



UNIVERSITAT DE
BARCELONA

Trabajo final de grado

GRADO DE MATEMÁTICAS

Facultad de Matemáticas e Informática

Universidad de Barcelona

**EL MOVIMIENTO
BROWNIANO COMO LÍMITE
DEL PASEO ALEATORIO:
EL TEOREMA DE DONSKER**

Autora: Marta Pitarch Ferreiro

Directora: Dra. Marta Sanz i Solé
Realizado en: Departamento de
Matemáticas e Informática

Barcelona, 17 de enero de 2019

Abstract

Brownian motion is a continuous time stochastic process with no memory, that is, the current state of the process is not influenced by its past. This property is named "Markov property". The main purpose of this paper is to obtain a Brownian motion from a discrete time stochastic process named Random Walk. The Random Walk is also a Markov process. To achieve this goal, we are going to study weak convergence on metric spaces and, in particular, on $C([0, 1])$. Brownian motion is obtained as a weak limit of a sequence of linear interpolations of Random Walk normalized in a suitable way. This is Donsker's theorem (1951).

Introducción

El objetivo principal de este trabajo es dar una construcción del movimiento browniano.

El movimiento browniano es un proceso estocástico que surgió a principios del s. XX como un intento de explicar los movimientos irregulares del polen suspendido en agua, observados por el botánico Robert Brown en 1827.

Fue Einstein en 1905 el primero en dar una explicación satisfactoria a este fenómeno, el cual afirmó que este movimiento era debido al continuo bombardeo de granos de polen por las moléculas del agua que rodeaban a dicha molécula, con sucesivos impactos moleculares en diferentes direcciones, los cuales proporcionaban diferentes impulsos a las partículas de polen. Los átomos y las moléculas habían sido teorizados por aquel entonces, pero estaban lejos de estar aceptadas universalmente, así que la explicación documentada de Einstein sobre el movimiento browniano ayudó a convencer a parte de la comunidad científica acerca de la teoría atómica.

El movimiento browniano es relativamente complicado así que se tardó más de una década en tener una imagen clara del mismo. No fue hasta 1923 cuando Norbert Wiener le dio un sólido fundamento matemático. Así pues, el movimiento browniano también es llamado *proceso de Wiener* o *proceso de Wiener-Einstein*.

Hoy en día el movimiento browniano existe en diversas áreas de ciencias puras y aplicadas. Tiene una aplicación especialmente fuerte en finanzas y también en física, donde los procesos de difusión y ósmosis están basados en este proceso estocástico.

A nivel de fundamento matemático, existen tres maneras principales de construir el movimiento browniano. Una de ellas es la llamada *construcción de Kolmogorov*, que nombraremos sin hacer demasiado hincapié en ella. Otra es la *construcción de Lévy's*, la cual sí que estudiaremos pero no en profundidad. Esta construcción nos da la continuidad de las trayectorias del movimiento browniano. Finalmente, tenemos la construcción del movimiento browniano como límite del paseo aleatorio. Esta última construcción será en la que nos centraremos, y la que desarrollaremos en este trabajo.

El trabajo está estructurado en tres capítulos. El primero trata de dar una idea general de los conceptos básicos de la probabilidad que usaremos, y luego tenemos

una pequeña introducción a los procesos estocásticos. En las dos siguientes secciones del capítulo nos centramos por separado en los procesos estocásticos a tiempo discreto y a tiempo continuo. En los procesos a tiempo discreto, estudiaremos a fondo las cadenas de Markov, y en particular un ejemplo muy relevante: el paseo aleatorio. En el grado de matemáticas no he podido cursar la asignatura de procesos estocásticos, y es por esto en parte que he aprovechado para profundizar en algunas propiedades y características que iban quizás más orientados al aprendizaje en sí, que no al objetivo concreto al que va dirigido este trabajo. La sección de procesos estocásticos a tiempo continuo empieza tratando los procesos gaussianos y su existencia, y sigue con el ejemplo más importante que es el movimiento browniano. Estudiamos como de dónde surge, y sus características y propiedades principales. Es en esta sección donde trataremos las construcciones de Lévy's y Kolmogorov

El segundo capítulo se centra en explicar la convergencia débil y algunos de sus resultados más interesantes para el enfoque de este trabajo. Trataremos la convergencia en espacios métricos, y daremos un resultado muy importante en la teoría de la medida: el teorema de Prohorov. La convergencia débil será lo que nos unirá el paseo aleatorio al movimiento browniano, mediante el teorema de Donsker (1951), que será lo que trataremos en el tercer y último capítulo. Así pues el tercer capítulo nos introduce en la parte de los fundamentos matemáticos de los procesos de Wiener. Estudiaremos algunos resultados de la convergencia en el espacio de las funciones continuas, que es el espacio donde haremos todo nuestro fundamento matemático, y procederemos a construir la medida de Wiener, que nos dará la función de distribución del movimiento browniano. Entonces podremos ver, mediante el teorema de Donsker, que un paseo aleatorio converge débilmente a un movimiento browniano.

Índice general

1. Procesos estocásticos	1
1.1. Preliminares	1
1.2. Procesos estocásticos a tiempo discreto: El paseo aleatorio	7
1.3. Procesos estocásticos a tiempo continuo. El movimiento browniano	18
2. Convergencia de medidas de probabilidad	28
2.1. Introducción	28
2.2. Convergencia débil en espacios métricos	29
2.3. Convergencia en distribución	32
2.4. Teorema de Prohorov	33
3. Teorema de Donsker	38
3.1. Introducción	38
3.2. Medidas de Wiener	40
3.3. Teorema de Donsker	45

Capítulo 1

Procesos estocásticos

En este capítulo nos centramos en dar las definiciones básicas necesarias de probabilidad, y una pequeña introducción de los procesos estocásticos. Luego trataremos en profundidad y de manera separada los procesos estocásticos a tiempo discreto y a tiempo continuo, con sus ejemplos más importantes correspondientes: el paseo aleatorio y el movimiento browniano.

1.1. Preliminares

Espacios de probabilidad

Hablamos de *experiencia aleatoria* cuando a pesar de conocer las condiciones en las que se realiza un suceso, no podemos predecir el resultado con exactitud. Recogemos y estudiamos estas experiencias en un objeto matemático llamado *espacio de probabilidad*.

Definición 1.1.1. *Un espacio de probabilidad es una terna (Ω, \mathcal{F}, P) donde*

- i) Ω es el **espacio muestral**, es decir, el conjunto de todos los resultados posibles de la experiencia aleatoria.*
- ii) $\mathcal{F} \in \mathcal{P}(\Omega)$ (conjunto de las partes de Ω) y tiene estructura de σ -álgebra, i.e.*
 - a) $\Omega \in \mathcal{F}$*
 - b) Si $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$*
 - c) \mathcal{F} es cerrado por uniones numerables. Esto significa que si $(A_n, n \geq 1)$ es una sucesión de subconjuntos de \mathcal{F} , entonces $\bigcup_{n \geq 1} A_n \in \mathcal{F}$.*
- iii) $P : \Omega \rightarrow [0, 1]$ es una **probabilidad**, es decir, cumple las dos siguientes condiciones:*
 - a) $P(\Omega) = 1$*

b) Para cualquier sucesión $(A_n, n \geq 1) \subset \mathcal{F}$ de conjuntos de \mathcal{F} disjuntos dos a dos,

$$P\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n \geq 1} P(A_n)$$

Esta propiedad se llama σ -aditividad. La probabilidad es una medida de la certidumbre de que un suceso ocurra, o bien se puede interpretar también como la frecuencia con la que se da el suceso.

Probabilidad condicionada e independencia

Ahora bien, una experiencia aleatoria puede ser única, en el sentido de que solo la tenemos en cuenta a ella y que no consideramos la dependencia con otros factores o experiencias, pero también puede darse el caso de que sí que esté sujeta a los resultados o realizaciones de otros eventos. Es lo que llamamos *probabilidad condicionada*.

Sea (Ω, \mathcal{F}, P) un espacio de probabilidad asociado a una cierta experiencia aleatoria. Fijamos un conjunto $B \in \mathcal{A}$ de probabilidad no nula.

Definición 1.1.2. La probabilidad de un conjunto $A \in \mathcal{A}$ condicionada por B se define por

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (1.1.1)$$

Se puede comprobar que $P(\cdot|B) : \mathcal{A} \rightarrow [0, 1]$ define también una probabilidad sobre (Ω, \mathcal{A}) . De hecho, $(\Omega, \mathcal{A}, P(\cdot|B))$ es el espacio de probabilidad que se obtiene de modificar el espacio inicial, añadiendo la información proporcionada por B .

La proposición siguiente nos proporciona una fórmula útil para el cálculo de la probabilidad a partir de las condicionadas, que usaremos con asiduidad más adelante.

Proposición 1.1.3. Consideramos una partición de Ω , $\{B_i, i \in I\}$ donde cada conjunto de la partición tiene una probabilidad estrictamente positiva. Entonces,

$$\forall A \in \mathcal{F}, P(A) = \sum_{i=1}^n P(A \cap A_i) = \sum_{i=1}^n P(A_i)P(A|A_i)$$

Hablamos ahora de la independencia de sucesos.

Definición 1.1.4. Los sucesos de una familia $(A_i, i \in I)$ son **independientes** si para todo subconjunto finito $(i_1, i_2, \dots, i_k) \subset I$

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k}) \quad (1.1.2)$$

En particular dos sucesos son independientes si $P(A \cap B) = P(A)P(B)$.

Observación 1.1.5. Si A y B son dos sucesos independientes y $P(B) > 0$, entonces $P(A|B) = P(A)$. Esto significa que conocer B no modifica la probabilidad del suceso A.

Si ahora combinamos la independencia con la probabilidad condicionada obtenemos lo siguiente

Definición 1.1.6. Los sucesos A y B son **condicionalmente independientes** dado un suceso C si

$$P(A \cap B|C) = P(A|C)P(B|C) \quad (1.1.3)$$

VARIABLES Y VECTORES ALEATORIOS

Desde el punto de vista matemático, el espacio muestral es incómodo de tratar ya que puede ser un conjunto de objetos bastante arbitrario. Para resolver esto, asociamos valores numéricos a los resultados de los fenómenos aleatorios, con unas aplicaciones que llamaremos *variables aleatorias*.

Para definir formalmente este concepto antes hemos de introducir la σ -álgebra de Borel (\mathcal{B}). Esta σ -álgebra es la generada por los conjuntos abiertos de \mathbb{R} , es decir la más pequeña de las σ -álgebras que contiene los conjuntos abiertos de \mathbb{R} .

Definición 1.1.7. Una **variable aleatoria** es una aplicación $X : \Omega \rightarrow \mathbb{R}$ de manera que $\forall B \in \mathcal{B}$, el conjunto $X^{-1}(B) \in \mathcal{F}$.

Con esta definición, la aplicación $P_X : \mathcal{B} \rightarrow [0, 1]$ definida como

$$P_X(B) = P(X^{-1}(B)) = P(\omega : X(\omega) \in B) \text{ con } B \in \mathcal{B}$$

es una probabilidad en \mathcal{B} , a la que llamamos *ley de X*. La terna $(\mathbb{R}, \mathcal{B}, P_X)$ es un espacio de probabilidad, y se puede considerar como una réplica numérica a través de X de (Ω, \mathcal{F}, P) .

Definición 1.1.8. La **función de distribución** asociada una v.a. X es la función $F : \mathbb{R} \rightarrow [0, 1]$ definida por

$$F(x) = P \circ X^{-1}((-\infty, x]). \quad (1.1.4)$$

La función de distribución cumple las siguientes propiedades:

- i) Es monótona creciente.
- ii) Es continua por la derecha.
- iii) $\lim_{x \rightarrow -\infty} F(x) = 0$ y $\lim_{x \rightarrow \infty} F(x) = 1$

Las variables aleatorias tienen dos subconjuntos importantes según sus leyes de distribución:

Una variable aleatoria X es *discreta* si su ley está concentrada en un conjunto numerable de \mathcal{B} . Decimos que es *continua* si su función de distribución es continua. Decimos que es *absolutamente continua* si su función de distribución puede escribirse como

$$F(x) = \int_{-\infty}^x f(y)dy \quad (1.1.5)$$

donde f es la *función densidad* de F y cumple

- i) $f \geq 0$
- ii) f es integrable en el sentido Riemann en \mathbb{R}
- iii) $\int_{-\infty}^{\infty} f(x) = 1$

Definición 1.1.9. Sean X_1, \dots, X_n , variables aleatorias definidas en un espacio de probabilidad (Ω, \mathcal{F}, P) . Diremos que son **independientes** si para conjuntos cualesquiera $B_1, \dots, B_m \in \mathcal{B}$, se cumple que

$$P\{X_1 \in B_1, \dots, X_m \in B_m\} = \prod_{i=1}^m P\{X_i \in B_i\} \quad (1.1.6)$$

Definición 1.1.10. Un **vector aleatorio** m -dimensional es una aplicación $X : \Omega \rightarrow \mathbb{R}$, $X = (X_1, \dots, X_n)$, tal que cada uno de los componentes $X_i : \Omega \rightarrow \mathbb{R}$ es una variable aleatoria.

Momentos de variables aleatorias

En este apartado nos centraremos en introducir parámetros relacionados con la ley de una variable aleatoria que nos proporcionan información sobre esta. Empezaremos hablando de la *esperanza matemática*, que es el número que formaliza la idea de valor medio de un fenómeno aleatorio.

Definición 1.1.11. La **esperanza matemática** de una variable aleatoria $X : \Omega \rightarrow \mathbb{R}$ es la integral de Lebesgue

$$(X) = \int_{\Omega} X dP \quad (1.1.7)$$

Observación 1.1.12. No toda función medible es Lebesgue-integrable, con lo cual la esperanza matemática puede no estar definida para una variable aleatoria arbitraria.

Vamos a discutir cómo calcular la esperanza de una v.a.

- **Variables aleatorias discretas:** La esperanza matemática de $X = \sum_{n \in \mathbb{N}} a_n \mathbb{1}_{A_n}$ existe si y sólo si $\sum_{n \in \mathbb{N}} |a_n| \mathbb{1}_{A_n} < \infty$, y en este caso,

$$E(X) = \sum_{n \in \mathbb{N}} a_n P(A_n) \quad (1.1.8)$$

- **Variables aleatorias absolutamente continuas:** La esperanza matemática de una v.a. X absolutamente continua existe si y sólo si $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$, y en este caso,

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx \quad (1.1.9)$$

Definición 1.1.13. Dada una variable aleatoria X y $k \in \mathbb{N}$ diremos que X tiene un **momento de orden k finito** si $E[|X|^k] < \infty$.

Observación 1.1.14. El momento de orden $k = 1$ es la esperanza matemática.

Definición 1.1.15. Si X tiene esperanza finita, diremos que X tiene **momento centrado de orden k finito** si $X - E(X)$ tiene momento de orden k finito.

Un elemento de interés es el momento centrado de orden dos, que definiremos en lo que sigue

Definición 1.1.16. La **varianza** (σ^2) de una variable aleatoria es una medida de dispersión definida como la esperanza del cuadrado de la desviación de dicha variable respecto a su media. La varianza se define también como el **momento centrado de orden dos** de una variable aleatoria, es decir, $E[(X - E(X))^2]$.

Un elemento relacionado con la varianza, y que es muy relevante en estudios probabilísticos es la *desviación típica*.

Definición 1.1.17. La **desviación típica o desviación estándar** (σ) es una medida de dispersión para variables de razón (variables cuantitativas o cantidades racionales) y de intervalo. Se define como la raíz cuadrada de la varianza de la variable.

Otro concepto importante que deberíamos tener en cuenta es la *covarianza*, que se vuelve relevante cuando las variables aleatorias no son independientes.

Definición 1.1.18. Dadas dos v.a. X e Y con momentos de orden dos finitos, i.e. $E(|X|^2) < \infty$ y $E(|Y|^2) < \infty$, un parámetro que mide el grado de dependencia entre las dos variables se llama **covarianza** y se define como

$$Cov(X, Y) = E((X - EX)(Y - EY)) \quad (1.1.10)$$

Observación 1.1.19. Si X e Y son independientes, $Cov(X, Y) = 0$.

