

De Arzelà–Ascoli a Donsker: Compacidad y apretamiento en la convergencia débil.

Este documento es una versión preliminar del Trabajo Fin de Grado presentado en el Grado en Matemáticas cursado en la UNED. Curso académico 2025/26.

Autor del trabajo: Javier López-de-Lacalle.

Supervisor del trabajo: Dr. Tomás Prieto Rumeau.

Resumen ejecutivo (Español)

Contexto y objetivo.

En el grado se estudian resultados de convergencia con la topología euclídea y posteriormente se extienden a espacios de funciones de dimensión infinita. Uno de los retos al determinar la convergencia de sucesiones en estos espacios es que la convergencia de las distribuciones finito dimensionales no implica convergencia. El Teorema de Arzelà-Ascoli lo resuelve caracterizando la precompacidad relativa de familias de funciones continuas. La versión probabilística de este resultado requiere el estudio de una propiedad nueva: apretamiento. Con ella se obtienen resultados de convergencia débil y se demuestra la versión funcional del Teorema Central del Límite. El objetivo de este trabajo es estudiar esta propiedad y proponer soluciones desde la topología y el análisis a las cuestiones que surjan durante su estudio.

Metodología.

En este trabajo se han aplicado resultados de Topología y Análisis Funcional.

Resultados principales.

Tras estudiar y motivar la necesidad de la topología de Skorohod para el estudio de la convergencia débil, se obtienen resultados que muestran las analogías entre los resultados de convergencia en espacios euclídeos con la convergencia débil. Se proponen demostraciones alternativas a las mostradas en el texto de referencia que justifican: El teorema de la aplicación continua, la caracterización del apretamiento en el espacio de funciones càdlàg, el teorema funcional del límite de Donsker.

Conclusiones y líneas futuras.

Este trabajo muestra la relevancia del apretamiento para estudiar la convergencia débil de variables aleatorias. Una forma de caracterizar la convergencia débil es por medio del comportamiento de integrales sobre la familia de funciones continuas y acotadas. Esta convergencia sólo tiene sentido cuando la masa de probabilidad no se extiende al infinito. La propiedad de apretamiento captura esta característica global. Durante el estudio han surgido cuestiones que excedían los objetivos del trabajo y que son de interés para continuar con el estudio, como son: condiciones alternativas de tightness basadas en tiempos óptimos de parada; aplicaciones de procesos empíricos.

Executive summary (English)

Background and aim.

In the undergraduate program, results on convergence under the Euclidean topology have been studied and later extended to infinite-dimensional function spaces. One of the challenges in determining convergence of sequences in these spaces is that convergence of finite-dimensional distributions does not imply convergence. The Arzelà-Ascoli theorem addresses this by characterizing the precompactness of families of continuous functions. The probabilistic version of this result requires studying a new property: tightness. Upon this property results of weak convergence can be obtained and the functional version of the Central Limit Theorem can be proven. The aim of this work is to study tightness and to propose solutions based on topology and analysis to the questions that arise in its study.

Methodology.

In this work results from Topology and Functional Analysis have been applied.

Main results.

After studying and motivating the need for the Skorokhod topology for the study of weak convergence, results are obtained that show the analogies between convergence results in Euclidean spaces and weak convergence. Alternative proofs to those given in the reference text are proposed, justifying: the continuous mapping theorem, the characterization of tightness in the space of càdlàg functions, and Donsker's functional central limit theorem.

Conclusions and future work.

This work demonstrates the importance of tightness for studying weak convergence of random variables. One way to characterize weak convergence is via the behavior of integrals against the family of bounded continuous functions. Such convergence only makes sense when the probability mass does not escape to infinity. The property of tightness captures this global feature. During the study arose some questions that exceeded the scope of this work and are of interest for further research, such as: alternative conditions for tightness based on optimal stopping times; applications to empirical processes.

Índice

1. Introducción.	2
2. Espacio de funciones $D[0, 1]$.	3
2.1. Definición y propiedades	3
2.2. Distancia de Skorohod	4
2.2.1. Necesidad de una métrica específica para el espacio $D[0, 1]$	4
2.3. Módulos de continuidad	6
2.3.1. Desigualdades	8
2.4. El conjunto T_P	10
2.5. Observaciones	11
2.5.1. Identificación $(D[0, 1], d)$ con espacio cociente $(D[0, 1]^+, \rho)/\sim$	11
2.5.2. $C[0, 1]$ en nunca denso en $D[0, 1]$	13
2.5.3. El conjunto de puntos de discontinuidad de $D[0, 1]$ es <i>pequeño</i>	14
3. Algunos resultados técnicos.	15
3.1. Portmanteau: Condiciones equivalentes para convergencia débil	15
3.1.1. Observación sobre la convergencia <i>débil</i>	16
3.2. Estabilidad de la convergencia	17
3.3. Teorema de la aplicación continua	19
3.3.1. Demostración alternativa	19
3.4. Teorema de Prohorov	22
3.5. Aproximación mediante funciones escalonadas	23
4. Tres resultados fundamentales.	23
4.1. Caracterización de compacidad en espacios de funciones	23
4.1.1. Espacio $C[0, 1]$. Teorema de Arzelà-Ascoli	24
4.1.2. Espacio $D[0, 1]$	25
4.2. Caracterización del apretamiento en espacios de medidas de probabilidad	25
4.2.1. Preliminar: propiedades y analogías	25
4.2.2. Teorema principal de caracterización del apretamiento	30
4.3. Condiciones para la convergencia a partir de leyes finito dimensionales	30
5. Convergencia débil en espacio localmente compacto.	33
6. Un Teorema funcional del límite. El Teorema de Donsker.	37
6.1. Espacio $C[0, 1]$	37
6.2. Espacio $D[0, 1]$	40
6.2.1. Paso 1. Las distribuciones finito dimensionales son convergentes	40
6.2.2. Paso 2. Las funciones de distribución de $\{X^n\}$ son apretadas	42
6.2.3. Paso 3. Aplicación del Teorema de Prohorov	45
7. Conclusiones	46
A. Apéndice	46
Referencias	47

1. Introducción.

Muchos de los resultados del análisis matemático requieren la **existencia de puntos límites**. En Probabilidad surgen límites de sucesiones de distribuciones (leyes de procesos estocásticos), de variables aleatorias y de las medidas de probabilidad inducidas por estas variables aleatorias. Los teoremas de convergencia requieren: (1) probar la existencia de un punto límite de las sucesiones; (2) identificar dicho punto límite cuando existe. El concepto de **compacidad** es fundamental para el objetivo (1).

La noción inicial de compacidad pretendía recoger precisamente la idea de que en un conjunto compacto toda sucesión infinita tiene un punto límite. En la recta real esto se garantiza si se toma un conjunto cerrado y acotado (Bolzano-Weierstrass). Los teoremas de valores extremos o el teorema de continuidad uniforme son una consecuencia de la compacidad de los intervalos cerrados de la recta real $\mathbb{R}$. Con objeto de generalizar esta idea a espacios topológicos arbitrarios, Maurice Fréchet destaca que lo relevante para la convergencia de sucesiones es que toda subsucesión sea convergente. Esto dio lugar a distintas nociones del concepto de compacidad, que en espacios métricos resultan equivalentes. Finalmente se consideró que la definición de compacidad en los términos actuales¹ era la más apropiada, a pesar de que la idea intuitiva original de existencia de puntos límites no es tan transparente con la definición actual.

Además de las situaciones citadas al comienzo, en muchas ocasiones las soluciones a un problema, como puede ser un sistema de ecuaciones derivadas parciales, se plantean a partir de una aproximación inicial que se refina de forma reiterada, dando lugar a una sucesión de aproximaciones. En espacios compactos se garantiza la existencia de un punto límite de esa sucesión de aproximaciones y, con ello, la existencia de una solución. Situaciones similares se dan en Probabilidad, donde se requiere determinar y conocer la convergencia de sucesiones de variables aleatorias, de sus distribuciones y leyes de probabilidad asociadas.

La compacidad es por tanto una propiedad topológica conveniente en muchas situaciones. Sin embargo, demostrar esta propiedad a partir de su definición (basada en cubrimientos abiertos) no es en la práctica sencillo o viable. En función del espacio, se han obtenido **resultados que caracterizan la compacidad a partir de otras propiedades *chequeables*** en términos geométricos o numéricos. Por ejemplo:

- **En un espacio Euclídeo** basta comprobar si un conjunto es cerrado y acotado (nociones geométricas) para concluir si es compacto.
- **En espacios de funciones** el Teorema de Arzelà-Ascoli establece la equicontinuidad y acotación uniforme como las características a chequear para confirmar si un conjunto de funciones es compacto (clausura compacta).
- **En espacios de medidas de probabilidad** con la topología débil, la compacidad se comprueba a partir de la propiedad de **apretamiento (*tightness*)** de un conjunto, ya

¹Un espacio topológico X es compacto si todo recubrimiento de abiertos, i.e. toda familia de abiertos $\{A_i\}_{i \in I}$ tales que $X \subseteq \cup_{i \in I} A_i$, contiene un subrecubrimiento finito.

que en ese caso el Teorema de Prohorov asegura que el conjunto es precompacto (su clausura es compacta).

En este sentido, el apretamiento es para la teoría de la convergencia débil lo que la acotación y la equicontinuidad son para la compacidad relativa en espacios de funciones. El apretamiento proporciona un criterio práctico del que se deriva la compacidad relativa. Es por ello una propiedad fundamental en el estudio de teoremas del límite de medidas de probabilidad y de procesos estocásticos. El objetivo de este trabajo es analizar el papel del apretamiento en el contexto de la convergencia débil de medidas de probabilidad y resolver las cuestiones que surjan durante su estudio.

Notación para referencias bibliográficas: El texto de referencia para este trabajo es [Billingsley \(1999\)](#) (en adelante B) y será citado en varias ocasiones. Las referencias a este texto se darán añadiendo delante del número de sección, teorema, etc. la letra B, por ejemplo: Teorema B1.1, § B8 hacen referencia respectivamente al Teorema 1.1 y a la Sección 8 del texto B. También se hará referencia a resultados de Topología del libro [Munkres \(2000\)](#) estudiado en el grado. En este caso las referencias se denotarán con la letra M.

El trabajo se organiza de la siguiente manera: La § 2 introduce los conceptos y definiciones que se emplearán en este trabajo. En la § 3 se presentan varios resultados técnicos que serán de utilidad en algunos de los desarrollos. La § 4 trata dos teoremas fundamentales para la caracterización de la compacidad y del apretamiento (tightness). Se demuestra un teorema que establece condiciones en las que la convergencia finito dimensional implica la convergencia en todo el espacio $D[0, 1]$. La § 5 ilustra el carácter global de la propiedad de apretamiento y muestra las condiciones en las que las funciones continuas de soporte compacto caracterizan la convergencia débil. En la § 6 se presenta una demostración del Teorema funcional del límite de Donsker. La § 7 recoge las conclusiones del trabajo. A lo largo de la exposición surgirán cuestiones para las que se propondrá una respuesta.

2. Espacio de funciones $D[0, 1]$.

2.1. Definición y propiedades

Denotamos como $D[0, 1]$ (o simplemente D cuando no haya lugar a confusión) el espacio de funciones reales definidas en el intervalo $[0, 1]$ y que satisfacen:

- (i) Continuidad por la derecha: Para $t \in [0, 1)$ existe el límite $x(t^+) = \lim_{s \downarrow t} x(s)$, cumpliendo además $x(t^+) = x(t)$.
- (ii) Límite por la izquierda: Para $t \in (0, 1]$ existe el límite $x(t^-) = \lim_{s \uparrow t} x(s)$ y es finito.

Las funciones de este espacio se denominan a veces *càdlàg* (del francés *continue à droite, limites à gauche*) o como haremos aquí *cadlag*. El espacio $D[0, 1]$ extiende así el espacio $C[0, 1]$

de funciones reales continuas definidas en $[0, 1]$, permitiendo la existencia de discontinuidades de salto finito.

Se comprueba que una función cadlag x satisface las siguientes propiedades (Lema B1 p. 122):

- El conjunto de puntos de discontinuidad de x es como máximo numerable.
- x es acotada, i.e.: $\|x\| = \sup_t |x(t)| < \infty$.
- x se puede aproximar uniformemente por funciones constantes en intervalos.

2.2. Distancia de Skorohod

2.2.1. Necesidad de una métrica específica para el espacio $D[0, 1]$

La distancia del supremo (o métrica de la convergencia uniforme) $\rho_\infty(x, y) := \sup_{t \in [0, 1]} |x(t) - y(t)|$ es apropiada para trabajar en el espacio $C[0, 1]$ porque las realizaciones de este espacio son recorridos continuos. En el espacio $D[0, 1]$, en cambio, la topología que se deriva de la métrica del supremo no resulta apropiada por los siguientes motivos:

- **La distancia del supremo hace que la convergencia uniforme en $D[0, 1]$ requiera unas condiciones demasiado estrictas.** En particular, requeriría que los saltos en las realizaciones de las funciones cadlag se produjeran en el mismo instante de tiempo que en el proceso límite. Esto resulta demasiado estricto, o al menos poco realista en la práctica, ya que los saltos se producen en periodos de tiempo aleatorios.

Por ejemplo, para las funciones indicatrices $x_n(t) = \mathbf{1}\{t \geq 1/2 + 1/n\}$ con $n \in \mathbb{N}$ y $x(t) = \mathbf{1}\{t \geq 1/2\}$ se tiene $\sup_t |x_n(t) - x(t)| = 1$ para todo n . Por ello, a pesar de que x_n y x sólo se diferencian en que el mismo salto se produce en un momento diferente (tras un instante de duración $1/n$), $\{x_n\}$ no converge uniformemente a x .

