puede ser representado como unión a lo más numerable de elementos de $\mathscr{A}$, $U = \cup_{i \in \mathbb{N}} A_i$, $A_i \in \mathscr{A}$, entonces para cada $\varepsilon > 0$ existirá $n \in \mathbb{N}$ tal que

$$\mathbb{P}(U) - \varepsilon < \mathbb{P}\left(\bigcup_{i=1}^n A_i\right),$$

[5] Pues $\partial(A \cap B) = \overline{A \cap B} \cap \overline{A^c \cup B^c} = \overline{A \cap B} \cap (\overline{A^c} \cup \overline{B^c}) \subset (\overline{A} \cap \overline{A^c}) \cup (\overline{B} \cap \overline{B^c}) = \partial(A) \cup \partial(B)$.

por lo que

$$\liminf_{k\in\mathbb{N}} \mathbb{P}_k(U) \geq \liminf_{k\in\mathbb{N}} \mathbb{P}_k\left(\bigcup_{i=1}^n A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) > \mathbb{P}(U) - \varepsilon,$$

de donde se sigue que $\mathbb{P}(U) \leq \liminf_{k\in\mathbb{N}} \mathbb{P}_k(U)$ para cualquier U abierto y por el teorema Portmanteau, $\left(\mathbb{P}^k\right)_{k\in\mathbb{N}}$ convergerá en distribución a $\mathbb{P}$. Veamos pues que cada conjunto abierto $U \subset M_1 \times M_2$ puede ser representado como unión a lo más numerable de elementos de $\mathscr{A}$. Sea (x_1, x_2) cualquier elemento de $M_1 \times M_2$. Dado que para cualquier $r > 0$, $B_r(x_1, x_2) = B_r(x_1) \times B_r(x_2)$, entonces U tiene posibilidades de pertenecer a $\mathscr{A}$. Para ver que esto sucede, casi para cualquier $r > 0$, consideremos las asignaciones, de $[0, \infty)$ en $[0, 1]$

$$\begin{aligned} r &\longmapsto \mathbb{P}_1(B_r(x_1)) = \mathbb{P}(B_r(x_1) \times M_2) \\ r &\longmapsto \mathbb{P}_2(B_r(x_2)) = \mathbb{P}(M_1 \times B_r(x_2)). \end{aligned}$$

Dado que son crecientes, admiten una cantidad a lo más numerable de discontinuidades, por lo que el conjunto de puntos en $[0, \infty)$ en el cual ambas funciones son continuas es denso en $[0, \infty)$. En cualquier punto de continuidad r de ambas asignaciones,

$$\begin{aligned} \mathbb{P}_i(\partial(B_r(x_i))) &\leq \lim_{h\downarrow 0} \mathbb{P}_i(B_{r+h}(x_i) \setminus B_{r-h}(x_i)) \\ &= \lim_{h\downarrow 0} \mathbb{P}_i(B_{r+h}(x_i)) - \mathbb{P}_i(B_{r-h}(x_i)) = 0, \quad i = 1, 2, \end{aligned}$$

por lo que el conjuntos de puntos en $[0, \infty)$ tal que $B_r(x_1, x_2) \in \mathscr{A}$ es denso en $[0, \infty)$ y por lo tanto podemos encontrar un subconjunto $\mathscr{R}_{(x_1, x_2)}$ de éste que sea denso y a lo más numerable (basta considerar $q_{r,s} \in (s - r, s + r) \cap [0, \infty)$ para cada $r, s \in (0, \infty) \cap \mathbb{Q}$ tal que $B_{q_{r,s}}(x_1, x_2) \in \mathscr{A}$ y como $\mathbb{Q} \cap (0, \infty)$ es denso en $[0, \infty)$ y a lo más numerable, se sigue que $\mathscr{R}_{(x_1, x_2)} = \{q_{r,s} : r, s \in (0, \infty) \cap \mathbb{Q}\}$ es denso y a lo más numerable.) Sea D_i un subconjunto denso y a lo más numerable de M_i, $i = 1, 2$, por lo que $D = D_1 \times D_2$ es un subconjunto denso en $M_1 \times M_2$ y es a lo más numerable. Si $U \subset M_1 \times M_2$ es abierto y $z \in U$, entonces existe $r > 0$ tal que $B_r(z) \subset U$. Sea $d \in D \cap B_{r/2}(z)$, por lo que $z \in B_{r/2}(d) \subset B_r(z) \subset U$, donde la última contención es consecuencia de la desigualdad del triángulo. Si $r' = \rho(d, z) \in [0, r/2)$, como $(r', r/2) \subset (0, \infty)$ es abierto en $[0, \infty)$ y no vacío, entonces podemos considerar a $q \in (r', r/2) \cap \mathscr{R}_d$. Dado que $q < r/2$, entonces $B_q(d) \subset U$ y como $q > r'$, entonces $z \in B_q(d)$. Así, para cada $z \in U$, existen $d_z \in D$ y $q_z \in \mathscr{R}_{d_z}$ y tal que $z \in B_{q_z}(d_z) \subset U$ y como $q_z \in \mathscr{R}_{d_z}$, entonces $B_{q_z}(d_z) \in \mathscr{A}$. Dado que la familia

$$\{B_r(d) : d \in D, r \in \mathscr{R}_d\} \subset \mathscr{A}$$

es a lo más numerable, y cada abierto $U \subset M_1 \times M_2$ puede ser representado como la unión de elementos de $\{B_r(d) : d \in D, r \in \mathscr{R}_d\}$,

$$U = \bigcup_{z \in U} B_{r_z}(d_z) ,$$

entonces U puede ser representado como la unión numerable de elementos de $\mathscr{A}$, con lo que concluimos la demostración.

b) Es claro que si $\left(\mathbb{P}_1^k \otimes \mathbb{P}_2^k\right)_{k \in \mathbb{N}}$ converge débilmente a $\mathbb{P}_1 \otimes \mathbb{P}_2$ entonces $\left(\mathbb{P}_i^k\right)_{k \in \mathbb{N}}$ converge débilmente a $\mathbb{P}_i$ para $i = 1, 2$, puesto que las proyecciones con continuas y $\mathbb{P}_i^k = \left(\mathbb{P}_1^k \otimes \mathbb{P}_2^k\right) \pi_i^{-1}$, $i = 1, 2$. Por otro lado, si $\left(\mathbb{P}_i^k\right)_{k \in \mathbb{N}}$ converge débilmente a $\mathbb{P}_i$, $i = 1, 2$, entonces para cada $A_1 \times A_2 \in \mathscr{B}_{M_1} \times \mathscr{B}_{M_2}$ tal que $\mathbb{P}_i(\partial(A_i)) = 0$ se tiene que

$$\mathbb{P}^k(A_1 \times A_2) = \mathbb{P}_1^k(A_1)\, \mathbb{P}_2^k(A_2) \longrightarrow \mathbb{P}_1(A_1)\, \mathbb{P}_2(A_2) = \mathbb{P}(A_1 \times A_2) ,$$

por lo que el inciso anterior nos permite concluir que $\left(\mathbb{P}_1^k \otimes \mathbb{P}_2^k\right)_{k \in \mathbb{N}}$ converge débilemente a $\mathbb{P}_1 \otimes \mathbb{P}_2$.

□

A continuación, daremos un ejemplo de una sucesión de medidas de probabilidad definidas sobre $\mathscr{B}_{\mathbb{R}^2}$ que no converge débilmente y para la cual las sucesiones de marginales convergen débilmente: Consideremos un espacio de probabilidad $(\Omega, \mathscr{F}, \mathbb{P})$ en el cual están definidas dos variables aleatorias independientes X y Y con con distribución normal estándar. Para $k \geq 0$, sean $Z_{2k} = (X, Y)$, $Z_{2k+1} = (X, X)$ y $\mathbb{P}^k(B) = \mathbb{P}(Z^k \in B)$ para $B \in \mathscr{B}_{\mathbb{R}^2}$. Entonces $\mathbb{P}^{2k} = \mathbb{P}_1^k \otimes \mathbb{P}_1^k$ y $\mathbb{P}^{2k+1}$ está concentrada en la diagonal $D = \{(x, y) \in \mathbb{R}^2 : x = y\}$, por lo que en particular $\mathbb{P}^{2k} \neq \mathbb{P}^{2k+1}$ y como $\mathbb{P}^{2k} = \mathbb{P}^{2l}$ y $\mathbb{P}^{2k+1} = \mathbb{P}^{2l+1}$, $k, l \geq 0$, la sucesión de distribuciones $\left(\mathbb{P}^k\right)_{k \in \mathbb{N}}$ no puede converger débilmente. Para ver esto, basta considerar la función continua y acotada f cuya regla de correspondencia es $(x, y) \longmapsto |x - y| \wedge 1$, que integrada respecto a $\mathbb{P}^{2k}$ da una cantidad positiva independientemente de k e integrada respecto a $\mathbb{P}^{2k+1}$ es cero, por lo que $\left(\int f \, d\mathbb{P}^k\right)_{k \in \mathbb{N}}$ no es convergente. Sin embargo, notemos que las sucesiones de marginales asociadas a la sucesion $\left(\mathbb{P}^k\right)_{k \in \mathbb{N}}$ son iguales a $(\mathbb{P}_1^1)_{k \in \mathbb{N}}$, por lo que convergen débilmente.

Los Teoremas 2.6 y 2.7 se pueden extender sin mucho problema para que abarquen al producto de n espacios métricos separables y dichas extensiones serán útiles a continuación.

Consideremos ahora n sucesiones de variables aleatorias $\left(X_i^k\right)_{k \in \mathbb{N}}$, para $i = 1, \ldots, n$, con valores en $\mathbb{R}$ y definidas sobre un espacio de probabilidad $(\Omega, \mathscr{F}, \mathbb{P})$ y tales que $X_1^k, \ldots, X_l^k$ son independientes para cada $k \in \mathbb{N}$. Si $X_1, \ldots, X_n$ son variables aleatorias reales independientes, entonces $\left(X_1^k, \ldots, X_n^k\right)$ y $(X_1, \ldots, X_n)$ son vectores aleatorios n-dimensionales y $\left((X_1^k, \ldots, X_n^k)\right)_{k \in \mathbb{N}}$ converge en distribución a $(X_1, \ldots, X_n)$ si y sólo si $\left(X_i^k\right)_{k \in \mathbb{N}}$ converge en distribución a X_i para $i = 1, \ldots n$. Si en particular, consideramos

$(\alpha_k)_{k\in\mathbb{N}}$ una sucesión real que converge a α y $(X_k)_{k\in\mathbb{N}}$ una sucesión de variables aleatorias reales con valores en $\mathbb{R}$ que convergen en distribución a X, entonces α_k y X_k son independientes,[6] pues las variables aleatorias constantes son independientes de cualquier variable, así como α y X y dado que $(\alpha_k)_{k\in\mathbb{N}}$ converge en distribución a α, entonces $((\alpha_k, X_k))_{k\in\mathbb{N}}$ converge en distribución a (α, X) y dado que la asignación $(x, y) \mapsto xy$ de $\mathbb{R}^2$ en $\mathbb{R}$ es continua, podemos concluir que $\alpha_k X_k$ converge en distribución a αX.

2.3.3. Tensión en el espacio S. El teorema de Prohorov nos permitirá, bajo la hipótesis de tensión, concluir la compacidad secuencial de una familia de medidas de probabilidad en la σ-álgebra de Borel generada por la métrica definida anteriormente en $S = \mathscr{C}([0,\infty))$, que nos será de gran utilidad para verificar la existencia de límites débiles. Para poder utilizar el concepto de tensión, necesitamos poder reconocer a los subconjuntos compactos de S, situación análoga a la presentada en el Teorema de Arzela-Ascoli[7], que nos dá condiciones necesarias y suficientes para que un subconjunto del espacio de funciones continuas de un espacio métrico compacto K en $\mathbb{R}$ sea compacto, cuando lo dotamos de la norma uniforme. Necesitamos generalizar este teorema puesto que $[0,\infty)$ no es compacto con la métrica usual para este conjunto, y sin embargo $S = \mathscr{C}([0,\infty))$ es el conjunto sobre el cual trabajaremos. Para esto, será de utilidad el concepto de módulo de continuidad, que procedemos a definir:

DEFINICIÓN. Si $f \in S$, $T > 0$ y $\delta > 0$, definimos el módulo de continuidad de f en $[0, T]$, denotado por $\omega^T(f, \delta)$, como sigue:

$$\omega^T(f,\delta) = \max_{\substack{|s-t|\le\delta \\ 0\le s,t\le T}} |f(s) - f(t)|.$$

La definición anterior tiene sentido puesto que la asignación $(s,t) \mapsto |f(s) - f(t)|$ de $[0,\infty)^2$ en $\mathbb{R}$ es continua para cualquier $f \in S$ y como el conjunto

$$\{|f(s) - f(t)| : 0 \le s, t \le T, |s-t| \le \delta\}$$

es la imagen del compacto $\{(s,t) : 0 \le s, t \le T, |s-t| \le \delta\}$ bajo la asignación $(s,t) \mapsto |f(s) - f(t)|$, entonces es compacto y por lo tanto cerrado y acotado, de donde existe el supremo (al ser trivialmente no vacío y acotado) y es de hecho un máximo pues el conjunto es cerrado.

LEMA 2.2. *La función $f \mapsto \omega^T(f,\delta)$ es continua en (S, d), la función $\delta \mapsto \omega^T(f,\delta)$ es no-decreciente y $\lim_{\delta\downarrow 0}\omega^T(f,\delta) = 0$ para cada $f \in S$.*

DEMOSTRACIÓN. Verifiquemos la primera afirmación: Sea $\varepsilon > 0$, $T > 0$, $\delta > 0$ y consideremos a $\varepsilon' = \varepsilon \wedge 1$ y a $N \in \mathbb{N}$ tal que $T \le N$. Si $d(w_1, w_2) < \varepsilon'/2^{N+1}$, entonces las

[6] En realidad, necesitamos considerar sobre el mismo espacio de probabilidad a la asignación constante con valor α_k.

[7]

cantidades $\max_{t\in[0,n]}|f_1(t)-f_2(t)|\wedge 1$, para $n=1,\ldots,N$ deben ser menores a $\varepsilon/2$, pues en caso contrario se tendría que

$$d(f_1,f_2)\geq\frac{1}{2^N}\max_{t\in[0,N]}|f_1(t)-f_2(t)|\wedge 1\geq\varepsilon'/2^{N+1},$$

que es una contradicción. Así, si $d(f_1,f_2)<\varepsilon/2^{N+1}$, entonces $\max_{t\in[0,T]}|f_1(t)-f_2(t)|\wedge 1<\varepsilon'/2$ y ya que $\varepsilon'/2<1$, entonces $\max_{t\in[0,T]}|f_1(t)-f_2(t)|<\varepsilon'/2\leq\varepsilon/2$. Al utilizar la desigualdad del triángulo, obtenemos la relación

$$|f_2(s)-f_1(t)|\leq|f_2(s)-f_1(s)|+|f_1(s)-f_1(t)|+|f_1(t)-f_2(t)|,$$

de donde, al tomar el máximo sobre el conjunto $|s-t|\leq\delta, 0\leq s,t\leq T$ y utilizar el hecho de que $\sup_{x\in X}(f(x)+g(x))\leq\sup_{x\in X}f(x)+\sup_{x\in X}g(x)$ (para $f,g:X\to\mathbb{R}$ acotadas), obtenemos:

$$\omega^T(f_2,\delta)\leq 2\max_{s\in[0,T]}|f_1(s)-f_2(s)|+\omega^T(f_1,\delta)<\omega^T(f_1,\delta)+\varepsilon$$

y como los papeles de f_1 y f_2 son simétricos en el anterior argumento, se sigue que

$$\omega^T(f_1,\delta)-\varepsilon<\omega^T(f_2,\delta)<\omega^T(f_1,\delta)+\varepsilon,$$

esto es, si $\varepsilon>0$ y $d(f_1,f_2)<1/2^{[T]+1}$, entonces $|\omega^T(f_1,\delta)-\omega^T(f_2,\delta)|<\varepsilon$, por lo que de hecho, la aplicación $f\mapsto\omega^T(f,\delta)$ es uniformemente continua bajo la métrica d.

Para verificar la segunda afirmación, notemos que si $0<\delta_1\leq\delta_2$, entonces

$$\{(s,t):|s-t|\leq\delta_1, s,t\in[0,T]\}\subset\{(s,t):|s-t|\leq\delta_2, s,t\in[0,T]\},$$

por lo que

$$\omega^T(f,\delta_1)=\max_{\substack{|s-t|\leq\delta_1\\0\leq s,t\leq T}}|f(s)-f(t)|\leq\max_{\substack{|s-t|\leq\delta_2\\0\leq s,t\leq T}}|f(s)-f(t)|=\omega^T(f,\delta_2).$$

La tercera afirmación es inmediata de la continuidad uniforme de cualquier elemento f de S en $[0,T]$, para cualquier $T>0$, puesto que para cada $\varepsilon>0$ existe $\delta>0$ tal que $|t-s|<\delta$, $s,t\in[0,T]$ implica que $|f(s)-f(t)|<\varepsilon$. Esto implica que para cada $\varepsilon>0$ existe $\delta>0$ tal que $\omega^T(f,\delta')<\varepsilon$ para $\delta'\in[0,\delta]$. En otras palabras, $\lim_{\delta\downarrow 0}\omega^T(f,\delta)=0$ para $f\in S$ y $T>0$. □

TEOREMA 2.8. *Para que un conjunto $A\subset S$ tenga cerradura compacta es necesario y suficiente que se cumplan las siguientes dos condiciones:*

a) $\sup_{f\in A}|f(0)|<\infty$.

b) $\lim_{\delta\downarrow 0}\sup_{f\in A}\omega^T(f,\delta)=0$ *para cada* $T>0$.