Procesos estocásticos

Definición 1.1.20. Un **proceso estocástico** con un espacio de estados S es una familia $\{X_i, i \in I\}$ de variables aleatorias $X_i : \Omega \rightarrow S$ indexadas por un conjunto I .

Es interesante considerar el conjunto I como un conjunto de tiempo, con lo cual un proceso estocástico nos sirve para modelizar fenómenos del mundo real con una evolución temporal. Los casos interesantes de procesos estocásticos son aquellos en los que las variables que los componen no son independientes.

En lo siguiente supondremos $I \subset \mathbb{R}_+$ y $S \subset \mathbb{R}$.

Definición 1.1.21. Las distribuciones conjuntas de dimensión finita del proceso $\{X_t, t \in T\}$ son leyes de probabilidad multidimensionales de familias finitas de vectores aleatorios X_{t_1}, \dots, X_{t_m} , donde $t_1, \dots, t_m \in T$ y $m \geq 1$ es arbitrario.

Definición 1.1.22. Las **trayectorias** de un proceso estocástico $\{X_t, t \in I\}$ son la familia de funciones indexadas por $\omega \in \Omega$, $X(\omega) : I \rightarrow S$, definidas por $X(\omega)(t) = X_t(\omega)$.

Es decir, una trayectoria proporciona el comportamiento del proceso estocástico a lo largo del tiempo t , una vez fijado $\omega \in \Omega$, i .e. de valores fijados de las v.a.

1.2. Procesos estocásticos a tiempo discreto: El paseo aleatorio

En esta sección trataremos principalmente cadenas de Markov, daremos la definición, y algunas de sus propiedades y luego nos centraremos en un ejemplo concreto: el paseo aleatorio.

Un *proceso de Markov* $\{X_t, t \in I\}$ es un proceso estocástico con la propiedad de que, dado un valor de X_t , los valores de $X_s, s > t$ no se ven influenciados por los valores de $X_u, u < t$. Es decir, el proceso estocástico “carece de memoria”, lo que significa que la distribución de probabilidad del valor futuro de una variable aleatoria depende únicamente de su valor presente, siendo independiente de la historia de dicha variable. Esta propiedad es la llamada *propiedad de Markov*

Una *cadena de Markov de tiempo discreto* es un proceso de Markov donde el espacio de estados es finito o numerable, y el conjunto de índices temporales es \mathbb{N}

La probabilidad de que X_{n+1} esté en el estado j , sabiendo que X_n está en el estado i se llama *probabilidad de transición en un paso* y se denota por $P_{ij}^{n, n+1}$. Cuando esta probabilidad es independiente de la variable temporal n , decimos que la cadena de Markov tiene *probabilidades de transición estacionarias* o bien que la cadena es *homogénea*. Nos centraremos exclusivamente en este caso. Así pues, podemos reunir todas las probabilidades de transición en un solo paso en una matriz que llamaremos *matriz de probabilidad de transición* del proceso.

Observación 1.2.1. Claramente, la matriz P_{ij} satisface:

- i) $P_{ij} \geq 0$
- ii) $\sum_{j=0}^{\infty} P_{ij} = 1$, por probabilidades totales.

Un proceso de Markov queda completamente definido con su matriz de probabilidad de transición y su estado inicial X_0 , o más generalmente la distribución de probabilidad de X_0 , con lo cual, podemos ahora sí definir:

Definición 1.2.2. Un proceso estocástico $\{X_n, n \geq 0\}$ es una ***cadena de Markov*** homogénea con distribución inicial p_i y matriz de transición de probabilidad $\mathbb{P} = (P_{ij})_{i, j \in I}$ si se cumple:

- i) $P(X_0 = i) = p_i, \forall i \in I$, i.e. conocemos la distribución de probabilidad inicial.
- ii) $P(X_{n+1} = j \mid X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i) = P(X_{n+1} = j \mid X_n = i)$. Esta propiedad es la llamada *propiedad de Markov*.

Observación 1.2.3. La propiedad de Markov se puede escribir como

$$\begin{aligned} P(X_{n+1} = j_1, \dots, X_{n+m} = j_m \mid X_0 = i_0, \dots, X_n = i_n) = \\ P(X_{n+1} = j_1, \dots, X_{n+m} = j_m \mid X_n = i_n) \end{aligned}$$

Es decir, una vez la propiedad de Markov queda establecida para $m=1$, se mantiene $\forall m \geq 1$

Matrices de probabilidad de transición de una cadena de Markov

Ahora vamos a considerar *la matriz de transición en n pasos*, es decir la matriz que recoge las probabilidades de que el proceso vaya del estado i al estado j en n transiciones. Formalmente $\mathbb{P}^{(n)} = (P_{ij}^{(n)})$, donde $P_{ij}^{(n)} = P(X_{n+m} = j \mid X_m = i)$.

Teorema 1.2.4. *La matriz de transición de probabilidades en n pasos de una cadena de Markov cumple:*

$$P_{ij} = \sum_{k=0}^{\infty} P_{ik} P_{kj}^{(n-1)} \quad (1.2.1)$$

donde

$$P_{ij}^{(0)} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases} \quad (1.2.2)$$

Demostración.

$$\begin{aligned} P_{ij}^{(n)} &= P(X_n = j \mid X_0 = i) \\ &= \sum_{k=0}^{\infty} P(X_n = j, X_1 = k \mid X_0 = i) \\ &= \sum_{k=0}^{\infty} \frac{P(X_n = j, X_1 = k, X_0 = i)}{P(X_0 = i)} \\ &= \sum_{k=0}^{\infty} \frac{P(X_n = j \mid X_1 = k, X_0 = i) P(X_1 = k, X_0 = i)}{P(X_0 = i)} \\ &= \sum_{k=0}^{\infty} \frac{P(X_n = j \mid X_1 = k, X_0 = i) P(X_1 = k \mid X_0 = i) P(X_0 = i)}{P(X_0 = i)} \\ &= \sum_{k=0}^{\infty} P_{ik} P(X_n = j \mid X_1 = k) \\ &= \sum_{k=0}^{\infty} P_{ik} P_{kj}^{(n-1)} \end{aligned}$$

donde hemos usado la fórmula de la probabilidad total y la propiedad de Markov.

□

Observación 1.2.5. Por la teoría de matrices, tenemos por iteración que

$$\mathbb{P}^{(n)} = \mathbb{P} \times \mathbb{P}^{(n-1)} = \dots = \mathbb{P} \times \mathbb{P} \times \dots \times \mathbb{P} = \mathbb{P}^n. \quad (1.2.3)$$

Es decir, las probabilidades de transición de n pasos, $P_{ij}^{(n)}$, son los coeficientes de la matriz \mathbb{P}^n , la n -ésima potencia de \mathbb{P} .

Análisis de primer orden

Estudiaremos ahora un procedimiento llamado *análisis de primer orden* o *análisis en una transición*, consistente en analizar los posibles casos que pueden surgir al finalizar una primera transición, y entonces, haciendo uso de la ley de la probabilidad total junto con la propiedad de Markov, obtener los datos que buscábamos.

Empezaremos tratando el *análisis de primer orden simple*. Veremos el caso tri-dimensional inicialmente, y luego generalizaremos para todo n . Para ello, consideramos una cadena de Markov $\{X_n\}$ cuya matriz de transición de probabilidades es

$$\mathbb{P} = \begin{pmatrix} 1 & 0 & 0 \\ \alpha & \beta & \gamma \\ 0 & 0 & 1 \end{pmatrix} \quad (1.2.4)$$

con α, β y γ estrictamente positivas y $\alpha + \beta + \gamma = 1$. Dada la estructura de la matriz, podemos observar que si la cadena de Markov empieza en el estado 1, se mantendrá en dicho estado durante un tiempo aleatorio y acabará por ir al estado 0 o al 2, donde será atrapado y absorbido. Estos estados son los llamados *estados absorbentes*. Es decir, una vez la cadena llega al estado 0, o al 2, permanecerá ahí durante el resto del tiempo. Esto nos plantea dos cuestiones: ¿En cuál de los dos estados (0 o 2) el proceso será absorbido y cuál es el tiempo medio que tarda en alcanzar dichos estados?

Necesitaremos una definición previa.

Definición 1.2.6. Llamamos *tiempo de absorción* del proceso al tiempo mínimo que tarda la cadena en llegar a un estado absorbente. Formalmente,

$$T = \min\{n \geq 0; X_n \text{ es un estado absorbente}\}$$

En nuestro caso particular $T = \min\{n \geq 0; X_n = 0 \vee X_n = 2\}$.

En términos formales, nuestras dos cuestiones planteadas son

$$u = P(X_T = 0 | X_0 = 1) \text{ y } v = E(T | X_0 = 1).$$

Observación 1.2.7. Es indiferente tratar $P(X_T = 0 | X_0 = 1)$ o $P(X_T = 2 | X_0 = 1)$, ya que estamos buscando la probabilidad con la que ocurre cada uno de los casos.

De la matriz de transición podemos obtener que, si iniciamos el proceso en el estado 1, $X_1 = 0$ ocurre con una probabilidad α (y en este caso $T=1$ y $X_T = 0$),

$X_1 = 2$ tiene probabilidad γ (con $T=1$ y $X_T = 2$). Si $X_1 = 1$, lo cual ocurre con una probabilidad β , entonces el proceso permanece en el estado 1 y el problema se repite desde el mismo punto que antes. Por tanto, con esto y teniendo en cuenta que 0 y 2 son estados absorbentes:

$$\begin{aligned} P(X_T = 0 | X_1 = 0) &= 1 \\ P(X_T = 0 | X_1 = 2) &= 0. \\ P(X_T = 0 | X_1 = 1) &= u \end{aligned}$$

Haciendo uso de la ley de la probabilidad total y de la propiedad de Markov, obtenemos:

$$\begin{aligned} u &= P(X_T = 0 | X_0 = 1) = \sum_{k=0}^2 P(X_T = 0 | X_0 = 1, X_1 = k) P(X_1 = k | X_0 = 1) = \\ &= \sum_{k=0}^2 P(X_T = 0 | X_1 = k) P(X_1 = k | X_0 = 1) = 1 \cdot \alpha + u \cdot \beta + 0 \cdot \gamma \end{aligned}$$

Con lo cual;

$$u = \alpha + \beta u \implies u = \frac{\alpha}{1 - \beta} = \frac{\alpha}{\alpha + \gamma}$$

donde hemos usado $\alpha + \beta + \gamma = 1$.

El resultado obtenido es la probabilidad condicionada de una transición al estado 0, sabiendo que se da una transición a 0 o a 2.

Ahora nos queda por determinar el tiempo medio de absorción, analizando, como antes, los casos posibles que se dan en la primera transición. Como preámbulo, debemos tener en cuenta que el tiempo de absorción es siempre al menos 1, ya que empezamos en un estado transitorio. Si $X_1 = 0$ o bien $X_1 = 2$, entonces no se requieren más transiciones y $T=1$. Por otra parte, si $X_1 = 1$, entonces el proceso vuelve a su punto inicial, y entonces habremos de añadir pasos adicionales. Es decir, obtenemos $v = 1 + \alpha \cdot 0 + \beta \cdot v + \gamma \cdot 0 = 1 + \beta v$, lo cual nos da

$$v = \frac{1}{1 - \beta}$$

.

De forma general, la fórmula para obtener el tiempo de absorción medio en una transición es

$$E[T | X_0 = i] = 1 + \sum_{j \notin A} E[T | X_0 = j] P_{ij} \quad (1.2.5)$$

donde A es el conjunto de los estados absorbentes.

Procedemos ahora a considerar el caso general. Formalizaremos previamente una definición:

Definición 1.2.8. Sea i un estado fijado.

Llamamos a j **estado transitorio** cuando la probabilidad $P_{ij}^{(n)}$ tiende a cero con $n \rightarrow \infty$.

Llamamos a j **estado absorbente** cuando $\exists n \mid P_{ij}^{(n)} = 1$.

Sea $\{X_n\}$ una cadena de Markov cuyos estados están numerados por $0, 1, \dots, N$. Supongamos que los estados $0, 1, \dots, r-1$ son estados transitorios, mientras que los estados r, \dots, N son estados absorbentes. La matriz de transición es de la forma

$$\mathbb{P} = \begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$$

donde $\mathbf{0}$ es una matriz $(N-r+1) \times r$ llena de ceros, \mathbf{I} es la identidad de dimensión $(N-r+1) \times (N-r+1)$, $Q_{ij} = P_{ij}$ para $0 \leq i, j < r$ y $R_{ij} = P_{ij}$ para $r \leq i, j \leq N$

Empezando el proceso en un estado transitorio $X_0 = i$ donde $0 \leq i < r$, el proceso permanecerá en ese estado durante un tiempo aleatorio, pero en última instancia quedará atrapado en un estado absorbente $i = r, \dots, N$. Como antes, nos interesa estudiar el tiempo medio que tarda el proceso en ser absorbido y la función de probabilidad de los estados en los que la absorción tiene lugar.

Consideremos la distribución de probabilidad de $u_i = P(X_T = k \mid X_0 = i)$, donde k es un estado absorbente fijado, e i el estado inicial, que será transitorio, también fijado.

Iniciamos el análisis de primer orden enumerando las posibilidades en la primera transición. Empezando por el estado i , el proceso va al estado k , para quedarse allí, con una probabilidad P_{ik} . El proceso también puede caer en un estado absorbente $j \neq k$ ($r \leq j \leq N$), y en tal caso la posibilidad de que la absorción se de en el estado k queda excluida. También puede ser que la primera transición derive en que el proceso caiga en otro estado transitorio $j < r$. Debido a la propiedad de Markov, una vez en el estado j , la probabilidad de que el estado sea absorbido en el estado k es $U_{jk} = u_j$ por definición. Obtenemos, pues, aplicando como siempre la ley de la probabilidad total, la ley de la probabilidad condicionada y la propiedad de Markov:

$$\begin{aligned} u_i &= P(\text{absorción en } k \mid X_0 = i) = \\ &= \sum_{j=0}^N P(\text{absorción en } k \mid X_0 = i, X_1 = j) P_{ij} = \\ &= P_{ik} + \sum_{j=r, j \neq k}^N P_{ij} \times 0 + \sum_{j=0}^{r-1} P_{ij} u_j \end{aligned}$$

Como esto funciona para cualquier estado i fijado, tenemos pues:

$$u_i = P_{ik} + \sum_{j=0}^{r-1} P_{ij} u_j \quad i = 0, 1, \dots, r-1. \quad (1.2.6)$$

Resolviendo el sistema de ecuaciones obtenemos los resultados requeridos.

Pasamos a la otra cuestión, el tiempo de absorción. Tenemos

$$T = \min\{n \geq 0; X_n \geq r\} \quad (1.2.7)$$

Suponemos que asociado a cada estado transitorio tenemos un valor $g(i)$, y queremos determinar el valor medio acumulado hasta la absorción del proceso. Sea w_i la media total, donde el subíndice i denota el estado inicial $X_0 = i$. Formalmente tenemos;

$$w_i = E \left[\sum_{n=0}^{T-1} g(X_n) \mid X_0 = i \right] \quad (1.2.8)$$

Con esta definición, podemos estudiar dos casos especiales:

(a) Si escogemos $g(i) = 1 \quad \forall i$ obtenemos

$$\sum_{n=0}^{T-1} g(X_n) = \sum_{n=0}^{T-1} 1 = T \quad (1.2.9)$$

con lo cual, $w_i = v_i = E(T \mid X_0 = i)$, la media del tiempo que tarda el proceso en ser absorbido.

(b) Si para un estado transitorio k tomamos

$$g(i) = \begin{cases} 1 & \text{si } i = k \\ 0 & \text{si } i \neq k \end{cases} \quad (1.2.10)$$

nos da w_i la media de visitas al estado k ($0 \leq k < r$) previas a la absorción. Denotaremos $g(i) = \delta_{ik}$ en este caso.