- **El espacio de funciones cadlag $D[0, 1]$ con la distancia del supremo no es separable.** Con ello se pierde una característica importante empleada para la justificación de algunos resultados de convergencia y teoremas del límite.

De la misma forma que en [García y Muñoz \(2014\)](#) comprobábamos que el espacio $(\ell^\infty(\mathbb{N}), \rho_\infty)$ no es separable, observamos que $(D[0, 1], \rho_\infty)$ tampoco lo es. Supongamos que fuera separable y sea $E = \{x^{(1)}, x^{(2)}, \dots\}$ el subconjunto numerable de funciones cadlag supuestamente denso en $D[0, 1]$. Consideramos el conjunto $A = \{x_a(t) = \mathbf{1}\{t \geq a\} : a \in [0, 1]\}$. Este conjunto no es numerable ($a \in [0, 1]$) y además en este conjunto $\rho_\infty(x_a, x_b) = 1$ siempre que $a \neq b$. Cada bola abierta $B(x^{(n)}, 1/2)$ en torno a cada elemento de E contiene como mucho un elemento de A . Así, existirá $x \in A$ tal que $x \notin B(x^{(n)}, 1/2)$ para todo $n \in \mathbb{N}$. Para ese $x \in A \subset D[0, 1]$ se tiene $\{x^{(1)}, x^{(2)}, \dots\} \cap B(x, 1/2) = \emptyset$ y por tanto E no puede ser denso en $D[0, 1]$.

La métrica de Skorohod ([Skorohod, 1956](#)) resuelve la circunstancia anterior para el espacio $D[0, 1]$. Para ello, considera *pequeñas* deformaciones del tiempo que permiten que las traslaciones en los saltos no impidan la convergencia.

Definición formal de la distancia de Skorohod: Sea Λ el conjunto de aplicaciones continuas y estrictamente crecientes en $[0, 1]$ a $[0, 1]$. Si $\lambda \in \Lambda$, entonces $\lambda(0) = 0$ y $\lambda(1) = 1$. Para x e y en $D[0, 1]$, se define $d(x, y)$ como el ínfimo de aquellos $\epsilon > 0$ para los que existe una $\lambda \in \Lambda$ que satisfaga:

$$\begin{aligned} \sup_t |\lambda(t) - t| &= \sup_t |t - \lambda^{-1}(t)| < \epsilon \\ \sup_t |x(t) - y(\lambda(t))| &= \sup_t |x(\lambda^{-1}(t)) - y(t)| < \epsilon. \end{aligned} \quad (1)$$

De forma más concisa, definiendo I como la aplicación identidad en $[0, 1]$ y $\|x\| = \sup_t |x(t)|$, se define la distancia de Skorohod:²

$$d(x, y) = \inf_{\lambda \in \Lambda} \{\|\lambda - I\| \vee \|x - y \circ \lambda\|\} \quad (2)$$

Distancia de Skorohod completa: Aunque se comprueba que el espacio $D[0, 1]$ es separable con la distancia d definida en (2) (Teorema B12.2), esta distancia no es completa. Se define en el espacio $D[0, 1]$ la métrica d° completa y equivalente a d de la siguiente forma. Para una aplicación λ no decreciente en $[0, 1]$ tal que $\lambda(0) = 0$ y $\lambda(1) = 1$, se define:

$$\|\lambda\|^\circ = \sup_{s < t} \left| \log \frac{\lambda t - \lambda s}{t - s} \right|.$$

Si $\|\lambda\|^\circ$ es finito, entonces las pendientes de las cuerdas entre dos puntos no toman valores nulo o infinito y, por tanto, esta λ es continua y estrictamente creciente, es decir $\lambda \in \Lambda$. Se define $d^\circ(x, y)$ como el ínfimo de aquellos $\epsilon > 0$ para los que Λ contiene alguna λ para la que $\|\lambda\|^\circ < \epsilon$ (además de la condición anterior (1)):

$$d^\circ(x, y) = \inf_{\lambda \in \Lambda} \{\|\lambda\|^\circ \vee \|x - y\lambda\|\}. \quad (3)$$

Se comprueba que d° es una métrica (Bp. 126). Intuitivamente, esta métrica permite una deformación del tiempo λ que es más próxima a la identidad que lo considerado en la métrica inicial d dada en (2). En concreto, se requiere que la pendiente $(\lambda t - \lambda s)/(t - s)$ en cada cuerda sea próxima a 1 (o equivalentemente, según se expresa en la definición, que su logaritmo sea próximo a 0). El espacio $D[0, 1]$ con la métrica d° es separable y completo (Teorema B12.2, p. 128).

Ilustración: Para no extender la exposición, no se reproducen aquí las demostraciones dadas en B de que d° es una métrica que convierte a $D[0, 1]$ en un espacio separable y completo. En su lugar, se muestran algunas ilustraciones que permiten observar y familiarizarse con la idea de esta métrica.

Se ha implementado la métrica d en scripts de R (R Core Team, 2025). Sólo es posible obtener una aproximación numérica a la distancia de Skorohod ya que sólo se pueden considerar un número finito de particiones del intervalo $[0, 1]$ y un conjunto representativo de algunas

²En el texto se denotará $a \vee b = \max\{a, b\}$ y $a \wedge b = \min\{a, b\}$.

aplicaciones λ crecientes de $[0, 1]$ a $[0, 1]$. En concreto se obtiene un límite superior de la distancia de Skorohod, ya que esta distancia es el ínfimo del máximo de dos valores sobre las posibles particiones de $[0, 1]$ y aplicaciones λ . La Figura 1 ilustra cómo interviene λ al calcular la distancia entre dos funciones.

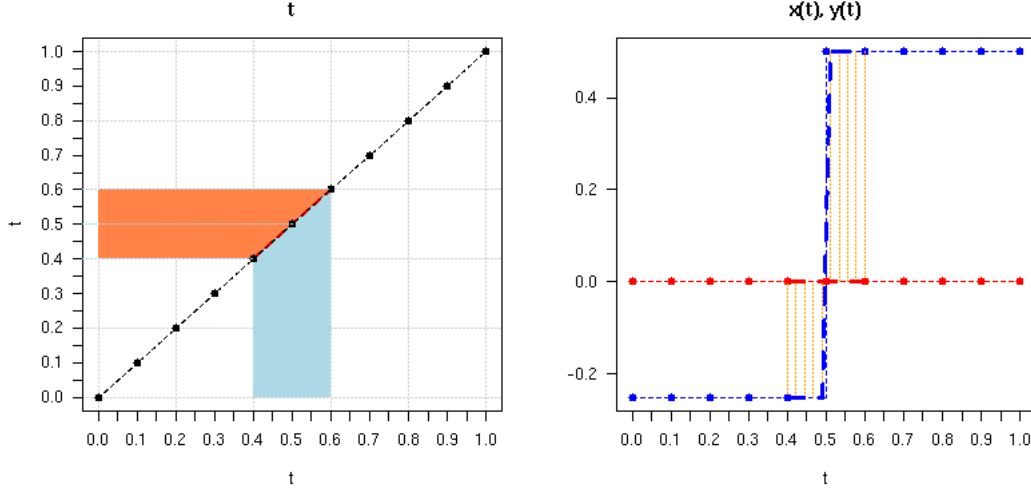


Figura 1: Distancia de Skorohod. Deformación del tiempo mediante λ .

El gráfico de la izquierda muestra cómo aplica λ un entorno de $t = 0,5$. El gráfico de la derecha muestrea dos funciones discretizadas en $[0, 1]$: la roja es una función constante y la azul es salto. Las líneas de puntos naranja ilustran las correspondencias de valores entre las dos funciones de acuerdo con $\lambda(t)$ mostrado en el gráfico de la izquierda.

2.3. Módulos de continuidad

El módulo de continuidad de una función es una aplicación $w_x(\delta) : (0, 1] \rightarrow (0, \infty)$ que proporciona una forma de cuantificar el grado de continuidad de funciones de $C[0, 1]$ y $D[0, 1]$. Permitirá caracterizar la compacidad relativa en los espacios de funciones $C[0, 1]$ y $D[0, 1]$. En particular, en el espacio $D[0, 1]$ da lugar a una variante de la idea de equicontinuidad, adaptada a funciones que pueden presentar discontinuidades.

Espacio $C[0, 1]$: Dado $x \in C[0, 1]$ o $x \in D[0, 1]$ se define:

$$w(x, \delta) \equiv w_x(\delta) = \sup_{0 \leq t \leq 1-\delta} w_x[t, t + \delta], \quad 0 < \delta \leq 1,$$

donde $w_x(T)$ se define para $T \subset [0, 1]$ como:

$$w_x(T) = \sup_{s, t \in T} |x(s) - x(t)|.$$

Módulo de continuidad w : Dado $x \in C[0, 1]$, la definición anterior es equivalente a:

$$w(x, \delta) \equiv w_x(\delta) = \sup_{|s-t| \leq \delta} |x(s) - x(t)|, \quad 0 < \delta \leq 1. \quad (4)$$

La siguiente condición es condición necesaria y suficiente para que la función x sea uniformemente continua en $[0, 1]$:

$$\lim_{\delta \rightarrow 0} w(x, \delta) = 0.$$

Asimismo, una familia de funciones de $C[0, 1]$ es uniformemente equicontinua si y sólo si:

$$\lim_{\delta \rightarrow 0} \sup_{x \in A} w_x(\delta) = 0.$$

En § 4.1.1 veremos que el Teorema de Arzelà-Ascoli se puede enunciar en términos del módulo de continuidad (4).

Espacio $D[0, 1]$: Los saltos que caracterizan las funciones cadlag no pueden ser controlados por el módulo de continuidad (4) en el siguiente sentido: Dada una función $x \in D[0, 1]$ con un salto de tamaño $a > 0$ en el punto $t_0 \in [0, 1]$, el $\sup_{|s-t| \leq \delta} |x(s) - x(t)|$ no puede tender a 0: para todo $\delta > 0$ se puede elegir $s \uparrow t_0$ y $t = t_0$ tales que $|s - t| \leq \delta$, pero $|x(s) - x(t)| \approx a > 0$. En consecuencia, el salto impone una cota inferior positiva y $w_x(\delta)$ no puede decrecer a 0. Esto hace que el módulo de continuidad (4) no sea útil para caracterizar la compacidad en el espacio $D[0, 1]$. Se definen en su lugar dos módulos de continuidad en el espacio $D[0, 1]$, w' y w'' .

Módulo de continuidad w' :

$$w'(x, \delta) \equiv w'_x(\delta) = \inf_{\{t_i\} \delta\text{-sparse}} \max_{1 \leq i \leq v} w_x[t_{i-1}, t_i], \quad 0 < \delta < 1, \quad (5)$$

donde el ínfimo se toma sobre todas las particiones δ -sparse, definidas como aquellas particiones $\{t_i\}_{i=0,1,\dots,v}$ de $[0, 1]$ para las que $\min_{1 \leq i \leq v} (t_i - t_{i-1}) > \delta$.

w' caracteriza las funciones cadlag. En concreto,

$$\lim_{\delta \rightarrow 0} w'(x, \delta) = 0 \quad \text{si y sólo si } x \in D[0, 1].$$

Así, esta condición implica que x satisface las propiedades mencionadas anteriormente en § 2.1. El Lema B1 p. 122 caracteriza esas condiciones de la siguiente forma: Dado $\epsilon > 0$ existe una partición $0 = t_0 < t_1 < \dots < t_v = 1$ tal que

$$w_x[t_{i-1}, t_i] < \epsilon \quad \text{para } i = 1, \dots, v.$$

La utilidad de $w'_x(\delta)$ se deberá a que éste puede tomar valores próximos a cero incluso cuando x tiene discontinuidades. Intuitivamente, el motivo es que la restricción a particiones δ -sparse impide generar una sucesión de particiones en las que el intervalo que contiene el salto de la discontinuidad (y que podemos suponer que contiene el supremo de todos los intervalos de la partición) contenga el ínfimo de los saltos sobre todas las particiones. De otro modo, la restricción a particiones δ -sparse hace que los saltos sean ignorados por el ínfimo.

En § 4.1.2 veremos que la compacidad relativa de una familia de funciones de $D[0, 1]$ se puede caracterizar en términos del módulo de continuidad (5).

Módulo de continuidad w'' :

$$w''(x, \delta) \equiv w''_x(\delta) = \sup_{\substack{t_1 \leq t \leq t_2 \\ t_2 - t_1 \leq \delta}} \{|x(t) - x(t_1)| \wedge |x(t_2) - x(t)|\}, \quad 0 < \delta < 1. \quad (6)$$

Aunque para $x \in D[0, 1]$ se cumple $w''(x, \delta) \downarrow 0$ cuando $\delta \downarrow 0$, por sí misma, w'' no caracteriza las funciones cadlag. Por ejemplo, para las funciones $\mathbf{1}_{[0, 1/n)}$ o $\mathbf{1}_{[1-1/n, 1]}$ se tiene $w'' = 0$, pero $w' = 1$ para $n \geq 1/\delta$. No obstante, en § 4.1.2 veremos que la compacidad relativa de una familia de funciones de $D[0, 1]$ se puede caracterizar en términos del módulo de continuidad (6) si se añaden condiciones que controlen las oscilaciones en torno a $t = 0$ y a $t = 1$.