DEMOSTRACIÓN. Supongamos primero que A tiene cerradura compacta. Entonces, dado que la función $f\mapsto f(0)$ definida sobre S es continua, entonces

$$\sup_{f\in\overline{A}}f(0)=\max_{f\in\overline{A}}f(0)<\infty,$$

pues una función continua sobre un compacto es acotada y alcanza el máximo. Por otro lado, para cada $\varepsilon > 0$, definimos

$$K_\delta = \left\{ f \in \overline{A} : \omega^T(f,\delta) \geq \varepsilon \right\}.$$

Por la proposición anterior, sabemos que $\cap_{\delta>0} K_\delta = \emptyset$ y que K_δ es cerrado (pues se trata de la intersección de un cerrado con la imagen inversa de un cerrado por una aplicación continua) y como está contenido en un compacto, se sigue que K_δ es compacto para cada $\delta > 0$. Entonces existen $\delta_1, \ldots, \delta_n > 0$ tal que $K_{\delta_1} \cap \cdots \cap K_{\delta_n} = \emptyset$, pues si las intersecciones finitas fueran no vacias, $\cap_{\delta>0} K_\delta$ sería no vacio, que no es el caso. Pero como $\delta_1 \leq \delta_2$ implica $K_{\delta_1} \subset K_{\delta_2}$, si $\delta = \min_i \delta_i$, se sigue que $K_{\delta'} = \emptyset$ para $\delta' \leq \delta$, de donde se sigue que $\lim_{\delta \downarrow 0} \sup_{f \in \overline{A}} \omega^T(f,\delta) = 0$ y por lo tanto lo mismo sucede si el supremo es sobre A en vez de sobre $\overline{A}$.

Supongamos que se cumplen las condiciones a) y b) del teorema. Para verificar que la cerradura de A es compacta, recordemos que en un espacio métrico, la compacidad es equivalente a la compacidad por sucesiones, significando que $\overline{A}$ es compacto si y sólo si cualquier sucesión con valores en $\overline{A}$ admite una subsucesión convergente, siendo esto último a su vez equivalente a que cualquier sucesión con valores en A admita una subsucesión convergente (en $\overline{A}$).

Una vez hecha esta aclaración, consideremos una sucesión de funciones $\{f_n\}_{n\in\mathbb{N}} \subset A$. Sean $T > 0$, $M = \sup_{n\in\mathbb{N}} |f_n(0)|$ y $\delta > 0$ tal que si $|s-t| \leq \delta$ y $s,t \in [0,T]$ entonces $|f_n(s) - f_n(t)| \leq 1$ (existe por la condición a). Si $m \in \mathbb{N}$ es tal que $1/m \leq \delta/T$, notemos que $\{kT/m : k = 0,\ldots,m\}$ es una partición de $[0,T]$ de norma menor o igual a δ. En consecuencia, si $k \in \{0,\ldots,m\}$ es el mayor entero que satisface $kT/m \leq t$, para $t \in [0,T]$:

$$\begin{aligned}
|f_n(t)| &\leq |f_n(0)| + |f_n(T/m) - f_n(0)| + \cdots + |f_n(kT/m) - f_n((k-1)T/m)| \\
&\quad + |f_n(t) - f_n(kT/m)| \\
&\leq M + k + 1 \\
&\leq M + m + 1.
\end{aligned}$$

Esto nos dice que la sucesión numérica $(f_n(t))_{n\in\mathbb{N}}$ es acotada (digamos por $C = C(T)$) para cualquier $t \in [0,T]$ y $T > 0$. Si $\{r_i : i \in \mathbb{N}\}$ es una enumeración de $\mathbb{Q} \cap [0,T]$, por la compacidad de $[-C,C]$ se sigue que existe una subsucesión convergente de $(f_n(r_1))_{n\in\mathbb{N}}$, llamémosle $(f_n^{(1)}(r_1))_{n\in\mathbb{N}}$. Por la misma razón existe una subsucesión de $(f_n^{(1)}(r_2))_{n\in\mathbb{N}}$, digamos $(f_n^{(2)}(r_2))_{n\in\mathbb{N}}$, que converge, de donde $(f_n^{(2)})_{n\in\mathbb{N}}$ converge tanto en r_1 como en r_2. Procediendo de manera recursiva, podemos obtener una subsucesión $(f_n^l)_{n\in\mathbb{N}}$ de $(f_n^{l-1})_{n\in\mathbb{N}}$ que converga en $r_1,\ldots,r_l$, para cualquier $l \in \mathbb{N}$. Si consideramos la subsucesión $(f_n^{(n)})_{n\in\mathbb{N}}$, verifiquemos que converge en $\mathbb{Q} \cap [0,T]$: Si $l \in \mathbb{N}$, entonces $(f_n^l(r_l))_{n\in\mathbb{N}}$ converge, y como $(f_n^n(r_l))_{n\in\mathbb{N}}$ es subsucesión de $(f_n^l(r_l))_{n\in\mathbb{N}}$, converge también.

Así, para cada $T > 0$ (y para cualquier sucesión $(f_n)_{n\in\mathbb{N}}$,) hemos encontrado una subsucesión $(f_n')_{n\in\mathbb{N}}$ que converge en $\mathbb{Q} \cap [0,T]$. Veamos que, de hecho, esta subsucesión converge en $[0,T]$ y que lo hace de manera uniforme. Sea $\varepsilon > 0$, $\delta > 0$ tal que si $s,t \in [0,T]$

y $|s-t| \leq \delta$ entonces $|f_n(t) - f_n(s)| \leq \varepsilon$ para toda $n \in \mathbb{N}$ (que existe por la condición b.) Si $n \in \mathbb{N}$ es tal que $1/n \leq \delta$, consideremos k el máximo natural tal que $k/n \leq T$, por lo que $0 \leq T - k/n \leq \delta$. Además, dado $t \in [0,T]$, existe $j \in \{0,\ldots,k\}$ tal que $|t - j/n| \leq \delta$. Como $\{j/n : j = 0,\ldots,k\} \subset \mathbb{Q} \cap [0,T]$, se sigue que existe $l \in \mathbb{N}$ tal que $\{j/n : j = 0,\ldots,k\} \subset \{r_0,\ldots,r_l\}$, por lo que dado $t \in [0,T]$ existe $j \in \{0,\ldots,l\}$ tal que $|t - r_j| \leq \delta$. Si utilizamos la convergencia de $(f'_n)_{n\in\mathbb{N}}$ en r_j, se sigue que existe $I \in \mathbb{N}$ tal que si $i_1, i_2 \geq I$ entonces $|f'_{i_1}(r_j) - |f'_{i_2}|(r_j)| < \varepsilon$, para $j = 0,\ldots,l$. Entonces, si $i_1, i_2 \geq I$ y $t \in [0,T]$, consideremos $j \in \{0,\ldots,l\}$ tal que $|t - r_j| \leq \delta$, por lo que

$$\begin{aligned} |f'_{i_1}(t) - f'_{i_2}(t)| &\leq |f'_{i_1}(t) - f'_{i_1}(r_j)| + |f'_{i_1}(r_j) - f'_{i_2}(r_j)| + |f'_{i_2}(r_j) - f'_{i_2}(t)| \\ &< 3\varepsilon \quad \text{para toda } t \in [0,T], \end{aligned}$$

de donde podemos concluir que existe el límite $f_T(t)$ de $f'_n(t)$ cuando $n \to \infty$. De hecho, si tomamos el límite cuando $i_2 \to \infty$, en la desigualdad anterior, obtenemos que para $i_1 \geq I$,

$$|f'_{i_1}(t) - f'_T(t)| \leq 3\varepsilon < 4\varepsilon \quad \forall t \in [0,T],$$

por lo que la convergencia es uniforme en $[0,T]$.

Finalmente, si consideramos la sucesión $(f_n)_{n\in\mathbb{N}}$, el razonamiento anterior muestra que existe una subsucesión $(f_n^{(1)})_{n\in\mathbb{N}}$ de la anterior que converge uniformemente en $[0,1]$. A esta última, le podemos extraer una subsucesión $(f_n^{(2)})_{n\in\mathbb{N}}$ que converga en $[0,2]$ y mediante recursión, podemos obtener para cada $m \in \mathbb{N}$ una sucesión $(f_n^{(m)})_{n\in\mathbb{N}}$ que converga uniformemente en $[0,n]$ y tal que $(f_n^{(m+1)})_{n\in\mathbb{N}}$ sea subsucesión de $(f_n^{m})_{n\in\mathbb{N}}$ para toda $m \in \mathbb{N}$. Si, finalmente, consideramos la subsucesión $(f_n^{(n)})_{n\in\mathbb{N}}$ de $(f_n)_{n\in\mathbb{N}}$, ésta converge uniformemente en $[0,N]$ para toda $N \in \mathbb{N}$, de donde converge en (S,d). □

Armados con la caracterización del Teorema 2.8, podemos dar condiciones necesarias y suficientes para garantizar la tensión de una familia de medidas de probabilidad sobre $\mathscr{B}_S$

TEOREMA 2.9. *Para que una sucesión $(\mathbb{P}_k)_{k\in\mathbb{N}}$ de medidas de probabilidad sobre $\mathscr{B}_S$ sea tensa son necesarias y suficientes las siguientes dos condiciones*

a)

$$\lim_{\lambda\to\infty} \sup_{k\in\mathbb{N}} \mathbb{P}_k(\{f \in S : |f(0)| > \lambda\}) = 0.$$

b) Para toda $\varepsilon > 0$ y para toda $T > 0$,

$$\lim_{\delta\downarrow 0} \sup_{k\in\mathbb{N}} \mathbb{P}_k\big(\{f \in S : \omega^T(f,\delta) \geq \varepsilon\}\big) = 0.$$

DEMOSTRACIÓN. Supongamos primero que la sucesión de medidas de probabilidad es tensa. Sea $K \subset S$ un conjunto compacto tal que $\mathbb{P}_k(K^c) \leq \varepsilon$. Como la asignación $f \mapsto |f(0)|$ de S en $\mathbb{R}$ es continua, se sigue que la imagen de K bajo ella es acotada, o dicho de otra manera, que existe $\lambda > 0$ tal que $|f(0)| < \lambda$ para cualquier $f \in K$. Si $\lambda' \geq \lambda$, entonces

$$\{f \in S : |f(0)| > \lambda'\} \subset \{f \in S : |f(0)| > \lambda\} \subset K^c,$$

por lo que

$$\mathbb{P}_k(\{f \in S : |f(0)|\} > \lambda') \leq \varepsilon \quad \text{para toda } k \in \mathbb{N} \text{ y para toda } \lambda' \geq \lambda,$$

de donde se concluye la necesidad de la primera condición. Para la segunda condición, consideremos $T, \varepsilon, \varepsilon' > 0$. Sea $K \subset S$ un conjunto compacto tal que $\mathbb{P}_k(K^c) \leq \varepsilon'$ para cualquier $k \in \mathbb{N}$ y consideremos los conjuntos

$$K_\delta = K \cap \{f \in S : \omega^T(f, \delta) \geq \varepsilon\}.$$

Dado que la asignación $f \mapsto \omega^T(f, \delta)$ es continua para cada $T, \delta > 0$, entonces los conjuntos K_δ son cerrados como la intersección de dos cerrados y están contenidos en el compacto K. Además, sabemos que para cada $f \in S$ y para cada $T > 0$, $\lim_{\delta \downarrow 0} \omega^T(f, \delta) = 0$, por lo que $\cap_{\delta > 0} K_\delta = \emptyset$ (pues $\varepsilon > 0$) y entonces las intersecciones en las cuales aparecen solamente una cantidad finita los conjuntos K_δ no pueden ser todas distintas del vacio. Pero como $\delta' \leq \delta$ implica que $K_{\delta'} \subset K_\delta$ (por la monotonía de $\delta \mapsto \omega^T(f, \delta)$), la observación anterior nos permite afirmar que existe $\delta > 0$ tal que $K_{\delta'} = \emptyset$ para cualquier $\delta' \in (0, \delta)$. Así, para cualquier $\delta' \in (0, \delta)$,

$$\begin{aligned}
\sup_{k \in \mathbb{N}} \mathbb{P}_k(\{f \in S : \omega^T(f, \delta') \geq \varepsilon\}) &\leq \sup_{k \in \mathbb{N}} \mathbb{P}_k(\{f \in S : \omega^T(f, \delta) \geq \varepsilon\}) \\
&\leq \sup_{k \in \mathbb{N}} \mathbb{P}_k(K^c) \\
&\leq \varepsilon',
\end{aligned}$$

de donde podemos deducir la necesidad de la segunda condición.

Supongamos ahora que se cumplen las condiciones a y b y verifiquemos que la sucesión de medidas de probabilidad es tensa. Sea $\eta > 0$. Por hipótesis, para cada $T > 0$ existe $\lambda_T > 0$ tal que

$$\sup_{k \in \mathbb{N}} \mathbb{P}_k(\{f \in S : |f(0)| > \lambda_T\}) \leq \eta / 2^{T+1}$$

así como $\delta_k^T > 0$ tal que

$$\sup_{k \in \mathbb{N}} \mathbb{P}_k(\{f \in S : \omega^T(f, \delta_k^T) > 1/k\}) \leq \eta / 2^{k+T+1}.$$

Consideremos a los conjuntos

$$A_T = \{f \in S : |f(0)| \leq \lambda_T, \omega^T(f, \delta_k^T) \leq 1/k, k = 1, 2, \ldots\}$$

y

$$A = \bigcap_{T \geq 1, T \in \mathbb{N}} A_T.$$

Dado que las asignaciones $f \mapsto |f(0)|$ y $f \mapsto \omega^T(f,\delta)$ son continuas, se sigue que los conjuntos A_T son cerrados, puesto que constan de intersecciones de imágenes inversas de cerrados bajo estas asignaciones y por lo tanto el conjunto A es cerrado. Además,

$$\sup_{f\in A}|f(0)| \leq \lambda_1 < \infty$$

y

$$\sup_{f\in A}\omega^T(f,\delta) \leq 1/k \qquad \text{para toda}\delta \leq \delta_k^T,$$

por lo que

$$\lim_{\delta\downarrow 0}\sup_{f\in A}\omega^T(f,\delta) = 0 \quad \text{para toda}T > 0.$$

Así, se cumplen las condiciones del Teorema 2.8 y podemos concluir que A tiene cerradura compacta y como A es cerrado, A es compacto. Finalmente,

$$\begin{aligned}
\mathbb{P}_k(A^c) &\leq \sum_{T=1}^{\infty}\mathbb{P}_k(|f(0)| > \lambda_T) + \mathbb{P}_k(\text{Existe } k \geq 1 \text{ tal que } \omega^T(f,\delta_k^T) > 1/k) \\
&\leq \sum_{T=1}^{\infty}\left(\eta/2^{T+1} + \sum_{k=1}^{\infty}\eta/2^{T+k+1}\right) \\
&\leq \sum_{T=1}^{\infty}\eta/2^T = \eta,
\end{aligned}$$

por lo que la sucesión $(\mathbb{P}_k)_{k\in\mathbb{N}}$ es tensa. □

2.3.4. Distribuciones finito dimensionales. Sea $\mathbb{P}$ una medida de probabilidad sobre $\mathscr{B}_S$. Como la asignación $f \mapsto (\pi_{t_1}(f),\ldots,\pi_{t_n}(f))$ es continua, entonces es $(\mathscr{B}_S,\mathscr{B}_{\mathbb{R}^n})$- medible, por lo que inmediatamente podemos considerar la medida de probabilidad inducida por esta asignación, que denotaremos por $\mathbb{P}_{t_1,\ldots,t_n}$:

$$\mathbb{P}_{t_1,\ldots,t_n}(A) = \mathbb{P}(\{f \in S : (\pi_{t_1}(f),\ldots,\pi_{t_n}(f)) \in A\}) \qquad A \in \mathscr{B}_{\mathbb{R}^n}.$$

A estas distribuciones se les conoce como distribuciones finito dimensionales. Anteriormente verificamos que si dos medidas de probabilidad sobre $\mathscr{B}_S$ coinciden en la clase de cilindros finito dimensionales, entonces coinciden en $\mathscr{B}_S$. Esto quiere decir que dos medidas de probabilidad sobre $\mathscr{B}_S$ tienen las mismas distribuciones finito dimensionales, entonces son iguales. Si $(\mathbb{P}^k)_{k\in\mathbb{N}}$ es una sucesión de medidas de probabilidad sobre $\mathscr{B}_S$ y converge débilmente a $\mathbb{P}$, entonces para cada $n \geq 1$ y $t_1,\ldots,t_n \in [0,\infty)$, la sucesión de medidas de probabilidad $(\mathbb{P}^k_{t_1,\ldots,t_n})_{k\in\mathbb{N}}$ converge débilmente a $\mathbb{P}_{t_1,\ldots,t_n}$, puesto que si $g:\mathbb{R}^n \to \mathbb{R}$ es una función continua y acotada, entonces la asignación $g\circ(\pi_{t_1},\ldots,\pi_{t_n})$ de S en $\mathbb{R}$ es continua (como composición de dos aplicaciones continuas) y acotada, por lo que

$$\int g\,d\mathbb{P}^k_{t_1,\ldots,t_n} = \int g\circ(\pi_{t_1},\ldots,\pi_{t_n})\,d\mathbb{P}^k \to \int g\circ(\pi_{t_1},\ldots,\pi_{t_n})\,d\mathbb{P} = \int g\,d\mathbb{P}_{t_1,\ldots,t_n},$$

conforme $k \to \infty$. Si por otro lado, sólamente suponemos que $(\mathbb{P}^k_{t_1,\ldots,t_n})_{k\in\mathbb{N}}$ converge débilmente, no podemos garantizar en general que $(\mathbb{P}^k)$ converge, pues en principio podría no existir una medida de probabilidad $\mathbb{P}$ en $\mathscr{B}_S$ que tuviera las distribuciones finito dimensionales que debería tener si $\mathbb{P}^k \Rightarrow \mathbb{P}$ conforme $k \to \infty$. Este es el motivo fundamental por el cual se introduce el concepto de tensión, cuyo uso se ilustra a continuación.