Procediendo una vez más por el método del análisis del primer paso obtenemos:

$$w_i = g(i) + \sum_{j=0}^{r-1} P_{ij} w_j \quad \text{para } i = 0, \dots, r-1. \quad (1.2.11)$$

Esto nos da los sistemas de ecuaciones siguientes:

$$(a) \quad v_i = 1 + \sum_{j=0}^{r-1} P_{ij} v_j \quad \text{para } i = 0, \dots, r-1..$$

$$(b) \quad W_{ik} = \delta_{ik} + \sum_{j=0}^{r-1} P_{ij} W_{jk} \quad \text{para } i = 0, \dots, r-1., \text{ donde } W_{ik} \text{ es el número medio de visitas al estado } k \text{ previo a la absorción, con estado inicial } i.$$

Cadenas de Markov especiales: el paseo aleatorio unidimensional

El paseo aleatorio es una formalización matemática de la trayectoria que resulta de hacer sucesivos pasos aleatorios. El recorrido que sigue una molécula al viajar por un líquido o gas, el camino que sigue un animal en su búsqueda de comida, el precio de una acción fluctuante y la situación financiera de un jugador son ejemplos de experiencias que pueden ser tratadas como un paseo aleatorio. El término fue introducido por Karl Pearson en 1905. Los resultados proporcionados por el estudio de este proceso de Markov han sido aplicados a múltiples campos tales como la computación, la economía, la física, la química, la psicología, la biología o la ecología.

Matemáticamente hablando, un *paseo aleatorio unidimensional* es una cadena de Markov cuyo espacio de estados es un subconjunto (finito o infinito) $a, a + 1, \dots, b$ $a, b \in \mathbb{Z}$ en el cual el proceso, con estado inicial $X_0 = i$, puede en una transición o bien quedarse en el mismo estado, o bien moverse a uno de los estados vecinos $i + 1$ o $i - 1$. Si tomamos el espacio de estados en los enteros no negativos, la matriz de transición es de la forma

$$\begin{pmatrix} r_0 & p_0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots \\ q_1 & r_1 & p_1 & \cdots & 0 & 0 & 0 & 0 & \cdots \\ 0 & q_2 & r_2 & \cdots & 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & \cdots & q_i & r_i & p_i & 0 & \cdots \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix} \quad (1.2.12)$$

donde $p_i > 0, q_i > 0, r_i \geq 0$ y $p_i + q_i + r_i = 1$ para $i \geq 1$ y $p_0 \geq 0, r_0 \geq 0$ $p_0 + r_0 = 1$. En particular, si $X_n = i$ entonces, para $i \geq 1$

$$P(X_{n+1} = i + 1 \mid X_n = i) = p_i \quad (1.2.13)$$

$$P(X_{n+1} = i \mid X_n = i) = r_i \quad (1.2.14)$$

$$P(X_{n+1} = i - 1 \mid X_n = i) = q_i \quad (1.2.15)$$

con las correspondientes modificaciones para $i = 0$.

Vamos a modelizar ahora el caso de la situación financiera de un jugador de juegos de azar. Supongamos un jugador A con una fortuna inicial k que juega contra un adversario infinitamente rico y tiene una probabilidad p_k de ganar una unidad en la contienda, probabilidad $q_k = 1 - p_k$ ($k \geq 1$) de perderla (la elección de la apuesta en cada etapa puede depender del dinero del jugador) y $r_0 = 1$ (es decir, una vez el jugador está arruinado, ya no puede seguir jugando). El proceso $\{X_n\}$, donde X_n representa la fortuna del jugador después de n asaltos es claramente un paseo aleatorio.

Si el adversario (el jugador B) no es infinitamente rico, sino que empieza con una fortuna $l < \infty$, y el jugador A inicia con un capital k , ($k + l = N$), entonces podemos

considerar también el proceso de Markov $\{X_n\}$ representando el dinero del jugador A , pero en este caso los valores del proceso quedan restringidos a $0, 1, \dots, N$. En cada enfrentamiento, $N - X_n$ será el dinero del jugador B . Si admitimos la posibilidad de que haya empate (el caso más general), la matriz de transición de probabilidades toma la forma

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots \\ q_1 & r_1 & p_1 & 0 & \cdots & \cdots & \cdots & \cdots \\ 0 & q_2 & r_2 & p_2 & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & q_{N-1} & r_{N-1} & p_{N-1} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix} \quad (1.2.16)$$

donde $p_i(q_i), i = 1, \dots, N - 1$ denota la probabilidad del incremento (disminución) de una unidad de la fortuna del jugador A en la jugada, cuando su dinero actual. r_i es interpretado como la probabilidad de un empate.

Observación 1.2.9. Cuando la fortuna del jugador A alcanza el 0 o la N , implica la ruina de uno de los dos jugadores (0 la del jugador A y N la del jugador B). Son los estados absorbentes de la cadena.

Procedemos ahora a estudiar la probabilidad de la ruina del jugador A mediante el análisis a primer orden.

El caso en que $p_k = p, q_k = 1 - p = q \quad \forall k \geq 1, r_0 = 1$ describe la situación en la que todos los enfrentamientos durante el juego son idénticos, y no dependen del dinero de cada jugador. Si $p < q$ existe una ventaja para el jugador B , y si $p > q$ la hay para el jugador A . Un juego justo corresponde al caso $p = q = \frac{1}{2}$.

Supongamos que, como antes, $N = k+l$ es la suma del dinero de ambos jugadores, que estamos en una situación de juego en la que no se puede empatar y que el juego es justo ($p = q = \frac{1}{2}$). Este caso se llama *paseo aleatorio simple*. El proceso de Markov $\{X_n\}$, donde X_n es el dinero del jugador A en la etapa n , tiene por matriz de transición de probabilidad

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ q & 0 & p & 0 & \cdots & 0 & 0 \\ 0 & q & 0 & p & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & p \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \quad (1.2.17)$$

Vemos que el estado 0 y el estado N son estados absorbentes. Sea u_k la probabilidad de la ruina del jugador A , con fortuna inicial k . Aplicando el análisis de primer paso obtenemos

$$\begin{cases} u_k = pu_{k+1} + qu_{k-1} & \text{para } k = 1, \dots, N-1 \\ u_0 = 1 \\ u_N = 0 \end{cases} \quad (1.2.18)$$

Resolviendo estas ecuaciones podemos ver

$$u_i = \begin{cases} \frac{N-k}{N} & \text{si } p = q = \frac{1}{2} \\ \frac{(q/p)^k - (q/p)^N}{1 - (q/p)^N} & \text{si } p \neq q \end{cases} \quad (1.2.19)$$

Si pasamos al límite en que el jugador B es infinitamente rico ($N \rightarrow \infty$), entonces $(q/p)^N \rightarrow \infty$ si $p < q$ y $(q/p)^N \rightarrow 0$ si $p > q$, con lo cual nuestra solución queda

$$u_k = \begin{cases} 1 & \text{si } p \leq q \\ \left(\frac{q}{p}\right)^k & \text{si } p > q \end{cases} \quad (1.2.20)$$

Observamos que en este último caso, la ruina del jugador A se da siempre en caso de que el juego sea justo, y en caso de desventaja para el jugador A , mientras que en caso de un juego favorable para el jugador A , la probabilidad de arruinarse es $\left(\frac{q}{p}\right)^k$.

Estudiamos ahora el tiempo medio que tarda el jugador en arruinarse. Para ello daremos la noción de *tiempo entrada en un conjunto*. Observamos que el tiempo de absorción del que hemos hablado en apartados anteriores es un caso particular.

Definición 1.2.10. Para un proceso estocástico $X = \{X_n, n \geq 0\}$ y para un conjunto $A \subset I$, el **tiempo de entrada en el conjunto** A por X está definido por

$$T_A = \inf\{n \geq 0; X_n \in A\} \quad (1.2.21)$$

Por convención, si el conjunto $\{n \geq 0; X_n \in A\}$ es vacío, $T_A = \infty$

En nuestro caso particular de la ruina del jugador,

$$T = \min\{n \geq 0; X_n = 0 \text{ ó } X_n = N\} \quad (1.2.22)$$

Queremos evaluar $v_k = E[T \mid X_0 = k]$. Por análisis del primer paso obtenemos

$$\begin{cases} v_k = 1 + pv_{k+1} + qv_{k-1} & \text{para } k = 1, \dots, N-1 \\ v_0 = 0 \\ v_N = 0 \end{cases} \quad (1.2.23)$$

Resolviendo para el caso $p = q = \frac{1}{2}$ obtenemos

$$v_k = k(N - k) \quad (1.2.24)$$

Observación 1.2.11. La duración del tiempo del juego es máxima para una fortuna inicial $k = N/2$. Los tiempos medios más altos están alrededor de esta cifra.

Generalización

Ahora vamos a enunciar una generalización del paseo aleatorio que es interesante porque será la que desarrollaremos para construir el movimiento browniano.

Sea ξ_1, \dots, ξ_n variables aleatorias independientes e idénticamente distribuidas. Suponemos que $E[\xi_i] = 0$, $E[\xi_i^2] = \sigma^2$. Definimos como $\{S_t, t \geq 0\}$, el proceso aleatorio a tiempo discreto que viene dado por

$$S_n = \sum_{j=0}^n \xi_j \quad (1.2.25)$$

con $\xi_0 = 0$ por convenio. Como podemos ver, este proceso es un paseo aleatorio.

A partir de $\{S_n, n \geq 1\}$ se puede definir el proceso $\{B_n(t)\}$ a tiempo continuo obtenido mediante interpolación lineal, de la forma

$$B_n(t) = (S_{n+1} - S_n)t \quad (1.2.26)$$

La siguiente gráfica nos muestra la trayectoria de un ejemplo particular del proceso.

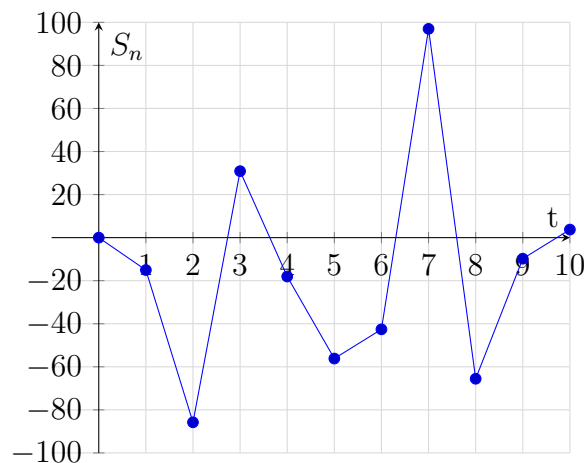


Figura 1.2.1: Ejemplo de la interpolación del paseo aleatorio generalizado explicado previamente.

Observación 1.2.12. Podemos observar que el paseo aleatorio dado por (1.2.13),

(1.2.14) y (1.2.15), con sus modificaciones para $i = 0$, se puede entender como un paseo con v.a.i.i.d. ξ'_1, ξ'_2, \dots donde

$$\xi'_j = \begin{cases} 1 & \text{con probabilidad } p_j \\ 0 & \text{con probabilidad } r_j \\ -1 & \text{con probabilidad } q_j \end{cases} \quad \xi'_0 = \begin{cases} 1 & \text{con probabilidad } p_0 \\ 0 & \text{con probabilidad } r_0 \end{cases} \quad (1.2.27)$$

Es decir, es un caso particular de la generalización que acabamos de dar, puesto que en la generalización, las variables ξ_i pueden tomar un valor cualquiera.

1.3. Procesos estocásticos a tiempo continuo. El movimiento browniano

Procesos gaussianos a tiempo continuo

Definición 1.3.1. Un vector aleatorio X_1, \dots, X_n se dice que tiene una **distribución multivariante normal** o **distribución conjunta normal** si toda combinación lineal $\alpha_1 X_1 + \dots + \alpha_n X_n$, con $\alpha_i \in \mathbb{R} \quad \forall i = 1, \dots, n$ tiene una distribución normal univariante.

Observación 1.3.2. Obviamente si X_1, \dots, X_n tiene una distribución conjunta normal, entonces también la tiene el vector aleatorio Y_1, \dots, Y_m definido por la transformación lineal

$$Y_j = \alpha_{j1} X_1, \dots, \alpha_{jn} X_n \text{ para } j = 1, \dots, m .$$

para constantes arbitrarias α_{ji}

La distribución normal multivariante viene especificada por dos parámetros: los valores medios $\mu_i = E[X_i]$ y la matriz de covarianza cuyos coeficientes son $\Gamma_{ij} = Cov[X_i, X_j]$. En esta distribución $\Gamma_{ij} = 0$ implica que X_i y X_j son variables independientes, lo cual no es cierto para distribuciones multidimensionales arbitrarias.

Daremos ahora un teorema, debido a Kolmogorov, que es el teorema de existencia de procesos gaussianos. Es el teorema en el que se basa la construcción del movimiento browniano de Kolmogorov.

Teorema 1.3.3. . Consideremos una familia

$$\{P_{t_1, \dots, t_n}; t_1 < \dots < t_n, n \geq 1, t_i \in T\} \quad (1.3.1)$$

donde

- a) P_{t_1, \dots, t_n} es una probabilidad en \mathbb{R}^n
- b) Si $\{t_{i_1} < \dots < t_{i_m}\} \subset \{t_1 < \dots < t_m\}$, la ley de probabilidad $\{P_{t_{i_1}, \dots, t_{i_m}}\}$ es la distribución marginal de P_{t_1, \dots, t_n} .

existe un proceso estocástico $\{X_t, t \in T\}$ definido en algún espacio de probabilidad tal que las distribuciones conjuntas finitas vienen dadas por 1.3.1. Es decir, la ley del vector aleatorio $(X_{t_1}, \dots, X_{t_n})$ es P_{t_1, \dots, t_n} .

Definición 1.3.4. Sea T un conjunto abstracto y $\{X(t); t \in T\}$ un proceso estocástico. Decimos que $\{X(t); t \in T\}$ un **proceso gaussiano** si para cada $n = 1, 2, \dots$ y para cada subconjunto finito (t_1, \dots, t_n) de T , el vector aleatorio $(X(t_1), \dots, X(t_n))$ tiene una distribución normal multivariante. Equivalentemente, un proceso es gaussiano si para toda combinación lineal

$$\alpha_1 X(t_1) + \dots + \alpha_n X(t_n), \quad \alpha_i \in \mathbb{R}$$

tiene una distribución lineal univariante.

Cada proceso gaussiano queda definido por sus dos parámetros, la media y la covarianza, dados por:

$$\mu(t_1, \dots, t_n) = E[X_{t_1}, \dots, X_{t_n}] = (E[X_{t_1}], \dots, E[X_{t_n}]) \quad (1.3.2)$$

$$\Gamma(t_i, t_j)_{1 \leq i, j \leq n} = E[\{X_{t_i} - \mu(X_{t_i})\}\{X_{t_j} - \mu(X_{t_j})\}]_{1 \leq i, j \leq n} \quad (1.3.3)$$

La función covarianza es definida no negativa, es decir, para todo $n = 1, 2, \dots$, para todo $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ y para todo t_1, \dots, t_n de T ,

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \Gamma(t_i, t_j) \geq 0 \quad (1.3.4)$$

Efectivamente,

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \Gamma(t_i, t_j) &= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \{E[X_{t_i} X_{t_j}] - E[X_{t_i}]E[X_{t_j}]\} \\ &= E \left[\sum_{i=1}^n \alpha_i \{X_{t_i} - E[X_{t_i}]\} \right]^2 \geq 0 \end{aligned}$$

A su vez, dada una media arbitraria y una covarianza definida no negativa, existe un proceso gaussiano correspondiente. Ilustraremos esta última parte con el siguiente ejemplo.

Ejemplo 1.3.5. Sea (Y_1, \dots, Y_n) un vector aleatorio cuyas componentes tienen media cero y momento finito de orden dos. Entonces existe un proceso gaussiano $\{X_t, t \geq 0\}$ tal que $E[X_t] = 0$ para cualquier $t \in T$ y $\text{Cov}(X_{t_i}, X_{t_j}) = \Gamma(t_i, t_j)$ para cualquier $t_i, t_j \in T$.