2.3.1. Desigualdades

Las siguientes desigualdades son de utilidad (se demuestran de forma relativamente inmediata de las definiciones de ínfimo/supremo y del tipo de particiones):

- $w'(x, \delta) \leq w(x, 2\delta)$, para $\delta < 1/2$.
- $w(x, \delta) \leq 2w'(x, \delta) + j(x)$, con $j(x) = \sup_{0 < t \leq 1} |x(t) - x(t^-)|$.
- $w(x, \delta) \leq 2w'(x, \delta)$, para $x \in C$.
- $w''(x, \delta) \leq w'(x, \delta)$.

La siguiente desigualdad no se menciona en Billingsley, pero resultará necesaria para justificar la demostración en § 4.3 p. 31.

$$w'(x, \delta) \leq 2w''(x, \delta) \vee a_x(\delta) \vee b_x(\delta) \quad \text{para } \delta \leq \frac{1}{2} \quad \text{y} \quad (7)$$

$$a_x(\delta) = \sup_{0 \leq t \leq \delta} |x(t) - x(0)|,$$

$$b_x(\delta) = \sup_{1-\delta \leq t \leq 1} |x(1) - x(t)|.$$

En principio, podríamos considerar la desigualdad

$$w'(x, \delta) \leq 2w''(x, \delta) + \sup_{t \in (0, 1]} |x(t) - x(t-)|,$$

análoga a la segunda desigualdad para w y w' . A pesar de que esta desigualdad se cumple, veremos que en las condiciones del Teorema 4.5 esta desigualdad no es suficiente para demostrar lo que se requerirá en ese apartado. De hecho, el siguiente ejemplo muestra que la cota dada en (7) es una mejor aproximación.

$$x(t) = \begin{cases} 0, & 0 \leq t < \frac{1}{2}, \\ 1, & \frac{1}{2} \leq t \leq 1. \end{cases}$$

Calculamos el valor de los módulos de oscilación para este ejemplo:

- $w''(x) = 0$: Si $\frac{1}{2} \notin [t_1, t_2]$, entonces x es constante y la contribución de ese intervalo es cero. Si $\frac{1}{2} \in [t_1, t_2]$, el salto mínimo contra los laterales será con t_2 y resultará también nulo. De manera que w'' no detecta el salto en esta trayectoria.

- $a_x(\delta) = 0, b_x(\delta) = 0$: Cerca de los extremos la trayectoria es constante.
- $w'(x, \delta) = 0$: Si tomamos una partición $\{t_i\}$ de $[0, 1]$ que incluya $t_j = \frac{1}{2}$, entonces para $\delta < \frac{1}{2}$ en cualquier intervalo de longitud superior a δ del tipo $[\cdot, \frac{1}{2})$ o $(\frac{1}{2}, \cdot]$, la trayectoria es constante. Por ser w' el ínfimo del mayor salto sobre las particiones de tamaño superior a δ , $w'(x, \delta) = 0$.

Ambas desigualdades se satisfacen, pero la segunda resulta poco informativa por coincidir la cota con el mayor salto en cualquier punto de la trayectoria:

$$\begin{aligned} 0 &= w'(x, \delta) \leq 2w''(x, \delta) \vee a_x(\delta) \vee b_x(\delta) = 0, \\ 0 &= w'(x, \delta) \leq 2w''(x, \delta) + \sup_{t \in (0, 1]} |x(t) - x(t-)| = 0 + 1 = 1. \end{aligned}$$

Revisando las hipótesis del Teorema 4.5 que veremos en su momento, se observa que estas permiten controlar: oscilaciones rápidas de ida y vuelta³ en el interior de $[0, 1]$ a través de w'' ; los saltos próximos al extremo $t = 0$ por la continuidad por la derecha de las funciones cadlag; los saltos próximos al extremo $t = 1$ por una condición requerida expresamente en el Teorema. Esta observación lleva a pensar en la necesidad de incluir en la desigualdad las tres zonas de las particiones de $[0, 1]$ (interior y entornos de ambos extremos).

Comprobamos que se cumple la desigualdad propuesta en (7): Dado $\delta \in (0, \frac{1}{2})$ denotamos $w'' \equiv w''(x, \delta)$ y construimos una partición de $[0, 1]$ denotada

$$\{t_0 = 0, t_1 = \delta, t_2, \dots, t_j, \dots, t_{m-1} = 1 - \delta, t_m = 1\},$$

de forma que para todo $2 \leq j \leq m - 1$ se cumpla

$$w_x[t_j, t_{j+1}) \leq 2w'' + \epsilon$$

y para $1 < j < m - 1$ se satisfaga $t_{j+1} - t_j > \delta$, es decir, δ -sparse.

- Construcción de la partición: Dado $t_j \leq 1 - \delta$, si para todo $t \in [t_j, 1 - \delta]$ resulta

$$\sup_{t_j \leq s \leq u \leq t} |x(u) - x(s)| \leq 2w'' + \epsilon,$$

entonces se define $t_{j+1} = 1 - \delta$ y se finaliza. En otro caso, se define t_{j+1} como el primer $t > t_j$ para el que $\sup_{t_j \leq s \leq u \leq t} |x(u) - x(s)| > 2w'' + \epsilon$.

- Los intervalos interiores son δ -sparse: Supongamos por contradicción que existe j tal que $t_{j+1} < 1 - \delta$ y $t_{j+1} - t_j \leq w''$. Por construcción de t_{j+1} , existirían s y u tales que:

$$t_j \leq s \leq u \leq t_{j+1} \quad \text{y} \quad |x(u) - x(s)| > 2w'' + \epsilon.$$

³Oscilaciones causadas por único salto son obviadas por w'' , como se ha visto en el ejemplo. w'' es distinto de cero cuando en algún intervalo corto $[t_1, t_2]$ hay un punto t para el que $x(t)$ está lejos de ambos extremos. Esto requiere una oscilación de ida y vuelta en un intervalo corto.

Consideramos el caso $x(u) - x(s) > 2w'' + \epsilon$ y localizamos el primer $t \in [s, u]$ para el que

$$x(t) - x(s) > w'' + \epsilon/2,$$

es decir

$$|x(t) - x(s)| \geq w'' + \epsilon/2.$$

Se tiene así

$$x(u) - x(t) = (x(u) - x(s)) - (x(t) - x(s)) > (2w'' + \epsilon) - (w'' + \epsilon/2) = w'' + \epsilon/2,$$

Así,

$$|x(u) - x(t)| > w'' + \epsilon/2 \quad \text{y} \quad u - s \leq t_{j+1} - t_j \leq \delta.$$

Dado que $w''(x, \delta)$ se define como el supremo sobre las particiones, se tiene

$$w''(x, \delta) \geq |x(t) - x(s)| \wedge |x(u) - x(t)| > w'',$$

que contradice la elección inicial $w'' = w''(x, \delta)$.

- Evaluación: Por definición $w_x[0, \delta] \leq a_x(\delta)$ y $w_x[1 - \delta, 1] \leq b_x(\delta)$ y por construcción de cada subintervalo interior $[t_j, t_{j+1})$ se tiene $w_x[t_j, t_{j+1}) \leq 2w'' + \epsilon$. Así, hemos comprobado que para $\epsilon > 0$ arbitrario:

$$w'(x, \delta) \leq (2w'' + \epsilon) \vee a_x(\delta) \vee b_x(\delta).$$

Tomando $\epsilon \downarrow 0$, se confirma la desigualdad:

$$w'(x, \delta) \leq 2w''(x, \delta) \vee a_x(\delta) \vee b_x(\delta).$$

2.4. El conjunto T_P

El conjunto T_P se define como el conjunto de puntos de $[0, 1]$ para los que la proyección $\pi_t : D[0, 1] \rightarrow S$ que asigna $\pi_t(x) = x(t)$ es continua excepto en un conjunto no-numerable de puntos. Es importante aclarar que T_P se define no a nivel de una trayectoria $x \in D[0, 1]$, sino a nivel probabilístico en el espacio $D[0, 1]$. Es decir, lo que se requiere para que t pertenezca a T_P no es que una trayectoria se continua en t (a pesar de que esto asegura que π_t es continua en x) sino que la probabilidad del conjunto de trayectorias de D para las que $x(t) \neq x(t^-)$ es nula. Definiendo

$$J_t = \{x \in D[0, 1] : x(t) \neq x(t^-)\},$$

el punto $t \in [0, 1]$ pertenecerá a T_P si y sólo si $P(J_t) = 0$.

Se tiene el siguiente resultado (Bp. 138): El conjunto T_P contiene a $t = 0, 1$ y su complemento es como mucho numerable. Aunque no se detalla aquí, una demostración posible puede verse como una versión probabilística de la demostración de que el conjunto de puntos de discontinuidad de una función monótona real es como mucho numerable. Dado que, como se ha comentado en la definición, las trayectorias de $D[0, 1]$ son acotadas, cada trayectoria sólo puede tener un número finito de saltos superiores a $\epsilon \equiv \frac{1}{k}$. Esta es la base para concluir que el complementario de T_P es como mucho numerable, o como nos interesará más adelante, T_P es denso en $[0, 1]$.

2.5. Observaciones

2.5.1. Identificación de $(D[0, 1], d)$ con espacio cociente $(D[0, 1]^+, \rho)/\sim$

Se define D^+ como el espacio de funciones reales definidas en $[0, 1]$ cuyos límites por la derecha y por la izquierda existen para todo $t \in [0, 1]$, pero el valor de la función en t no tiene por qué coincidir con ninguno de estos límites. A diferencia del espacio D , en D^+ no se requiere continuidad por la derecha.

El Ejercicio B12.1 propone definir una pseudométrica en D^+ de forma que dos funciones de D^+ están a distancia cero si coinciden en todos los puntos $t \in [0, 1]$ en los que ambas funciones son continuas.

1. **Imposición de límites derechos en D^+ :** Dado $x \in D^+$ definimos $\bar{x}$:

$$\bar{x}(t) = \begin{cases} \lim_{s \rightarrow t^+} x(s), & t \in [0, 1) \\ x(1), & t = 1 \end{cases}$$

De esta forma, en los puntos de continuidad $\bar{x}$ coincide con x , mientras que en los puntos de discontinuidad toma el valor del límite derecho (que existe por definición de D^+ , aunque no tiene por qué coincidir con el valor de x en ese punto de discontinuidad). Así, conseguimos que $\bar{x} \in D$, donde tenemos definida la distancia d de Skorohod.

2. **Construcción de una medida ρ en D^+ :** Definimos la distancia ρ de D^+ a partir de la distancia de Skorohod, d :

$$\rho(x, y) = d(\bar{x}, \bar{y}).$$

Se comprueba que ρ es una pseudo-métrica .

- No-negatividad: $\rho(x, y) \geq 0$ para todo $x, y \in D^+$.
- Simétrica: $\rho(x, y) = \rho(y, x)$ para todo $x, y \in D^+$.
- Pseudo-métrica: $\rho(x, x) = 0$ para todo $x \in D^+$, pero $\rho(x, y) = 0$ no implica $x = y$, ya $x \neq y \in D^+$ pueden dar lugar a $\bar{x} = \bar{y}$. Como veremos, el espacio cociente se construye a partir de esta observación.
- Desigualdad triangular: $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$ para todo $x, y, z \in D^+$:

$$\rho(x, y) = d(\bar{x}, \bar{y}) \leq d(\bar{x}, \bar{z}) + d(\bar{z}, \bar{y}) = \rho(x, z) + \rho(z, y).$$

3. **Construcción del espacio cociente:** Sea la siguiente relación de equivalencia en D^+ :

$$x \sim y \iff \rho(x, y) = 0.$$

Definimos así el espacio cociente $D^+/\sim$ cuyas clases de equivalencia son los elementos de D^+ que están a distancia cero para la pseudo-métrica ρ .

- **Observación:** Cada clase de equivalencia contiene una única función cadlag; i.e., un único elemento de D , que podemos tomar como el representante de la clase.

- Existencia: Dado $x \in D^+$, imponiendo continuidad por la derecha a través de $\bar{x} \in D[0, 1]$ definida anteriormente, se tiene que en todo punto $t \in [0, 1)$ de continuidad de x se cumple $\bar{x}(t) = x(t)$ (también para $t = 1$ por la definición de $\bar{x}(1)$). Así $\bar{x} \sim x$ y todo clase contiene al menos una función cadlag de $D[0, 1]$.
- Unicidad: Sean $f, g \in D[0, 1]$ tales que $f \sim g$ (pertenecen a la misma clase de equivalencia). Por definición f y g coinciden en los puntos comunes de continuidad y en $t = 1$. Por la observación que veremos en § 2.5.3, los puntos de continuidad determinan la trayectoria, lo que implica que debe ser $f = g$ en todo $t \in [0, 1]$. En consecuencia, una misma clase de equivalencia no puede contener dos funciones cadlag diferentes.

4. **El espacio (D, d) es isométricamente equivalente a $(D^+ / \sim, \tilde{\rho})$:** Definimos la aplicación:

$$\Phi : D^+ / \sim \rightarrow D[0, 1], \quad \Phi([x]) = \bar{x}$$

y dotamos el espacio cociente $D^+ / \sim$ con una métrica, denotada $\tilde{\rho}$, a partir de la pseudo-métrica ρ de D^+ :

$$\tilde{\rho}([x], [y]) = \rho(x, y), \quad \text{para } [x], [y] \in D^+ / \sim.$$

- Φ está bien definida: Si $[x] = [y]$, entonces $\rho(x, y) = 0$ y por tanto $\bar{x} = \bar{y}$. De manera que $\Phi([x])$ no depende del representante de clase elegido.
- Se comprueba que $\tilde{\rho}$ es una métrica.
- Φ es una isometría: $d(\Phi([x]), \Phi([y])) = d(\bar{x}, \bar{y}) = \rho(x, y) = \tilde{\rho}([x], [y])$.
- Por la observación en § 2.5.3 (el conjunto de puntos de continuidad es denso) y dado que se ha visto que cada clase de equivalencia contiene una única función cadlag, se concluye que Φ es una biyección.