TEOREMA 2.10. *Si $(\mathbb{P}^k)_{k\in\mathbb{N}}$ es una sucesión tensa de medidas de probabilidad sobre $\mathscr{B}_S$ tal que para cualquier $n \geq 1$ y $t_1,\ldots,t^n$, $(\mathbb{P}^k_{t_1,\ldots,t_n})_{k\in\mathbb{N}}$ converge débilmente, entonces existe una medida de probabilidad $\mathbb{P}$ sobre $\mathscr{B}_S$ tal que $\mathbb{P}^k \Rightarrow \mathbb{P}$.*

DEMOSTRACIÓN. Dado que $(\mathbb{P}^k)_{k\in\mathbb{N}}$ es tensa, es secuencialmente compacta (Teorema de Prohorov) por lo que existe una medida de probabilidad $\mathbb{P}$ y una subsucesión $(\mathbb{P}^{k_l})_{l\in\mathbb{N}}$ de $(\mathbb{P}^k)_{k\in\mathbb{N}}$ que converge débilmente a $\mathbb{P}$. Si $(\mathbb{P}^{k'_l})_{l\in\mathbb{N}}$ fuera otra subsucesión $(\mathbb{P}^k)_{k\in\mathbb{N}}$ que converge débilmente a una medida de probabilidad $\mathbb{Q}$, entonces $\mathbb{P}$ y $\mathbb{Q}$ tendrían las mismas distribuciones finito dimensionales, por la segunda hipótesis del teorema, de donde $\mathbb{P} = \mathbb{Q}$. Veamos ahora que de hecho la sucesión original, $(\mathbb{P}^k)_{k\in\mathbb{N}}$ converge a $\mathbb{P}$. Por contradicción: Si suponemos que $(\mathbb{P}^k)_{k\in\mathbb{N}}$ no converge a $\mathbb{P}$, entonces existe una función $g : S \to \mathbb{R}$ continua y acotada tal que

$$\int g\, d\mathbb{P}^k \not\to \int g\, d\mathbb{P} \text{ cuando } k \to \infty.$$

Dado que g es acotada, la sucesión $\left(\int g\, d\mathbb{P}^k\right)_{k\in\mathbb{N}}$ es acotada, por lo que cualquier subsucesión de ella admite a su vez una subsucesión convergente. Sabemos que en un espacio métrico, una sucesión converge a x si y sólo si cualquier subsucesión admite a su vez una subsucesión que converge a x (Teorema A.1,) por lo que la cantidad $\int g\, d\mathbb{P}$ no puede ser el único límite de subsucesiones de la sucesión anterior, pues de otra manera sí tendríamos la convergencia. Esto implica que podemos escoger una subsucesión $(\mathbb{P}^{k_l})_{l\in\mathbb{N}}$ de $(\mathbb{P}^k)_{k\in\mathbb{N}}$ tal que $\int g\, d\mathbb{P}^{k_l}$ converja a un valor distinto a $\int g\, d\mathbb{P}$ cuando $k \to \infty$. Esto es una contradicción por la siguiente razón: la subsucesión que acabamos de construir es tensa, las distribuciones finito dimensionales convergen a las de $\mathbb{P}$, pero no puede tener ninguna subsucesión que converga débilmente a $\mathbb{P}$. Por otro lado, por el mismo argumento que utilizamos para la sucesión original, podemos asegurar la existencia de una subsucesión que converga débilmente a $\mathbb{P}$. Esto demuestra que $(\mathbb{P}^k)_{k\in\mathbb{N}}$ converge débilmente a $\mathbb{P}$. □

2.3.5. El principio de invariancia de Donsker. Sean $(X_i)_{i\geq 1}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con varianza $0 < \sigma^2 < \infty$ definidas en un espacio de probabilidad $(\Omega, \mathscr{F}, \mathbb{P})$ y $(S_n)_{n\in\mathbb{N}}$ la sucesión de sumas parciales asociada, donde $S_0 = 0$ y para $n \geq 1$, $S_n = \sum_{i=1}^n X_i$. Definimos a $X : \Omega \to S$ mediante

$$X_t(\omega) = \pi_t \circ X(\omega) = S_{[t]}(\omega) + (t - [t])\, X_{[t]+1}(\omega)\,,$$

geométricamente, la gráfica de X se obtiene interpolando linealmente entre los puntos (n, S_n), para $n \in \mathbb{N}$. La primera pregunta que es natural hacerse es si X es $(\mathscr{F}, \mathscr{B}_S)$-medible, proposición que resulta ser verdadera pues $\pi_t \circ X$ es $(\mathscr{F}, \mathscr{B}_S)$-medible para cada

$t \in [0, \infty)$. Mediante X, podemos definir a otra sucesión de variables aleatorias con valores en S que serían el análogo de las variables aleatorias que intervienen en el Teorema Central del Límite:

$$Y_t^k(\omega) = \pi_t \circ Y^k(\omega) = \frac{\pi_{kt} \circ X(\omega)}{\sigma\sqrt{k}} = \frac{X_{kt}(\omega)}{\sigma\sqrt{k}}, \quad k \geq 1.$$

Dada la medibilidad de X, resulta que Y^k es $(\mathscr{F}, \mathscr{B}_S)$-medible. Como una generalización al Teorema Central del Límite, se presenta el siguiente resultado:

TEOREMA 2.11 (Principio de Invariancia de Donsker). *La sucesión de variables aleatorias con valores en S, $(Y^k)_{k\geq 1}$, converge en distribución a B, donde B es un Movimiento Browniano.*

Como por definición del Movimiento Browniano, $t \longmapsto \pi_t \circ B(\omega)$ es continua y $\pi_t \circ B$ es variable aleatoria, se sigue que B es medible respecto a la σ-álgebra del espacio de probabilidad en el cual está definido y a la σ-álgebra de Borel en S. Un asunto pendiente es que no hemos verificado la existencia del movimiento browniano, pero eso forma parte de la demostración y para eso se necesita el Teorema de Prohorov.

En lugar de considerar directamente a las variables Y^k, sea $\mathbb{P}^k$ la medida de probabilidad inducida por Y^k en S, dada por $\mathbb{P}^k(A) = \mathbb{P}(Y^k \in A)$, para $A \in \mathscr{B}_S$. Así, para demostrar el principio de invariancia de Donsker es suficiente verificar que la sucesión de medidas $(\mathbb{P}^k)_{k\geq 1}$ converge débilmente a una medida de probabilidad $\mathbb{W}$ cuyas distribuciones finito dimensionales $\mathbb{W}_{t_1,\ldots,t_n}$, $t_1 \leq \ldots \leq t_n$, son normales n-dimensionales, con media cero y matriz de varianza-covarianza

$$V = (\text{mín}\,(t_i, t_j))_{i,j=1}^n .$$

Es suficiente considerar el caso anterior, pues si $t_1, \ldots, t_n \in [0, \infty)$, podemos encontrar una permutación τ tal que $t_{\tau_1} \leq \cdots \leq t_{\tau_n}$ y como la asignación

$$f : (x_1, \ldots, x_n) \longmapsto (x_{\tau_1}, \ldots, x_{\tau_n})$$

es un homeomorfismo de $\mathbb{R}^n$ en $\mathbb{R}^n$, se sigue que dos medidas de probabilidad $\mathbb{P}_1$ y $\mathbb{P}_2$ sobre $\mathscr{B}_{\mathbb{R}^n}$ son iguales si y sólo si $\mathbb{P}_1 f^{-1} = \mathbb{P}_2 f^{-1}$. Empecemos por verificar que las distribuciones finito dimensionales de $(\mathbb{P}^k)_{k\geq 1}$ convergen, para lo cual necesitamos el siguiente resultado, que nos permitirá substituir una sucesión de variables aleatorias por otra que sea más fácil de analizar, en cuanto a la determinación de límites débiles. Aplica solamente a sucesiones de medidas de probabilidad inducidas por variables aleatorias.

LEMA 2.3. *Si $(X_i)_{i\in\mathbb{N}}$, $(Y_i)_{i\in\mathbb{N}}$ y Y son variables aleatorias definidas en $(\Omega, \mathscr{F}, \mathbb{P})$ y con valores en un espacio métrico separable (M, ρ), tales que $(Y_i)_{i\in\mathbb{N}}$ converge en distribución a Y y*

$$\rho(X_i, Y_i) \xrightarrow{\mathscr{P}} 0$$

entonces $(X_i)_{i\in\mathbb{N}}$ converge en distribución a Y.

DEMOSTRACIÓN. Como verificamos anteriormente, la separabilidad garantiza que la asignación $\omega \mapsto (X(\omega), Y(\omega))$ sea $(\mathscr{F}, \mathscr{B}_{M\times M})$-medible y esto implica que la composición de ρ con la anterior función sea $(\mathscr{F}, \mathscr{B}_{\mathbb{R}})$-medible. Por el teorema Portmanteau, es suficiente verificar que para cada $F \subset M$ cerrado,

$$\limsup_{i\to\infty} \mathbb{P}(X_i \in F) \leq \mathbb{P}(Y \in F).$$

Para esto, consideremos $\varepsilon > 0$ y $F_\varepsilon = \{x \in M : \rho(x, F) \leq \varepsilon\}$. Por la continuidad de la asignación $x \mapsto \rho(x, F)$, el conjunto F_ε es cerrado y además, tenemos la contención $F_\varepsilon \subset F_{\varepsilon'}$ si $0 < \varepsilon \leq \varepsilon'$ así como las igualdades

$$F = \bigcap_{\varepsilon>0} F_\varepsilon = \bigcap_{\substack{\varepsilon>0\\ \varepsilon\in\mathbb{Q}}} F_\varepsilon$$

de donde se sigue que $\mathbb{P}(Y \in F) = \lim_{\varepsilon\downarrow 0} \mathbb{P}(Y \in F_\varepsilon)$. Si $X_i \in F$ y $\rho(X_i, Y_i) \leq \varepsilon$, entonces $Y_i \in F_\varepsilon$, de donde

$$\begin{aligned}
\limsup_{i\in\mathbb{N}} \mathbb{P}(X_i \in F) &\leq \limsup_{i\in\mathbb{N}} \mathbb{P}(X_i \in F, \rho(X_i, Y_i) \leq \varepsilon) + \limsup_{i\in\mathbb{N}} \mathbb{P}(X_i \in F, \rho(X_i, Y_i) > \varepsilon)\\
&\leq \limsup_{i\in\mathbb{N}} \mathbb{P}(Y_i \in F_\varepsilon, \rho(X_i, Y_i) \leq \varepsilon)\\
&= \mathbb{P}(Y \in F_\varepsilon),
\end{aligned}$$

donde la segunda desigualdad se explica pues $\mathbb{P}(\rho(X_i, Y_i) > \varepsilon) \to 0$ para cualquier $\varepsilon > 0$. Como las desigualdades son válidas para cualquier $\varepsilon > 0$, se sigue que

$$\limsup_{i\in\mathbb{N}} \mathbb{P}(X_i \in F) \leq \mathbb{P}(Y \in F)$$

y por lo tanto, $(X_i)_{i\in\mathbb{N}}$ converge en distribución a Y. □

Volviendo al principio de invariancia de Donsker, veamos que pasa con las distribuciones finito dimensionales de las medidas $\mathbb{P}^k$:

LEMA 2.4. *Las distribuciónes finito dimensionales $\mathbb{P}^k_{t_1,\dots,t_n}$, $t_1 \leq \cdots \leq t_n$, convergen débilmente y la distribución límite es normal con media cero y matriz de varianza-covarianza*

$$V = (\min(t_i, t_j))_{i,j=1}^n .$$

DEMOSTRACIÓN. Consideremos la variable aleatoria

$$\psi_{k,t}(\omega) = (kt - [kt])\, X_{[kt]+1}/\sigma\sqrt{k},$$

cuya definición implica $Y_t^k = S_{[kt]}/\sigma\sqrt{k} + \psi_{k,t}$. Como

$$\mathbb{E}(|\psi_{k,t}|) \leq \mathbb{E}(|X_1|)\, \frac{kt - [kt]}{\sigma\sqrt{k}} \leq \mathbb{E}(|X_1|)\,/\sigma\sqrt{k} \to 0,$$

podemos concluir que $\psi_{k,t}$ converge en distribución a cero cuando $k \to \infty$ para cualquier $t \in [0, \infty)$. Esto sucede ya que si $f : \mathbb{R} \to \mathbb{R}$ es continua y acotada, para cada $\varepsilon > 0$ existe $\delta > 0$ tal que si $|x| < \delta$ entonces $|f(x) - f(0)| < \varepsilon$. Esto implica que si M es una cota para f,

$$\begin{aligned}
&\lim_{k\to\infty} |\mathbb{E}(f(\psi_{k,t})) - \mathbb{E}(f(0))| \\
&\qquad \le \lim_{k\to\infty} \mathbb{E}\big(|f(\psi_{k,t}) - f(0)| 1_{|\psi_{k,t}|<\delta}\big) + \mathbb{E}\big(|f(\psi_{k,t}) - f(0)| 1_{|\psi_{k,t}|\ge\delta}\big) \\
&\qquad \le \lim_{k\to\infty} \varepsilon \mathbb{P}(|\psi_{k,t}| < \delta) + 2M\mathbb{P}(|\psi_{k,t}| \ge \delta) \\
&\qquad \le \varepsilon + \lim_{k\to\infty} 2M \frac{\mathbb{E}(|\psi_{k,t}|)}{\delta} \\
&\qquad = \varepsilon
\end{aligned}$$

y como las desigualdades de arriba son válidas para cualquier $\varepsilon > 0$, se tiene que $\mathbb{E}(f(\psi_{k,t})) \to \mathbb{E}(f(0))$ conforme $k \to \infty$ para toda $t \in [0, \infty)$.

Mediante la misma técnica utilizada en el párrafo anterior, podemos verificar que si $t_1, \ldots, t_n \in [0, \infty)$, entonces el vector aleatorio $(\psi_{k,t_1}, \ldots, \psi_{k,t_n})$ converge en distribución a 0 cuando $k \to \infty$. Como consecuencia de lo anterior, vemos que el vector aleatorio

$$(Y^k_{t_1}, \ldots, Y^k_{t_n})$$

tiene la misma distribución límite que el vector aleatorio

$$(S_{[kt_1]}, \ldots, S_{[kt_n]}) \cdot \frac{1}{\sigma\sqrt{k}}.$$

Para encontrar esta última distribución límite, consideremos el vector aleatorio

$$(Z^k_1, \ldots, Z^k_n) = \frac{1}{\sigma\sqrt{k}} \left(S_{[kt_1]}, S_{[kt_2]} - S_{[kt_1]}, \ldots, S_{[kt_n]} - S_{[kt_{n-1}]}\right),$$

cuyas entradas son independientes. Así, para encontrar la distribución límite de este vector con entradas independientes, basta encontrar la distribución límite de cada una de las entradas, por el Teorema (2.7).

Empecemos por la primera entrada, Z^k_1, que podemos escribir como

$$\frac{S_{[kt_1]}}{\sigma\sqrt{[kt_1]}} \frac{\sqrt{[kt_1]}}{\sqrt{k}}.$$

Como $\sqrt{[kt_1]}/\sqrt{k} \to \sqrt{t_1}$ y dado que

$$\left(\frac{S_{[kt_1]}}{\sigma\sqrt{[kt_1]}}\right)_{k\in\mathbb{N}}$$

es subsucesión de

$$\left(\frac{S_k}{\sigma\sqrt{k}}\right)_{k\in\mathbb{N}},$$

entonces $(Z_1^k)_{k\in\mathbb{N}}$ converge en distribución a $\sqrt{t_1}X$, donde X es normal estándar[8], que es una variable con distribución $N(0, t_1)$.

Para las otras componentes, consideremos $i \geq 2$ y notemos que

$$[k(t_i - t_{i-1})] \leq [kt_i] - [kt_{i-1}] \leq [k(t_i - t_{i-1})] + 1,$$

se sigue que Z_i^k tiene la misma distribución que $S_{[k(t_i)-t_{i-1}]}/\sigma\sqrt{k}$ o que $S_{[k(t_i)-t_{i-1}]+1}\sigma\sqrt{k}$, puesto que las variables $(X_i)_{i\geq 1}$ son idénticamente distribuidas. Como $X_{[k(t_i)-t_{i-1}]+1}/\sigma\sqrt{k}$ converge a cero en distribución, utilizando la misma técnica que para $\psi_{k,t}$, entonces

$$\left(\frac{S_{[k(t_i)-t_{i-1}]}}{\sigma\sqrt{k}}\right)_{k\in\mathbb{N}}$$

y

$$\left(\frac{S_{[k(t_i)-t_{i-1}]+1}}{\sigma\sqrt{k}}\right)_{k\in\mathbb{N}}$$

convergen en distribución a la misma variable, que en completa analogía con el caso de la primera entrada de $(Z^k)_{k\in\mathbb{N}}$, se distribuye $N(0, t_i - t_{i-1})$. Así, la sucesión

$$((Z_1^k, \ldots, Z_n^k))_{k\in\mathbb{N}}$$

converge en distribución a $(Z_1, \ldots, Z_n)$, donde la distribución de Z_i es $N(0, t_i - t_{i-1})$ (con $t_0 = 0$) y donde las entradas del vector límite son independientes, por lo que este último es un vector gaussiano.

Finalmente, apliquemos el Teorema 2.5 a la función

$$(x_1, \ldots, x_n) \mapsto (x_1, x_1 + x_2, \ldots, x_1 + \cdots + x_n),$$

que es una función continua de $\mathbb{R}^n$ en $\mathbb{R}^n$, por lo que la ley límite del vector

$$(Y^k(t_1), \ldots, Y^k(t_n)),$$

que es precisamente $\mathbb{P}_{t_1,\ldots,t_n}$ es la ley vector

$$(Z_1, Z_1 + Z_2, \ldots, Z_1 + \cdots + Z_n),$$

que queda determinada por la media y la matriz de varianza-covarianza:

$$\mathbb{E}\left(\sum_{i=1}^{j} Z_i\right) = 0$$

y

$$\mathbb{E}\left(\left(\sum_{l=1}^{i} Z_l\right)\left(\sum_{l=1}^{j} Z_l\right)\right) = \operatorname{Var}\left(\sum_{l=1}^{i\wedge j} Z_l\right) = \sum_{l=1}^{i\wedge j} t_l - t_{l-1} = t_{i\wedge j} = t_i \wedge t_j.$$

□

[8]Utilizando el Teorema Central del Límite y los resultados que siguen al Teorema 2.7.