Para demostrar esto utilizaremos el teorema anterior. Fijamos $t_1, \dots, t_n \in T$ y tomamos $\mu(t_1, \dots, t_n) = (0, \dots, 0) \in \mathbb{R}^n$, $\Gamma(t_1, \dots, t_n) = \Gamma(t_i, t_j)_{1 \leq i, j \leq n}$ y

$$P_{t_1, \dots, t_n} = N(0, \Gamma_{t_1, \dots, t_n}) \quad (1.3.5)$$

Llamamos $(X_{t_1}, \dots, X_{t_n})$ al vector aleatorio con ley P_{t_1, \dots, t_n} . Para cualquier subconjunto $\{t_{i_1}, \dots, t_{i_m}\}$ de $\{t_1, \dots, t_n\}$ se cumple

$$A(X_{t_1}, \dots, X_{t_n}) = (X_{t_{i_1}}, \dots, X_{t_{i_m}}) \quad (1.3.6)$$

donde

$$A = \begin{pmatrix} \delta_{t_1, t_{i_1}} & \cdots & \delta_{t_n, t_{i_1}} \\ \cdots & \cdots & \cdots \\ \delta_{t_1, t_{i_m}} & \cdots & \delta_{t_n, t_{i_m}} \end{pmatrix} \quad (1.3.7)$$

y donde $\delta_{s,t}$ es la función delta de Kronecker.

Por las propiedades de los vectores gaussianos, el vector aleatorio $(X_{t_{i_1}}, \dots, X_{t_{i_m}})$ tiene función de distribución normal m -dimensional, media cero y covarianza $A \Gamma_{t_1, \dots, t_n} A^t = \Gamma(t_{i_l}, t_{i_k})_{1 \leq l, k \leq m}$.

Por tanto, las hipótesis del teorema se cumplen, y obtenemos el resultado que queríamos.

El movimiento browniano

El *movimiento browniano* es el movimiento aleatorio que se observa en las partículas que se hallan en un medio fluido (líquido o gas), que resulta de los choques contra las moléculas de dicho fluido. Hoy en día el movimiento browniano y sus muchas generalizaciones y extensiones son estudiadas en diversas áreas de ciencias puras y aplicadas, tales como económicas, teoría de la comunicación, biología o estadística.

En el campo de la probabilidad, este movimiento también es llamado *proceso de Wiener*. El movimiento browniano es un ejemplo de un proceso de Markov a tiempo continuo y espacio de estados continuo.

Sea $\{B_t\}$ un movimiento browniano. Sea $B(t)$ la componente y (respecto al tiempo) de una partícula en movimiento browniano, y sea x_0 la posición de la partícula en el instante inicial: $B(t_0) = x_0$. Sea $p(y, t | x_0)$ la densidad de probabilidad en y de $B(t_0 + t)$, dado $B(t_0) = x_0$. Suponemos un estado estacionario en el tiempo, con lo cual $p(y, t | x_0)$ no depende de t_0 .

Por ser $p(t, y | x_0)$ una función de densidad, sabemos que se cumple:

$$p(t, y | x_0) \geq 0 \quad (1.3.8)$$

$$\int_{-\infty}^{\infty} p(y, t | x_0) = 1 \quad (1.3.9)$$

Imponemos además que $\lim_{t \rightarrow 0} p(t, y | x_0) = 0$ para $y \neq x$.

Einstein demostró que $p(t, y | x_0)$ satisface la ecuación con derivadas parciales

$$\frac{\partial p}{\partial t} = \frac{1}{2} \sigma^2 \frac{\partial^2 p}{\partial y^2} \quad (1.3.10)$$

llamada *ecuación de difusión* y σ^2 el *coeficiente de difusión*. Tomando la escala

adecuada para poder usar $\sigma^2 = 1$ se puede demostrar que

$$p(t, y | x_0) = \frac{1}{\sqrt{2\pi t}} \exp \left\{ -\frac{(y - x_0)^2}{2t} \right\} \quad (1.3.11)$$

es la única solución a la ecuación de difusión que cumple las propiedades de función de distribución de probabilidad. Esta es la distribución de probabilidad de $B(t) - B(0)$.

Definición 1.3.6. Decimos que un proceso estocástico tiene **incrementos independientes** si para cualquier $t_1 < t_2 < \dots < t_k$ las variables aleatorias

$$X_{t_2} - X_{t_1}, \dots, X_{t_k} - X_{t_{k-1}} \quad (1.3.12)$$

son independientes

Definición 1.3.7. Un proceso estocástico tiene **incrementos estacionarios** si para cualquier $t_1 < t_2$ la ley de la variable aleatoria $X_{t_2} - X_{t_1}$ es la misma que la de $X_{t_2-t_1}$

Hemos dado la ley de probabilidad de $B(t) - B(0)$ del movimiento browniano. Para generalizar para cualquier tipo de incremento temporal, definimos:

Definición 1.3.8. El **movimiento browniano** con coeficiente de difusión σ^2 es un proceso estocástico $\{B(t); t \geq 0\}$ con las propiedades:

- (a) Todo incremento $B(s+t) - B(s)$ tiene una distribución normal con media cero y varianza $\sigma^2 t$, donde $\sigma^2 > 0$ es un parámetro fijo.
- (b) Para cada par de intervalos de tiempo disjuntos $(t_1, t_2], (t_3, t_4]$ con $0 \leq t_1 < t_2 \leq t_3 < t_4$, los incrementos $B(t_4) - B(t_3)$ y $B(t_2) - B(t_1)$ son variables aleatorias independientes. Esto se cumple para cada n intervalos de tiempo disjuntos, con $n \in \mathbb{N} - \{0\}$
- (c) $B(0) = 0$ y $B(t)$ es una función continua en t , casi seguramente.

Otras formas de definir el movimiento browniano son las que siguen.

Definición 1.3.9. El movimiento browniano es un proceso estocástico gaussiano $\{B_t, t \geq 0\}$ tal que

$$\mu(t) = 0 \quad (1.3.13)$$

$$\Gamma(s, t) = s \wedge t := \min(s, t) \quad (1.3.14)$$

Observación 1.3.10. Podemos observar que

$$s \wedge t = \int_0^\infty \mathbb{1}_{[0, s]}(r) \mathbb{1}_{[0, t]}(r) dr \quad (1.3.15)$$

Antes de dar la siguiente proposición, que nos dará otra definición equivalente del movimiento browniano, hemos de recordar la convergencia casi segura. Una sucesión de variables aleatorias $\{X_n, n \geq 1\}$ converge casi seguramente a una variable aleatoria X si existe un conjunto $\mathcal{N} \in \mathcal{F}$ de probabilidad cero tal que $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$ para todo $\omega \notin \mathcal{N}$

Proposición 1.3.11. *Un proceso estocástico $\{X_t, t \geq 0\}$ es un movimiento browniano si y solo si:*

- i) $X_0 = 0$ casi seguramente.
- ii) para cualquier $0 \leq s < t$, la variable aleatoria $X_t - X_s$ es independiente de la σ -álgebra generada por $X_r, 0 \leq r \leq s$, $\sigma(X_r, 0 \leq r \leq s)$ y $X_t - X_s$ es una variable aleatoria $N(0, t - s)$.

Observación 1.3.12. El movimiento browniano es el único proceso gaussiano que tiene trayectorias continuas, media cero y covarianza $\sigma^2 \min\{s, t\}$

Observación 1.3.13. La definición de movimiento browniano nos dice que el desplazamiento $B(t + s) - B(s)$ es independiente de su pasado, es decir, si conocemos $B(s) = x$, el conocimiento de los valores de $B(\tau)$ para tiempos anteriores $\tau < s$ no tienen ningún efecto en la ley de probabilidad que gobierna el comportamiento posterior del movimiento $B(s + t) - B(s)$. Esta característica es la propiedad de Markov. Hemos de tener en cuenta, sin embargo, que el apartado (b) de la definición del movimiento browniano es, de hecho, más restrictivo que la propiedad de Markov.

Observación 1.3.14. La elección $B(0) = 0$ es arbitraria, y de hecho, a menudo se considera el *movimiento browniano con inicio en x_0* , para el cual $B(0) = x_0$ para un punto fijado x_0 . En este caso, la varianza es $\sigma^2 t$ donde σ^2 es el llamado *parámetro de varianza*. El proceso $\tilde{B} = B(t)/\sigma$ es un movimiento browniano con parámetro de varianza 1, llamado *movimiento browniano estándar*. En general, reduciremos un movimiento browniano arbitrario a un movimiento browniano estándar para hacer los cálculos necesarios.

Observación 1.3.15. Del apartado (a) de la definición 1.3.8 sacamos que un movimiento browniano estándar cumple

$$\begin{aligned} P(B(s + t) \leq y \mid B(s) = x_0) &= P(B(s + t) - B(s) \leq y - x_0) \\ &= \int_{-\infty}^z \frac{1}{\sqrt{2\pi t}} \exp\left\{-\frac{(y - x_0)^2}{2t}\right\} dy \end{aligned}$$

Parámetros del movimiento browniano

- **Desviación típica y varianza** Queremos estudiar la variabilidad del desplazamiento del movimiento browniano $B(\Delta t)$ tras un pequeño periodo de tiempo Δt . Es usual trabajar con la desviación típica (σ) para estos casos, ya que ésta está en las mismas unidades que la medida original. La desviación

típica del desplazamiento browniano es $\sqrt{\Delta t}$, que es mucho mayor que Δt cuando éste es pequeño. De hecho,

$$\frac{\sigma(B(\Delta t))}{\Delta t} = \frac{\sqrt{\Delta t}}{\Delta t} = \frac{1}{\sqrt{\Delta t}} \rightarrow \infty \text{ cuando } \Delta t \rightarrow 0.$$

Recordemos que hemos supuesto que los desplazamientos en intervalos de tiempo disjuntos son estacionarios, y que por el apartado (b) de la definición de movimiento browniano son independientes. Entonces la varianza (σ^2), lineal con el tiempo, es la única posibilidad de medir la variación de los desplazamientos brownianos. Tenemos pues

$$Var[B(t+s)] = Var[B(t)] + Var[B(s)]$$

- **Covarianza.** Teniendo en cuenta el apartado (b) de la definición de movimiento browniano, y recordando que $E[B(t)] = 0$ y $E[B(t)^2] = \sigma^2 t$, para $0 \leq s < t$ tenemos:

$$\begin{aligned} Cov[B(s), B(t)] &= E[B(s)B(t)] \\ &= E[B(s)(B(t) - B(s) + B(s))] \\ &= E[B(s)^2] + E[B(s)(B(t) - B(s))] \\ &= \sigma^2 s + E[B(s)]E[B(t) - B(s)] = \sigma^2 s \end{aligned}$$

De manera idéntica, si $0 \leq t < s$, podemos obtener $Cov[B(t), B(s)] = \sigma^2 t$. Por tanto, podemos escribir, de forma general:

$$Cov[B(s), B(t)] = \sigma^2 \min\{s, t\} \quad (1.3.16)$$

Propiedades del movimiento browniano

El movimiento browniano posee varias propiedades de invariancia. Algunas de las más relevantes son:

- (1) Si $B = \{B_t, t \geq 0\}$ es un movimiento browniano, $-B = \{-B_t, t \geq 0\}$ también lo es.
- (2) Para cualquier $\lambda > 0$, el proceso $B^\lambda = \{\frac{1}{\lambda} B_{\lambda^2 t}, t \geq 0\}$ es también un movimiento browniano. Es la propiedad de invarianza por un cambio de escala del movimiento browniano.
- (3) Para cualquier $a > 0$, $B^{+a} = \{B_{t+a} - B_a, t \geq 0\}$ es un movimiento browniano.

Trayectorias del movimiento browniano

El movimiento browniano fue introducido como un modelo para las trayectorias erráticas de las partículas. En consecuencia, uno espera encontrar unas trayectorias casi seguramente continuas, pero con cambios bruscos en las direcciones.

Hay varias maneras de obtener un movimiento browniano. Este trabajo está enfocado a dar su construcción a partir del paseo aleatorio, por el teorema de Donsker, pero en esta sección trataremos la construcción de P. Lévy's, ya que esta nos proporciona la continuidad de sus trayectorias.

Esta construcción se da en $L^2([0, 1], \mathcal{B}([0, 1]), \lambda) = X$ que es un espacio vectorial normado, y donde λ es una medida de Lebesgue. Es el espacio de todas las funciones medibles que cumplen

$$\int_X |f|^2 d\mu < \infty$$

La norma de este espacio viene definida por $\|f\|_2 = \left(\int |f|^2 d\mu \right)^{1/2}$

Este espacio está dotado de un producto interno $\langle f, g \rangle = \int fg d\mu$

Para aplicar este procedimiento, debemos antes definir las *funciones Haar* como sigue:

$$h_0(t) = 1$$

$$h_n^k(t) = 2^{n/2} \mathbf{1}_{\left[\frac{2k}{2^{n+1}}, \frac{2k+1}{2^{n+1}}\right)} - 2^{n/2} \mathbf{1}_{\left[\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}}\right)}$$

con $n \geq 1$ y $k \in \{0, 1, \dots, 2^n - 1\}$.

El conjunto de funciones (h_0, h_n^k) es una base ortonormal del espacio X . Por tanto, para cualquier $f \in X$ podemos escribir el desarrollo:

$$f = \langle f, h_0 \rangle h_0 + \sum_{n=1}^{\infty} \sum_{k=0}^{2^n-1} \langle f, h_n^k \rangle h_n^k \quad (1.3.17)$$

Usando el desarrollo podemos definir una isometría entre X y $L^2(\Omega, \mathcal{F}, P)$ de la siguiente forma. Consideremos una familia de variables independientes (N_0, N_n^k) para $n \geq 1$ y $k \in \{0, 1, \dots, 2^n - 1\}$ con ley $N(0, 1)$. Entonces, para $f \in L^2([0, 1], \mathcal{B}([0, 1]))$, definiendo

$$I(f) = \langle f, h_0 \rangle N_0 + \sum_{n=1}^{\infty} \sum_{k=0}^{2^n-1} \langle f, h_n^k \rangle N_n^k \quad (1.3.18)$$

Claramente

$$E[I(f)^2] = \|f\|_2^2. \quad (1.3.19)$$

I define la isometría deseada. Además, dado que

$$I(f) = \lim_{m \rightarrow \infty} \langle f, h_0 \rangle N_0 + \sum_{n=1}^m \sum_{k=0}^{2^n-1} \langle f, h_n^k \rangle N_n^k, \quad (1.3.20)$$

en la convergencia de $L^2(\Omega)$, la variable aleatoria $I(f)$ es $N(0, \|f\|_2^2)$ y por la identidad de Parseval,

$$E(I(f)I(g)) = \langle f, g \rangle, \quad (1.3.21)$$

para cualquier $f, g \in L^2([0, 1], \mathcal{B}([0, 1]), \lambda)$.

Teorema 1.3.16. *El proceso $B = \{B_t = I(\mathbb{1}_{[0,t]}), t \in [0, 1]\}$ define un movimiento browniano indexado por $[0, 1]$. Además, las trayectorias son continuas casi seguramente.*

Demostración. Por construcción $B_0 = 0$. Además $B_t - B_s = I(\mathbb{1}_{(s,t]})$ para $0 \leq s \leq t \leq 1$. Como $E(I(f)I(g)) = \langle f, g \rangle$, entonces el proceso $B_t - B_s$ es independiente de cualquier B_r , ($0 < r < s$), ya que como $E[I(f)] = E[I(g)] = 0$, $\text{Cov}(I(f), I(g)) = E(I(f)I(g)) = 0$ y además tiene una ley $N(0, t - s)$. Usando la proposición 1.3.10 se cumple la primera parte del teorema.