De este ejercicio se concluye y plantea:

- La métrica de Skorohod no tiene en cuenta las diferencias que se producen en los puntos de discontinuidad de dos funciones. Sólo se tiene en cuenta el comportamiento de las funciones en los puntos de continuidad.
- ¿Podría inferirse una métrica completa de D a partir de una métrica completa en el espacio cociente donde se puede embeber el espacio D ? De acuerdo con el Teorema M43.7 todo espacio métrico se puede embeber isométricamente en un espacio métrico completo. Aunque se ha planteado alguna posibilidad (siguiendo por ejemplo una estrategia similar al Ejercicio M43.9), no se ha encontrado la construcción que permitiría definir tal embebimiento. Tras este intento, se ha revisado la literatura al respecto. En particular se ha revisado el trabajo de Lifshits y Vysotsky (2020), que se apoya en la definición de *turbofunciones*. Aunque interesante desde el punto de vista teórico, no se ha encontrado una aplicación directa en los desarrollos presentados en este trabajo. Como se observará en el planteamiento propuesto en este trabajo para justificar el Teorema de Donsker en el

espacio D , lo relevante de la métrica es que existe y que convierte al espacio en separable. Explicitar expresiones o definiciones alternativas para su cálculo no resultará necesario una vez observado el papel de la métrica en este contexto.

2.5.2. $C[0, 1]$ es nunca denso en $D[0, 1]$

El Ejercicio B12.3 propone demostrar que el espacio $C[0, 1]$ de funciones continuas en $[0, 1]$ es nunca denso en $D[0, 1]$.

La afirmación anterior puede resultar en primera instancia contraintuitiva. Parecería que cualquier función de D se puede aproximar por medio de funciones continuas. Suponiendo una discontinuidad en el punto t , bastaría definir una sucesión de funciones continuas que se aproximen al límite derecho de la función. En ese caso C sería en realidad denso en D . Por ejemplo, si $x \in D \setminus C$ tiene una discontinuidad en el punto t_0 , siendo $x(t) = 0$ para $t < t_0$ y $x(t) = 1$ para $t \geq t_0$, podemos definir la siguiente sucesión de funciones continuas ([García y Muñoz, 2014](#), Ejemplo 1.10):

$$x_n(t) = \begin{cases} 0 & \text{si } t \in [0, \frac{1}{2} - \frac{1}{n}] \\ nt - \frac{n}{2} + 1 & \text{si } t \in [\frac{1}{2} - \frac{1}{n}, \frac{1}{2}] \\ 1 & \text{si } t \in [\frac{1}{2}, 1] \end{cases}.$$

Sin embargo, esto sería contradictorio con que C es cerrado. Una sucesión convergente de funciones de C converge a una función C . De manera que el ejemplo anterior no es válido para contradecir el enunciado propuesto.

Demostración del enunciado: Veremos que el interior de C es vacío. Comprobando que C es además cerrado, se concluye que C es nunca denso.

- **El interior de C es vacío:** Ninguna bola abierta de D puede estar contenida en C : Para todo $x \in C$ y para cualquier radio $r > 0$, existe una función $y \in D \setminus C$ tal que $d(x, y) < r$. Dada una función $x \in C$ y un $r > 0$, elegimos $t_0 \in (0, 1)$, $\delta \in (0, r)$ y definimos $y(t)$ como

$$y(t) = \begin{cases} x(t), & t < t_0, \\ x(t) + \delta, & t \geq t_0. \end{cases}$$

Se comprueba que $y \in D \setminus C$:

- y es cadlag: En todo t es continua por la derecha y existe el límite por la izquierda.
- y no es continua: En t_0 hay un salto de tamaño δ : $y(t_0^-) = x(t_0)$, $y(t_0) = x(t_0) + \delta$.

Para la función identidad contenida en Λ , $\lambda(t) = t$, la distancia $d(x, y)$ de Skorohod es:

$$\sup_t |\lambda(t) - t| = 0, \quad \sup_t |x(t) - y(\lambda(t))| = \sup_t |x(t) - y(t)| = \delta.$$

$$d(x, y) \leq \max\left(\sup_t |\lambda(t) - t|, \sup_t |x(t) - y(\lambda(t))|\right) = \delta < r.$$

Vemos así que en el interior de cualquier bola abierta $B_d(x, r)$ alrededor de una función continua x , existe una función cadlag discontinua como y . Por tanto, ninguna bola abierta puede estar contenida en C , i.e.: $\text{Int}(C) = \emptyset$.

- **C es cerrado:** Sea una sucesión $x_n \in C[0, 1]$ tal que $x_n \rightarrow x$ con la métrica de Skorohod. Existe $\{\lambda_n\} \in \Lambda$ tal que

$$\sup_t |\lambda_n(t) - t| \rightarrow 0 \quad \text{y} \quad \sup_t |x_n(\lambda_n(t)) - x(t)| \rightarrow 0.$$

Cada x_n es continua y cada λ_n es continua y estrictamente creciente, de modo que la composición

$$z_n(t) \equiv x_n(\lambda_n(t))$$

es una función continua en $[0, 1]$. Además $z_n \rightarrow x$ uniformemente, ya que

$$\sup_t |z_n(t) - x(t)| \rightarrow 0,$$

Dado que el límite uniforme de una función continua es continua (Teorema del límite uniforme M21.6), se tiene que $x \in C[0, 1]$. El límite de una sucesión de funciones continuas convergente con la distancia de Skorohod es así una función continua. En consecuencia, C es cerrado en (D, d) , i.e.: $\overline{C} = C$.

De los dos puntos anteriores se deduce que la clausura de C tiene interior vacío: $\text{Int } \overline{C} = \text{Int } C = \emptyset$.

2.5.3. El conjunto de puntos de discontinuidad de $D[0, 1]$ es *pequeño*

Al comienzo de la Sección B12 (Lema B1) se concluye que el conjunto de puntos de discontinuidad de las funciones cadlag es **pequeño en términos de cardinalidad**, como mucho numerable. Veremos que este conjunto de puntos de discontinuidad es también **topológicamente pequeño** en el sentido de que su clausura no contiene abiertos.

Recordamos el siguiente Teorema M48.5: Dada una sucesión $f_n : X \rightarrow Y$ de funciones continuas tales que $f_n(x) \rightarrow f(x)$ para todo $x \in X$, con $f : X \rightarrow Y$, siendo Y un espacio métrico; si X es un espacio de Baire,⁴ entonces el conjunto de puntos de continuidad de f es denso en X .

En nuestro caso $X \equiv [0, 1]$ es un espacio de Baire y el conjunto de puntos de discontinuidad sería nunca denso, por ser el complementario de un conjunto denso.

Sabemos que las funciones cadlag se pueden aproximar por funciones escalonadas (como se hace en Lema B3 p. 127). A su vez, podemos aproximar estas funciones escalonadas por funciones continuas (por ejemplo mediante interpolación lineal en un intervalo pequeño alrededor de cada discontinuidad). Obtenemos una sucesión de funciones continuas que converge (por punto) a

⁴Un espacio de Baire es un espacio topológico en el que la unión numerable de conjuntos cerrados con interior vacío tiene también interior vacío. Por ejemplo: espacios de Hausdorff compactos o un espacios métricos completos.

la función cadlag.⁵ Se completa así la aplicabilidad del Teorema anterior, concluyendo que el conjunto de puntos de discontinuidad de una función cadlag es nunca denso.

3. Algunos resultados técnicos.

En esta sección se contextualizan varios resultados técnicos, que serán algunas de las herramientas empleadas para demostrar los teoremas principales o de más inmediata relevancia.

Conjunto de continuidad P : Es aquel conjunto A perteneciente a la σ -álgebra en la que está definida una probabilidad P y cuya frontera tiene probabilidad nula, es decir $P(\partial A) = 0$.

3.1. Portmanteau: Condiciones equivalentes para convergencia débil

Sea S un espacio métrico y $\mathcal{S}$ la σ -álgebra generada por los abiertos de S . El par $(S, \mathcal{S})$ se denomina espacio de los estados. Dadas las medidas de probabilidad P_n y P definidas en $(S, \mathcal{S})$, son equivalentes:

- (i) $P_n \Rightarrow P$, i.e.: P_n converge débilmente a P , definida en cualquiera de las siguientes condiciones (ii) – (v).
- (ii) $P_n f := \int_S f dP_n \rightarrow \int_S f dP := P f$ para toda función real f acotada y uniformemente continua.
- (iii) $\limsup_n P_n F \leq P F$ para todo conjunto $F \in \mathcal{S}$ cerrado.
- (iv) $\liminf_n P_n G \geq P G$ para todo conjunto $G \in \mathcal{S}$ abierto.
- (v) $P_n(A) \rightarrow P(A)$ para todo conjunto $A \in \mathcal{S}$ de continuidad P , i.e. $P(\partial A) = 0$.

En el apartado § 3.1.1 se dan algunas ideas intuitivas del tipo de convergencia que implican estas equivalencias y en qué sentido puede entenderse que es una convergencia débil.

Una variable aleatoria X que aplica el espacio de probabilidad $(\Omega, \mathcal{F}, P)$ en el espacio de los estados $(S, \mathcal{S})$ induce una medida de probabilidad en $(S, \mathcal{S})$ dada por (para $A \subset S$): $P(A) = P(\omega \in \Omega : X(\omega) \in A)$. Esta medida de probabilidad se denomina **distribución o ley de probabilidad de X** . Una sucesión de variables aleatorias $\{X_n\}$ converge a X si sus distribuciones convergen y se denota: $X_n \Rightarrow X$ si $P_n \Rightarrow P$. Las equivalencias anteriores se pueden expresar en términos de **convergencia en distribución** de variables aleatorias:

- (i) $X_n \Rightarrow X$.
- (ii) $E(f(X_n)) \rightarrow E(f(X))$ para toda f acotada y uniformemente continua.
- (iii) $\limsup_n P(X_n \in F) \leq P(X \in F)$ para todo $F \in \mathcal{S}$ cerrado.

⁵Obsérvese que esto no contradice lo comprobado anteriormente (C es nunca denso en D) ya que esta convergencia es por punto y no en la norma del espacio.

(iv) $\liminf_n P(X_n \in G) \geq P(X \in G)$ para todo $G \in \mathcal{S}$ abierto.

(v) $P(X_n \in A) \rightarrow P(X \in A)$ para todo conjunto $A \in \mathcal{S}$ de continuidad P , i.e. $P(\partial A) = 0$.

3.1.1. Observación sobre la convergencia *débil*

En el apartado anterior se han establecido las equivalencias formales para la convergencia, que en el contexto de medidas de probabilidad se denomina convergencia *débil*. En este apartado ilustraremos la idea intuitiva de convergencia y en qué sentido entiendo que esta convergencia es débil.

En la asignatura del grado Espacios Normados hemos visto que en el espacio de operadores lineales se define la topología débil como la topología menos fina para la que los funcionales lineales son continuos. El Teorema de representación de Riesz muestra que los funcionales lineales son de la forma $\int_S f d\mu$ para funciones f continuas y acotadas. Esto se correspondería con la equivalencia (ii) de Portmanteau. Sin embargo, la relación entre la convergencia de medidas de probabilidad y la convergencia en la topología débil no resulta del todo clara.

Es natural entender que la convergencia es débil en el sentido de que es menos estricta o fuerte que la convergencia con la medida del supremo. Esto se ilustra en el siguiente ejemplo con $S = \mathbb{R}$, $P_n = \delta_{1/n}$ y $P = \delta_0$. Consideramos la medida de probabilidad de Dirac δ_a , para la que $\delta_a(A) = 1$ cuando $a \in A$ y 0 en otro caso. En este caso se tiene convergencia en el sentido de la equivalencia (ii) de Portmanteau (denominada débil) ya que para toda función f continua y acotada se tiene:

$$\int_{\mathbb{R}} f d\delta_{1/n} = f(1/n) \rightarrow f(0) = \int_{\mathbb{R}} f d\delta_0.$$

Así, decimos que $\delta_{1/n} \Rightarrow \delta_0$.