Si verificamos que la sucesión de medidas $(\mathbb{P}^k)_{k\in\mathbb{N}}$ es tensa, puesto que ya demostramos que sus distribuciones finito dimensionales convergen, el Teorema 2.10 nos permitirá concluir que existe una medida de probabilidad sobre $\mathscr{B}_S$, que denotaremos por $\mathbb{W}$, tal que las distribuciones finito dimensionales, que son las leyes inducidas por las proyecciones, son las mismas que las de un Movimiento Browniano y además, las medidas $\mathbb{P}^k$ convergen débilmente a $\mathbb{W}$. Así, en el espacio de probabilidad $(S, \mathscr{B}_S, \mathbb{W})$, las proyecciones $(\pi_t)_{t\geq 0}$ constituyen un Movimiento Browniano que verifica la existencia del Movimiento Browniano y además podemos concluir algunos aspectos del Movimiento Browniano a través de la aproximación mediante caminatas aleatorias.

A continuación, veamos que la sucesión $(\mathbb{P}^k)_{k\in\mathbb{N}}$ es efectivamente tensa, que por el Teorema 2.9 es equivalente a verificar que

$$\lim_{\delta\downarrow 0}\sup_{k\in\mathbb{N}}\mathbb{P}^k\big(\{f\in S:\omega^T(f,\delta)\geq\varepsilon\}\big)=0 \qquad \text{para toda} T>0 \text{ y para toda} \varepsilon>0,$$

puesto que $\mathbb{P}^k(\{f\in S: f(0)=0\})=1$ para $k\in\mathbb{N}$. Si traducimos el significado de las medidas $\mathbb{P}^k$ y del módulo de continuidad ω, la condición que debemos verificar es la siguiente:

$$\lim_{\delta\downarrow 0}\sup_{k\in\mathbb{N}}\mathbb{P}\left(\max_{\substack{|t-s|\leq\delta\\0\leq s,t\leq T}}|Y_t^k-Y_s^k|\geq\varepsilon\right)=0 \qquad \text{para toda } T>0 \text{ y para toda } \varepsilon>0,$$

que si utilizamos la definición de Y^k, se convierte en

$$\lim_{\delta\downarrow 0}\sup_{k\in\mathbb{N}}\mathbb{P}\left(\max_{\substack{|t-s|\leq k\delta\\0\leq s,t\leq kT}}|X_t-X_s|\geq\varepsilon\sigma\sqrt{k}\right)=0 \qquad \text{para toda } T>0 \text{ y para toda } \varepsilon>0. \tag{17}$$

Dado que para cada k fija, si tomamos el límite cuando $\delta\downarrow 0$ en la expresión

$$\mathbb{P}\left(\max_{\substack{|t-s|\leq k\delta\\0\leq s,t\leq kT}}|X_t-X_s|\geq\varepsilon\sigma\sqrt{k}\right)$$

obtenemos como resultado 0 para cualquier $\varepsilon>0$ y $T>0$ (puesto que cualquier medida de probabilidad sobre $\mathscr{B}_S$ es tensa ya que S es completo y separable), el supremo en (17) lo podemos cambiar por un límite superior, esto es, verificar la tensión de la sucesión de medidas que nos interesa, es equivalente a demostrar que

$$\lim_{\delta\downarrow 0}\limsup_{k\in\mathbb{N}}\mathbb{P}\left(\max_{\substack{|t-s|\leq k\delta\\0\leq s,t\leq kT}}|X_t-X_s|\geq\varepsilon\sigma\sqrt{k}\right)=0 \qquad \text{para toda } T>0 \text{ y para toda} \varepsilon>0.$$

Si $s, t \in [0, kT]$ y $|s - t| \leq k\delta$, sean $i = [t]$, $j = [s]$, $\alpha = t - i \in [0, 1]$ y $\beta = s - j \in [0, 1]$, por lo que

$$\begin{aligned}|X_t - X_s| &= |(1-\alpha) S_i + \alpha S_{i+1} - (1-\beta) S_j - \beta S_{j+1}| \\ &\leq \max \{S_i - S_j, S_i - S_{j+1}, S_{i+1} - S_{j+1}, S_{i+1} - S_j\},\end{aligned}$$

y como $|s - t| \leq k\delta$ e $|t - s| \leq kT$, entonces $|i - j|, |i - j + 1|, |i - j - 1| \leq [k\delta] + 1$ y $i + 1, j + 1 \leq [kT] + 1$, de donde obtenemos las contenciónes

$$\begin{aligned}\left\{ \max_{\substack{|t-s|\leq k\delta \\ 0\leq s,t\leq kT}} |X_t - X_s| \geq \varepsilon\sigma\sqrt{k} \right\} &\subset \left\{ \max_{\substack{|i-j|\leq [k\delta]+1 \\ 0\leq i,j\leq [kT]+1}} |S_i - S_j| \geq \varepsilon\sigma\sqrt{k} \right\} \\ &\subset \left\{ \max_{\substack{0\leq j\leq [k\delta]+1 \\ 0\leq i\leq [kT]+1}} |S_{i+j} - S_i| \geq \varepsilon\sigma\sqrt{k} \right\}\end{aligned}$$

de donde la tensión de la sucesión $\left(\mathbb{P}^k\right)_{k\in\mathbb{N}}$ se seguirá de la relación

$$\lim_{\delta\downarrow 0} \sup_{k\in\mathbb{N}} \mathbb{P}\left(\max_{\substack{0\leq j\leq [k\delta]+1 \\ 0\leq i\leq [kT]+1}} |S_{i+j} - S_i| \geq \varepsilon\sigma\sqrt{k} \right) = 0 \quad \text{para toda } T > 0 \text{ y para toda } \varepsilon > 0,$$

en la que nos concentraremos a continuación.

Sea m el menor natural mayor que T/δ (que existe por el principio del buen orden), es decir, m es el único natural que satisface las desigualdades

$$m - 1 \leq T/\delta < m.$$

Dado que

$$\lim_{n\to\infty} \frac{([kT] + 1)}{([k\delta] + 1)} = T/\delta < m,$$

existe $k_0 \in \mathbb{N}$ tal que para cada $k \geq k_0$ tiene lugar la desigualdad

$$[kT] + 1 < m\,([k\delta] + 1).$$

Designemos por I a $[kT] + 1$ y por J a $[k\delta] + 1$ y supongamos que $k \geq k_0$ y

$$\max_{\substack{0\leq j\leq [k\delta]+1 \\ 0\leq i\leq [kT]+1}} |S_{i+j} - S_i| \geq \varepsilon\sigma\sqrt{k}.$$

Esto nos dice que existen $i \in \{0, \ldots, I\}$ y $j \in \{0, \ldots, J\}$ tal que $|S_i - S_{i+j}| > \varepsilon\sigma\sqrt{k}$. Dado que $I < mJ$ existe $p \in \{0, \ldots m - 1\}$ tal que $i \in \{(p + k) I : 0 \leq k \leq I\}$ y aquí tenemos dos opciones: $i + j \in \{(p + k) : 0 \leq k \leq I\}$ o $i + j \in \{(p + 1 + k)I : 0 \leq k \leq I\}$. En el primer caso,

$$|S_{pI} - S_i| > \varepsilon\sigma\sqrt{k}/3$$

o

$$|S_{pI} - S_{i+j}| > \varepsilon\sigma\sqrt{k}/3,$$

ya que en caso contrario, la desigualdad del triángulo nos obsequia una contradicción. En el segundo caso,

$$|S_{pI} - S_i| > \varepsilon\sigma\sqrt{k}/3$$

o

$$|S_{pI} - S_{(p+1)I}| > \varepsilon\sigma\sqrt{k}/3$$

o

$$|S_{(p+1)I} - S_{i+j}| > \varepsilon\sigma\sqrt{k}/3,$$

ya que, de nueva cuenta, el suponer lo contrario nos lleva a una contradicción. Así, en cualquier caso, observamos que existen $p \in \{0, \ldots, m-1\}$ y $j \in 0, \ldots, J$ tal que $|S_{pI} - S_{j+pI}| > \varepsilon\sigma\sqrt{k}/3$, lo cual nos lleva a la contención

$$\left\{ \max_{\substack{0 \le j \le [k\delta]+1 \\ 0 \le i \le [kT]+1}} |S_{i+j} - S_i| \ge \varepsilon\sigma\sqrt{k} \right\} \subset \bigcup_{p=0}^{m-1} \left\{ \max_{0 \le j \le J} |S_{pJ} - S_{j+pJ}| > \varepsilon\sigma\sqrt{k}/3 \right\}$$

y dado que para cualquier $p \in \{0, \ldots, m-1\}$, los conjuntos del lado derecho de la contención arriba tienen la misma probabilidad, la subaditividad finita de $\mathbb{P}$ nos dice que

$$\begin{aligned} \mathbb{P}\left(\max_{\substack{0 \le j \le [k\delta]+1 \\ 0 \le i \le [kT]+1}} |S_{i+j} - S_i| \ge \varepsilon\sigma\sqrt{k} \right) &\le m\mathbb{P}\left(\max_{0 \le j \le J} |S_j| > \varepsilon\sigma\sqrt{k}/3 \right) \\ &\le \left(1 + \frac{T}{\delta}\right) \mathbb{P}\left(\max_{0 \le j \le [k\delta]+1} |S_j| > \varepsilon\sigma\sqrt{k}/3 \right). \end{aligned} \tag{18}$$

Recordemos que para cada $\delta > 0$, existe $k_0 \in \mathbb{N}$ tal que para cualquier $k \ge k_0$ tiene lugar la desigualdad (18).

A continuación, demostraremos que tiene lugar la relación

$$\lim_{\delta \downarrow 0} \limsup_{k \to \infty} \frac{1}{\delta} \mathbb{P}\left(\max_{0 \le j \le [k\delta]+1} |S_j| > \varepsilon\sigma\sqrt{k} \right) = 0 \quad \forall \varepsilon > 0,$$

lo cual nos dice que el límite cuando $\delta \downarrow 0$ del límite superior (sobre k) de la última expresión que interviene en la última desigualdad del párrafo anterior es cero, lo cual implica que la sucesión $(\mathbb{P}^k)_{k \in \mathbb{N}}$ es tensa. Para esto, utilizaremos la desigualdad de Etmandi, la cual afirma que

$$\mathbb{P}\left(\max_{1 \le j \le k} |S_j| > 3x \right) \le 3 \max_{1 \le j \le k} \mathbb{P}(|S_j| > x).$$

Para verificar que esto sucede, sean $B_j = \{|S_i| \leq 3x, i < j, |S_j| > 3x\}$, por lo que estos conjuntos son ajenos para distintos valores de j. Entonces

$$\begin{aligned}
\mathbb{P}\left(\max_{1\leq j\leq k}|S_j| > 3x\right) &\leq \mathbb{P}(|S_k| > x) + \sum_{j=1}^{k}\mathbb{P}(|S_k| \leq x, B_j) \\
&\leq \mathbb{P}(|S_k| > x) + \sum_{j=1}^{k}\mathbb{P}(|S_k - S_j| > 2x, B_j) \\
&\leq \mathbb{P}(|S_k| > x) + \sum_{j=1}^{k}\mathbb{P}(|S_k - S_j| > 2x)\,\mathbb{P}(B_j) \\
&\leq \mathbb{P}(|S_k| > x) + \max_{1\leq j\leq k}\mathbb{P}(|S_k - S_j| > 2x) \\
&\leq \mathbb{P}(|S_k| > x) + \max_{1\leq j\leq k}\left(\mathbb{P}(|S_k| > x) + \mathbb{P}(|S_j| > x)\right) \\
&\leq 3\max_{1\leq j\leq k}\mathbb{P}(|S_j| > x)\,.
\end{aligned}$$

Para aplicarla, notemos que $\sqrt{k}/\sqrt{[k\delta]+2}$ converge a $1/\delta$ cuando $k \to \infty$, por lo que dada $\delta > 0$, existe $k_\delta \in \mathbb{N}$ tal que si $k \geq k_\delta$, entonces $\sqrt{k} > \sqrt{[k\delta]}/\sqrt{2\delta}$. Así, al utilizar la desigualdad de Etmandi y el comentario de la frase anterior:

$$\begin{aligned}
\limsup_{k\to\infty}\mathbb{P}\left(\max_{1\leq j\leq [k\delta]+2}|S_j| > \varepsilon\sigma\sqrt{k}\right) &\leq 3\limsup_{k\to\infty}\max_{1\leq j\leq [k\delta]+2}\mathbb{P}\left(|S_j| > \varepsilon\sigma\sqrt{k}/3\right) \\
&\leq 3\limsup_{k\to\infty}\max_{1\leq j\leq [k\delta]+2}\mathbb{P}\left(|S_j| > \varepsilon\sigma\frac{\sqrt{[k\delta]+2}}{\left(3\sqrt{2\delta}\right)}\right).
\end{aligned}$$

Por el Teorema Central del Límite, la sucesión $\left(S_j/\sigma\sqrt{j}\right)_{j\in\mathbb{N}}$ converge en distribución a Z, donde Z tiene distribución normal estándar. Como la ley normal le asigna cero a los conjuntos que constan de un sólo punto, se sigue que la distribución de $S_j/\sqrt{j}$ converge puntualmente a la de Z y como $\mathbb{P}(|Z| > x) \leq \mathbb{E}(Z^4)/x^4$ (desigualdad de Tchebyshev), entonces dada $\delta > 0$, existe $k_\delta \in \mathbb{N}$ (sólo depende de δ, no de k) tal que si $k_\delta \leq j \leq [k\delta]+2$, entonces

$$\begin{aligned}
\mathbb{P}\left(|S_j| > \frac{\varepsilon\sigma\sqrt{[k\delta]+2}}{\left(3\sqrt{2\delta}\right)}\right) &\leq \mathbb{P}\left(|S_j| > \frac{\varepsilon\sigma\sqrt{j}}{\left(3\sqrt{2\delta}\right)}\right) \\
&< \mathbb{E}(Z^4)\,4\cdot\frac{3^4\delta^2}{\varepsilon^4}.
\end{aligned}$$

Para $j \leq k_\delta$, podemos utilizar la desigualdad de Tchebyshev:

$$\mathbb{P}\Big(|S_j| > \varepsilon\sigma\sqrt{[k\delta]+2}/\left(3\sqrt{2\delta}\right)\Big) \leq j\frac{18\delta}{\varepsilon^2\left([k\delta]+2\right)}$$
$$\leq k_\delta\frac{18\delta}{\varepsilon^2\left([k\delta]+2\right)},$$

por lo que

$$\limsup_{k\to\infty} 3 \max_{1\leq j\leq [k\delta]+2} \mathbb{P}\left(|S_j| > \frac{\varepsilon\sigma\sqrt{j}}{\left(3\sqrt{2\delta}\right)}\right) \leq \limsup_{k\to\infty}\left(\mathbb{E}(Z^4)\,\frac{12\cdot 3^4\delta^2}{\varepsilon^4} \vee \frac{18k_\delta\delta}{\varepsilon^2\left([k\delta]+2\right)}\right)$$
$$= \mathbb{E}(Z^4)\,12\cdot\frac{3^4\delta^2}{\varepsilon^4}.$$

De esta manera,

$$\lim_{\delta\downarrow 0}\limsup_{k\to\infty}\frac{1}{\delta}\mathbb{P}\left(\max_{0\leq j\leq [k\delta]+1}|S_j| > \varepsilon\sigma\sqrt{k}\right) \leq \lim_{\delta\downarrow 0}\mathbb{E}(Z^4)\,12\cdot\frac{3^4\delta}{\varepsilon^4} = 0.$$

de donde podemos concluir que la sucesión $(\mathbb{P}_k)_{k\in\mathbb{N}}$ es tensa y por lo tanto, existe una medida de probabilidad $\mathbb{W}$ sobre $\mathscr{B}_S$ bajo la cual las proyecciones $(\pi_t)_{t\geq 0}$ constituyen un Movimiento Browniano y que la sucesión de medidas $\left(\mathbb{P}^k\right)_{k\in\mathbb{N}}$ definidas en términos de la caminata aleatoria $(S_n)_{n\in\mathbb{N}}$ converge débilmente a ella.

2.4. La ley arcoseno

Como un ejemplo de la aplicación del principio de invariancia de Donsker, encontraremos la distribución de la fracción de tiempo que el Movimiento Browniano es positivo durante un intervalo de tiempo. No sólo verificaremos de nueva cuenta el principio de invariancia de Erdös y Kac, sino que obtendremos un resultado sobre el Movimiento Browniano, que es más difícil de obtener sin el uso de convergencia en distribución.

En el espacio $S = \mathscr{C}_{[0,\infty)}$, tenemos la métrica d, bajo la cual, si $(f_n)_{n\in\mathbb{N}}$ es una sucesión de elementos de S, la relación $d(f_n, f) \to 0$ conforme $n \to \infty$ se cumple si y sólo si f_n converge uniformemente a f sobre cualquier compacto contenido en $[0,\infty)$. Además, comprobamos la existencia de una medida de probabilidad $\mathbb{W}$ sobre $\mathscr{B}_S$ tal que las proyecciones $(\pi_t)_{t\geq 0}$ constituyen un Movimiento Browniano definido en el espacio de probabilidad $(S, \mathscr{B}_S, \mathbb{W})$ y hemos exhibido a esta medida como el límite débil de medidas de probabilidad asociadas a caminatas aleatorias, con la única restricción de que las variables que intervienen en su definición tengan varianza finita. El problema que se plantea, en analogía con el desarrollado en el capítulo anterior, es calcular la distribución de la medida de Lebesgue del conjunto $\{t \in [0,1] : \pi_t(f) > 0\}$. A continuación se presenta un esbozo de la demostración: Si denotamos por h a la asignación que manda a $f \in S$ en la anterior integral, se verifica la medibilidad de h y después se demuestra que $\partial(\{f \in S : h(f) \le c\})$ tiene medida $\mathbf{W}$ cero, por lo que

$$\mathbf{W}(h \le c) = \lim_{k\to\infty} \mathbb{P}(h \circ Y^k \le c),$$

donde Y^k es el proceso que consideramos para el principio de invariancia de Donsker. Al escoger la distribución de Y_1^1 de manera conveniente, podemos calcular el límite que aparece en el lado derecho de la igualdad anterior, de donde obtenemos la distribución de h.