Nuestro siguiente objetivo es demostrar que la serie

$$B_t = I(\mathbb{1}_{[0,t]}) = \langle \mathbb{1}_{[0,t]}, h_0 \rangle N_0 + \sum_{n=1}^{\infty} \sum_{k=0}^{2^n-1} \langle \mathbb{1}_{[0,t]}, h_n^k \rangle N_n^k = g_0(t)N_0 + \sum_{n=1}^{\infty} \sum_{k=0}^{2^n-1} g_n^k(t)N_n^k$$

converge uniformemente casi seguramente. En el último término hemos introducido las *funciones Schauder* definidas como

$$\begin{aligned} g_0(t) &= \langle \mathbb{1}_{[0,t]}, h_0 \rangle = t \\ g_n^k(t) &= \langle \mathbb{1}_{[0,t]}, h_n^k \rangle = \int_0^t h_n^k(s) ds \end{aligned}$$

para cualquier $t \in [0, 1]$

Por construcción, para cualquier $n \geq 1$ fijado, las funciones $g_n^k(t)$, $k = 0, \dots, 2^n - 1$ son positivas, y $g_n^k(t) \leq 2^{-n/2}$. Así,

$$\sup_{t \in [0,1]} \left| \sum_{k=0}^{2^n-1} g_n^k(t)N_n^k \right| \leq 2^{-n/2} \sup_{0 \leq k \leq 2^n-1} |N_n^k|. \quad (1.3.22)$$

El siguiente paso consiste en probar que $|N_n^k|$ está acotada por alguna constante que dependa de n tal que al multiplicarla por $2^{-n/2}$ converja.

Para probar esto, usaremos el siguiente lema

Lema 1.3.17. *Para toda variable aleatoria X con ley $N(0, 1)$ y para toda $a \geq 1$*

$$P(|X| \geq a) \leq e^{-a^2/2} \quad (1.3.23)$$

Demostración. Tenemos

$$P(|X| \geq a) = \frac{2}{\sqrt{2\pi}} \int_a^\infty e^{-\frac{x^2}{2}} dx \leq \frac{2}{\sqrt{2\pi}} \int_a^\infty \frac{x}{a} e^{-\frac{x^2}{2}} dx = \frac{2}{a\sqrt{2\pi}} e^{-\frac{a^2}{2}} \leq e^{-\frac{a^2}{2}}$$

donde hemos usado $1 \leq \frac{x}{a}$ y $\frac{2}{a\sqrt{2\pi}} \leq 1$ □

Por el lema que acabamos de demostrar,

$$P\left(\sup_{0 \leq k \leq 2^n - 1} |N_n^k| > 2^{\frac{n}{4}}\right) \leq \sum_{k=0}^{2^n - 1} P(|N_n^k| > 2^{\frac{n}{4}}) \leq 2^n \exp(-2^{\frac{n}{2}-1})$$

Con lo cual

$$\sum_{n=1}^{\infty} P\left(\sup_{0 \leq k \leq 2^n - 1} |N_n^k| > 2^{\frac{n}{4}}\right) < \infty \quad (1.3.24)$$

y por el primer lema de Borel-Cantelli

$$P\left(\liminf_{n \rightarrow \infty} \left\{ \sup_{0 \leq k \leq 2^n - 1} |N_n^k| \leq 2^{\frac{n}{4}} \right\}\right) = 1 \quad (1.3.25)$$

Es decir, casi seguramente existe un n_0 tal que $\sup_{0 \leq k \leq 2^n - 1} |N_n^k| \leq 2^{\frac{n}{4}}$ para todo $n \geq n_0$.

Hemos probado pues, que

$$\sup_{t \in [0,1]} \left| \sum_{k=0}^{2^n - 1} g_n^k(t) N_n^k \right| \leq 2^{\frac{-n}{2}} \sup_{0 \leq k \leq 2^n - 1} |N_n^k| \leq 2^{\frac{-n}{4}} \quad (1.3.26)$$

casi seguramente para un n suficientemente grande. Esto prueba la convergencia uniforme casi seguramente que queríamos demostrar. □

Hemos visto que las trayectorias del movimiento browniano son continuas. Ahora veremos otro tipo de continuidad.

Definición 1.3.18. *Una función $f : \mathbb{R} \rightarrow \mathbb{R}$ es **Hölder-continua** si $\forall x, y \in \mathbb{R}$*

existen α, C constantes tales que

$$|f(x) - f(y)| \leq C\|x - y\|^\alpha \quad (1.3.27)$$

Proposición 1.3.19. Criterio de continuidad de Kolmogorov.

Sea $\{X_t, t \geq 0\}$ un proceso estocástico tal que para algunos reales positivos α, β y C ,

$$E(|X_t - X_s|^\alpha) \leq C|t - s|^{1+\beta}. \quad (1.3.28)$$

Entonces, casi seguramente, las trayectorias del proceso son γ -Hölder continuas con $\gamma < \frac{\beta}{\alpha}$

La ley de la variable aleatoria $B_t - B_s$ es $N(0, t - s)$. Entonces, podemos calcular

$$E((B_t - B_s)^{2k}) = \frac{(2k)!}{2^k k!} (t - s)^k \quad (1.3.29)$$

para cualquier $k \in \mathbb{N}$. La proposición nos dice que, casi seguramente, las trayectorias del movimiento browniano son γ -Holder continuas con $\gamma \in (0, \frac{1}{2})$

Respecto a la diferenciabilidad, el célebre resultado de Dvoretzky, Erdős y Kakutani nos dice que las trayectorias del movimiento browniano no son diferenciables casi seguramente en ningún punto.

Una vez dados todos los resultados sobre los movimientos brownianos, sería interesante ilustrar como es un movimiento browniano.

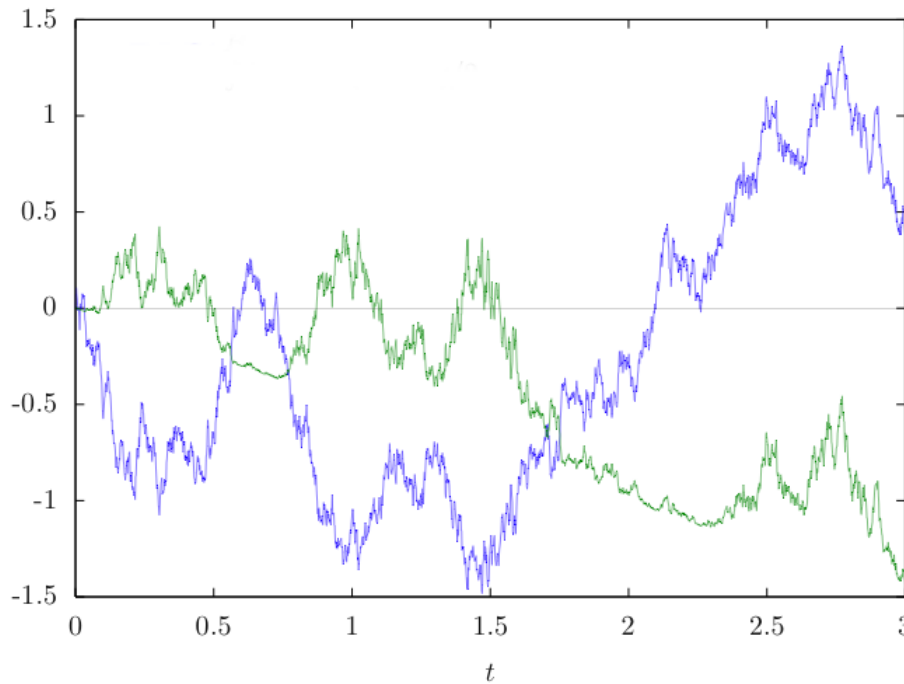


Figura 1.3.2: Trayectorias típicas del movimiento browniano.

Capítulo 2

Convergencia de medidas de probabilidad

El propósito de este capítulo es introducirnos en los conceptos y resultados principales sobre la convergencia de medidas de probabilidad sobre todo en espacios métricos. Empezaremos dando una idea general de qué es la convergencia débil, y trataremos más en profundidad la idea de convergencia débil en espacios métricos y la convergencia en distribución.

2.1. Introducción

El llamado *teorema del límite central de Moivre-Laplace* que nos dice que si

$$F_n(x) = P \left\{ \frac{S_n - np}{\sqrt{npq}} \leq x \right\} \quad (2.1.1)$$

es la función de distribución de variables aleatorias con distribución de Bernoulli con parámetro p y n tiradas, y

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du \quad (2.1.2)$$

es la función de distribución normal, entonces $F_n(x) \rightarrow F(x)$.

Decimos de las funciones de distribución anteriores que F_n converge débilmente a F ($F_n \Rightarrow F$) si $F_n \rightarrow F$ para todo punto de continuidad $x \in F$.

En el caso que acabamos de ver la condición de los puntos de continuidad es irrelevante puesto que la función de distribución normal es continua en todos los puntos. Sin embargo, si tomamos $F_n = \mathbb{1}_{[1/n, \infty)}(x)$ y $F = \mathbb{1}_{[0, \infty)}(x)$, vemos que esta vez sí que es necesaria la condición sobre los puntos de continuidad, puesto que en $x = 0$ vemos que $0 = F_n(0) \not\rightarrow F(0) = 1$. Como la función $F(x)$ no es continua en este punto, por la definición se cumple $F_n \Rightarrow F$.

Generalizando esta idea, consideremos las medidas de probabilidad P_n y P , de-

finidas en la clase de los subconjuntos de Borel, que vienen determinadas por las funciones de distribución $F_n(x)$ y $F(x)$ tal que

$$P_n(-\infty, x] = F_n(x), \quad P(-\infty, x] = F(x)$$

Como F es continua en x si y solo si el conjunto $\{x\}$ cumple $P(\{x\}) = 0$, $F_n \Rightarrow F$ significa que

$$P_n(-\infty, x] \rightarrow P(-\infty, x], \text{ si } P\{x\} = 0,$$

se cumple para cada x .

Sea ahora ∂A la frontera de un conjunto A de Borel. Recordemos que un conjunto de Borel es un conjunto que pertenece a la σ -álgebra de Borel (\mathcal{B}), aquella generada por los conjuntos abiertos. Entonces $F_n \Rightarrow F$ si y solo si

$$P_n(A) \rightarrow P(A), \text{ si } P(\partial A) = 0,$$

se cumple para todo conjunto de Borel A .

Hablamos de *conjunto P -continuo* para los conjuntos de Borel A tales que $P(\partial A) = 0$, y diremos que P_n converge débilmente a P ($P_n \Rightarrow P$) si $P_n(A) \rightarrow P(A)$ para todo conjunto P -continuo A . Tenemos pues :

Proposición 2.1.1. $P_n \Rightarrow P$ si y solo si $F_n \Rightarrow F$

Aunque el concepto de *convergencia débil de funciones de distribución* está ligado al espacio euclideo, el concepto de *convergencia débil de medidas de probabilidad* se puede formular en un espacio métrico cualquiera, por lo cual este último concepto es preferible en un contexto más general.

Observación 2.1.2. Hablamos de convergencia débil cuando tratamos con convergencia de medidas cualesquiera, y hablamos de convergencia en distribución cuando tratamos la convergencia de funciones de distribución. Si estamos tratando una medida que tiene asociada una función de distribución (recordemos que dada una función de distribución F automáticamente existe una única medida de probabilidad que la tiene por función de distribución) entonces estamos dando dos expresiones diferentes de un mismo hecho, como nos muestra la proposición anterior.

2.2. Convergencia débil en espacios métricos

Empezamos estudiando las medidas de probabilidad en el espacio métrico general, que llamaremos (S, d) , donde d es una distancia. Sea \mathcal{B} la σ -álgebra de Borel, aquella generada por los conjuntos abiertos definidos por esta distancia d . Sus elementos son los llamados *conjuntos de Borel*.

Recordemos que una medida de probabilidad en \mathcal{B} es una función P no negativa, σ -aditiva y que cumple $P(S) = 1$

Definición 2.2.1. Si las medidas de probabilidad P_n y P cumplen

$$\int_S f dP_n \rightarrow \int_S f dP \quad (2.2.1)$$

para cualquier función f real continua y acotada en S , decimos que P_n **converge débilmente** a P y escribimos $P_n \Rightarrow P$

Observación 2.2.2. Escribiremos Pf para referirnos a $\int_S f dP$

El siguiente teorema nos permitirá determinar completamente el valor de P por los valores de $P(F)$ para conjuntos F cerrados.

Teorema 2.2.3. Toda medida de probabilidad P en el espacio (S, \mathcal{B}) es **regular**, i.e. para todo conjunto $A \in \mathcal{B}$ y para todo ϵ , existe un conjunto cerrado F y un conjunto abierto G , tal que $F \subset A \subset G$ y $P(G \setminus F) < \epsilon$, donde $G \setminus F = G \cap F^c$.

Vamos a ver ahora que P también queda completamente determinado por los valores Pf para funciones continuas y acotadas f . Esto implica que una sucesión $\{P_n\}$ no puede converger débilmente a dos límites diferentes.

Teorema 2.2.4. Las medidas de probabilidad P y Q de \mathcal{B} coinciden si $Pf = Qf$ para toda función acotada, real y uniformemente continua f .

Observación 2.2.5. Teoremas como este permiten trabajar con medidas PA o con integrales Pf indistintamente.

El siguiente teorema, llamado *teorema de Portmanteau* nos proporciona definiciones equivalentes de la convergencia débil. Recordemos que un conjunto $A \in \mathcal{B}$ con frontera $\partial A = 0$ se denomina un conjunto P -continuo.

Teorema 2.2.6. Estas cinco condiciones son equivalentes:

- i) $P_n \Rightarrow P$
- ii) $P_n f \rightarrow P f$ para toda función acotada y uniformemente continua f
- iii) $\limsup_n P_n(F) \leq P(F)$ para todo conjunto cerrado F .
- iv) $\liminf_n P_n(G) \geq P(G)$ para todo conjunto abierto G .
- v) $P_n A \rightarrow P A$ para todo conjunto A P -continuo.

Otra condición para la convergencia débil nos la da el siguiente teorema

Teorema 2.2.7. Una condición necesaria y suficiente para $P_n \Rightarrow P$ es que cada subsucesión $\{P_{n_i}\}$ contiene a su vez otra subsucesión $\{P_{n_{i_m}}\}$ que converge débilmente ($m \rightarrow \infty$) a P .

Supongamos ahora que h es una función que aplica S en otro espacio métrico S' , con métrica ρ' y σ -álgebra de Borel \mathcal{B}' . Si h es medible (i.e. preserva la estructura entre dos espacios medibles), entonces cada probabilidad P en (S, \mathcal{B}) induce una probabilidad Ph^{-1} en (S', \mathcal{B}') definida como $Ph^{-1}(A) = P(h^{-1}A)$.

Necesitaremos condiciones bajo las cuales $P_n \Rightarrow P$ implique $P_nh^{-1} \Rightarrow Ph^{-1}$. Una de ellas es la continuidad de h ; si f es acotada y continua en S' , entonces fh es acotada y continua en S y aplicando un cambio de variable, $P_n \rightarrow P$ implica

$$\int_{S'} f(y)(P_nh^{-1})(dy) = \int_S f(h(x))P_n(dx) \rightarrow \int_S f(h(x))P(dx) = \int_{S'} f(y)(Ph^{-1})(dy)$$

Esto implica que $P_nh^{-1} \Rightarrow Ph^{-1}$. Pero ahora podemos incluso poner menos condiciones. Supongamos que h es medible y sea D_h el conjunto de sus discontinuidades, ($D_h \in \mathcal{B}$). El *teorema del mapeado* nos dice:

Teorema 2.2.8. *Si $P_n \Rightarrow P$ y $P(D_h) = 0$, entonces $P_nh^{-1} \Rightarrow Ph^{-1}$*

Este teorema nos da un resultado muy útil que utilizaremos varias veces a lo largo del trabajo.

Procedemos ahora a dar un par de ejemplos de espacios métricos, que son los espacios con los que trabajaremos principalmente.

Decimos que una clase \mathcal{A} de \mathcal{B} es una *clase de separación* si dos medidas de probabilidad que coinciden en \mathcal{A} , necesariamente coinciden en todo \mathcal{B} . Recordemos que una clase es un π -sistema si es cerrado por intersecciones finitas, y que \mathcal{A} es una clase separable si es un π -sistema que genera \mathcal{B}

Ejemplo 2.2.9. Sea \mathbb{R}^k el espacio euclídeo k -dimensional con la métrica habitual

$$|x - y| = \sqrt{\sum_{i=1}^k (x_i - y_i)^2},$$

y sea \mathcal{R}^k (σ -álgebra) la clase de los conjuntos k -dimensionales borelianos. La función de distribución que corresponde a la medida de probabilidad P en \mathcal{R}^k es

$$F(x_1, \dots, x_k) = P[y; y_i \leq x_i, i \leq k]$$

Como los conjuntos de la derecha de la igualdad forman un π -sistema que genera \mathcal{R}^k , forman una clase separación. Por tanto, F determina completamente P .