Sin embargo, si consideramos la distancia del supremo sobre todos los conjuntos de Borel $\mathcal{B}(\mathbb{R})$:

$$d_{\sup}(\delta_{1/n}, \delta_0) = \sup_{A \in \mathcal{B}(\mathbb{R})} |\delta_{1/n}(A) - \delta_0(A)|$$

y elegimos el conjunto $A = \{0\}$, entonces, dado que $\{1/n\} \notin \{0\}$, se tiene:

$$\begin{aligned} \delta_0(A) &= \delta_0(\{0\}) = 1, & \delta_{1/n}(A) &= \delta_{1/n}(\{0\}) = 0, \\ |\delta_{1/n}(A) - \delta_0(A)| &= |\delta_{1/n}(\{0\}) - \delta_0(\{0\})| = |0 - 1| = 1. \end{aligned}$$

En consecuencia la distancia del supremo será $d_{\sup}(\delta_{1/n}, \delta_0) \geq 1$. De hecho será igual a 1, ya que $\delta_{1/n}(A)$ y $\delta_0(A)$ sólo toma valores 0 ó 1 para cualquier $A \in \mathcal{B}(\mathbb{R})$. Así, para todo n se tiene:

$$d_{\sup}(\delta_{1/n}, \delta_0) = 1.$$

Intuitivamente: Si contrastamos la convergencia por medio de funciones continuas y acotadas, éstas no detectan la discontinuidad en $\{0\}$ y se concluye que existe convergencia. Sin embargo, si contrastamos convergencia por medio de conjuntos de Borel, dado que podemos encontrar dos conjuntos de Borel que separen el evento $\{0\}$ como en el ejemplo, la convergencia no es posible. La convergencia de medidas de probabilidad en los teoremas del límite no requerirá

una topología tan fuerte como la de la distancia del supremo, sino en los términos que se define en las equivalencias de Portmanteau. En este sentido puede entenderse la denominación de convergencia *débil*.

3.2. Estabilidad de la convergencia

Muchos de los resultados clásicos de convergencia en variables aleatorias reales ([Vélez Ibarrola, 2019](#), Capítulo 16) se extienden al caso de espacios métricos no euclídeos. En este apartado se presenta uno de estos resultados en espacios productos que resultará relevante.

El siguiente Teorema 3.1 se puede interpretar de la siguiente manera: La convergencia es estable ante pequeñas perturbaciones. La demostración dada en Teorema B3.1 p. 28 es muy escueta. Completaremos aquí los detalles de esa demostración.

Teorema 3.1 (La convergencia es estable ante pequeñas perturbaciones). *Sea S un espacio métrico con distancia ρ y sean (X_n, Y_n) elementos aleatorios de $S \times S$. Si $X_n \Rightarrow X$ y $\rho(X_n, Y_n) \Rightarrow 0$, entonces $Y_n \Rightarrow X$.*

Aclaraciones previas:

- El concepto *elemento aleatorio* extiende el de *variable aleatoria* al caso en el que el espacio de destino no es la recta real con la métrica usual. Es decir, X_n, Y_n son aplicaciones medibles de $\mathcal{F}$ a un espacio métrico general (S, ρ) , donde $\mathcal{F}$ es la σ -álgebra de un espacio de probabilidad $(\Omega, \mathcal{F}, P)$.
- Desarrollando la notación del segundo supuesto del teorema como:

$$P\left(\omega \in \Omega : \rho(X_n(\omega), Y_n(\omega))\right) \rightarrow 0,$$

resulta más transparente que:

- $\rho(X_n, Y_n)$ es una variable aleatoria.
- La convergencia es a una distribución degenerada en 0, en cuyo caso ([Vélez Ibarrola, 2019](#), Capítulo 16) es equivalente a convergencia en probabilidad:

$$P\left(\omega \in \Omega : \rho(X_n(\omega), Y_n(\omega))\right) < \epsilon \rightarrow 1, \quad \text{para } \epsilon > 0 \text{ arbitrario.}$$

Demostración. Sean $\epsilon > 0$ arbitrario y F un conjunto cerrado de S . Consideremos los conjuntos:

$$\begin{aligned} A &:= \{\omega \in \Omega : Y_n(\omega) \in F\} := \{Y_n \in F\}, \\ B &:= \{\omega \in \Omega : \rho(X_n(\omega), Y_n(\omega)) \geq \epsilon\} := \{\rho(X_n, Y_n) \geq \epsilon\}, \\ C &:= \{\omega \in \Omega : \rho(X_n(\omega), Y_n(\omega)) < \epsilon\} := \{\rho(X_n, Y_n) < \epsilon\}. \end{aligned}$$

Los conjuntos B y C son mutuamente excluyentes, luego obviamente $A \subset B \cup C$. Por la monotonidad y subaditividad de la medida de la probabilidad, se tiene la desigualdad:

$$P(\{Y_n \in F\}) \leq P(\{\rho(X_n, Y_n) \geq \epsilon\}) + P(\{\rho(X_n, Y_n) < \epsilon\}). \quad (8)$$

- Conjunto B : Por hipótesis se tiene convergencia en probabilidad $\rho(X_n, Y_n) \Rightarrow 0$ y por la condición (iii) de Portmanteau § 3.1 aplicada al conjunto cerrado $[\epsilon, \infty) \subset \mathbb{R}$ esto implica:

$$\limsup_n P(\rho(X_n, Y_n) \in [\epsilon, \infty)) \leq P(0 \in [\epsilon, \infty)) = 0. \quad (9)$$

- Conjunto C : No podemos aplicar la misma idea que para B porque $(0, \epsilon)$ es abierto y, en ese caso, la condición (iv) de Portmanteau conduce a una desigualdad $\geq$ que no resulta informativa en este caso. Sin embargo, podemos observar que el conjunto C se puede escribir en términos del conjunto cerrado $F_\epsilon = \{x \in S : \rho(x, F) \leq \epsilon\}$. En concreto veremos que:

para $\omega \in A := \{Y_n \in F\}$, $C := \{\rho(X_n, Y_n) < \epsilon\}$ coincide con $\{\omega \in \Omega : X_n(\omega) \in F_\epsilon\}$.

Sea $\omega \in \Omega$ tal que $Y_n(\omega) \in F$ y $\omega \in C$, es decir, $\rho(X_n(\omega), Y_n(\omega)) < \epsilon$. A partir de la definición de distancia de un punto a un conjunto:

$$\rho(X_n(\omega), F) := \inf_{z \in F} \rho(X_n(\omega), z) \leq \rho(X_n(\omega), Y_n(\omega)) < \epsilon.$$

Así, en este caso $\omega \in \{\omega' \in \Omega : X_n(\omega') \in F_\epsilon\}$.

Dado que por hipótesis $X_n \Rightarrow X$ y que F_ϵ es cerrado, de nuevo por la condición (iii) de Portmanteau se tiene:

$$\limsup_n P(X_n \in F_\epsilon) \leq P(X \in F_\epsilon). \quad (10)$$

Aplicando (9) y (10) en la desigualdad (8) se concluye:

$$\limsup_n P(Y_n \in F) \leq 0 + P(X \in F_\epsilon) = P(X \in F_\epsilon).$$

Por ser F cerrado, F_ϵ decrece a F cuando $\epsilon \rightarrow 0$ y dado que P es continua superiormente:

$$\limsup_n P(Y_n \in F) \leq \lim_{\epsilon \rightarrow 0} P(X \in F_\epsilon) = P(X \in F).$$

De nuevo por la condición (iii) de Portmanteau, la expresión anterior es equivalente a $Y_n \Rightarrow X$. $\square$

Corolario 3.2. *Sustituyendo en el Teorema 3.1 X_n por X se concluye: Convergencia en probabilidad de $\rho(X, Y_n)$ a 0, denotado $\rho(X, Y_n) \Rightarrow 0$, implica $Y_n \Rightarrow X$. Es decir, si para todo ϵ se tiene $P(\{\omega \in \Omega : |\rho(X(\omega), Y_n(\omega)) - 0| < \epsilon\}) \rightarrow 1$, entonces Y_n converge en distribución a X .*

Usando los mismos argumentos anteriores e imponiendo una condición adicional, el siguiente Teorema 3.3 (Teorema B3.2) justifica que es posible probar resultados de convergencia a partir de una aproximación X_{un} de X_n con la que resulte más sencillo trabajar, siempre y cuando el error de aproximación sea asintóticamente nulo.

Teorema 3.3. *Sean $\{(X_{un}, X_n)\}_n$ elementos aleatorios de $S \times S$ y u un entero positivo fijado de antemano. Si $X_{un} \Rightarrow_n Z_u \Rightarrow_u X$ y si para todo $\epsilon > 0$*

$$\lim_u \limsup_n P[\rho(X_{un}, X_n) \geq \epsilon] = 0,$$

entonces $X_n \Rightarrow_n X$.

3.3. Teorema de la aplicación continua

Dados dos espacios métricos S y S' con sus respectivas álgebras de Borel $\mathcal{S}$ y $\mathcal{S}'$ y dada una medida de probabilidad P definida sobre $\mathcal{S}$, una función medible $h : S \rightarrow S'$ induce una medida de probabilidad $P' \equiv (Ph^{-1})$ en $\mathcal{S}'$ de la siguiente forma. Sean $A' \in \mathcal{S}'$ y $h^{-1}(A') = \{a \in S : h(a) \in A'\} \in \mathcal{S}$,

$$P'(A') \equiv (Ph^{-1})(A') = P(h^{-1}(A')).$$

Una cuestión que resulta de utilidad en muchas aplicaciones es: ¿En qué condiciones la convergencia $P_n \Rightarrow P$ en $\mathcal{S}$ implica convergencia de las probabilidades inducidas $(P_n h^{-1}) \Rightarrow (Ph^{-1})$ en $\mathcal{S}'$?

A partir de la caracterización (ii) de Portmanteau, según la cual $P_n f \rightarrow P f$ para toda función real f acotada y uniformemente continua implica la convergencia débil $P_n \Rightarrow P$ en $\mathcal{S}$, un cambio de variable (Valdivia, 1988) permite comprobar que si h es continua entonces se tiene convergencia en las probabilidades inducidas por h sobre $\mathcal{S}'$.

$$\begin{aligned} P_n f \equiv \int_{S'} f(y) (P_n h^{-1})(dy) &= \int_S f(h(x)) P_n(dx) \\ &\longrightarrow \int_S f(h(x)) P(dx) \equiv P f = \int_{S'} f(y) (Ph^{-1})(dy) \equiv (Ph^{-1}) f. \end{aligned}$$

Lo anterior podía ser esperable, ya que en el grado hemos visto que las funciones continuas preservan muchas propiedades topológicas del espacio de origen, como son los límites de sucesiones. Sabemos, por ejemplo, que si f es una función continua y $x_n \rightarrow x$, entonces $f(x_n) \rightarrow f(x)$.

La convergencia débil de las medidas de probabilidad permite relajar el requerimiento de continuidad. En concreto se tiene el siguiente resultado, según el cual la continuidad casi segura de h es suficiente para que las discontinuidades no afecten a la distribución límite. Denotamos como D_h el conjunto de puntos de discontinuidad de h .

Teorema 3.4. Teorema de la aplicación continua. *Supongamos que $P_n \Rightarrow P$ en $\mathcal{S}$ y $h : S \rightarrow S'$ es medible (no necesariamente continua). Sea D_h el conjunto de puntos de discontinuidad de h . Si $P(D_h) = 0$, entonces $(Ph^{-1})_n \Rightarrow (Ph^{-1})$ en $\mathcal{S}'$. Equivalentemente, si $X_n \Rightarrow X$ y $P(X \in D_h) = 0$, entonces $h(X_n) \Rightarrow h(X)$.*

Aunque en § 3.1.1 se discutirá en qué sentido es débil la convergencia, lo anterior muestra que el requerimiento para la convergencia en espacios de medidas de probabilidad no es tan fuerte como en otras situaciones.

3.3.1. Demostración alternativa

Billingsley demuestra el Teorema 3.4 anterior a partir de la equivalencia (iii) de Portmanteau § 3.1. Siguiendo las indicaciones del Ejercicio B2.10, aquí lo demostraremos a partir de la equivalencia (v). El Ejercicio B2.10 contiene varias erratas. Curiosamente, determinar la corrección de esas erratas es lo que ha dado la pista para encontrar la línea argumental adecuada en la

demostración alternativa propuesta en el ejercicio. El enunciado original del primer apartado es:⁶

Without using the mapping theorem, show that $P_n \Rightarrow Pg$ if $P_n g \rightarrow P$ for all bounded, continuous g such that $P(D_g) = 0$,

donde D_g denota el conjunto de puntos de discontinuidad de g .

- Se entiende que g es medible de Borel, no continua, ya que en este último caso se tendría trivialmente $P(D_g) = 0$.
- En lugar de ' $P_n \Rightarrow Pg$ si $P_n g \rightarrow P$ ', debería ser:

$$\text{bien } P_n \Rightarrow P \text{ si } P_n g \rightarrow Pg, \quad (11)$$

$$\text{o bien } P_n g \rightarrow Pg \text{ si } P_n \Rightarrow P. \quad (12)$$

Como se describe a continuación, el trabajo de este ejercicio ha conducido a comprobar que se satisfacen las dos implicaciones del segundo punto. De hecho, ambas serán necesarias para completar la demostración dada en este apartado.

- **Si $P_n g \rightarrow Pg$ entonces $P_n \Rightarrow P$:** Intuitivamente: el conjunto de funciones continuas en casi todo punto es suficientemente grande como para llegar a la condición (v) de las equivalencias de Portmanteau § 3.1.

Del mismo modo que los conjuntos cerrados se representan por medio de funciones indicatrices para demostrar que una medida de probabilidad está completamente determinada por funciones acotadas y uniformemente continuas, representamos por medio de $f = \mathbf{1}_A$ a un conjunto de continuidad P .⁷

Los puntos de discontinuidad de $\mathbf{1}_A$ son precisamente la frontera de A , $D_{\mathbf{1}_A} = \partial A$ y dado que $\mathbf{1}_A$ es medible y acotada, podemos aplicar la hipótesis de este punto para $g \equiv \mathbf{1}_A$:

$$P_n(A) = P_n \mathbf{1}_A \rightarrow P \mathbf{1}_A = P(A).$$

Se cumple así la equivalencia (v) de Portmanteau § 3.1, i.e. $P_n(A) \rightarrow P(A)$ para todo conjunto de continuidad P . En consecuencia $P_n \Rightarrow P$.