Empecemos por la medibilidad de h:

LEMA 2.5. *La función $h : S \to \mathbb{R}$ es $(\mathscr{B}_S, \mathscr{B}_\mathbf{R})$-medible.*

DEMOSTRACIÓN. Si consideramos al espacio $[0,\infty) \times S$ con la topología producto, inducida por la métrica $((t_1, f_1), (t_2, f_2)) \mapsto |t_1 - t_2| \vee d(f_1, f_2)$, dado que ambos espacios son separables, se sigue que la σ-álgebra Borel en el producto es el producto de las σ-álgebras $\mathscr{B}_{[0,\infty)} \otimes \mathscr{B}_S$. Sea $f : [0,\infty) \times S \to \mathbb{R}$ la función cuya regla de correspondencia es $(t, f) \mapsto \pi_t \circ f$, que es continua puesto que si $(t_n, f_n) \to (t, f)$ cuando $n \to \infty$, entonces f_n converge uniformemente a f en $[0, t+1]$ y $t_n \to t \in [0, t+1]$ cuando $n \to \infty$, por lo cual $f_n(t_n) \to f(t)$ cuando $n \to \infty$. La continuidad de f implica que es $(\mathscr{B}_{[0,\infty)\times S}, \mathscr{B}_\mathbf{R})$-medible. Esto nos dice que $\{(t, f) \in [0,\infty) \times S : \pi_t(f) > 0\} \in \mathscr{B}_{[0,\infty)\times S}$, por lo que si denotamos por λ a la medida de Lebesgue restringida a $[0,\infty)$, la asignación

$$h : f \mapsto \int \mathbf{1}_{(\pi_t(f)>0)} \mathbf{1}_{[0,1]\times S}\, d\lambda$$

es $(\mathscr{B}_S, \mathscr{B}_\mathbf{R})$-medible. □

Los siguientes dos resultados nos ayudarán a verificar que podemos calcular la distribución de h mediante aproximaciones por caminatas aleatorias:

LEMA 2.6. *Si*

$$A = \{f \in S : \lambda(\{t \in [0,1] : f(t) = 0\}) = 0\},$$

entonces $A \in \mathscr{B}_S$ y $\mathbf{W}(A) = 1$.

DEMOSTRACIÓN. Si utilizamos la misma técnica que en el lema anterior, podemos verificar que $g : (t, f) \mapsto \mathbf{1}_{(f(t)=0)}$ es $(\mathscr{B}_{[0,\infty)\times S}, \mathscr{B}_\mathbf{R})$-medible, por lo que la asignación

$$f \mapsto \lambda(\{t \in [0,1] : f(t) = 0\}) = \int g \mathbf{1}_{[0,1]\times S}\, d\lambda$$

es $(\mathscr{B}_S, \mathscr{B}_\mathbf{R})$-medible. Por otro lado, como $(t, f) \mapsto \mathbf{1}_{(f(t)=0)}$ es $(\mathscr{B}_{[0,\infty)\times S}, \mathscr{B}_\mathbf{R})$-medible y no-negativa, podemos aplicar el teorema de Tonelli[9] para concluir que

$$\mathbb{E}\left(\int_{[0,1]} \mathbf{1}_{(f(t)=0)}\, d\lambda\right) = \int_{[0,1]} \mathbb{E}\left(\mathbf{1}_{(f(t)=0)}\right)\, d\lambda = \int_{[0,1]} \mathbf{W}(f(t) = 0)\, d\lambda = 0,$$

[9]Ver [2].

donde la última igualdad es consecuencia de que $f(t)$ tenga distribución normal no-degenerada para $t > 0$. Pero la variable $\lambda(\{t \in [0,1] : f(t) = 0\})$ es no-negativa y el que tenga esperanza cero nos dice que es cero (c.s. rel. $\mathbf{W}$) y como

$$A = \{f \in S : \lambda(\{t \in [0,1] : f(t) = 0\}) = 0\},$$

entonces $\mathbf{W}(A) = 1$. □

LEMA 2.7. *Si*

$$B = \{f \in S : h(f) \leq c\},$$

entonces $\partial(B) \subset \left\{f \in S : h(f) f \leq c \leq \hat{h}(f)\right\}$, *donde* $\hat{h}(f) = \int_{[0,1]} \mathbf{1}_{(f(t) \geq 0)}\, d\lambda$.

DEMOSTRACIÓN. Si $f \in \partial(B)$, entonces existe una sucesión $\left(f_1^k\right)_{k \in \mathbf{N}}$ contenida en B tal que $f_1^n \longrightarrow f$ cuando $n \longrightarrow \infty$ y existe una sucesión $\left(f_2^k\right)_{k \in \mathbf{N}}$ contenida en B^c tal que $f_2^k \longrightarrow f$ cuando $k \longrightarrow \infty$. Si $f(t) > 0$, entonces a partir de cierta $k_1 \in \mathbb{N}$, $f_1^k(t) > 0$, por lo que

$$1 = \mathbf{1}_{(f(t) > 0)} = \liminf_{k \to \infty} \mathbf{1}_{\left(f_1^k(t) > 0\right)}.$$

Si $f(t) \leq 0$, entonces

$$0 = \mathbf{1}_{(f(t) > 0)} \leq \liminf_{k \to \infty} \mathbf{1}_{\left(f_1^k(t) > 0\right)},$$

por lo que tenemos la desigualdad $\mathbf{1}_{(f(t)>0)} \leq \liminf_{k \to \infty} \mathbf{1}_{(f(t)>0)}$ y por el lema de Fatou,

$$h(f) = \int_{[0,1]} \mathbf{1}_{(f(t)>0)}\, d\lambda \leq \liminf_{k \to \infty} \int_{[0,1]} \mathbf{1}_{\left(f_1^k(t)>0\right)}\, d\lambda \leq c,$$

puesto que $f_1^k \in B$. Por otro lado, si $f(t) \geq 0$, entonces

$$1 = \mathbf{1}_{(f(t) \geq 0)} \geq \limsup_{k \to \infty} \mathbf{1}_{\left(f_2^k(t) \geq 0\right)}$$

y si $f(t) < 0$, entonces $f_2^k(t) < 0$ a partir de un cierta $k_2 \in \mathbb{N}$, por lo que

$$0 = \mathbf{1}_{(f(t) \geq 0)} = \limsup_{k \to \infty} \mathbf{1}_{\left(f_2^k(t) \geq 0\right)},$$

de donde $\mathbf{1}_{(f(t) \geq 0)} \geq \limsup_{k \to \infty} \mathbf{1}_{\left(f_2^k(t) \geq 0\right)}$ y dado que $\mathbf{1}_{\left(f_2^k(t) \geq 0\right)} \leq 1$, y 1 es integrable sobre $[0,1]$ respecto a la medida de Lebesgue, entonces

$$\hat{h}(f) = \int_{[0,1]} \mathbf{1}_{(f(t) \geq 0)}\, d\lambda \geq \limsup_{k \to \infty} \int_{[0,1]} \mathbf{1}_{\left(f_2^k(t) \geq 0\right)}\, d\lambda \geq c.$$

□

NOTA. De la misma manera que verificamos la medibilidad de h, se puede probar la de $\hat{h}$. Así, podemos considerar a ambas como variables aleatorias reales definidas en el espacio de probabilidad $(S, \mathscr{B}_S, \mathbb{W})$.

En el siguiente teorema, denotaremos por A a la distribución arcoseno que nos hemos encontrado anteriormente:

$$A(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x \geq 1 \\ \int_{[0,x]} \frac{1}{\pi\sqrt{t(1-t)}}\, dt & x \in (0,1) \end{cases}.$$

TEOREMA 2.12. *La distribución de la variables aleatoria h definida sobre el espacio de probabilidad $(S, \mathscr{B}_S, \mathbb{W})$, dada por*

$$F_h(x) = \mathbb{W}(\{f \in S : h(f) \leq c\}),$$

es igual a A.

DEMOSTRACIÓN. Si $\mathscr{D}$ es el conjunto de puntos en el cual F_h no es continua, entonces dado que esta función es creciente, $\mathscr{D}$ es a lo más numerable y por lo tanto su complemento es denso. Si consideramos $c \in \mathscr{D}^c$, entonces

$$\begin{aligned} \mathbb{W}(\partial(\{f \in S : h(f) \leq c\})) &\leq \mathbb{W}\Big(\Big\{f \in S : h(f) \leq c \leq \hat{h}(f)\Big\}\Big) \\ &\leq \mathbb{W}(\{f \in S : h(f) = c\}) \\ &= 0, \end{aligned}$$

por lo que al utilizar el principio de invariancia de Donsker y el Teorema Portmanteau

$$F_h(c) = \lim_{k\to\infty} \mathbb{P}(h \circ Y^k \leq c).$$

Como

$$h \circ Y^k = \int_{[0,1]} \mathbf{1}_{(Y_t^k > 0)}\, d\lambda = \int_{[0,1]} \mathbf{1}_{(X_{kt} > 0)}\, d\lambda = \frac{1}{k}\int_{[0,k]} \mathbf{1}_{(X_s > 0)}\, d\lambda$$

vemos que si X_1 es tal que $\mathbb{P}(X_1 = -1) = \mathbb{P}(X_1 = 1) = 1/2$, entonces $h \circ Y^{2k}$ es la fracción de tiempo en que la caminata aleatoria simple de $2k$ pasos es positiva y por lo tanto

$$F_h(c) = \lim_{k\to\infty} \mathbb{P}(h \circ Y^k \leq c) = \lim_{k\to\infty} \mathbb{P}(h \circ Y^{2k} \leq c) = A(c), \quad c \in \mathscr{D}^c,$$

como vimos anteriormente. Pero si F_h y A coinciden en un conjunto denso, entonces son iguales, puesto que ambas son continuas por la derecha. □

Notas

El desarrollo de éste capítulo se basa principalmente en el material contenido en las referencias [4], [8] y [10]. El artículo [4] nos ha servido como uns introducción al principio de invariancia de Donsker, mismo que se ha probado cuando el espacio de funciones es $\mathscr{C}_{[0,\infty)}$ y no $\mathscr{C}_{[0,1]}$ como se hace usualmente, basándonos en [10]. Finalmente, para el cálculo de la distribución de h bajo la medida de Wiener, el material contenido en [8] fué de gran ayuda y sorprende que se pueda obtener este resultado por aproximación cuando es tan difícil de obtener utilizando otros métodos, por ejemplo, mediante el uso de la fórmula de Feynman-Kac o mediante el estudio del tiempo local.

CAPÍTULO 3

El proceso Poisson compuesto y la ley arcoseno

3.1. Convergencia débil en el espacio de Skorohod

Con el objeto de estudiar la convergencia débil de sucesiones de medidas de probabilidad inducidas por procesos estocásticos con valores en $\mathbb{R}$ que no sean necesariamente continuos es que se introduce el espacio de Skorohod D. Los elementos de D son las funciones $f : [0, 1] \longrightarrow \mathbb{R}$ tal que:

a) Para cada $t \in (0, 1]$, el límite $\lim_{s \uparrow t} f(s)$ existe, esto es, f admite límites por la izquierda.
b) Para cada $t \in [0, 1)$, $\lim_{s \downarrow t} f(s) = f(t)$, o dicho de otra manera, f es continua por la derecha.

Es importante notar que estas funciones son acotadas, que la cardinalidad de

$$\{t \in [0, 1] : |f(t) - f(t-)| > \varepsilon\}$$

es finita, de lo cual se sigue que cada elemento de D tiene a lo más una cantidad numerable de discontinuidades, y que son $(\mathscr{B}_{[0,1]}, \mathscr{B}_{\mathbb{R}})$-medibles,

Para poder hablar de convergencia en distribución en D, primero debemos dotarlo de una métrica. Como D es subconjunto del espacio de funciones acotadas de $[0, 1]$ en $\mathbb{R}$, es natural pensar en la métrica inducida por la norma uniforme en ese espacio, dada por

$$\|f\| = \sup_{t \in [0,1]} |f(t)|,$$

para metrizar a D; sin embargo, dicha métrica resulta limitada en cuanto a la convergencia se refiere: la sucesión de funciones $\mathbf{1}_{[0,\alpha+1/n)} \in D$ ($\alpha \in (0, 1)$ y $n \geq 1/(1-\alpha)$) no converge a $\mathbf{1}_{[0,\alpha)} \in D$ con la norma uniforme y con la métrica que definiremos en ese espacio sí lo hace. Para definir a la métrica de Skorohod d en $D \times D$, sea Λ el conjunto que consta de aquellas funciones $\lambda : [0, 1] \longrightarrow [0, 1]$ que son estrictamente crecientes, sobreyectivas y continuas (con la topología usual,) por lo que cada una de ellas es homeomorfismo, puesto que $[0, 1]$ es compacto y Hausdorff y $\mathbb{R}$ es Hausdorff, de donde resulta que si $\lambda \in \Lambda$, entonces $\lambda^{-1} \in \Lambda$[1]. La métrica d está dada por

$$d(f, g) = \inf_{\lambda \in \Lambda} \|f - g \circ \lambda\| \vee \|\lambda - \mathrm{Id}\|,$$

[1]Ver [11].

donde Id $\in \Lambda$ es la función identidad y $\|\cdot\|$ denota a la norma uniforme. Bajo la métrica d, el espacio D es separable y cuando restringimos d a $\mathscr{C}([0,1])$, el espacio de funciones continuas de $[0,1]$ en $\mathbb{R}$, resulta equivalente a la métrica inducida por la norma uniforme; sin embargo, utilizando d, el espacio de Skorohod D no es completo. Para resolver este problema, se introduce una métrica más complicada que es equivalente a d en el sentido de inducir la misma topolgía (por lo que la convergencia en distribución en cualquiera de los dos espacios métricos es la misma,) pero que convierte a D en completo y separable. La completez se utiliza en la caracterización de los conjuntos compactos de (D,d) y nos permite afirmar que una familia de medidas de probabilidad definidas en los Borelianos de D (al utilizar cualquiera de las dos métricas) es tensa si y sólo si es secuencialmente compacta (por el Teorema de Prohorov.)

Para caracterizar a los conjuntos compactos en el espacio (D,d), se introduce una función que jugará el mismo papel que el módulo de continuidad en $\mathscr{C}_{[0,\infty)}$.

DEFINICIÓN. Decimos que el conjunto $\{t_i\}_{i=0}^n$ se encuentra δ-esparcido si $t_0 = 0 < t_1 < \cdots < t_n = 1$ y

$$\min_{i=1\dots n} t_i - t_{i-1} \geq \delta.$$

Para cada $A \subset [0,1]$, sea

$$m_A(f) = \sup_{s,t\in A} |f(t) - f(s)|$$

y definimos $m' : D \times [0,\infty) \to \mathbb{R}$ como sigue:

$$m'(f,\delta) = \inf\left\{\max_{1\leq i\leq n} m_{[t_{i-1},t_i)}(f) : n \geq 1 \text{ y } \{t_i\}_{i=0}^n \text{ es } \delta\text{-esparcido}\right\}.$$

La caracterización de los conjuntos compactos de (D,d) es como sigue:

TEOREMA 3.1. *Un subconjunto A de D tiene cerradura compacta si y sólo si*

a) $\sup_{f\in A} \|f\| < \infty$.
b) $\lim_{\delta\downarrow 0} \sup_{x\in A} m'(f,\delta) = 0$.

Armados con la caracterización del Teorema 3.1 se puede obtener una condición necesaria y suficiente para poder concluir la tensión de una sucesión de medidas de probabilidad:

TEOREMA 3.2. *Una sucesión de medidas de probabilidad $(\mathbb{P}_k)_{k\in\mathbb{N}}$ definidas en $\mathscr{B}_D$ es tensa si y sólo si*

a) $\lim_{a\to\infty} \sup_{k\in\mathbb{N}} \mathbb{P}_k(\{f \in D : \|f\| > a\}) = 0$.
b) $\lim_{\delta\downarrow 0} \sup_{k\in\mathbb{N}} \mathbb{P}_k(\{f \in D : m'(f,\delta) > \varepsilon\}) = 0$ *para toda* $\varepsilon > 0$.

Consideremos a las proyecciones $\pi_{t_1,\dots,t_n} : D \to \mathbb{R}^n$ definidas mediante

$$\pi_{t_1,\dots,t_n}(f) = (f(t_1),\dots,f(t_n)),$$

para $t_1,\dots,t_n \in [0,1]$, como lo hemos hecho anteriormente en el caso de S. Estas funciones son $(\mathscr{B}_D, \mathscr{B}_{\mathbb{R}^n})$-medibles y además, la σ-álgebra generada por $\{\pi_t : t \in [0,1]\}$ es igual a la σ-álgebra de Borel en D, lo que garantiza que una función $f : X \to D$ definida en un espacio medible $(X, \mathscr{X})$ es $(\mathscr{X}, \mathscr{B}_D)$- medible si y sólo si $\pi_t \circ f$ es $(\mathscr{X}, \mathscr{B}_{\mathbb{R}})$-medible para toda $t \in [0, 1]$. Mediante las proyecciones podemos definir a las distribuciones finito-dimensionales de una medida de probabilidad $\mathbb{P}$ sobre $\mathscr{B}_D$, como en el caso de S, mediante la asignación

$$\mathbb{P}_{t_1,\dots,t_n}(A) = \mathbb{P}(\{f \in D : \pi_{t_1,\dots,t_n} \circ f \in A\}), \quad A \in \mathscr{B}_{\mathbb{R}^n}.$$

Si $\mathbb{P}^k \Rightarrow \mathbb{P}$, no podemos, como en el caso de S, deducir que $\mathbb{P}^k_{t_1,\dots,t_n} \Rightarrow \mathbb{P}_{t_1,\dots,t_n}$, puesto que las proyecciones no son necesariamente continuas, como lo muestra el siguiente resultado:

TEOREMA 3.3. *π_t es continua en f si y sólo si f es continua en t.*

Las distribuciones finito dimensionales se utilizan posteriormente para verificar la convergencia débil, con la ayuda del concepto de tensión.