Ejemplo 2.2.10. Sea $C = C[0, 1]$ el *espacio de las funciones continuas en $[0, 1]$* . Definimos una norma en C del modo siguiente:

$$\|x\|_\infty = \sup_t |x(t)|, \quad x \in C \tag{2.2.2}$$

A su vez definimos la *métrica uniforme*

$$\rho(x, y) = \|x - y\|_\infty = \sup_t |x(t) - y(t)|. \tag{2.2.3}$$

Llamaremos \mathcal{C} a la topología asociada a esta métrica.

Como $\rho(x_n, x) \rightarrow 0$ significa que x_n converge uniformemente a x , implica la convergencia puntual. Por supuesto, la otra implicación es falsa: Consideremos la función que crece linealmente de 0 a 1 en el intervalo $[0, 1/n]$, decrece linealmente de 1 a 0 en el intervalo $[1/n, 2/n]$ y es constante 0 a la derecha de $2/n$, es decir

$$z_n(t) = nt\mathbb{1}_{[0,1/n]}(t) + (2 - nt)\mathbb{1}_{(1/n,2/n]}(t)$$

z_n converge puntualmente a la función 0, mientras que $\rho(z_n, 0) = 1$

También es interesante ver que, por la continuidad de las proyecciones naturales π_{t_1, \dots, t_k} de C a \mathbb{R}^k , si $P_n \Rightarrow P$ para probabilidades en C , entonces por el teorema del mapeado, $P_n \pi_{t_1, \dots, t_k}^{-1} \Rightarrow P \pi_{t_1, \dots, t_k}^{-1}$ para todo k y para toda k -tupla t_1, \dots, t_k

2.3. Convergencia en distribución

La teoría de la convergencia en distribución (o convergencia en ley) es análoga a la de la convergencia débil. Lo interesante es ver como al estudiar estas mismas ideas en términos de convergencia en distribución, muchos resultados adquieren una forma más concisa y clarificadora.

Como dijimos en la introducción la principal diferencia entre la convergencia débil y la convergencia en distribución es que la segunda aplica sólo para leyes de elementos aleatorios.

Definición 2.3.1. Sean (Ω, \mathcal{A}) y (E, \mathcal{F}) dos espacios medibles. Diremos que una aplicación $X : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{F})$ es medible si $\forall B \in \mathcal{F}, X^{-1}(B) \in \mathcal{A}$.

Sea X una aplicación de un espacio de probabilidad $(\Omega, \mathcal{F}, \mathcal{P})$ a un espacio métrico (S, \mathcal{B}) . Decimos que $X : \Omega \rightarrow S$ es un *elemento aleatorio* de S si es medible (i.e. preserva la estructura entre dos espacios medibles). Llamamos a X variable aleatoria si $S = \mathbb{R}$, *vector aleatorio* si $S = \mathbb{R}^k, k \in \mathbb{N}, k > 1$ y *sucesión aleatoria* si $S = \mathbb{R}^\infty$

La *función de distribución (o ley)* de X es la medida de probabilidad $P = \mathcal{P}X^{-1}$ en (S, \mathcal{B}) , definida por

$$P(A) = \mathcal{P}(X^{-1}A) = \mathcal{P}[\omega : X(\omega) \in A] = \mathcal{P}[X \in A] \tag{2.3.1}$$

Observación 2.3.2. \mathcal{P} es una medida de probabilidad en un espacio medible arbitrario, mientras que P está siempre definida en la σ -álgebra de Borel de un espacio métrico.

Observación 2.3.3. Si f es una función real medible en S , por cambio de variable;

$$E[f(X)] = \int_{\Omega} f(X)d\mathcal{P} = \int_S f dP = Pf \tag{2.3.2}$$

en el sentido de que ambas integrales existen, o ninguna de las dos, y en caso de existencia, ambas tienen el mismo valor.

Definición 2.3.4. Decimos que una sucesión $\{X_n\}$ de elementos aleatorios **converge en distribución** hacia el elemento aleatorio X si $P_n \Rightarrow P$, donde P_n y P son funciones de distribución de X_n y X , respectivamente.

Hablemos ahora de terminología. Por nomenclatura usaremos $X_n \Rightarrow X$ para referirnos a que $\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X)$ donde $\mathcal{L}(X)$ es la ley de X y $\mathcal{L}(X_n)$ la ley de X_n . Es decir $P_n \Rightarrow P$ podemos representarlo de las cuatro formas siguientes indistintamente.

$$\left\{ \begin{array}{l} P_n \Rightarrow P \\ X_n \Rightarrow X \\ X_n \Rightarrow P \\ P_n \Rightarrow X \end{array} \right.$$

Ahora enunciaremos el teorema de Portmanteau readaptado a la convergencia en distribución. Llamaremos *conjunto X -continuo* a un conjunto $A \in \mathcal{B}$ si $\mathcal{P}(X \in \partial A) = 0$.

Teorema 2.3.5. *Tenemos que las cinco siguientes condiciones son equivalentes:*

- i) $X_n \Rightarrow X$
- ii) $E[F(X_n)] \rightarrow E[f(X)]$ para toda función continua y acotada f .
- iii) $\limsup_n \mathcal{P}(X_n \in F) \leq \mathcal{P}(X \in F)$ para todo conjunto cerrado F
- iv) $\liminf_n \mathcal{P}(X_n \in G) \geq \mathcal{P}(X \in G)$ para todo conjunto abierto G .
- v) $\mathcal{P}(X_n \in A) \rightarrow \mathcal{P}(X \in A)$ para todo conjunto A X -continuo.

Supongamos ahora $h : S \rightarrow S'$ medible y sea D_h el conjunto de los puntos de discontinuidad. Si X tiene ley P , entonces $h(X)$ tiene ley Ph^{-1} . El teorema del mapeo se transforma en:

Teorema 2.3.6. $X_n \Rightarrow X$ implica $h(X_n) \Rightarrow h(X)$ si $\mathcal{P}(X \in D_h) = 0$.

2.4. Teorema de Prohorov

Dedicaremos esta sección a enunciar y demostrar este importante teorema de convergencia. Para ello daremos antes unas nociones necesarias.

La noción siguiente juega un papel fundamental en la teoría de la convergencia débil y sus aplicaciones:

Definición 2.4.1. Una medida de probabilidad P en (S, \mathcal{B}) es **ajustada** si para todo ϵ existe un conjunto compacto $K(\epsilon)$ tal que $P(K) > 1 - \epsilon$.

Antes de estudiar el siguiente teorema, recordemos que un espacio métrico es *separable* si incluye un subconjunto denso numerable. A su vez un conjunto es *denso* en un espacio métrico si su adherencia es el total. También necesitamos recordar que un espacio métrico es completo si toda sucesión de Cauchy contenida en éste es convergente. Tenemos entonces:

Teorema 2.4.2. *Si el espacio métrico S es separable y completo, entonces toda medida de probabilidad en (S, \mathcal{B}) es ajustada.*

Observación 2.4.3. El teorema nos dice que cada medida de probabilidad en el espacio euclideo k -dimensional con la métrica ordinaria $(\mathbb{R}^k, \mathcal{R}^k)$ es ajustada. Lo mismo ocurre en el espacio de las funciones continuas $(C[0, 1], \mathcal{C})$.

El concepto de medida de probabilidad ajustada se puede generalizar a una familia de medidas de probabilidad. Veamos:

Definición 2.4.4. *Una familia M de medidas de probabilidad en (S, \mathcal{B}) es **ajustada** si para cada ϵ existe un compacto $K(\epsilon)$ tal que $P(K) > 1 - \epsilon$ para cada $P \in M$.*

Observación 2.4.5. Notemos que en esta definición, el compacto K depende de ϵ , pero no de los elementos P de la familia M .

Definición 2.4.6. *Sea M una familia de medidas de probabilidad en (S, \mathcal{B}) . Decimos que M es **relativamente compacta** si toda sucesión de elementos de M contiene una subsucesión débilmente convergente. Es decir, si para cada sucesión $\{P_n\}$ en M , existe una subsucesión $\{P_{n_i}\}$ y una medida de probabilidad Q (definida en (S, \mathcal{B}) , pero no necesariamente un elemento de M), tal que $P_{n_i} \Rightarrow_i Q$.*

Lo que realmente nos interesa es la compacidad relativa de sucesiones $\{P_n\}$. Una sucesión $\{P_n\}$ es relativamente compacta si toda subsucesión $\{P_{n_i}\}$ contiene a su vez una subsucesión $\{P_{n_{i_m}}\}$ tal que $\{P_{n_{i_m}}\} \Rightarrow_m Q$ donde Q es una medida de probabilidad.

Ejemplo 2.4.7. Vamos a dar ahora dos resultados sobre compacidad relativa en el espacio de las funciones continuas. Sean P_n y P dos medidas de probabilidad en (C, \mathcal{C}) , que recordamos que es el espacio de las funciones continuas con la topología uniforme.

Si $\{P_n\}$ es relativamente compacta y las distribuciones de dimensión finita de P_n convergen débilmente a las de P : $P_n \pi_{t_1, \dots, t_k}^{-1} \Rightarrow_n P \pi_{t_1, \dots, t_k}^{-1}$ para todo k y para todo t_1, \dots, t_k . Entonces $P_n \Rightarrow_n P$.

Esta idea nos da un método potente para probar la convergencia débil en (C, \mathcal{C}) y otros espacios de funciones. Suponiendo como antes que P_n y P son medidas de probabilidad en (C, \mathcal{C})

Si $\{P_n\}$ es relativamente compacta y $P_n \pi_{t_1, \dots, t_k}^{-1} \Rightarrow_n \mu_{t_1, \dots, t_k}$ donde μ_{t_1, \dots, t_k} es una medida de probabilidad en $(\mathbb{R}^k, \mathcal{R}^k)$, entonces P cumple $P \pi_{t_1, \dots, t_k}^{-1} = \mu_{t_1, \dots, t_k}$ para todo t_1, \dots, t_k .

Sin más preámbulos, procedemos con el teorema de Prohorov.

Teorema 2.4.8. *Sea S un espacio completo y separable y sea M una familia de medidas de probabilidad definida en (S, \mathcal{B}) . Entonces M es ajustada si y sólo si es relativamente compacta.*

Demostración. Empezaremos demostrando la implicación hacia la izquierda, ya que es significativamente más corta y sencilla.

\Leftarrow) Consideremos la sucesión de conjuntos abiertos G_n crecientes ($G_{n-1} \subset G_n$) con $G_n \nearrow S$. Para cada ϵ existe una n tal que $P(G_n) > 1 - \epsilon$ para toda medida de probabilidad $P \in M$, ya que en caso contrario, para cada n tendríamos $P_n(G_n) \leq 1 - \epsilon$ para algún $P_n \in M$, y por compacidad relativa $P_{n_i} \Rightarrow_i Q$ para alguna subsucesión y para alguna medida de probabilidad Q , lo cual es imposible porque en ese caso $QG_n \leq \liminf_i P_{n_i}G_n \leq P_{n_i}G_{n_i} \leq 1 - \epsilon$, cuando $G_n \nearrow S$.

Esto implica que si A_{k_1}, A_{k_2}, \dots es una sucesión de bolas abiertas de radio $1/k$ recubriendo S , entonces (por separabilidad) existe un n_k tal que $P(\cup_{i \leq n_k} A_{k_i}) > 1 - \frac{\epsilon}{2^k}$ para toda P en M . Si K es la clausura del conjunto acotado $\cap_{k \geq 1} \cup_{i \leq n_k} A_{k_i}$, entonces (por ser S completo) K es compacto y $P(K) > 1 - \epsilon$ para todo P en M .

\Rightarrow) Antes de empezar con la demostración, necesitamos dar una definición previa.

Definición. Una *medida exterior* es una función γ definida para todos los subconjuntos del espacio Ω , cumpliendo las siguientes cuatro propiedades:

- 1) $\gamma(A) \in [0, \infty]$ para todo $A \subset \Omega$
- 2) $\gamma(\emptyset) = 0$
- 3) γ es monótona, i.e. $A \subset B \Rightarrow \gamma(A) \leq \gamma(B)$
- 4) $\gamma(\cup_n A_n) \leq \sum_n \gamma(A_n)$

Procedemos ahora con la demostración. Supongamos que $\{P_n\}$ es una sucesión en la familia ajustada M . Tenemos que encontrar una subsucesión $\{P_{n_i}\}$ y una medida de probabilidad P tal que $P_{n_i} \Rightarrow_i P$. Lo haremos por construcción.

Elegimos conjuntos compactos K_u de manera que $K_1 \subset K_2 \subset \dots$ y $P_n(K_u) > 1 - 1/u$ para todo u y n . El conjunto $\cup_u K_u$ es separable. Entonces, existe $\mathcal{A} \in \mathcal{B}$ numerable de conjuntos abiertos que cumplen que si $x \in \cup_u K_u \cap G$ y G es abierto, entonces $x \in A \subset \bar{A} \subset G$ para algún $A \in \mathcal{A}$. Sea \mathcal{H} la unión de \emptyset y las uniones finitas de conjuntos de la forma $\bar{A} \cap K_u$ para $A \in \mathcal{A}$ y $u \geq 1$.

Escogemos la subsucesión $\{P_{n_i}\} \subset \{P_n\}$ en la cual el límite $\alpha(H) := \lim_i P_{n_i}H$ existe para cada $H \in \mathcal{H}$. El objetivo es construir una medida de probabilidad $P \in \mathcal{B}$ tal que $P(G) = \sup_{H \subset G} \alpha(H)$ para todos los conjuntos abiertos G . Si P existe entonces la demostración acaba: Si $H \subset G$, entonces $\alpha(H) = \lim_i P_{n_i}(H) \leq \liminf_i P_{n_i}G$, con lo cual $P(G) \leq \liminf_i P_{n_i}(G)$ y por tanto $P_{n_i} \Rightarrow_i P$

Para construir nuestra P , observamos que \mathcal{H} es cerrada por uniones finitas y que $\alpha(H)$ cumple las siguientes propiedades:

- $\alpha(H_1) \leq \alpha(H_2)$ si $H_1 \subset H_2$
- $\alpha(H_1 \cup H_2) = \alpha(H_1) + \alpha(H_2)$ si $H_1 \cap H_2 = \emptyset$
- $\alpha(H_1 \cup H_2) \leq \alpha(H_1) + \alpha(H_2)$
- $\alpha(\emptyset) = 0$

Definimos $\beta(G) := \sup_{H \subset G} \alpha(H)$ para G abierto. Entonces β es monótona y $\beta(\emptyset) = \alpha(\emptyset) = 0$.

Definimos también $\gamma(D) := \inf_{D \subset G} \beta(G)$ para un conjunto arbitrario $D \subset S$.

Observación. $\gamma(G) = \beta(G)$ para G abierto.

Supongamos que γ es una medida exterior. Recordemos que D es γ -medible si $\gamma(D) \geq \gamma(D \cap L) + \gamma(D^c \cap L)$ para todo $L \subset S$, que la clase \mathcal{M} de conjuntos γ -medibles es una σ -álgebra y que la restricción de γ a \mathcal{M} es una medida. Supongamos también que cada conjunto cerrado pertenece a \mathcal{M} . Entonces $\mathcal{B} \subset \mathcal{M}$ y que la restricción $\gamma|_{\mathcal{B}} = P$ es una medida que satisface $P(G) = \gamma(G) = \beta(G)$, con lo cual $P(G) = \sup_{H \subset G} \alpha(H)$ se cumple para G abierto. P será una medida de probabilidad porque

$$\geq P(S) = \beta(S) \geq \sup_u \alpha(K_u) \geq \sup_u (1 - 1/u) = 1 \quad (2.4.1)$$

Observación. Cada $K_u \in \mathcal{H}$, y tiene un recubrimiento finito de \mathcal{A} -conjuntos.