- **Si $P_n \Rightarrow P$ entonces $P_n g \rightarrow Pg$:**

- Por hipótesis g es acotada. Podemos suponer que en particular $0 \leq g(x) \leq 1$ (para otro rango $[a, b] \in \mathbb{R}$ podríamos aplicar lo siguiente a la transformación $\frac{g(x)-a}{b-a}$).

⁶Modifico la notación original del texto empleando g en lugar de f , ya que f se ha identificado en varias ocasiones con una función real acotada uniformemente continua.

⁷Es decir, $A \in \mathcal{S}$ y $P(\partial A) = 0$ según definición en p. 15.

- El ejercicio sugiere aplicar la misma idea que en la demostración $(v) \Rightarrow (i)$ de las equivalencias de Portmanteau, que se apoya en el siguiente resultado:⁸

$$P_n g = \int_0^1 P_n(\{g > t\}) dt, \quad P g = \int_0^1 P(\{g > t\}) dt.$$

donde $\{g > t\}$ denota *conjuntos de supernivel* $\{s \in S : g(s) > t\}$. Lo anterior se cumple si $P_n(\{g > t\}) \rightarrow P(\{g > t\})$ para casi todo $t \in (0, 1)$.

- Para casi todo t Los conjuntos de supernivel $\{g > t\}$ son conjuntos de continuidad P : Comprobamos primero la siguiente inclusión:

$$\partial\{g > t\} \subset D_g \cup \{g = t\}.$$

Supongamos que $x \notin D_g \cup \{g = t\}$. Tenemos entonces que g es continua en x y $g(x) \neq t$. Si además fuera $g(x) > t$, por la continuidad en ese punto existiría un entorno U de x en el que $g(y) > t$ para todo $y \in U$. Por tanto se tendría $U \subset \{g > t\}$ y debería ser $x \notin \partial\{g > t\}$. De forma similar, si $g(x) < t$, debería ser de nuevo $x \notin \partial\{g > t\}$. De modo que se satisface la inclusión anterior.

- Por lo anterior: $P(\partial\{g > t\}) \leq P(D_g) + P(\{g = t\}) = P(\{g = t\})$. Observamos que los conjuntos $\{g = t\}$ son disjuntos para diferentes valores de t , de manera que sólo puede haber como mucho un conjunto numerable de valores t en el rango $[0, 1]$ para los que $P(\{g = t\}) > 0$. Es decir, para todo t excepto un conjunto como mucho numerable de $[0, 1]$ debe ser $P(\{g = t\}) = 0$. De otro modo, para casi todo t los conjuntos $\{g > t\}$ son puntos de continuidad P , es decir $P(\partial\{g > t\}) = 0$ en casi todo t .
- Aplicando ahora la hipótesis $P_n \Rightarrow P$ a los conjuntos $\{g > t\}$, por la equivalencia (v) de Portmanteau se obtiene:

$$P_n(\{g > t\}) \rightarrow P(\{g > t\}) \quad \text{para casi todo } t \in (0, 1).$$

Por este resultado y el Teorema de convergencia dominada (aplicable $0 \leq P_n(\cdot) \leq 1$):

$$P_n g := \int_0^1 P_n(\{g > t\}) dt \rightarrow \int_0^1 P(\{g > t\}) dt := P g.$$

Demostración alternativa del Teorema de la aplicación continua: Podemos ahora proponer la siguiente demostración alternativa del Teorema 3.4 enunciado en p. 19.

Sean (S, d) y (S', d') espacios métricos con sus respectivas álgebras de Borel $\mathcal{S}$ y $\mathcal{S}'$. Sean P_n, P medidas de probabilidad definidas en $\mathcal{S}$ y $h : S \rightarrow S'$ una aplicación medible tal que $P(D_h) = 0$.⁹

Suponiendo $P_n \Rightarrow P$ veremos que $(P_n h^{-1}) \Rightarrow (P h^{-1})$.

⁸En el Apéndice § 7 se enuncia este resultado relacionado con el Principio de Cavalieri.

⁹Para que la condiciones del tipo $P(D_h) = 0$ que se hacen aquí tengan sentido, se requiere $D_h \in \mathcal{S}$. Billingsley p. 243 muestra que el conjunto D_h se puede expresar como unión de intersección de abiertos.

Sea $f : S' \rightarrow \mathbb{R}$ una función acotada y medible tal que sus puntos de discontinuidad tienen probabilidad nula para la probabilidad inducida, es decir, $(P_n h^{-1})(D_f) = 0$.

Definimos la composición $g \equiv f \circ h : S \rightarrow \mathbb{R}$, que es por construcción acotada y medible. Estudiamos los puntos de discontinuidad de $g \equiv f \circ h$. Observamos que

$$D_{f \circ h} \subset D_h \cup h^{-1}(D_f) \quad \text{y con ello} \quad P(D_{f \circ h}) \leq P(D_h) + P(h^{-1}(D_f)).$$

Por la definición de la probabilidad inducida y la definición anterior de D_f ,

$$P(h^{-1}(D_f)) = (Ph^{-1})(D_f) = 0.$$

Además, por hipótesis $P(D_h) = 0$. En consecuencia, los puntos de discontinuidad D_g tienen probabilidad nula,

$$P(D_{f \circ h}) = 0.$$

Retomamos ahora la condición (12) justificada anteriormente, según la cual si $P_n \Rightarrow P$, entonces para toda función acotada y medible g tal que $P(D_g) = 0$, se cumple $P_n g \rightarrow P g$. Aplicándolo a $g \equiv f \circ h$, se tiene

$$P_n(f \circ h) \rightarrow P(f \circ h). \quad (13)$$

Por último, aplicamos sobre P_n y P el mismo cambio de variable que en la demostración original (B p. 20):

$$\begin{aligned} (P_n h^{-1}) f &:= \int_{S'} f(y) (P_n h^{-1})(dy) = \int_S f(h(x)) P_n(dx) := P_n(f \circ h), \\ (Ph^{-1}) f &:= \int_{S'} f(y) (Ph^{-1})(dy) = \int_S f(h(x)) P(dx) := P(f \circ h), \end{aligned}$$

y con ello, por el resultado (13), $(P_n h^{-1})^{-1} f \rightarrow (Ph^{-1}) f$.

Como este resultado se cumple para toda función $f : \mathbb{S} \rightarrow \mathbb{R}$ acotada y medible para la que $(Ph^{-1})(D_f) = 0$, por la condición (11) justificada anteriormente se concluye:

$$(P_n h^{-1}) \Rightarrow (Ph^{-1}),$$

que es el resultado del Teorema de la aplicación continua.

3.4. Teorema de Prohorov

Billingsley detalla en la § 5 una demostración del Teorema de Prohorov (1956) y en las notas bibliográficas cita referencias complementarias. Aquí nos limitaremos a enunciar el Teorema, por motivos de espacio y porque las técnicas de la demostración no se utilizan en los desarrollos realizados en este trabajo. No obstante, por ser un resultado fundamental para la temática de este trabajo, se ha optado por enunciarlo en el texto y no en el apéndice.

Se dice que una familia Π de medidas de probabilidad definidas en $(S, \mathcal{S})$ es **relativamente compacta** si toda sucesión de elementos de Π contiene una subsucesión que converge débilmente (no necesariamente a un elemento de Π). Se dice que la **familia Π es apretada (*tight*)** cuando para todo $\epsilon > 0$ y para toda $P \in \Pi$ existe un conjunto compacto K tal que $P(K) > 1 - \epsilon$.

Teorema 3.5 (Teorema de Prohorov. Teorema B5.1 p. 59.). *Si una familia de medidas de probabilidad Π es apretada (tight), entonces Π es relativamente compacta.*

El resultado recíproco no se satisface en cualquier espacio métrico.

Teorema 3.6 (Recíproco del Teorema de Prohorov. Teorema B5.2 p. 60.). *Sea S un espacio separable y completo. Si Π es relativamente compacta, entonces Π es apretada (tight).*

3.5. Aproximación mediante funciones escalonadas

En ocasiones es útil aproximar una función del espacio $D[0, 1]$ por medio de funciones escalonadas de la siguiente forma. Se toma una partición $\sigma = \{0 = s_0 < s_1 < \dots < s_k = 1\}$ y se define $A_\sigma : D \rightarrow D$ como:

$$(A_\sigma(x))(t) = x(s_{u-1}) \quad \text{para } t \in [s_{u-1}, s_u), \quad (A_\sigma(x))(1) = x(1).$$

Esta aproximación fija el valor de la trayectoria durante la longitud de cada subintervalo en el valor que toma la trayectoria original x en el extremo izquierdo del subintervalo.

Si las particiones se restringen de forma que la longitud del mayor subintervalo no supera δ :

$$\|\sigma\| := \max_i (s_i - s_{i-1}) \leq \delta,$$

se obtiene la siguiente desigualdad para la distancia de Skorohod:

$$d(A_\sigma(x), x) := \inf_{\lambda \in \Lambda} \left\{ \sup_t |\lambda(t) - t| \vee \sup_t |x(t) - y(\lambda(t))| \right\} \leq \delta \vee w'_x(\delta). \quad (14)$$

Billingsley (Lema B3, p. 127) construye una partición que demuestra la desigualdad anterior. Básicamente, se trata de definir una partición tal que:

$$\begin{aligned} \sup_t |\lambda(t) - t| &\leq \delta, \\ \sup_t |(A_\sigma(x))(t) - x(\lambda(t))| &\leq w'_x(\delta). \end{aligned}$$

De esta forma, se observa que:

- La aproximación desplaza los tiempos de saltos como mucho un periodo de longitud δ .
- Dado un desplazamiento o deformación del tiempo λ , la diferencia máxima entre el valor original y la aproximación de la trayectoria está acotada por $w'_x(\delta)$.

4. Tres resultados fundamentales.

4.1. Caracterización de compacidad en espacios de funciones

En el grado se ha estudiado el Teorema de Arzelà-Ascoli, que caracteriza la compacidad relativa en espacios de funciones continuas. En esta sección se retoma este resultado y se enuncia el teorema análogo para el espacio $D[0, 1]$.

4.1.1. Espacio $C[0, 1]$. Teorema de Arzelà-Ascoli

Sabemos (Vera y Alegría, 1997, § I.5) que la clausura de conjunto $A \subset C[0, 1]$ es compacta si y sólo si es *puntualmente acotado* y *equicontinuo*. Estas condiciones son equivalentes a:

$$\begin{aligned} \sup_{x \in A} |x(0)| &< \infty, & \text{órbitas en } t = 0 \text{ acotadas,} \\ \lim_{\delta \rightarrow 0} \sup_{x \in A} w_x(\delta) &= 0, & \text{equicontinuidad uniforme.} \end{aligned}$$

Aparentemente la primera condición es más débil, ya que requiere acotación sólo en el punto $t = 0$. Veremos que esta condición, junto con la condición de equicontinuidad uniforme, implican que las funciones son uniformemente acotadas (de hecho se requiere para la demostración de suficiencia de estas condiciones para compacidad).

Supongamos que se satisfacen las dos condiciones anteriores. Para un $\epsilon > 0$, por la hipótesis de equicontinuidad uniforme existe $\delta_0 > 0$ tal que $w_x(\delta_0) \leq \epsilon$ para todo $x \in A$, es decir, para todo $x \in A$ se cumple

$$|x(t) - x(s)| \leq \epsilon \quad \text{siempre que } |t - s| \leq \delta_0.$$

Definimos una partición finita $\{t_i\}_{i=0, \dots, N}$ de $[0, 1]$ de tamaño no superior a $N \geq \frac{1}{\delta_0}$, por ejemplo:

$$0 = t_0 < t_1 = \frac{1}{N} < t_i = \frac{i}{N} < \dots < t_N = 1.$$

Así, para cualquier $t \in [0, 1]$ y $x \in A$ existe t_k tal que

$$|x(t) - x(t_k)| \leq \epsilon, \quad (\text{basta elegir } k \text{ que satisfaga } |t - t_k| \leq 1/N \leq \delta_0).$$

Observamos que de $t_0 = 0$ a t_k no hay más de k pasos de longitud no superior a δ_0 . En cada paso las oscilaciones de todo $x \in A$ no superan el valor $w_A(\delta_0) \leq \epsilon$. En concreto, considerando la suma telescópica:

$$x(t) = x(0) + \sum_{i=1}^k (x(it/k) - x((i-1)/k))$$

se tiene:

$$|x(t)| \leq |x(0)| + kw_A(\delta_0) \leq |x(0)| + k\epsilon.$$

Tomando el supremo

$$\sup_{x \in A} \sup_{t \in [0, 1]} |x(t)| \leq \sup_{x \in A} |x(0)| + k\epsilon < \infty \text{ por la condición } \sup_{x \in A} |x(0)| < \infty.$$

Lo anterior es aplicable para cualquier $\epsilon > 0$ (basta redefinir la partición de N intervalos con δ_0 apropiado). Tomando $\epsilon \rightarrow 0$ se concluye que

$$\sup_{x \in A} \sup_{t \in [0, 1]} |x(t)| < \infty,$$

es decir, en presencia de equicontinuidad uniforme, es suficiente con acotar las órbitas en un punto $t \in [0, 1]$ (podría elegirse otro punto $t_0 \in [0, 1]$) para asegurar que las oscilaciones de

un conjunto de funciones son también acotadas. Las condiciones anteriores son, por tanto, equivalentes a la caracterización habitual de la compacidad relativa en $C[0, 1]$.