Si consideramos una función $f \in D$, entonces existe la cantidad $j(f) = \text{máx}_{t\in(0,1]}|f(t) - f(t-)|$, el máximo salto de la función f, y la asignación $f \mapsto j(f)$ es continua, por lo que $C = \mathscr{C}_{[0,1]} = j^{-1}(0)$ es un elemento de $\mathscr{B}_D$.

TEOREMA 3.4. *Sea $(\Omega, \mathscr{F}, \mathbb{P})$ un espacio de probabilidad y $(X_n)_{n\in\mathbb{N}}$ una sucesión de variables aleatorias con valores en D que convergen en distribución a X. Una condición necesaria y suficiente para que $\mathbb{P}(X \in C) = 1$ es que $j(X_n) \xrightarrow{\mathscr{D}} 0$.*

El siguiente resultado nos permite probar la convergencia débil en D utilizando las condiciones que garantizan convergencia débil en C. Las primeras dos condiciones garantizan la tensión (en D) de la sucesión de medidas de probabilidad, mientras que la tercera nos permite identificar al límite débil.

TEOREMA 3.5. *Si $(\mathbb{P}^k)_{k\in\mathbb{N}}$ es una sucesión de medidas de probabilidad definidas en $\mathscr{B}_D$ tal que*

a) $\lim_{a\to\infty} \limsup_{k\to\infty} \mathbb{P}^k(\{f \in D : |f(0)| > a\}) = 0,$
b) $\lim_{\delta\downarrow 0} \limsup_{k\to\infty} \mathbb{P}^k(\{f \in D : \omega^1(f, \delta)\} > \varepsilon) = 0$ *para toda* $\varepsilon > 0$ *y*
c) $\mathbb{P}^k_{t_1,\dots,t_n} \Rightarrow \mathbb{P}_{t_1,\dots,t_n}$ *si* $t_1, \dots, t_n \in [0, 1]$,

donde $\mathbb{P}$ es una medida de probabilidad sobre $\mathscr{B}_D$, entonces $\mathbb{P}^k \Rightarrow \mathbb{P}$ y $\mathbb{P}(C) = 1$.

Notemos que en el teorema anterior, se utiliza a la función ω^1, definida en 2.3.3, y no a m'.

3.2. Procesos de Lévy

Un proceso de Levy es una extensión de lo que representa la caminata aleatoria como proceso a tiempo discreto. Las propiedades que nos interesan de esta última son la independencia y la estacionariedad de los incrementos. Hay otras dos condiciones que ayudan a que el proceso sea más fácil de estudiar.

DEFINICIÓN. Un proceso de Levy $X = (X_t)_{t\geq 0}$ es una colección de variables aleatorias reales indicadas por los elementos de $[0,\infty)$ definidas sobre un espacio de probabilidad $(\Omega, \mathscr{F}, \mathbb{P})$ tales que

a) $X_0 = 0$.
b) La asignación $t \mapsto X_t(\omega)$ es continua por la derecha en $[0,\infty)$ y admite límites por la izquierda en $(0,\infty)$ para cualquier $\omega \in \Omega$.
c) Si $0 = t_0 \leq t_1 \leq \cdots \leq t_n$, entonces $\{X_{t_i} - X_{t_{i-1}}\}_{i=1}^n$ son variables aleatorias independientes.
d) Si $s \leq t$, entonces $X_t - X_s$ tiene la misma distribución que X_{t-s}.

Como un primer ejemplo de un proceso de Levy, tenemos al Movimiento Browniano. Otro ejemplo importante lo encontramos en el proceso Poisson:

DEFINICIÓN. Un proceso de Levy $X = (X_t)_{t\geq 0}$ tal que X_t tiene distribución Poisson de parámetro λt, con λ un real positivo, es un proceso Poisson de intensidad λ.

La existencia del proceso Poisson se demuestra verificando que si $(S_n)_{n=0}^{\infty}$ es una caminata aleatoria tal que S_1 tiene distribución exponencial de parámetro $\lambda > 0$, entonces el proceso estocástico $(N_t)_{t\geq 0}$ dado por

$$N_t = \sum_{k=1}^{\infty} k\mathbf{1}_{t\in[S_k,S_{k+1})},$$

es un proceso Poisson de parámetro λ. De hecho, el proceso Poisson es el único proceso de conteo que es además un proceso de Lévy y esta propiedad es la que justifica su aplicación en ciertos modelos actuariales.

A través del proceso Poisson podemos construir una clase de procesos de Levy conocidos como procesos Poisson compuestos. Para esto, consideramos $N = (N_t)_{t\geq 0}$ un Proceso Poisson de intensidad λ y $(X_i)_{i=1}^n$ una sucesión de variables aleatorias independientes e idénticamente distribuidas tal que las dos familias de variables aleatorias $\{N_t : t \geq 0\}$ y $\{X_i : i \geq 1\}$ son independientes. Sean $S_0 = 0$ y $S_n = \sum_{i=1}^n X_i$ para $n \geq 1$. Si $Y_t = S_{N_t}$, entonces al proceso $(Y_t)_{t\geq 0}$ lo llamamos proceso Poisson compuesto. Veamos que es un proceso de Levy:

a) Como $N_0 = 0$ y $S_0 = 0$, entonces $Y_0 = 0$.
b) Como N es càdlàg y con valores enteros, se sigue que $\lim_{s\downarrow t} Y_s = Y_t$ y que $\lim_{s\uparrow t} Y_s$ existe (y es igual a $S_{N_{t-}}$.)

c) Si $0 = t_0 \leq t_1 \leq \cdots \leq t_n$ y $A_i \in \mathscr{B}_{\mathbb{R}}$ para $i = 1, \ldots, n$, entonces

$$\begin{aligned}
&\mathbb{P}(Y_{t_i} - Y_{t_{i-1}} \in A_i, i = 1, \ldots, n) \\
&= \sum_{0=k_0 \leq k_1 \leq \cdots \leq k_n} \mathbb{P}(S_{k_i} - S_{k_{i-1}} \in A_i, N_{t_i} = k_i, i = 1, \ldots, n) \\
&= \sum_{0=k_0 \leq k_1 \leq \cdots \leq k_n} \prod_{i=1}^{n} \mathbb{P}(S_{k_i} - S_{k_{i-1}} \in A_i)\, \mathbb{P}(N_{t_i} = k_i, i = 1, \ldots, n),
\end{aligned}$$

puesto que las variables $\{S_i : i = 0, 1, \cdots\}$ y $\{N_t : t \geq 0\}$ son independientes. Si $k_2 \geq k_1$, entonces $S_{k_2} - S_{k_1}$ tiene la misma distribución que $S_{k_1 - k_2}$, entonces

$$\begin{aligned}
&\sum_{0=k_0 \leq k_1 \leq \cdots \leq k_n} \prod_{i=1}^{n} \mathbb{P}(S_{k_i} - S_{k_{i-1}} \in A_i)\, \mathbb{P}(N_{t_i} = k_i, i = 1, \ldots, n) \\
&= \sum_{0=k_0 \leq k_1 \leq \cdots \leq k_n} \prod_{i=1}^{n} \mathbb{P}(S_{k_i - k_{i-1}} \in A_i) \prod_{i=1}^{n} \mathbb{P}(N_{t_i} - N_{t_{i-1}} = k_i - k_{i-1}) \\
&= \sum_{0 \leq j_1, \ldots, j_n} \prod_{i=1}^{n} \mathbb{P}(S_{j_i} \in A_i, N_{t_i - t_{i-1}} = j_i) \\
&= \prod_{i=1}^{n} \mathbb{P}(Y_{t_i - t_{i-1}} \in A_i)\,.
\end{aligned}$$

Si utilizamos el caso $n = 2$, con $A_1 = \mathbb{R}$, podemos concluir que $Y_{t_2} - Y_{t_1}$ y $Y_{t_2 - t_1}$ tienen la misma distribución, de donde la igualdad

$$\mathbb{P}(Y_{t_i} - Y_{t_{i-1}} \in A_i, i = 1, \ldots, n) = \prod_{i=1}^{n} \mathbb{P}(Y_{t_i - t_{i-1}} \in A_i)$$

nos permite concluir la independencia de los incrementos y por lo tanto, que el proceso Poisson compuesto es un proceso de Lévy.

En la siguientes dos secciones, responderemos a la pregunta planteada en la introducción de la tesis, pero en vez de considerar un proceso Poisson compuesto con saltos acotados, consideraremos un caso más general, el de procesos de Lévy con varianza finita. A continuación veremos que un proceso Poisson compuesto con saltos acotados tiene momentos de segundo orden. Esto sucede pues como $\mu = \mathbb{E}(S_1)$ y $\sigma^2 = \mathrm{Var}(S_1)$ son finitas al ser S_1 acotada (digamos por M), entonces

$$\begin{aligned}\mathbb{E}(Y_t^2) &= \sum_{n=0}^{\infty} \mathbb{E}(S_n^2 \mathbf{1}_{(N_t=n)}) \\ &= \sum_{n=0}^{\infty} (n\sigma^2 + n^2\mu^2)\, \mathbb{P}(N_t = n) \\ &= \mathbb{E}(\sigma^2 N_t + \mu^2 N_t^2) < \infty,\end{aligned}$$

ya que las variables Poisson tienen momentos de cualquier orden.

Si X es un proceso de Lévy y $\mathrm{Var}(X_1) < \infty$, vemos que

$$\mathrm{Var}(X_{t+s}) = \mathrm{Var}(X_{t+s} - X_t + X_t) = \mathrm{Var}(X_{t+s} - X_t) + \mathrm{Var}(X_t)\,,$$

por lo que la asignación $t \mapsto \mathrm{Var}(X_t)$ (que podría en principio tomar el valor $= \infty$) es creciente y por lo tanto basta que exista $s > 0$ tal que $\mathrm{Var}(X_s) < \infty$ para que $\mathrm{Var}(X_t) < \infty$ para toda $t \geq 0$, ya que $\mathrm{Var}(X_{ns}) = n\mathrm{Var}(X_s) < \infty$ para toda $n \in \mathbb{N}$.

De hecho, podemos expresar a la varianza al tiempo t en términos de la varianza al tiempo 1 para un proceso de Lévy cualquiera, al dar una expresión para la transformada de Fourier, o función característica asociada a las variables X_t. Sea $f_u : [0, \infty) \to \mathbb{C}$ para $u \in \mathbb{R}$ dada por

$$f_u(t) = \mathbb{E}(e^{iuX_t})\,.$$

Por la independencia y estacionaridad de los incrementos de X, se sigue que si $s, t \geq 0$, entonces

$$\begin{aligned}f_u(s+t) &= \mathbb{E}(e^{iuX_{s+t}}) \\ &= \mathbb{E}(e^{iu(X_{s+t}-X_s)}e^{iuX_s}) \\ &= \mathbb{E}(e^{iu(X_{s+t}-X_s)})\, \mathbb{E}(e^{iuX_s}) \\ &= \mathbb{E}(e^{iuX_t})\, \mathbb{E}(e^{iuX_s}) \\ &= f_u(t)\, f_u(s)\,.\end{aligned}$$

Utilizaremos la anterior propiedad para obtener una relación entre las funciones características de las variables, pero necesitamos el siguiente resultado:

LEMA 3.1. *La función $f : [0, \infty) \times \mathbb{R} \to \mathbb{C}$ dada por $f(t, u) = f_u(t)$ jamás se anula.*

DEMOSTRACIÓN. Basta verificar que para cada $u \in \mathbb{R}$, la función $f_u : [0, \infty) \to \mathbb{C}$ no se anula. Como X_s tiende a 0 conforme $s \to 0^+$, debido a la continuidad por la derecha de X, el teorema de convergencia acotada nos permite concluir que $f_u(s)$ tiende a $1 = f_u(0)$ conforme $s \to 0^+$. Sea entonces $\delta > 0$ tal que si $0 \leq s < \delta$, $f_u(s) \neq 0$. Para cualquier $t \geq 0$, el principio Arquimediano del orden[2] nos permite encontrar $n \in \mathbb{Z}^+$ tal

[2]Ver [1]. Ahí, se refieren al principio Arquimediano del orden como la propiedad Arquimediana del sistema de los números reales.

que $0 \leq t/n < \delta$, por lo que $f_u(t/n) \neq 0$. Como $f_u(t+s) = f_u(t)\, f_u(s)$ para $s,t \geq 0$, entonces $0 \neq f_u(t/n)^n = f_u(t)$. □

Al considerar un proceso de Lévy $X' = (X'_t)_{t\geq 0}$ con la misma distribución que X pero independiente de él, vemos que, para g definida de manera análoga a f pero asociada al proceso $X - X'$,

$$g(t,u) = \mathbb{E}\Big(e^{iu(X_t - X'_t)}\Big) = \mathbb{E}(e^{iu}X_t)\,\mathbb{E}\Big(e^{-iuX'_t}\Big) = f(t,u)\overline{f(t,u)} = |f(u,t)|.$$

Así, la función g jamás se anula y vemos que para $s,t \geq 0$ se tiene que $g(t+s,u) = g(t,u)\,g(s,u)$, ya que $X - X'$ es un proceso de Lévy. Consideremos para cada $u \in \mathbb{R}$ a la función $g_u : [0,\infty) \to (0,1]$ dada por $g_u(t) = g(t,u)$. Esta función es continua, pues satisface las dos pripiedades:

1. $\lim_{s\to o^+} g_u(s) = g_u(0)$,
2. $g_u(t+s) = g_u(t)\,g_u(s)$,

por lo cual

$$\lim_{s\uparrow t} g_u(s) = \lim_{s\uparrow t} \frac{g_u(t)}{g_u(t-s)} = g_u(t) \quad \text{y}$$
$$\lim_{s\downarrow t} g_u(s) = \lim_{s\downarrow t} g_u(t)\,g_u(s-t) = g_u(t)\,.$$

Además, el logaritmo de g_u, satisface

$$\log g_u(s+t) = \log(g_u(t)) + \log(g_u(s))\,,$$

por lo que $\log(g_u(r)) = \log(g_u(1))\,r$ para cualquier $r \in \mathbb{Q}\cap[0,\infty)$, de donde la continuidad nos permite afirmar que al considerar $x_u = \log(g_u(1))$, se tiene $g_u(t) = \exp x_u t$ para $t \geq 0$.

Estamos en posición de considerar a la función $h : [0,\infty) \times \mathbb{R} \to S^1$, con $S^1 = \{z \in \mathbb{C} : |z| = 1\}$, dada por

$$h(t,u) = \frac{f(t,u)}{g(t,u)} = e^{-x_u t} f(t,u)\,.$$

Esta función también satisface la relación $h(t+s,u) = h(t,u)\,h(s,u)$, puesto que tanto f como g la satisfacen. Mediante el siguiente lema, podemos dar una expresión para h y así, finalmente llegar a una para f:

LEMA 3.2. *Sea $\gamma : [0,\infty) \to S^1$ continua tal que $\gamma(0) = 1$. Entonces existe una función continua $\beta : [0,\infty) \to \mathbb{R}$ tal que $\beta(0) = 0$ y $\gamma(t) = \exp i\beta(t)$ para toda $t \geq 0$.*

Al considerar la función $h_u : [0,\infty) \to S^1$ dada por $h_u(t) = h(t,u)$, vemos que satisface las mismas dos condiciones que g_u y por lo tanto podemos concluir que es continua. Además, $h_u(0) = 1$, por lo que existe una función continua $\beta_u : [0,\infty) \to \mathbb{R}$ tal que $\beta_u(0) = 0$ y $h_u(t) = \exp i\beta_u(t)$. Como $h(t+s,u) = h(t,u)\,h(s,u)$, se sigue que para cada $s,t \geq 0$ existe un único $k = k(s,t) \in \mathbb{Z}$ tal que $\beta_u(t+s) = \beta_u(t) + \beta_u(s) + 2k\pi$. Como la asignación $s \mapsto \beta_u(t+s) - \beta_u(s) - \beta_u(t) = 2k\pi$ es continua, se sigue que k no depende de

s (pues toma valores enteros) y por el mismo argumento, intercambiando los papeles de s y t, tampoco depende de t. Así, $k = k(0,0) = 0$ y por lo tanto, $\beta_u(t+s) = \beta_u(t) + \beta_u(s)$. Como β_u es continua, entonces $\beta_u(t) = y_u t$, con $y_u = \beta_u(1)$. Al definir $z_u = x_u + iy_u$, vemos que

$$f(t,u) = g(t,u)\,h(t,u) = e^{x_u t}e^{iy_u t} = e^{z_u t}.$$

Mediante la igualdad $f(t,u) = e^{z_u t}$, si X es un proceso de Lévy con $\mathrm{Var}(X_1) < \infty$, entonces $\mathrm{Var}(X_t) < \infty$ para toda $t \geq 0$, por lo que como función de u, f es derivable 2 veces con continuidad, de donde $u \longmapsto z_u$ es derivable 2 veces con continuidad, además de que

$$\mathbb{E}(X_t) = -i\frac{\partial}{\partial u}f(t,u)\bigg|_{u=0} = -i\,t f(t,u)\,\frac{\partial}{\partial u}z_u\bigg|_{u=0} = t\,(-iz_0') = t\mathbb{E}(X_1)$$

y

$$\mathbb{E}(X_t^2) = -\frac{\partial^2}{\partial u^2}f(t,u)\bigg|_{u=0} = tf(t,u)\left(-z_u'' + t\,(-iz_u')^2\right)\Big|_{u=0} = t\left(-z_0'' + t\,(-iz_0')^2\right),$$

de donde

$$\mathrm{Var}(X_t)^2 = t\,(-z_0'') = t\mathrm{Var}(X_1)\,.$$

3.3. Un teorema límite para ciertos procesos de Lévy

Esta sección se dedica a demostrar un teorema límite válido para ciertos procesos de Levy, entre los cuales se encuentra incluido el proceso Poisson compuesto con saltos acotados, cuando las variables que intervienen tienen varianza finita. Al utilizar este teorema, finalmente podremos entender la aproximación a la solución del problema presentado en la introducción, en términos de convergencia débil. Como los procesos de Levy son càdlàg y hemos visto que las proyecciones unidimensionales generan la σ-álgebra de Borel en D, entonces un proceso de este tipo se puede entender como una variable aleatoria con valores en el espacio de Skorohod. Este es el punto de vista que se adopta. Además, si consideramos un Movimiento Browniano, entonces podemos inducir mediante este, restringido al intervalo $[0, 1]$, una medida, a la cual seguiremos llamando medida de Wiener y denotando por el mismo símbolo, en el espacio D, con el objeto de estudiar la convergencia débil de una sucesión de procesos estocásticos al Movimiento Browniano. Nos ocuparemos de la verificación del siguiente resultado.