Procedemos, ahora sí, con la construcción.

- Si $F \subset G$ donde F es cerrado y G abierto, y si $F \subset H$ para algún $H \in \mathcal{H}$, entonces $F \subset H_0 \subset G$ para algún $H_0 \in \mathcal{H}$.

Veamos esto; sea para cada $x \in F$, $A_x \in \mathcal{A}$ tal que $x \in A_x \subset \overline{A_x} \subset G$. Los A_x forman un recubrimiento de F y como F es compacto (por ser subconjunto de H), entonces existe un subrecubrimiento finito A_{x_1}, \dots, A_{x_k} . Como $F \subset K_u$ para algún u , podemos tomar $H_0 = \bigcup_{i=1}^k (\overline{A_{x_i}} \cap K_u)$.

- β es finitamente subaditiva en los conjuntos abiertos.

Supongamos que $H \subset G_1 \subset G_2$, donde $H \in \mathcal{H}$ y G_1, G_2 son dos abiertos. Definimos

$$\begin{aligned} F_1 &= \{x \in H; \rho(x, G_1^c) \geq \rho(x, G_2^c)\} \\ F_2 &= \{x \in H; \rho(x, G_2^c) \geq \rho(x, G_1^c)\} \end{aligned}$$

Supongamos $x \in F_1$ y $x \notin G_1$, entonces $x \in G_2$, así que, como G_2^c es cerrado, $\rho(x, G_1^c) = 0 < \rho(x, G_2^c)$ que es una contradicción. Por tanto $F_1 \subset G_1$ y de

manera análoga $F_2 \subset G_2$. Como $F_1 \subset H$ y $H \in \mathcal{H}$, por el punto anterior $F_1 \subset H_1 \subset G_1$ para algún $H_1 \in \mathcal{H}$. De la misma forma, sacamos $F_2 \subset H_2 \subset G_2$ para algún $H_2 \in \mathcal{H}$. Por las propiedades antes nombradas $\alpha(H) \leq \alpha(H_1 \cup H_2) \leq \alpha(H_1) + \alpha(H_2) \leq \beta(G_1) + \beta(G_2)$. Como esto funciona para toda $H \in G_1 \cup G_2$ entonces

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

- β es numerablemente subaditiva en los conjuntos abiertos.

Si $H \subset \bigcup_n G_n$, entonces como H es compacto, $H \subset \bigcup_{n \leq n_0} G_n$ para algún n_0 , y la subaditividad finita implica $\alpha(H) \leq \beta(\bigcup_{n \leq n_0} G_n) \leq \sum_{n \leq n_0} \beta(G_n) \leq \sum_n \beta(G_n)$. Tomando el supremo de H contenido en $\bigcup_n G_n$ obtenemos

$$\beta\left(\bigcup_n G_n\right) \leq \sum_n \beta(G_n)$$

- γ es una medida exterior.

Como γ es monótona y cumple $\gamma(\emptyset) = 0$, solo necesitamos probar que es numerablemente subaditiva. Dado un $\epsilon > 0$ y los subconjuntos arbitrarios $D_n \in S$, escogemos conjuntos abiertos G_n tal que $D_n \subset G_n$ y $\beta(G_n) < \gamma(D_n) + \epsilon/2^n$. Como β es numerablemente subaditiva tenemos $\gamma(\bigcup_n D_n) \leq \beta(\bigcup_n G_n) < \sum_n \gamma(D_n) + \epsilon$. Como ϵ es arbitrario, podemos escogerlo de manera que

$$\gamma\left(\bigcup_n D_n\right) \leq \sum_n \gamma(D_n)$$

- $\beta(G) \geq \gamma(F \cap G) + \gamma(F^c \cap G)$ para F cerrado y G abierto.

Para un ϵ dado, cogemos $H_1 \in \mathcal{H}$ para el cual se cumple $H_1 \subset (F^c \cap G)$ y $\alpha(H_1) > \beta(F^c \cap G) - \epsilon$. Escogemos ahora $H_0 \in \mathcal{H}$ tal que $H_0 \subset H_1^c \cap G$ y $\alpha(H_0) > \beta(H_1^c \cap G) - \epsilon$. Como H_0 y H_1 son disjuntos y ambos están contenidos en G , por las propiedades antes dadas de α , tenemos

$$\beta(G) \geq \alpha(H_0 \cup H_1) = \alpha(H_0) + \alpha(H_1) > \beta(H_1^c \cap G) + \beta(F^c \cap G) - 2\epsilon \geq \gamma(F \cap G) + \gamma(F^c \cap G) - 2\epsilon.$$

Como ϵ es arbitrario, podemos escogerlo tal que

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

- $F \in \mathcal{D}$ si F es un conjunto cerrado.

Por la desigualdad del punto anterior $\beta(G) \geq \gamma(F \cap T) + \gamma(F^c \cap T)$ si G es abierto y $T \subset G$. Tomando el ínfimo de este G obtenemos que F es γ -medible. Esto completa nuestra construcción. \square

Corolario 2.4.9. Si $\{P_n\}$ es ajustada y cada subsucesión que converge débilmente, lo hace a P , entonces toda la sucesión converge débilmente a P : $P_n \Rightarrow P$

Capítulo 3

Teorema de Donsker

El tercer capítulo trata de relacionar el primero y el segundo en un resultado muy importante: el teorema de Donsker. Se estudia la construcción de la medida de Wiener en el espacio de las funciones continuas $C([0, 1])$, y una vez tenemos el fundamento hecho, demostramos el teorema de Donsker, el cual nos proporciona el resultado que buscamos a lo largo de este trabajo: obtener el movimiento browniano como límite del paseo aleatorio, usando la convergencia débil en el espacio de las funciones continuas.

3.1. Introducción

Aunque ya hemos introducido este tema con ejemplos en secciones previas, el espacio C es muy importante en todo lo que sigue. De hecho todos los resultados siguientes pertenecen a este espacio. Por ello, además de lo ya explicado, introduciremos algunos resultados que necesitaremos para más adelante tratar con propiedad las medidas de Wiener y el teorema de Donsker.

Recordemos que el espacio $C = C[0, 1]$ es el llamado espacio de las funciones continuas en el intervalo unidad, donde damos al espacio C la llamada topología uniforme, que define la distancia entre dos puntos x, y (dos funciones continuas $x(\cdot), y(\cdot)$ en $[0, 1]$) como

$$\rho(x, y) = \|x - y\| = \sup_t |x(t) - y(t)|.$$

Teorema 3.1.1. *Sean P_n, P dos medidas de probabilidad en (C, \mathcal{C}) . Si las distribuciones finitas de P_n convergen débilmente a P y si $\{P_n\}$ es ajustada, entonces $P_n \Rightarrow P$.*

Definición 3.1.2. *El **módulo de continuidad** de una función arbitraria $x(\cdot)$ en $[0, 1]$ viene definido por*

$$\omega_x(\delta) = \omega(x, \delta) = \sup_{|s-t| \leq \delta} |x(s) - x(t)|, \quad 0 < \delta \leq 1$$

Una condición necesaria y suficiente para asegurar la continuidad uniforme de x es $\lim_{\delta \rightarrow 0} \omega_x(\delta) = 0$

Funciones aleatorias

Sea $(\Omega, \mathcal{F}, \mathcal{P})$ un espacio de probabilidad y sea $X : \Omega \rightarrow C$. $X(\omega)$ es un elemento de C con valor $X_t(\omega) = X(t, \omega)$. Para un t fijado, sea $X_t = X(t)$ la función real en Ω con valor $X_t(\omega)$ en ω . Entonces, X_t es la composición $\pi_t X$, donde π_t es la proyección natural. Siguiendo esta línea, $(X_{t_1}(\omega), \dots, X_{t_k}(\omega))$ es la aplicación que envía ω a $(X_{t_1}(\omega), \dots, X_{t_k}(\omega)) = \pi_{t_1 \dots t_k}(X(\omega))$ en \mathbb{R}^k .

Si X es una función aleatoria (es decir, si es medible), entonces la composición $\pi_{t_1 \dots t_k} X$ es medible, así que $(X_{t_1}, \dots, X_{t_k})$ es un vector aleatorio. A su vez, un conjunto finito general tiene la forma $A = \pi_{t_1 \dots t_k}^{-1} H$, $H \in \mathcal{R}^k$. Si $\pi_{t_1 \dots t_k} X$ es medible, entonces $X^{-1}A = (\pi_{t_1 \dots t_k} H)^{-1} \in \mathcal{F}$. Como la clase de los conjuntos finitos genera \mathcal{C} , entonces X es medible. Obtenemos que X es una función aleatoria si y solo si cada $(X_{t_1}, \dots, X_{t_k})$ es un vector aleatorio.

Sea $P = \mathcal{P}X^{-1}$ la función de distribución de X . Entonces, $\mathcal{P}((X_{t_1}, \dots, X_{t_k}) \in H) = P\pi_{t_1 \dots t_k}^{-1} H$.

Daremos ahora un teorema que sobre condiciones de ajuste en C que nos será muy útil en lo que sigue. Sean P_n medidas de probabilidad en (C, \mathcal{C}) , y x una función continua.

Teorema 3.1.3. *La sucesión $\{P_n\}$ es ajustada si y solo si se cumplen las dos siguientes condiciones:*

i) *Para cada η positivo, existe una a y un n_0 tal que,*

$$P_n(x : |x(0)| \geq a) \leq \eta, \quad n \geq n_0 \quad (3.1.1)$$

ii) *Para cada positivo ϵ y η , existe un $\delta, 0 < \delta < 1$, y un n_0 tal que*

$$P_n(x : \omega_x(\delta) \geq \epsilon) \leq \eta, \quad n \geq n_0 \quad (3.1.2)$$

Observación. Observemos que la segunda condición puede expresarse como

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} P_n(x : \omega_x(\delta) \geq \epsilon) = 0. \quad (3.1.3)$$

El teorema siguiente será muy importante porque es la base de la demostración del teorema de Donsker. Supongamos que $X, X^2, X^3 \dots$ son funciones aleatorias.

Teorema 3.1.4. *Si*

$$(X_{t_1}^n, \dots, X_{t_k}^n) \Rightarrow_n (X_{t_1}, \dots, X_{t_k}) \quad (3.1.4)$$

se cumple para todo t_1, \dots, t_k , y si

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathcal{P}(\omega(X^n, \delta) \geq \epsilon) = 0 \quad (3.1.5)$$

para cada $\epsilon > 0$, entonces $X^n \Rightarrow_n X$.

Demostración. Sean P y P_n las distribuciones de X y X_n . Podemos reescribir (3.1.4) como $P_n \pi_{t_1 \dots t_k}^{-1} \Rightarrow_n P \pi_{t_1 \dots t_k}^{-1}$, y (3.1.5) como $P_n \Rightarrow_n P$, ya que son equivalentes. Así pues, el teorema (3.1.1) se cumplirá si demostramos que $\{X^n\}$ es ajustada, en el sentido de que $\{P_n\}$ es ajustada. Para esto usaremos el teorema (3.1.3). Vemos que $X_0^n \Rightarrow_n X_0$ implica que $\{P_n \pi_0^{-1}\}$ es ajustada, lo cual nos da la condición *i*) del lema. Por otra parte, como (3.1.5) se puede traducir en (3.1.3), la segunda condición del lema se cumple, y por tanto P_n es ajustada, que es lo que queríamos ver. \square

3.2. Medidas de Wiener

La medida de Wiener es una medida de probabilidad W en el espacio de funciones continuas inducidas por el proceso de Wiener o movimiento browniano. Una integral que usa la medida de Wiener se denomina integral de Wiener. Nos centraremos en la construcción de esta medida, que describe la distribución de probabilidad de un camino trazado por una partícula en movimiento browniano.

La *medida de Wiener*, denotada por W , es una medida de probabilidad en (C, \mathcal{C}) que cumple las siguientes propiedades:

- i) Cada X_t tiene una distribución normal respecto W , con media 0 y varianza t , i.e.

$$W(X_t \leq \alpha) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\alpha} e^{-\frac{u^2}{2t}} du \quad (3.2.1)$$

- Para $t = 0$ tenemos $W(x_0 = 0) = 1$

- ii) El proceso estocástico $\{X_t; 0 \leq t \leq 1\}$ tiene incrementos estacionarios respecto W , i.e, si

$$0 \leq t_0 \leq t_1 \leq \dots \leq t_k = 1$$

entonces

$$X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \dots, X_{t_k} - X_{t_{k-1}}$$

son independientes respecto W .

Si W cumple estas dos premisas, y dado $s \leq t$, entonces X_t es la suma de las v.a. independientes X_s y $X_t - X_s$, así que $X_t - X_s$ tiene una distribución normal con media 0 y varianza $t - s$. Entonces, para $0 \leq t_0 \leq t_1 \leq \dots \leq t_k = 1$,

$$W(X_{t_i} - X_{t_{i-1}} \leq \alpha_i, i = 1, \dots, k) = \prod_{i=1}^k \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \int_{-\infty}^{\alpha_i} e^{\frac{-u^2}{2(t_i - t_{i-1})}} du \quad (3.2.2)$$

Es decir, los incrementos son estacionarios e independientes. Esta expresión es la forma más clara de explicitar las distribuciones finitas de esta medida de probabilidad.

Además, como $X_s X_t = X_s^2 + X_s(X_t - X_s)$, por la independencia de los incrementos, la $\text{Cov}(X_s, X_t) = s$ si $s \leq t$. Con un cambio de variables lineal la distribución conjunta $(X_{t_1}, \dots, X_{t_k})$ es una distribución normal centrada con $\text{Cov}(X_{t_i}, X_{t_j}) = t_i \wedge t_j$.

Con todo esto, podemos ver que la medida de Wiener es la ley del movimiento browniano. Ahora vamos a probar la existencia de esta medida por construcción.

Observación 3.2.1. En este capítulo usamos la nomenclatura W para referirnos a la medida de Wiener, y a la función aleatoria que la tiene por función de distribución. Esta W es equivalente a la ley del movimiento browniano B que hemos tratado en capítulos anteriores.

Construcción de la medida de Wiener

Antes de empezar con la construcción, daremos un teorema previo, que necesitaremos para demostrar la existencia de esta medida.

Teorema 3.2.2. *Supongamos que $0 = t_0 < t_1 < \dots < t_v = 1$ y que*

$$\min_{1 \leq i < v} (t_i - t_{i-1}) \geq \delta. \quad (3.2.3)$$

Entonces, para una x arbitraria,

$$\omega_x(\delta) \leq 3 \max_{1 \leq i \leq v} \sup_{t_{i-1} \leq s \leq t_i} |x(s) - x(t_{i-1})|, \quad (3.2.4)$$

y, para una P arbitraria

$$P(x : \omega_x(\delta) \geq 3\epsilon) \leq \sum_{i=1}^v P(x : \sup_{t_{i-1} \leq s \leq t_i} |x(s) - x(t_{i-1})| \geq \epsilon). \quad (3.2.5)$$

Observación 3.2.3. Observemos que (3.2.3) no necesita que $t_i - t_{i-1} \geq \delta$ para $i = 1$ o $i = v$.

Ahora procedemos con el teorema que nos da la existencia de la medida de Wiener. Es en la demostración de este teorema donde la construiremos.

Teorema 3.2.4. *Existe en (C, \mathcal{C}) una medida de probabilidad W cuyas distribuciones de dimensión finita vienen dadas por (3.2.2)*

Demostración. Para demostrar la existencia procedemos a construir la medida de Wiener.

Empezamos con una secuencia ξ_1, ξ_2, \dots de variables aleatorias independientes idénticamente distribuidas en un espacio de probabilidad, con media 0 y covarianza $\sigma^2 > 0$. Sea $S_n = \xi_1 + \dots + \xi_n$, y por convenio $S_0 = 0$. Sea $X^n(\omega)$ la función continua

$$X_t^n(\omega) = \frac{1}{\sigma\sqrt{n}}S_{[nt]}(\omega) + (nt - [nt])\frac{1}{\sigma\sqrt{n}}\xi_{[nt]+1}(\omega) \quad (3.2.6)$$

La función $X^n(\omega)$ es la función definida por la interpolación lineal entre los valores $X_{i/n}^n(\omega) = S_i(\omega)/\sigma\sqrt{n}$ en los puntos i/n . X_t^n define una variable aleatoria para cada t , así que X^n es una función aleatoria.