A partir de aquí, comprobando que A es totalmente acotado (en espacio C completo) se completa la justificación de suficiencia en el Teorema de Arzelà-Ascoli. No se reproducen aquí los detalles de la demostración dada en Teorema B7.2.

4.1.2. Espacio $D[0, 1]$

Los siguientes Teoremas extienden la noción de compacidad al espacio de funciones $D[0, 1]$. Lo novedoso es que la compacidad se enuncia en términos del módulo de continuidad presentado anteriormente en § 2.3. En el texto de referencia pueden encontrarse las demostraciones en detalle.

Teorema 4.1 (Teorema B12.3. Primera caracterización de compacidad.). *Un conjunto $A \subseteq D[0, 1]$ es relativamente compacto (clausura compacta) con la topología de Skorohod si y sólo si:*

$$\begin{aligned} \sup_{x \in A} \|x\| &< \infty, \\ \lim_{\delta \rightarrow 0} \sup_{x \in A} w'_x(\delta) &= 0. \end{aligned} \tag{15}$$

En la práctica puede ser más sencillo trabajar con el módulo de continuidad w'' en lugar de w' . Controlando las oscilaciones cerca de los extremos de $[0, 1]$, se tiene la siguiente caracterización.

Teorema 4.2 (Teorema B12.4. Segunda caracterización de compacidad.). *Un conjunto $A \subseteq D[0, 1]$ es relativamente compacto con la topología de Skorohod si y sólo si, además de (15), se satisface:*

$$\begin{cases} \lim_{\delta \rightarrow 0} \sup_{x \in A} w''(\delta) = 0, \\ \lim_{\delta \rightarrow 0} \sup_{x \in A} |x(\delta) - x(0)| = 0, \\ \lim_{\delta \rightarrow 0} \sup_{x \in A} |x(1) - x(1 - \delta)| = 0. \end{cases}$$

4.2. Caracterización del apretamiento en espacios de medidas de probabilidad

Se recuerda de la § 3.4 que una medida de probabilidad P en el espacio de los estados $(S, \mathcal{S})$ es apretada o tight si para cada $\epsilon > 0$ existe un conjunto compacto K tal que $P(K) > 1 - \epsilon$.

4.2.1. Preliminar: propiedades y analogías

En este apartado se trabajan varios ejercicios propuestos en § B5 que permiten tratar el concepto de apretamiento desde varias perspectivas.

1. Caracterización mediante las funciones de distribución:

Una sucesión de medidas de probabilidad $\{P_n\}$ sobre la recta real es apretada si y sólo si la sucesión de sus funciones de distribución asociadas $\{F_n\}$ satisfacen $\lim_{x \rightarrow \infty} F_n(x) = 1$ y $\lim_{x \rightarrow -\infty} F_n(x) = 0$ uniformemente.

- **Tightness $\Rightarrow$ límites uniformes en las colas:** Por hipótesis existe M finito y un compacto $K \subset [-M, M]$ tal que $P_n(K) \geq 1 - \epsilon$. Por la monotonicidad de la probabilidad $P_n([-M, M]) \geq 1 - \epsilon$. Se tiene entonces:

$$P_n((M, \infty)) \leq \epsilon \iff F_n(M) \geq 1 - \epsilon, \quad P_n((-\infty, -M]) \leq \epsilon \iff F_n(-M) \leq \epsilon,$$

uniformemente en n . Es decir, $F_n(x) \rightarrow 1$ cuando $x \rightarrow \infty$ y $F_n(x) \rightarrow 0$ cuando $x \rightarrow -\infty$.

- **Límites uniformes en las colas $\Rightarrow$ apretamiento:** Por hipótesis podemos elegir $M^+ > 0$ y $M^- < 0$ tales que:

$$\inf_n F_n(M^+) \geq 1 - \epsilon/2, \quad \sup_n F_n(M^-) \leq \epsilon/2.$$

Sea $M = \max\{-M^-, M^+\}$. Para todo n y dado que F_n es no decreciente se cumple:

$$P_n([-M, M]) = F_n(M) - \lim_{x \uparrow -M} F_n(x) \geq F_n(M) - F_n(-M) \geq (1 - \epsilon/2) - (\epsilon/2) = 1 - \epsilon.$$

Así, para un $\epsilon > 0$ arbitrario, es posible encontrar un compacto de $\mathbb{R}$, $K = [-M, M]$, que satisface el requerimiento de apretamiento $P_n(K) \geq 1 - \epsilon$ para todo n .

2. Caracterización en distribuciones normales:

Una clase de variables aleatorias normales $\mathcal{N}$ es apretada si y sólo si sus medias y varianzas son acotadas.

- **Medias y varianzas acotadas $\Rightarrow$ apretamiento:** Sea la variable aleatoria $X : (\Omega, \mathcal{F}, P) \rightarrow \mathbb{R}$ con función de distribución sea Normal de media $\mu < M < \infty$ y varianza $\sigma^2 < V < \infty$. Denotamos el conjunto $\{\omega \in \Omega : |X(\omega)| > A\}$ simplemente como $\{|X| > A\}$. Si expresamos este conjunto como desviaciones respecto de la media μ podremos aplicar la desigualdad de Chebyshev.

Dado un ω , por la desigualdad triangular en $\mathbb{R}$:

$$|X(\omega)| = |(X(\omega) - \mu) + \mu| \leq |X(\omega) - \mu| + |\mu|.$$

Para $A > |\mu|$ se tiene la desigualdad:

$$|X(\omega)| > A \Rightarrow |X(\omega) - \mu| + |\mu| > A \Rightarrow |X(\omega) - \mu| > A - |\mu|.$$

Esta desigualdad *por punto* implica la siguiente inclusión de conjuntos:

$$\{|X| > A\} \subset \{|X - \mu| > A - |\mu|\}.$$

Como $|\mu| \leq M$, $A - |\mu| \geq A - M$ (tanto si $|\mu| < A < M$ como cuando $A > M$), se tiene

$$\{|X - \mu| > A - |\mu|\} \subset \{|X - \mu| > A - M\},$$

es decir (por la desigualdad anterior),

$$\{|X| > A\} \subset \{|X - \mu| > A - M\}.$$

Por la monotonicidad de la probabilidad:

$$P(|X| > A) \leq P(|X - \mu| > A - M).$$

Aplicando la desigualdad de Chebyshev con $E[X - \mu] = 0$ y $E[(X - \mu)^2] = \sigma^2 < V$:

$$P(|X| \geq A) \leq P(|X - \mu| \geq A - M) \leq \frac{\sigma^2}{(A - M)^2} \leq \frac{V}{(A - M)^2}.$$

Para $\epsilon > 0$ arbitrario, basta ahora elegir A tal que $V/(A - M)^2 \leq \epsilon$, ya que entonces la desigualdad anterior implica la condición de tightness para el conjunto $[-A, A]$ acotado de $\mathbb{R}$:

$$P(X \in [-A, A]) = 1 - P(|X| > A) \geq 1 - \epsilon.$$

- **Apretamiento $\Rightarrow$ medias y varianzas acotadas:** Sea por ejemplo $\epsilon = \frac{1}{4}$. Por hipótesis existe $M > 0$ tal que para toda $X \in \mathcal{N}$

$$P(\{\omega \in \Omega : |X(\omega)| \leq M\}) \equiv P(|X| \leq M) \geq \frac{3}{4}. \quad (16)$$

Si $\mu > M$ (similar si $\mu < -M$), entonces el intervalo acotado $[-M, M]$ queda a la derecha (izquierda) de μ y se tienen las desigualdades

$$P(|X| \leq M) \leq P(X \leq M) = P\left(\frac{X - \mu}{\sigma} \leq \frac{M - \mu}{\sigma}\right) \leq P(Z \leq 0) = \frac{1}{2},$$

donde Z es la distribución Normal estándar. La anterior desigualdad está en contradicción con (16) (tanto para $\mu > M$ como $\mu < -M$). Por tanto, debe ser $\mu \in [-M, M]$, es decir, las medias de la clase $\mathcal{N}$ son acotadas.

Sabiendo que $|\mu| < M$ para toda $X \in \mathcal{N}$, supongamos por contradicción que la varianzas no son acotadas. Podemos seleccionar entonces una sucesión $\{\sigma_k^2\}$ que tienda a ∞ . Si definimos $X_k = \mu_k + \sigma_k Z$, se tiene:

$$\begin{aligned} P(\{\omega \in \Omega : |X(\omega)| \leq M\}) &\equiv P(|X_k| \leq M) = P(-M \leq \mu_k + \sigma_k Z \leq M) = \\ &= P\left(\frac{-M - \mu_k}{\sigma_k} \leq Z \leq \frac{M - \mu_k}{\sigma_k}\right). \end{aligned}$$

Pero la longitud del intervalo $\left[\frac{-M - \mu_k}{\sigma_k}, \frac{M - \mu_k}{\sigma_k}\right]$ es $2M/\sigma_k$, que tiende a 0 para la sucesión $\{\sigma_k^2\} \rightarrow \infty$. De manera que $P(|X_k| \leq M)$ tiende a 0 (a la probabilidad de Z en un punto), en contradicción la hipótesis de apretamiento, que implica la desigualdad (16) para el M elegido. Las varianzas de $\mathcal{N}$ deben estar, por tanto, acotadas.

3. Suficiencia de integrabilidad en $\mathbb{R}$:

Una sucesión de distribuciones de variables aleatorias $\{X_n : (\Omega, \mathcal{F}, P) \rightarrow \mathbb{R}\}$ es apretada si es uniformemente integrable en $\mathbb{R}$.

- **Integrabilidad uniforme $\Rightarrow$ apretamiento:** Por hipótesis existe una constante $C < \infty$ tal que:

$$\sup_n E|X_n| \leq C.$$

Dado $\epsilon > 0$ arbitrario elegimos $M > 0$ suficientemente grande para que $\frac{C}{M} < \epsilon$. Por la desigualdad de Markov, para todo n se cumple

$$P(|X_n| > M) \leq \frac{E|X_n|}{M} \leq \frac{C}{M} < \epsilon.$$

Es decir, para todo n

$$P(X_n \in [-M, M]) \geq 1 - \epsilon.$$

Así, para todo n la función de distribución de X_n , F_n , satisface:

$$F_n([-M, M]) \geq 1 - \epsilon,$$

que es la condición de apretamiento para las sucesión $\{X_n\}$.

- **Apretamiento $\nRightarrow$ integrabilidad uniforme. Contraejemplo:** Usando la idea opuesta a los ejemplos que muestran *filtraciones* de la masa de probabilidad, podemos construir el siguiente contraejemplo. Sea $\{X_n\}$ una sucesión de variables aleatorias definida por:

$$P(X_n = n) = \frac{1}{n}, \quad P(X_n = 0) = 1 - \frac{1}{n}.$$

- $\{X_n\}$ es apretada: Dado $M > 0$, si $n \leq M$ entonces $P(|X_n| > M) = 0$, y si $n > M$ entonces $P(|X_n| > M) = 1/n$. Así,

$$\sup_n P(|X_n| > M) = \frac{1}{M+1} \xrightarrow{M \rightarrow \infty} 0.$$

Así, para cualquier ϵ podemos encontrar M tal que el compacto $[-M, M]$ acumula $1 - \epsilon$ de la probabilidad.

- No se cumple la condición de integrabilidad uniforme: Dado $M > 0$ y $n > M$ se tiene $\{|X_n| > M\} = \{X_n = n\}$. Así,

$$\mathbb{E}(|X_n| \mathbf{1}_{\{|X_n| > M\}}) = n \cdot \frac{1}{n} = 1.$$

De manera que para cualquier M

$$\sup_n \mathbb{E}(|X_n| \mathbf{1}_{\{|X_n| > M\}}) \geq 1, \quad \text{no tiende a 0 con } M.$$

En conclusión, vemos que aunque la condición de apretamiento asegura que la mayor parte de la masa de probabilidad $1 - \epsilon$ se concentra en un conjunto compacto. esto no es suficiente para garantizar la integrabilidad uniforme. La condición de integrabilidad uniforme es más fuerte que la condición de apretamiento porque requiere controlar globalmente la probabilidad de todos los conjuntos del espacio.

4. Caracterización mediante marginales:

Una familia $\mathcal{P}$ de medidas de probabilidad sobre el espacio producto $S' \times S''$ es apretada si y sólo si las distribuciones marginales lo son en S' y S'' respectivamente.

- **Apretamiento en $S' \times S'' \Rightarrow$ apretamiento de las marginales:** Dado $\epsilon > 0$, por hipótesis existe un compacto $K \subset S' \times S''$ tal que:

$$\inf_{P \in \mathcal{P}} P(K) \geq 1 - \epsilon.$$

La proyección en la primera coordenada $\pi : S' \times S'' \rightarrow S'$ es continua y, por tanto, aplica compactos en compactos (Teorema M26.5). Así $K' \equiv \pi(K) \subset S'$ es compacto. Además, como $K \subset K' \times S''$, para todo $P \in \mathcal{P}$ se tiene:

$$P'(K') = P(K' \times S'') \geq P(K) \geq 1 - \epsilon.$$

Así, $\inf_{P \in \mathcal{P}} P'(K') \geq 1 - \epsilon$, que es la condición de apretamiento para $\mathcal{P}'$. El mismo argumento se aplica a $\mathcal{P}''$.