TEOREMA 3.6. *Sea $X = (X_t)_{t\in[0,\infty)}$ un proceso de Levy tal que $0 < \sigma^2 = Var(X_1) < \infty$. Entonces la sucesión de procesos definida mediante*

$$X_t^k = \frac{X_{kt} - \mu kt}{\sigma\sqrt{k}}, \quad t \in [0,1]$$

donde μ denota a $\mathbb{E}(X_1)$, converge débilmente a un Movimiento Browniano en $[0, 1]$.

Más especificamente, si denotamos por $B = (B_t)_{t\in[0,1]}$ a la restricción al intervalo $[0,1]$ de un Movimiento Browniano, entonces para cualquier función $f : D \to \mathbb{R}$ continua y acotada, $\mathbb{E}(f \circ X^k) \to \mathbb{E}(f \circ B)$ cuando $k \to \infty$. La demostración sigue la misma idea que la del principio de invariancia de Donsker, se prueba la convergencia de las distribuciones finito-dimensionales inducidas por la sucesión de procesos y luego se verifica la tensión de ésta. De hecho, para verificar la tensión, se utiliza una extensión de la desigualdad de Etmandi utilizada previamente. Así, la demostración es completamente paralela a la del principio de invariancia de Donsker, cosa que no resulta tan extraña si se piensa que los procesos de Levy son la versión a tiempo continuo de lo que representa la caminata aleatoria como proceso de incrementos independientes y estacionarios a tiempo discreto. Para verificar el Teorema 3.6, nos ayudaremos de los resultados sobre convergencia débil en el espacio de Skorohod, para lo cual necesitamos algunos lemas previos.

LEMA 3.3. *Si $f : [0,a] \to \mathbb{R}$ es una función continua por la derecha en $[0,a)$ y con límites por la izquierda en $(0,a]$, entonces*

$$\max_{0\le i\le 2^n} f\left(\frac{i}{2^n}a\right) \uparrow \sup_{t\in[0,a]} f(t) \quad \textit{conforme } n \to \infty.$$

DEMOSTRACIÓN. Sea

$$M = \sup_{t\in[0,a]} f(t)\,,$$

que existe por ser f acotada,

$$M_n = \max_{1\le i\le 2^n} f\left(\frac{i}{2^n}a\right)$$

y para cada $n \in \mathbb{N}$, consideremos $t'_n \in [0,a]$ tal que

$$M - 1/n < f(t'_n) \le M.$$

Entonces $f(t'_n)$ converge a M conforme $n \to \infty$. De la compacidad secuencial de $[0,a]$[3], se sigue que existe una subsucesión $(t_n)_{n\in\mathbb{N}}$ de $(t'_n)_{n\in\mathbb{N}}$ convergente, digamos a un punto $t \in [0,a]$. Pueden suceder dos cosas: que el conjunto $\{n \in \mathbb{N} : t_n \ge t\}$ sea infinito o que su complemento lo sea. En el primer caso, por la continuidad a la derecha de f, $M = f(t)$ y en el otro, por la existencia del límite por la izquierda, $M = f(t-)$. Consideremos ahora una sucesión s_k de puntos de la forma $i/2^n$ que converja a t, por la derecha en el primer caso, y por la izquierda en el segundo. Entonces $\lim_{k\to\infty} f(s_k) = M$, de donde se sigue que $\lim_{n\to\infty} M_n \ge \lim_{k\to\infty} f(s_k) = M$ (el primer límite existe por ser una sucesión creciente y acotada) y la desigualdad contraria es evidente. $\square$

LEMA 3.4. *Sea $X = (X_t)_{t\in[0,\infty)}$ un proceso de Levy. Entonces, para cada $\lambda > 0$,*

$$\mathbb{P}\left(\sup_{t\in[0,a]} |X_t| > \lambda\right) \le 3 \sup_{t\in[0,a]} \mathbb{P}(|X_t| > \lambda/3)\,.$$

[3] Ver [1].

DEMOSTRACIÓN. Como X es càdlàg, se sigue del lema anterior que

$$\sup_{t\in[0,a]}|X_t| = \lim_{n\to\infty}\max_{0\leq i\leq 2^n}|X_{\frac{i}{2^n}a}|,$$

de donde se sigue que $\sup_{t\in[0,a]}|X_t|$ es una variable aleatoria, por lo que la probabilidad del lado izquierdo de la desigualdad anterior tiene sentido. Notemos que

$$\begin{aligned}\mathbb{P}\left(\sup_{t\in[0,a]}|X_t| > \lambda\right) &= \mathbb{P}\left(\lim_{n\to\infty}\max_{0\leq i\leq 2^n}|X_{\frac{i}{2^n}a}| > \lambda\right)\\ &= \lim_{n\to\infty}\mathbb{P}\left(\max_{0\leq i\leq 2^n}|X_{\frac{i}{2^n}a}| > \lambda\right).\end{aligned}$$

Donda la última igualdad se sigue pues la sucesión

$$\left(\max_{0\leq i\leq 2^n}|X_{\frac{i}{2^n}a}|\right)_{n\geq 1}$$

es creciente. Al utilizar la desigualdad de Etmandi, probada como parte del principio de invariancia de Donsker, se tiene que

$$\mathbb{P}\left(\max_{0\leq i\leq 2^n}|X_{\frac{i}{2^n}a}| > \lambda\right) \leq 3\max_{0\leq i\leq 2^n}\mathbb{P}\left(|X_{\frac{i}{2^n}a}| > \lambda/3\right),$$

por lo que

$$\begin{aligned}\lim_{n\to\infty}\mathbb{P}\left(\max_{0\leq i\leq 2^n}|X_{\frac{i}{2^n}a}| > \lambda\right) &\leq 3\lim_{n\to\infty}\max_{0\leq i\leq 2^n}\mathbb{P}\left(|X_{\frac{i}{2^n}a}| > \lambda/3\right)\\ &\leq 3\sup_{t\in[0,a]}\mathbb{P}(|X_t| > \lambda/3).\end{aligned}$$

□

LEMA 3.5. *Si $X = (X_t)_{t\geq 0}$ es un proceso de Levy con $\mathbf{E}(X_1) = 0$ y $0 < Var(X_1) = \sigma^2 < \infty$, entonces*

$$\lim_{t\to\infty}\mathbb{P}\left(|X_t| > \lambda\sigma\sqrt{t}\right) = \mathbb{P}(|B_1| > \lambda).$$

DEMOSTRACIÓN. Basta probar que si $(t_n)_{n\in\mathbb{N}}$ es una sucesión de reales no negativos tal que $t_n \to \infty$ cuando $n \to \infty$, entonces $X_{t_n}/\sigma\sqrt{t_n}$ converge en distribución a B_1, puesto que la función $|\cdot|$ y la distribución normal son continuas. Esto es cierto cuando $t_n = n$, por el teorema límite central y la independencia y estacionariedad de los incrementos de X (así como el que tenga varianza finita y positiva). Si $(t_n)_{n\in\mathbb{N}}$ es una sucesión para la cual $t_n \to \infty$ cuando $n \to \infty$, entonces $s_n = [t_n] \to \infty$ conforme $n \to \infty$, por lo que $X_{s_n}/\sigma\sqrt{s_n}$ converge en distribución a B_1 cuando $n \to \infty$. Pero como $\sqrt{s_n}/\sqrt{t_n} \to 1$, se sigue que

$X_{s_n}/\sqrt{t_n}$ converge en distribución a B_1. Finalmente, notemos que $(X_{t_n} - X_{s_n})/\sigma\sqrt{t_n}$ converge en probabilidad (o en distribución) a 0, puesto que si $\varepsilon > 0$, entonces

$$\mathbb{P}(|X_{t_n} - X_{s_n}| > \varepsilon\sigma\sqrt{t_n}) \leq \frac{t_n - s_n}{\varepsilon^2 t_n} \to 0.$$

□

Finalmente, daremos la prueba del Teorema 3.6 a continuación. Durante toda la demostración consideraremos a los procesos dados en el Teorema 3.6 y al proceso $Y = (Y_t)_{t\geq 0}$ definido por $Y_t = X_t - \mu t$, que es un proceso de Levy con la misma varianza que X pero con media 0. Así, la sucesión de procesos a los que se refiere el Teorema 3.6, $(X^k)_{k\in\mathbb{N}}$, se obtienen a partir de Y de la siguiente manera: $X_t^k = \frac{Y_{kt}}{\sigma\sqrt{k}}$ con $t \geq 0$. Para cada $t > 0$, notemos que

$$X_t^k = \frac{1}{\sigma\sqrt{k}} \sum_{i=1}^{k} Y_{it} - Y_{(i-1)t},$$

que por la independencia y estacionariedad de los incrementos, es una suma de k variables aleatorias independientes e idénticamente distribuidas de media cero y varianza t, de donde convergen en distribución a una variable cuya distribución límite es $N(0,t)$. El argumento no es válido para $t = 0$, sin embargo la conclusión es cierta. Si

$$0 = t_0 \leq t_1 \leq \cdots \leq t_n,$$

entonces las variables

$$X_{t_1}^k - X_{t_0}, X_{t_2}^k - X_{t_1}^k, \ldots, X_{t_n}^k - X_{t_{n-1}}^k$$

son independientes y tienen la misma distribución que las variables

$$X_{t_1}^k, X_{t_2-t_1}^k, \ldots, X_{t_n-t_{n-1}}^k$$

respectivamente, por lo que la distribución límite del vector aleatorio

$$(X_{t_i} - X_{t_{i-1}})_{i=1}^n$$

es normal con media cero y vector de varianza-covarianza $((t_i - t_{i-1})\,\delta_{ij})_{i,j=1}^n$, de donde,la distribución límite del vector aleatorio

$$(X_{t_i})_{i=1}^n$$

es normal con media cero y matriz de varianza-covarianza $(t_i \wedge t_j)_{i,j=1}^n$, puesto que la asignación

$$(x_1, \ldots, x_n) \mapsto (x_1, x_1 + x_2, \ldots, x_1 + \cdots + x_n)$$

es continua. Esto significa que las distribuciones finito-dimensionales de X^k convergen a las de B. Faltan ahora las dos primeras condiciones del Teorema 3.5 para poder concluir que

X^k converge débilmente a B, siendo la primera evidente pues $X_0^k = 0$ y nos ocuparemos de la segunda, que es verificar igualdad

$$\lim_{\delta \downarrow 0} \limsup_{k \to \infty} \mathbb{P}\left(\left\{\omega \in \Omega : \omega^1(X^k, \delta) > \varepsilon\right\}\right) = 0$$

para cualquier $\varepsilon > 0$, para lo cual es suficiente considerar el primer límite para una sucesión que converja a cero, por ejemplo $\delta_n = 1/n$, puesto que como función de delta, el límite superior sobre k de las probabilidades es una función decreciente Sea $\varepsilon > 0$ y notemos que

$$\mathbb{P}(\omega^1(X^k, 1/n) > \varepsilon) = \mathbb{P}\left(\sup_{\substack{s,t \in [0,k] \\ |s-t| \leq k/n}} |Y_t - Y_s| > \varepsilon\sigma\sqrt{k}\right).$$

Sea $\omega \in \Omega$ tal que $\omega^1(X^k(\omega), 1/n) > \varepsilon$. Entonces existen $s, t \in [0, n]$ tal que $|s - t| \leq k/n$ y $|Y_s - Y_t| > \varepsilon\sigma\sqrt{k}$, donde podemos suponer sin pérdida de generalidad que $s < t$. Pero entonces, consideremos a $j \in \{1, \ldots, n\}$ tal que $s \in [\frac{k(j-1)}{n}, \frac{kj}{n})$, por lo que tenemos dos opciones para t, o pertenece a $(\frac{k(j-1)}{n}, \frac{kj}{n})$ o a $[\frac{kj}{n}, k(j+1)n)$. En el primer caso, notemos que alguna de las 2 cantidades siguientes debe ser mayor a $\varepsilon\sigma\sqrt{k}/3$,

$$\begin{aligned} &|Y_s - Y_{\frac{k(j-1)}{n}}| \quad \text{o} \\ &|Y_t - Y_{\frac{k(j-1)}{n}}|, \end{aligned}$$

pues de suponer lo contrario, la desigualdad del tríangulo nos regala una contradicción. En el segundo caso, alguna de las siguientes 3 cantidades debe ser mayor a $\varepsilon\sigma\sqrt{k}/3$,

$$\begin{aligned} &|Y_s - Y_{\frac{k(j-1)}{n}}| \quad \text{o} \\ &|Y_{\frac{kj}{n}} - Y_{\frac{k(j-1)}{n}}| \quad \text{o} \\ &|Y_t - Y_{\frac{kj}{n}}|, \end{aligned}$$

pues de nueva cuenta, el suponer lo contrario nos lleva a una contradicción al utilizar la desigualdad del triángulo. En cualquier caso, nos damos cuenta de que, si $\omega \in \Omega$ es tal que $\omega^1(X^k(\omega), 1/n) > \varepsilon$ entonces existe un natural j entre 1 y n tal que

$$\sup_{t \in [\frac{k(j-1)}{n}, \frac{kj}{n}]} |Y_t - Y_{\frac{k(j-1)}{n}}| > \varepsilon\sigma\sqrt{k}/3,$$

de donde obtenemos la contención

$$\{\omega^1(X^k, 1/n) > \varepsilon\} \subset \bigcup_{j=1}^{n} \left\{\sup_{s \in [0, k/n]} |Y_{\frac{k(j-1)}{n}+s} - Y_{\frac{k(j-1)}{n}}| > \varepsilon\sigma\sqrt{k}/3\right\}$$

y por la estacionariedad de los incrementos, la probabilidad de los eventos del lado derecho de la anterior contención no depende de j, por lo que la subaditividad de $\mathbb{P}$ nos permite afirmar que

$$\mathbb{P}\big(\omega^1(X^k, 1/n) > \varepsilon\big) \leq n\mathbb{P}\left(\sup_{t\in[0,k/n]} |Y_t| > \varepsilon\sigma\sqrt{k}/3\right),$$

donde al utilizar el Lema 3.4, que es la generalización de la desigualdad de Etmandi a procesos de Levy, acotamos el lado derecho de la desigualdad anterior por

$$3n \sup_{t\in[0,k/n]} \mathbb{P}\Big(|Y_t| > \varepsilon\sigma\sqrt{k}/9\Big)$$

y para acotar última expresión, utilizaremos el Lema 3.5, que nos permite afirmar la existencia de $t^* = t^*(\lambda)$ y $a > 0$ tal que para toda $t \geq t^*$,

$$\mathbb{P}\Big(|Y_t| > \lambda\sigma\sqrt{t}\Big) < \frac{a}{\lambda^4},$$

al considerar $a > \mathbb{E}(B_1^4)$. Consideremos a $t^* = t^*(\varepsilon/9)$ y si k es tal que $k/n \geq t^*$ y $t \in [t^*, k/n]$, entonces

$$\mathbb{P}\Big(|Y_t| > \varepsilon\sigma\sqrt{k}/9\Big) \leq \mathbb{P}\Big(|Y_t| > \varepsilon\sigma\sqrt{nt}/9\Big) \leq \frac{9^4 a}{\varepsilon^4 n^2}.$$

Por otro lado si $t \leq t^* \leq k/n$, entonces la desigualdad de Tchebyshev nos permite afirmar que

$$\mathbb{P}\Big(|Y_t| > \varepsilon\sigma\sqrt{k}/9\Big) \leq \frac{9^2\sqrt{t}}{\varepsilon^2 k} \leq \frac{9^2\sqrt{t^*}}{\varepsilon^2 k}.$$

Así, vemos que si $k/n \geq t^*$, se tiene que

$$3n \sup_{t\in[0,k/n]} \mathbb{P}\Big(|Y_t| > \varepsilon\sigma\sqrt{k}/9\Big) \leq \frac{3\cdot 9^4 a}{\varepsilon^4 n} \vee \frac{3\cdot 9^2 n\sqrt{t^*}}{\varepsilon^2 k},$$

de donde

$$\lim_{n\to\infty} \limsup_{k\to\infty} \mathbb{P}\big(\omega^1(X^k, 1/n) > \varepsilon\big) \leq 3 \lim_{n\to\infty} \limsup_{k\to\infty} \frac{9^4 a}{\varepsilon^4 n} \vee \frac{9^2 n\sqrt{t^*}}{\varepsilon^2 k} \leq 3 \lim_{n\to\infty} \frac{9^4 a}{\varepsilon^4 n} = 0,$$

lo cual establece la tensión de la sucesión de medidas de probabilidad inducidas por la sucesión de procesos $\left(X^k\right)_{k\geq 1}$ y termina la demostración del Teorema 3.6.