Sea

$$\psi_{n,t} = (nt - [nt])\frac{1}{\sigma\sqrt{n}}\xi_{[nt]+1}(\omega) \quad (3.2.7)$$

Tenemos que $\psi_{n,t} \Rightarrow 0$ por la desigualdad de Chebyshev. Usando el teorema central del límite, podemos llegar a $X_t^n \Rightarrow_n \sqrt{t}N(0, 1)$. De manera similar podemos obtener que si $s \leq t$

$$(X_s^n, X_t^n - X_s^n) = \frac{1}{\sigma\sqrt{n}}(S_{[ns]}, S_{[nt]} - X_{[ns]}) + (\psi_{n,s}, \psi_{n,t} - \psi_{n,s}) \Rightarrow_n (N_1, N_2) \quad (3.2.8)$$

con N_1, N_2 funciones de distribución normales independientes con varianzas s , y $t - s$. Por el teorema del mapeado,

$$(X_s^n, X_t^n) \Rightarrow_n (N_1, N_1 + N_2) \quad (3.2.9)$$

. La extensión trivial nos muestra que las distribuciones límite de los vectores aleatorios $(X_{t_1}^n, \dots, X_{t_k}^n)$ son exactamente aquellas especificadas para las distribuciones finitas de la medida W que estamos construyendo. Dicho de otra forma, si P_n es la distribución de X^n en C , entonces para cada t_1, \dots, t_k ,

$$P_n \pi_{t_1, \dots, t_k}^{-1} \Rightarrow W \pi_{t_1, \dots, t_k}^{-1}. \quad (3.2.10)$$

Supongamos ahora que $\{P_n\}$ es ajustada. Por el teorema de Prohorov podemos afirmar que alguna subsucesión $\{P_{n_i}\}$ convergerá débilmente a un límite que podemos llamar W . En este caso,

$$P_{n_i} \pi_{t_1, \dots, t_k}^{-1} \Rightarrow_i W \pi_{t_1, \dots, t_k}^{-1} \quad (3.2.11)$$

y por tanto, con esto ya hemos probado que $W \pi_{t_1, \dots, t_k}^{-1}$ es la medida de probabilidad en \mathbb{R}^k que queríamos, y esto probaría lo que queremos demostrar. Por tanto, nos

falta ver que efectivamente $\{P_n\}$ es ajustada.

Para ver esto necesitamos el lema siguiente que demostraremos.

Lema 3.2.5. *Sea X^n definido por (3.2.6) de manera que $\{\xi_n\}$ es estacionario, i.e la distribución de $(\xi_k, \dots, \xi_{k+j})$ es la misma para toda k , y que*

$$\lim_{\lambda \rightarrow \infty} \limsup_{n \rightarrow \infty} \lambda^2 \mathcal{P}(\max_{k \leq n} |S_k| \geq \lambda \sigma \sqrt{n}) = 0 \quad (3.2.12)$$

Entonces $\{X^n\}$ es ajustada.

Demostración. Usaremos el teorema (3.1.3). Como $X_0^n = 0$, entonces $\{P_n\}$ satisface la primera hipótesis del teorema. Para ver que se cumple la segunda hipótesis, veremos que se cumple (3.1.3). Esta condición es equivalente a

$$\lim_{\lambda \rightarrow \infty} \limsup \mathcal{P}(\omega(X^n, \delta) \geq \epsilon) = 0 \quad (3.2.13)$$

para cada ϵ . Para ver que esto se cumple, usaremos el teorema (3.2.2). Si se cumple (3.2.3), entonces también se cumplen (3.2.4) y (3.2.5). Por tanto,

$$\mathcal{P}(\omega(X^n, \delta) \geq 3\epsilon) \leq \sum_{i=1}^v \mathcal{P}(\sup_{t_{i-1} \leq s \leq t_i} |X_s^n - X_{t_{i-1}}^n| \geq \epsilon), \quad (3.2.14)$$

si $\min_{1 < i < v} (t_i - t_{i-1}) \geq \delta$.

Si tomamos $t_i = m_i/n$ con n y $m_i, i = 0, \dots, v$ enteros que cumplan $0 = m_0 < m_1 < \dots < m_v = n$, la expresión (3.2.14) es más sencilla de analizar. Esto es porque debido al carácter poligonal de la función X^n , si t_i toma la forma que hemos explicitado, entonces el supremo en (3.2.14) se convierte en un máximo de las diferencias $|S_k - S_{m_{i-1}}| / \sigma \sqrt{n}$ como sigue:

$$\begin{aligned} \mathcal{P}(\omega(X^n, \delta) \geq 3\epsilon) &\leq \sum_{i=1}^v \mathcal{P}\left(\max_{m_{i-1} \leq k \leq m_i} \frac{|S_k - S_{m_{i-1}}|}{\sigma \sqrt{n}} \geq \epsilon\right) \\ &= \sum_{i=1}^v \mathcal{P}\left(\max_{k \leq m_i - m_{i-1}} |S_k| \geq \epsilon \sigma \sqrt{n}\right), \end{aligned} \quad (3.2.15)$$

donde la última igualdad se debe a que hemos asumido estacionariedad. La desigualdad se cumple si se cumple (3.2.14), para lo cual necesitamos

$$\frac{m_i}{n} - \frac{m_{i-1}}{n} \geq \delta, \quad 1 < i < v \quad (3.2.16)$$

Para simplificar aún más, tomamos $m_i = im$ para $0 \leq i < v$ (y $m_v = n$) donde m es un entero (que depende de n y de δ), escogido de acuerdo al siguiente criterio:

como necesitamos $m_i - m_{i-1} = m \geq n\delta$ para $i < v$, tomamos $m = [n\delta]$. Como también necesitamos $(v-1)m < n \leq vm$, tomamos $v = [n/m]$. Entonces,

$$m_v - m_{v-1} \leq m, \quad v = [n/m] \rightarrow_n \frac{1}{\delta} < \frac{2}{\delta}, \quad \frac{n}{m} \rightarrow_n \frac{1}{\delta} > \frac{1}{2\delta} \quad (3.2.17)$$

y tomando (3.2.15) para n grande tenemos,

$$\mathcal{P}(\omega(X^n, \delta) \geq 3\epsilon) \leq v \cdot \mathcal{P}\left(\max_{k \leq m} |S_k| \geq \epsilon\sigma\sqrt{n}\right) \leq \frac{2}{\delta} \mathcal{P}\left(\max_{k \leq m} |S_k| \geq \frac{\epsilon}{\sqrt{2\delta}}\sigma\sqrt{m}\right) \quad (3.2.18)$$

Tomamos λ y δ de manera que $\lambda = \frac{\epsilon}{\sqrt{2\delta}}$. En términos de λ , (3.2.18) queda de la forma:

$$\mathcal{P}(\omega(X^n, \delta) \geq 3\epsilon) \leq \frac{4\lambda^2}{\epsilon^2} \cdot \mathcal{P}\left(\max_{k \leq m} |S_k| \geq \lambda\sigma\sqrt{m}\right). \quad (3.2.19)$$

Para un ϵ y una η positivos dados, usando la hipótesis del lema (3.2.12), tenemos una λ tal que

$$\frac{4\lambda^2}{\epsilon^2} \limsup_m \mathcal{P}\left(\max_{k \leq m} |S_k| \geq \lambda\sigma\sqrt{m}\right) < \eta. \quad (3.2.20)$$

Una vez λ y δ están fijados, m va a infinito junto con n , y entonces (3.2.13) se cumple. □

Para completar la construcción de la medida de Wiener podemos usar la independencia de las v.a. ξ_n . Por esta independencia, la desigualdad de Etemadi [2] implica que

$$\mathcal{P}\left(\max_{u \leq m} |S_u| \geq \alpha\right) \leq 3 \max_{u \leq m} \mathcal{P}\left(|S_u| \geq \alpha/3\right) \quad (3.2.21)$$

y por tanto, la segunda hipótesis del lema se cumple si

$$\lim_{\lambda \rightarrow \infty} \limsup_{n \rightarrow \infty} \lambda^2 \max_{k \leq n} \mathcal{P}\left(|S_k| \geq \lambda\sigma\sqrt{n}\right) = 0. \quad (3.2.22)$$

Para nuestra construcción de la medida de Wiener podemos usar cualquier secuencia $\{\xi_i\}$ conveniente, por tanto supongamos que ξ_i son independientes con función de distribución normal ($\sigma = 1$), con lo cual S_k/\sqrt{k} tiene una distribución

normal estándar. Como

$$\mathcal{P}(|N| \geq \lambda) < E[N^4 \lambda^{-4}] = 3\lambda^{-4} \quad (3.2.23)$$

tenemos que $\mathcal{P}(|S_k| \geq \lambda \sigma \sqrt{n}) = \mathcal{P}(\sqrt{k} |N| \geq \lambda \sigma \sqrt{n}) < 3/\lambda^4 \sigma^4$ para $k \leq n$, lo cual implica directamente la condición que necesitábamos.

Con esto, tenemos que $\{X_n\}$ es ajustada, y por tanto hemos probado la existencia de la medida de Wiener. \square

3.3. Teorema de Donsker

Introducción

El teorema de Donsker se puede entender como una extensión del teorema central del límite.

Como hemos visto al inicio del capítulo dos, sean ξ_1, \dots, ξ_n una sucesión de v.a.i.i.d. definidas en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathcal{P})$. Si $\xi_i, i = 1, \dots, n$ tiene media 0 y varianza σ^2 , el teorema central del límite de Lindenberg-Lévy nos dice que la distribución de la suma normalizada

$$\frac{1}{\sigma \sqrt{n}} S_n = \frac{1}{\sigma \sqrt{n}} (\xi_1 + \dots + \xi_n) \quad (3.3.1)$$

converge débilmente ($n \rightarrow \infty$) a $N(0, 1)$.

Podemos formular un refinamiento o extensión demostrando la convergencia débil de las distribuciones de ciertas funciones aleatorias construidas a partir de las sumas parciales S_n . Para cada entero n y cada punto de la muestra ω , construimos en una unidad de intervalo la función poligonal, que es lineal en cada uno de los subintervalos $[\frac{i-1}{n}, \frac{i}{n}]$, $i = 1, \dots, n$ y tiene valor $\frac{S_i(\omega)}{\sigma \sqrt{n}}$ en el punto i/n . $S_0(\omega) = 0$ por convenio. En otras palabras, construimos la función $X^n(\omega)$ cuyo valor en un punto $t \in [0, 1]$ es

$$X_t^n(\omega) = \frac{1}{\sigma \sqrt{n}} S_{i-1}(\omega) + \frac{t - (i-1)/n}{1/n} \frac{1}{\sigma \sqrt{n}} \xi_i(\omega), \quad t \in \left[\frac{i-1}{n}, \frac{i}{n}\right].$$

Observación 3.3.1. Esta fórmula es equivalente a la (3.2.6) dada anteriormente.

Para cada ω , $X^n(\omega)$ es un elemento del espacio C , es decir, una función aleatoria. Sea P_n la distribución de probabilidad de $X^n(\omega)$ en C . El teorema de Donsker nos dice entonces que

$$P_n \Rightarrow W$$

donde W es la medida de Wiener, y como hemos dicho ya anteriormente, es la

distribución de probabilidad de una partícula en movimiento browniano.

El proceso estocástico S_n puede ser interpretado como la posición n en un paseo aleatorio. El teorema central del límite nos dice que esta posición, normalizada por \sqrt{n} ($\sigma = 1$), está, para $n \rightarrow \infty$ aproximadamente distribuido como la posición a tiempo $t = 1$ de una partícula en movimiento browniano. Por su parte, el teorema de Donsker nos dice que todo el camino del paseo aleatorio durante los n primeros pasos está, para $n \rightarrow \infty$, aproximadamente distribuido como el camino a tiempo $t = 1$ de una partícula en movimiento browniano.

Teorema de Donsker

Teorema 3.3.2. *Si ξ_1, ξ_2, \dots son v.a.i.i.d. con media 0 y varianza σ^2 , y si X^n es la función aleatoria definida por (3.2.6). Entonces $X^n \Rightarrow_n W$.*

Demostración. Basaremos la demostración del teorema de Donsker en el teorema 3.1.4. Veamos que se cumplen las dos hipótesis:

- (i) Como la existencia de la medida de Wiener ha quedado establecida, y también su correspondiente función aleatoria W , podemos reescribir (3.2.8) como

$$(X_s^n, X_t^n - X_s^n) \Rightarrow_n (W_s, W_t - W_s) \quad (3.3.2)$$

lo cual implica $(X_s^n, X_t^n) \Rightarrow_n (W_s, W_t)$. Una extensión simple nos da

$$(X_{t_1}^n, \dots, X_{t_k}^n) \Rightarrow_n (W_{t_1}, \dots, W_{t_k}) \quad (3.3.3)$$

- (ii) Procederemos considerando de manera separada los valores de k pequeños y grandes en el máximo de (3.2.22).

- Por el teorema central del límite, si k_λ es suficientemente pequeño, y $k_\lambda < k \leq n$, entonces, usando (3.2.23) tenemos

$$\mathcal{P}(|S_k| \geq \lambda\sigma\sqrt{n}) \leq \mathcal{P}(|S_k| \geq \lambda\sigma\sqrt{k}) < \frac{3}{\lambda^4}$$

- Si $k \leq k_\lambda$, podemos usar la desigualdad de Chebyshev y obtenemos

$$\mathcal{P}(|S_k| \geq \lambda\sigma\sqrt{n}) \leq \frac{k_\lambda}{\lambda^2 n}$$

El máximo de $\mathcal{P}(|S_k| \geq \lambda\sigma\sqrt{k})$ viene dado pues por $(3/\lambda^4) \vee (k_\lambda/\lambda^2 n)$ y se cumple entonces (3.2.22). Como hemos visto anteriormente en la construcción de la medida de Wiener, como los ξ_i son independientes, (3.2.22) implica (3.2.12), con lo cual el lema (3.2.5) nos dice que $\{X_n\}$ es ajustada. Sabemos que si $\{X_n\}$ es ajustada, entonces $\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathcal{P}(\omega(X^n, \delta) \geq \epsilon) = 0$ ([2]). Entonces, se cumple la segunda hipótesis.

Con todo, vemos que aplicando pues el teorema 3.1.4, $X^n \Rightarrow_n W$. □

Esta demostración puede encontrarse en [2].

Hemos construido pues un proceso de Wiener (o movimiento browniano) a partir de un proceso estocástico que cumplía las condiciones de un paseo aleatorio mediante la convergencia débil.

Bibliografía

- [1] Karatzas, Ioannis; Shreve, Steven E.: “*Brownian motion and stochastic calculus*”. Springer-Verlag New York Inc. (1988)
- [2] Billingsley, Patrick: “*Convergence of Probability Measures*”. John Wiley & Sons, Inc. New York (1999).
- [3] Pinsky, Marc A.; Karlin, Samuel: “*An introduction to stochastic modelling*”. Elsevier Inc. (2011)
- [4] Sanz i Solé, Marta: “*Probabilitats*”. Edicions de la Universitat de Barcelona (1999).
- [5] Nualart, David; Sanz i Solé, Marta: “*Curs de probabilitats*”. Poblagràfic S.A. Lleida (1990).
- [6] Billingsley, Patrick: “*Probability and Measure*”. John Wiley & Sons, Inc. New York (1995).
- [7] Sanz i Solé, Marta: “*An introduction to stochastic calculus*”. Universitat de Barcelona, 2017.
- [8] Sanz i Solé, Marta: “*An elementary course on stochastic processes*”. Universitat de Barcelona, 2008.
- [9] Corcuera, José Manuel. “*A course in stochastic processes*”. Universitat de Barcelona, 2017.
- [10] Donsker, M. “*An invariance principle for certain probability limit theorems*”. Mem. Amer. Math. Soc. 6(1951).