3.4. La ley arcoseno

Como una aplicación del teorema límite presentado en la sección anterior (el Teorema 3.6), calcularemos la distribución límite conforme $t \to \infty$ de

$$Y_t = \frac{1}{t}\lambda(\{s \in [0,t] : X_s > \mu s\}),$$

donde λ es la medida de Lebesgue en $\mathbb{R}$ y $X = (X_t)_{t\geq 0}$ es un proceso de Lévy tal que $\mathrm{Var}(X_1) < \infty$ y $\mathbb{E}(X_1) = \mu$. Todavía no se verifica que Y_t está bien definida (no se ha probado que el conjunto $\{s \in [0,t] : X_s > \mu s\}$ sea Boreliano) y que es variable aleatoria, además de verificar que la distribución de Y_t converge puntualmente a la distribución arcoseno conforme $t \to \infty$. Este teorema límite es de carácter distinto a los anteriores pues en ellos utilizábamos sucesiones de variables aleatorias y no colecciones indicadas por un subconjunto de los reales, pero las técnicas que utilizaremos serán muy parecidas.

Antes que nada, notemos que las asignaciónes

$$(t,\omega) \mapsto X_s(\omega)$$
$$(t,\omega) \mapsto \mathbf{1}_{s\in[a,b]}$$

de $[0,\infty) \times \Omega$ en $\mathbb{R}$ son $(\mathscr{B}_{[0,\infty)} \otimes \mathscr{F}, \mathscr{B}_{\mathbf{R}})$-medibles, por lo que la asignación

$$(t,\omega) \mapsto X_{\frac{[kt]+1}{k}}(\omega) = \sum_{i=1}^{\infty} X_{\frac{i+1}{k}}(\omega)\, \mathbf{1}_{t\in[\frac{i}{k},\frac{i+1}{k})},$$

que denotaremos por X_t^k, también lo es. Como X continuo por la derecha, se sigue que X_t^k converge a X_t para todo (t,ω) conforme $k \to \infty$, por lo que la asignación $(t,\omega) \mapsto X_t(\omega)$ es medible respecto a $\mathscr{B}_{[0,\infty)} \otimes \mathscr{F}$ y $\mathscr{B}_{\mathbf{R}}$. Esto nos dice dos cosas, que el conjunto

$$\{s \in [0,t] : X_s > \mu s\}$$

pertenece a $\mathscr{F}$ para cada $\omega \in \Omega$ y que la asignación

$$\omega \mapsto \lambda(\{s \in [0,t] : X_s > \mu s\})$$

es $(\mathscr{F}, \mathscr{B}_{\mathbf{R}})$-medible, por lo que Y_t está bien definida y es medible respecto a $\mathscr{F}$ y $\mathscr{B}_{\mathbf{R}}$. En cuanto a la convergencia de las distribuciones de Y_t a la distribución arcoseno conforme $t \to \infty$, basta probar que la convergencia en distribución de $(Y_k)_{k\in\mathbf{N}}$ a una variable con distribución arcoseno, puesto que en ese caso si $c \in \mathbb{R}$ y $\alpha > 0$, entonces

$$F_{Y_{[t]}}(c-\alpha) - \mathbb{P}\big(|Y_{[t]} - Y_t| \geq \alpha\big) \leq F_{Y_t}(c) \leq F_{Y_{[t]}}(c+\alpha) + \mathbb{P}\big(|Y_{[t]} - Y_t| \geq \alpha\big),$$

de donde al considerar límites inferiores y superiores y utilizar la continuidad de la distribución arcoseno, se sigue que Y_t tiene distribución límite y que coincide con la distribución arcoseno siempre y cuando la sucesión $\{Y_k\}_{k\in\mathbf{N}}$ converga en distribución a una variable con distribución arcoseno, todo esto gracias a la desigualdad

$$|Y_t - Y_{[t]}| \leq 2[t]\frac{t-[t]}{[t]t} \leq \frac{2}{t},$$

de la cual se desprende que

$$\mathbb{P}\big(|Y_{[t]} - Y_t| \geq \alpha\big) \leq \mathbf{1}_{\alpha\leq\frac{2}{t}},$$

que converge a 0 conforme $t \to \infty$ para cualquier α positiva.

Ahora nos encargaremos de verificar que $(Y_k)_{k\in\mathbb{N}}$ converge en distribución a una variable con distribución arcoseno. Para esto, consideremos la aplicación $h : D \longrightarrow \mathbb{R}$ definida mediante la igualdad

$$h(x) = \int_{[0,1]} x \, d\lambda,$$

que tiene sentido pues los elementos de D son $(\mathscr{B}_D, \mathscr{B}_{\mathbb{R}})$-medibles. Primero, notemos que la asignación

$$(t, x) \longmapsto x(t) = x_t$$

es $(\mathscr{B}_{[0,1]} \otimes \mathscr{B}_D, \mathscr{B}_{\mathbb{R}})$-medible, puesto que las asignaciones

$$(t, x) \longmapsto x_s$$
$$(t, x) \longmapsto \mathbf{1}_{t\in[a,b]},$$

donde $s \in [0, 1]$, lo son, de donde la aplicación

$$(t, x) \longmapsto x\left(\frac{[kt]+1}{k}\right) = x_1 \mathbf{1}_{t=1} + \sum_{i=1}^{k} x_{\frac{i}{k}} \mathbf{1}_{t\in[\frac{i-1}{k}, \frac{i}{k})}$$

también lo es, de donde el límite cuando $k \longrightarrow \infty$ es $(\mathscr{B}_{[0,1]} \otimes \mathscr{B}_D, \mathscr{B}_{\mathbb{R}})$-medible y éste es precisamente la aplicación $(t, x) \longmapsto x_t$. Esto nos dice que h es $(\mathscr{B}_D, \mathscr{B}_{\mathbb{R}})$-medible. Por otro lado, en el Capítulo 2 hemos visto que

$$\mathbb{W}(C \cap \partial(\{x \in D : h(x) \leq c\})) = \mathbb{W}(\partial(\{x \in C : h(x) \leq c\})) = 0$$

para cualquier $c \in \mathbb{R}$ que pertenezca al complemento de un conjunto a lo más numerable, por lo cual, en el complemento de dicho conjunto se cumple que

$$\mathbb{P}(h \circ X^k \leq c) \longrightarrow A(c)\,,$$

donde

$$X_t^k = \frac{X_{kt} - \mu kt}{k\sigma^2}$$

y A denota a la distribución arcoseno. Pero como A es continua, el límite también es $A(c)$ para cualquier $c \in \mathbb{R}$. Finalmente, basta notar que

$$h \circ X^k = \frac{1}{k} \int_{[0,k]} \mathbf{1}_{X_t > \mu t} \, d\lambda = Y_k,$$

de donde hemos verificado que $\{Y_k\}_{k\in\mathbb{N}}$ converge en distribución a una variable con distribución arcoseno conforme $k \longrightarrow \infty$.

Notas

Con el objeto de resolver el problema planteado en la introducción de la tesis es que se presenta un esbozo de la teoría de la convergencia en distribución en el espacio de Skorohod. Las demostraciones de estos resultados se pueden consultar en [3]. En la Sección 3.3, se da una demostración de un teorema límite para ciertos procesos de Lévy mediante una adaptación de la prueba del principio de invariancia de Donsker presentada en el Capítulo 2. Vale la pena mencionar que ésta no corresponde a un primer intento de solución. De hecho, la primera prueba para este teorema límite utilizaba resultados un poco más finos sobre convergencia en distribución en el espacio de Skorohod, por ejemplo, resultados concercientes a cambios de tiempo aleatorios.

Conclusión

El problema que se planteó en la introducción fué el de aproximar la distribución de

$$h_T = \frac{1}{T}\int_{[0,T]} \mathbf{1}_{(Y_t > \lambda\mu t)}\, dt$$

donde $(Y_t)_{t\geq 0} = (S_{N_t})_{t\geq 0}$ es un proceso Poisson compuesto tal que $\mathbb{E}(S_1) = \mu$ y $\mathbb{E}(N_1) = \lambda$, bajo la hipótesis adicional de que la variable S_n fuese acotada. Esto implica que el proceso $(Y_t)_{t\geq 0}$ es un proceso de Lévy con varianza finita, y como $\mathbb{E}(Y_t) = \lambda\mu t$, se sigue que el Teorema 3.6 aplica y así como la ley límite para h_T, esto es, para cada $x \in \mathbb{R}$,

$$\lim_{T\to\infty} \mathbb{P}(h_T \leq x) = A(x)\,,$$

donde A es la distribución arcoseno introducida anteriormente, es decir, para T grande, la distribución de h_T se puede aproximar mediante la distribución arcoseno. Como los cuantiles 1/4 y 3/4 para la distribución A son 0,15 y 0,85 aproximadamente, se concluye que, al menos para T grande, con probabilidad 1/2, menos del 15 % del tiempo o más del 85 % del tiempo la aseguradora tiene que aportar capital adicional para la constitución de la reserva de riesgos en curso. Esto es, con probabilidad 1/2, la compañía de seguros tiene muy buena suerte o muy mala suerte, siendo esto consecuencia de que la distribución arcoseno se encuentre concentrada alrededor de sus valores extremos, 0 y 1, en vez de alrededor de la media, que es 1/2.

Las compañías de seguro deben considerar el anterior resultado al garantizar (en la medida de lo posible) que tienen el capital suficiente para poder constituir la reserva de riesgos en curso en caso de que la prima cobrada no sea suficiente para tal efecto durante la vigencia de la póliza, que como ya vimos, no es una situación anómala, sino mas bien, un tanto común.

APÉNDICE A

Algunas propiedades de los espacios métricos

El motivo de este apéndice es verificar tres resultados concernientes a propiedades de los espacios métricos que fueron utilizados en el Capítulo 2. Empecemos con el más sencillo, que nos habla acerca de convergencia de sucesiones:

TEOREMA A.1. *Sea (M,ρ) un espacio métrico y $(x_n)_{n\in\mathbb{N}}$ una sucesión de elementos de M. Entonces dicha sucesión converge a x si y sólo si cualquier subsucesión $(x'_n)_{n\in\mathbb{N}}$ de $(x_n)_{n\in\mathbb{N}}$ admite a su vez una subsucesión $(x''_n)_{n\in\mathbb{N}}$ que converge a x.*

DEMOSTRACIÓN. Si la sucesión $(x_n)_{n\in\mathbb{N}}$ converge a x, entonces toda subsucesión converge a x, por lo que toda subsucesión de cualquier subsucesión converge a x.

Por otro lado, si la sucesión $(x_n)_{n\in\mathbb{N}}$ no converge a x, entonces existe $\varepsilon > 0$ tal que para cualquier $n \in \mathbb{N}$ existe $m \geq n$ tal que $\rho(x_m, x) \geq \varepsilon$. Así, podemos considerar $m_1 \geq 1$ tal que $\rho(x_{m_1}, x) \geq \varepsilon$ y a partir de m_1, asegurar la existencia de $m_2 \geq m_1 + 1$ tal que $\rho(x_{m_2}, x) \geq \varepsilon$. De manera recursiva, consideremos una sucesión estrictamente creciente de naturales $(m_n)_{n\in\mathbb{N}}$ tal que $\rho(x_{m_n}, x) \geq \varepsilon$. Claramente, la subsucesión $(x_{m_n})_{n\in\mathbb{N}}$ de $(x_n)_{n\in\mathbb{N}}$ no admite subsucesiones convergentes a x. $\square$

El siguiente resultado nos habla acerca de cierta propiedad de los subespacios separables de los espacios métricos y fué utilizado en la demostración del Teorema de Prohorov:

TEOREMA A.2. *Sea (M,ρ) un espacio métrico y $N \subset M$ separable. Entonces existe una familia a lo más numerable $\mathscr{U}$ de conjuntos abiertos con la siguiente propiedad: si*

$$x \in G \cap N,$$

donde $G \subset M$ es abierto, entonces existe $U \in \mathscr{U}$ tal que

$$x \in U \subset \overline{U} \subset G.$$

DEMOSTRACIÓN. Sea $D \subset N$ un conjunto numerable tal que $N \subset \overline{D}$. Entonces la familia

$$\mathscr{U} = \{B_q(d) : q \in \mathbb{Q} \cap (0,\infty), d \in D\}$$

es numerable. Si $G \subset M$ es abierto y $x \in N \cap G$, entonces existe $r > 0$ tal que $B_r(x) \subset G$. Por otro lado, sabemos que existe $q \in \mathbb{Q} \cap (0, r/2)$, y como $x \in N \subset \overline{D}$, entonces existe $d \in B_q(x)$. Sea $U = B_q(d)$, por lo que $x \in U$ y además, si $z \in U$, entonces

$$\rho(z,x) \leq \rho(z,d) + \rho(d,x) < 2q < r,$$

de donde $\overline{U} \subset B_r(x) \subset G$. Así, la familia $\mathscr{U}$ satisface las condiciones deseadas. □

El último resultado tiene que ver con una caracterización de la compacidad y también se utilizó en la demostración del Teorema de Prohorov. Recordemos que en un espacio métrico (M, ρ), un subconjunto C de M es compacto si y sólo si es compacto por sucesiones, es decir, si y sólo si cualquier sucesión con valores en C admite a una subsucesión convergente a un punto de C. Decimos que $C \subset M$ es totalmente acotado si para cada $\varepsilon > 0$ existen $n \geq 1$ y $x_1, \ldots, x_n \in M$ tal que

$$C \subset \bigcup_{i=1}^{n} B_\varepsilon(x_i)\,.$$

El siguiente teorema nos proporciona otra caracterización de la compacidad en espacios métricos:

TEOREMA A.3. *Si (M, ρ) es un espacio métrico y $C \subset M$, las siguientes condiciones son equivalentes*

a) C es compacto.
b) C es totalmente acotado y completo.

DEMOSTRACIÓN.

a)⇒ b): Si C no es totalmente acotado, entonces existe $\varepsilon > 0$ tal que si $x_1, \ldots, x_n \in M$, entonces existe

$$x \in C \setminus \bigcup_{i=1}^{n} B_\varepsilon(x_i)\,.$$

Como C es no vacío, sea $x_1 \in C$, por lo que existe $x_2 \in C \setminus B_\varepsilon(x_1)$, esto es, $\rho(x_1, x_2) \geq \varepsilon$. Además, existe

$$x_3 \in C \setminus \bigcup_{i=1}^{2} B_\varepsilon(x_i)\,,$$

por lo que $\rho(x_3, x_i) \geq \varepsilon$ para $i = 1, 2$. Continuando recursivamente, podemos construir una sucesión $(x_n)_{n=1}^{\infty}$ tal que si $n \neq m$ entonces $\rho(x_n, x_m) \geq \varepsilon$. Claramente esta sucesión no admite ninguna subsucesión convergente a un punto de C, puesto que no es de Cauchy, por lo que C no es compacto. Esto nos dice que si C es compacto, entonces C es totalmente acotado. Además, cualquier conjunto compacto es completo, pues si $(x_n)_{n\in\mathbb{N}}$ es una sucesión de Cauchy con valores en C, entonces admite una subsucesión convergente a un punto de C, por lo que la sucesión original es convergente a un punto de C.

b)⇒ a): Para cada $n \geq 1$, existe $k_n \geq 1$ tal que C está contenido en

$$\bigcup_{i=1}^{k_n} B_{n,i},$$

donde $B_{n,i}$ es una bola de radio $1/n$ con centro en algún punto de M. Si $(x_n)_{n\in\mathbb{N}}$ es una sucesión de valores en C, entonces para $n = 1$, existe un conjunto $B_{1,i}$ que contiene a una infinidad de elementos de la sucesión, digamos que $x_{m_{1,j}} \in B_{1,i}$ para $j = 1, 2, \cdots$, donde la sucesión $(m_{1,j})_{j=1}^{\infty}$ es estrictamente creciente. De la misma manera, para $n = 2$, podemos extraer de la subsucesión $\left(x_{m_{1,j}}\right)_{j=1}^{\infty}$ una subsucesión $\left(x_{m_{2,j}}\right)_{j=1}^{\infty}$ tal que todos los elementos de dicha subsucesión pertenecen a la misma bola de radio $1/2$, digamos $B_{2,i}$. Procediendo de manera recursiva, construimos para cada $k \geq 1$ una subsucesión $\left(x_{m_{k,j}}\right)_{j\geq 1}$ de tal manera que $\left(x_{m_{k+1,j}}\right)_{j\geq 1}$ sea subsucesión de $\left(x_{m_{k,j}}\right)_{j\geq 1}$ y que $x_{m_{k,j}}$ pertenezca a la misma bola de radio $1/k$. Esto nos dice que la subsucesión $\left(x_{m_{j,j}}\right)_{j=1}^{\infty}$ de $(x_n)_{n\in\mathbb{N}}$ es de Cauchy, puesto que si $i < j$, $\rho\left(x_{m_{i,i}}, x_{m_{j,j}}\right) < 2/i$, y por la completez de C, converge a un punto de C, de donde C es secuencialmente compacto y por lo tanto compacto.

$\square$

Bibliografía

[1] T.M. Apostol, *Análisis Matemático*, segunda edición, Editorial Reverté S.A., 1988
[2] R. G. Bartle, *The elements of integration and Lebesgue Measure*, Wiley, 1995.
[3] P. Billingsley, *Convergence of Probability Measures*, segunda edición, John Wiley & Sons Inc., 1999
[4] P. Erdös & M. Kac, *On the number of positive sums of independent random variables*, Bulletin of the American Mathematical Society, 1947, Vol. 53, p.1011-1020
[5] W. Feller, *An introduction to Probability Theory and its Applications*, Vol. 1 John Wiley & Sons Inc., 1950
[6] W. Feller, *An introduction to Probability Theory and its Applications*, Vol. 1, tercera edición revisada, John Wiley & Sons Inc., 1970
[7] W. Feller, *An introduction to Probability Theory and its Applications*, Vol. 2, segunda edición, John Wiley & Sons Inc., 1971
[8] B. Fristedt & L. Gray, *A Modern Aproach to Probability Theory*, Birkhäuser, 1997
[9] T. Garza, *Técnicas Modernas para el Actuario*, Fondo de Cultura Económica, 1998
[10] I. Karatzas & S.E. Shreve, *Brownian Motion and Stochastic Calculus*, segunda edición, Springer, 1988
[11] G. Salicrup, *Introducción a la topología*, Sociedad Matemática Mexicana, 1993
[12] C. Tudor, *Procesos estocásticos*, segunda edición, Sociedad Matemática Mexicana, 1997