- **Apretamiento de las marginales S' , $S'' \Rightarrow$ apretamiento en $S' \times S''$:** Dado $\epsilon > 0$, por hipótesis existen compactos $K' \subset S'$ y $K'' \subset S''$ tales que:

$$\begin{aligned} \inf_{P \in \mathcal{P}} P'(K') &\geq 1 - \epsilon/2. \\ \inf_{P \in \mathcal{P}} P''(K'') &\geq 1 - \epsilon/2. \end{aligned} \tag{17}$$

Sabemos que el producto numerable de compactos es compacto (Teorema M26.7). De modo que $K' \times K''$ es compacto en $S' \times S''$. Para todo $P \in \mathcal{P}$ consideramos los siguientes conjuntos complementarios:

$$(S' \times S'') \setminus (K' \times K'') \equiv (K' \times K'')^c \subset ((S' \setminus K') \times S'') \cup (S' \times (S'' \setminus K'')).$$

Por monotonicidad y subaditividad de la probabilidad:

$$P((K' \times K'')^c) \leq P((S' \setminus K') \times S'') + P(S' \times (S'' \setminus K'')) = P'(S' \setminus K') + P''(S'' \setminus K'').$$

Aplicando las cotas anteriores (17): $P'(S' \setminus K') \leq \epsilon/2$ y $P''(S'' \setminus K'') \leq \epsilon/2$, de manera que $P((K' \times K'')^c) \leq \epsilon$, cumpliéndose la condición de apretamiento $S' \times S''$:

$$P(K' \times K'') > 1 - \epsilon \quad \forall P \in \mathcal{P}.$$

El Teorema B2.8 pp. 23 que no presentamos aquí muestra una idea que ya sabemos: las distribuciones marginales no determinan la distribución conjunta (salvo independencia). En la demostración anterior vemos que podemos construir compactos en el espacio producto a partir de compactos en cada una de las coordenadas. Esto hace que la propiedad de apretamiento se pueda analizar desde las distribuciones marginales sin necesidad de considerar la posible dependencia de estas distribuciones.

4.2.2. Teorema principal de caracterización del apretamiento

De la definición de apretamiento se deduce que esta propiedad requiere que exista un conjunto compacto que acumule casi toda la probabilidad. Cuando la masa de probabilidad se *filtra* fuera de cualquier conjunto compacto, la familia de probabilidades no es apretada. Un ejemplo sencillo es la sucesión $\{\delta_n\}$ donde la probabilidad se concentra para cada elemento de la sucesión en n .

Los siguientes teoremas establecen condiciones suficientes y necesarias para que una sucesión de medidas de probabilidad sea apretada. Estas condiciones se pueden interpretar como la *traducción* probabilística de las condiciones de Arzelà-Ascoli (se satisfacen en todo punto del espacio salvo en un conjunto de medida cero). En el texto de referencia pueden encontrarse las demostraciones en detalle.

Teorema 4.3 (Caracterización en espacio $C[0, 1]$. Teorema B7.3.). *Sea $\{P_n\}$ una sucesión de medidas de probabilidad en $(C, \mathcal{C})$. La sucesión $\{P_n\}$ es apretada si y sólo si se satisfacen:*

- (i) $\forall \eta > 0$, existen a y n_0 tales que
$$P_n[x : |x(0)| \geq a] \leq \eta, \quad n \geq n_0.$$
- (ii) $\forall \epsilon > 0$ y $\eta > 0$, existen $\delta \in [0, 1]$ y n_0 tales que
$$P_n[x : w_x(\delta) \geq \epsilon] \leq \eta, \quad n \geq n_0.$$

Teorema 4.4 (Caracterización en espacio $D[0, 1]$. Teorema B13.2.). *Sea $\{P_n\}$ una sucesión de medidas de probabilidad en $(D, \mathcal{D})$. La sucesión $\{P_n\}$ es apretada si y sólo si se satisfacen:*

- (i) $\lim_{a \rightarrow \infty} \limsup_n P_n(x : \|x\| \geq a) = 0.$
- (ii) $\forall \epsilon > 0 \quad \lim_{\delta} \limsup_n P_n(x : w'_x(\delta) \geq \epsilon) = 0.$

4.3. Condiciones para la convergencia a partir de leyes finito dimensionales

Dado un conjunto finito $0 \leq t_1 < \dots < t_k \leq 1$, se define la **proyección** de $C[0, 1]$ o $D[0, 1]$ a $\mathbb{R}^k$ como $\pi_{t_1 \dots t_k}(x) = (x(t_1), \dots, x(t_k))$. En C o D los **conjuntos finito dimensionales** son de la forma $\pi_{t_1 \dots t_k}^{-1} H$ con $H \in \mathcal{R}^k$ (la álgebra de Borel de $\mathbb{R}^k$). Dado que las proyecciones son continuas, estos conjuntos pertenecen a la álgebra de Borel de cada espacio.

A diferencia del espacio de sucesiones infinitas $\mathbb{R}^\infty$, donde la familia de conjuntos finito dimensionales determina la convergencia (Ejemplo B2.4 p. 19), en los espacios de funciones $C[0, 1]$ y $D[0, 1]$ la convergencia no está determinada por las distribuciones finito dimensionales (Ejemplo B2.5 p. 19).

Comprobar si las distribuciones finito dimensionales convergen suele ser en la práctica más sencillo que comprobar la convergencia en el espacio de trayectorias completo. Es por esto que resulta de gran utilidad conocer condiciones bajo las cuales la convergencia de las leyes de probabilidad finito dimensionales asegura la convergencia de la sucesión de medidas de probabilidad. En este sentido, el siguiente teorema es de especial relevancia.

Teorema 4.5 (Teorema B13.3). *Si se satisfacen:*

$$(i) \quad P_n \pi_{t_1 \dots t_k}^{-1} \Rightarrow P \pi_{t_1 \dots t_k}^{-1} \quad \text{para toda partición tal que } \{t_i\} \in T_P \text{ definido en § 2.4, p.10.} \quad (18)$$

$$(ii) \quad \forall \epsilon > 0, \quad \lim_{\delta \rightarrow 0} P(x : |x(1) - x(1 - \delta)| \geq \epsilon) = 0. \quad (19)$$

$$(iii) \quad \forall \epsilon > 0 \quad \forall \eta > 0 \quad \exists \delta \in (0, 1) \quad \exists n_0 \text{ tales que } P_n(x : w_x''(\delta) \geq \epsilon) \leq \eta \text{ para } n \geq n_0. \quad (20)$$

Entonces $P_n \Rightarrow P$.

Demostración 1 En la demostración de Billingsley se comprueba que las condiciones de este teorema son equivalentes a tightness. Previamente se ha demostrado que bajo la propiedad de tightness, la convergencia de las distribuciones finito-dimensionales implica convergencia débil de la sucesión (Teorema B13.1).¹⁰

Demostración 2 Después de todo, este Teorema también se puede demostrar sin referencia expresa a la propiedad de tightness o al Teorema de Prohorov. Siguiendo la sugerencia del Ejercicio B13.1 y la demostración del Teorema análogo para el espacio $C[0, 1]$, adaptaré al espacio $D[0, 1]$ esta demostración alternativa.

- **Aproximación por medio de funciones escalonadas A_σ :** Como se describen en § 3.5, tomamos una partición $\sigma = \{0 = s_0 < s_1 < \dots < s_k = 1\}$ para aproximar $x \in D$ mediante $A_\sigma : D \rightarrow D$:

$$(A_\sigma(x))(t) = x(s_{u-1}) \quad \text{para } t \in [s_{u-1}, s_u), \quad (A_\sigma(x))(1) = x(1).$$

En general, para $\alpha \in \mathbb{R}^{k+1}$, $\alpha = (\alpha_0, \dots, \alpha_k)$, definimos la aplicación $L_\sigma : \mathbb{R}^{k+1} \rightarrow D$ como:

$$(L_\sigma \alpha)(t) = \alpha_{u-1} \quad \text{en } [s_{u-1}, s_u), \quad (L_\sigma \alpha)(1) = \alpha_k.$$

La aproximación es así la composición:

$$A_\sigma = L_\sigma \circ \pi_{s_0, \dots, s_k}.$$

Sean $\alpha, \beta \in \mathbb{R}^{k+1}$, tomando $\lambda(t) = t$ (la identidad) en la métrica de Skorohod se tiene:

$$d(L_\sigma \alpha, L_\sigma \beta) \leq \sup_{t \in [0, 1]} |(L_\sigma \alpha)(t) - (L_\sigma \beta)(t)| = \max_{0 \leq i \leq k} |\alpha_i - \beta_i|,$$

y por tanto L_σ es continua con la métrica de Skorohod.

¹⁰En el espacio $C[0, 1]$ el teorema análogo es el Teorema B7.5. Pedagógicamente, su demostración resulta más fiel a la sucesión de resultados presentados hasta entonces en el texto de Billingsley: Se comprueba la propiedad de tightness. Esto permite aplicar el Teorema de Prohorov (enunciado aquí en § 3.6, p. 23), que asegura que la propiedad de tightness implica compacidad relativa. Bajo esta última propiedad, la convergencia finito dimensional implica convergencia en todo el espacio (Ejemplo B5.1).

- **Sucesión de aproximaciones:** Definimos una sucesión de particiones $\sigma(u) = \{s_0^{(u)}, \dots, s_{k(u)}^{(u)}\}$ con $s_i^{(u)} \in T_P$ y $s_0^{(u)} = 0$, $s_{k(u)}^{(u)} = 1$, tales que su tamaño tiende a cero:¹¹

$$\|\sigma(u)\| = \max_i (s_i^{(u)} - s_{i-1}^{(u)}) \rightarrow 0.$$

Tenemos así una sucesión de aproximaciones $\{A_{\sigma(u)}\}_u$.

- **$P_{un} \Rightarrow_n P_u$: Aplicación de hipótesis de convergencia finito-dimensional:** Por la hipótesis *i*) de convergencia finito-dimensional, las particiones finitas $\sigma(u) \subset T_P$ satisfacen:

$$P_n \circ \pi_{\sigma(u)}^{-1} \Rightarrow P \circ \pi_{\sigma(u)}^{-1} \quad \text{en } \mathbb{R}^{k(u)+1}.$$

Se ha visto que $L_{\sigma(u)}$ es continua. Por el teorema de continuidad:

$$P_n \circ A_{\sigma(u)}^{-1} := (P_n \circ \pi_{\sigma(u)}^{-1}) \circ L_{\sigma(u)}^{-1} \Rightarrow (P \circ \pi_{\sigma(u)}^{-1}) \circ L_{\sigma(u)}^{-1} := P \circ A_{\sigma(u)}^{-1}.$$

Se cumple así la primera condición del Teorema 3.3 p. 18, es decir, para u fijo se tiene convergencia cuando $n \rightarrow \infty$:

$$P_n \circ A_{\sigma(u)}^{-1} \Rightarrow_n P \circ A_{\sigma(u)}^{-1}.$$

Equivalentemente, en términos de distribuciones:

$$A_{\sigma(u)} X^n \Rightarrow_n A_{\sigma(u)} X.$$

- **$P_u \Rightarrow_u P$:** Para $\|\sigma(u)\| \leq \delta$ se satisface la siguiente desigualdad (§ 3.5 p. 23):

$$d(A_{\sigma(u)}(x), x) \leq \delta \vee w'_x(\delta).$$

Así, para todo $x \in D$ (caracterizado por $w'_x(\delta) \rightarrow 0$ cuando $\delta \rightarrow 0$) se satisface:

$$d(A_{\sigma(u)}(x), x) \rightarrow 0 \quad \text{cuando } \|\sigma(u)\| \rightarrow 0.$$

Por la estabilidad de la convergencia ante pequeñas perturbaciones (Teorema 3.1, p. 17):

$$P \circ A_{\sigma(u)}^{-1} \Rightarrow_u P, \quad \text{equivalentemente} \quad A_{\sigma(u)} X \Rightarrow_u X.$$

- **Reformulación de la hipótesis *iii*) del Teorema:** Se puede expresar en términos de la última condición del Teorema 3.3 que queremos validar y que en el contexto actual denotamos:

$$\lim_u \limsup_n P(d(A_{\sigma(u)}(X^n), X^n) \geq \epsilon) = 0.$$

De nuevo, por el Lema B3 p. 127:

$$d(A_{\sigma_u}(x), x) \leq \delta_u \vee w'(x, \delta_u).$$

¹¹Esto es posible porque el conjunto T_P introducido en § 2.4 p. 10 es denso en $[0, 1]$.

Empleando la desigualdad justificada en § 2.3.1:

$$d(A_{\sigma_u}(x), x) \leq \delta_u 2w''(x, \delta) \vee \sup_{0 \leq t \leq \delta} |x(t) - x(0)| \vee \sup_{1-\delta \leq t \leq 1} |x(1) - x(t)|.$$

Así, $(a_{X^n}$ y b_{X^n} representan el supremo en los extremos, como en § 2.3.1 8):

$$\begin{aligned} P[d(A_{\sigma_u}(X^n), X^n) \geq \epsilon] &\leq P[\delta_u \geq \epsilon] + P[2w''(X^n, \delta_u) \geq \epsilon] \\ &\quad + P[a_{X^n}(\delta_u) \geq \epsilon] + P[b_{X^n}(\delta_u) \geq \epsilon]. \end{aligned}$$

Tomando $u \rightarrow 0$ y $n \rightarrow \infty$, la anterior probabilidad tiende a cero, ya que: (1) $\delta_u \rightarrow 0$; (2) la probabilidad dependiente de w'' tiende a 0 precisamente por la condición *iii*) p. 31 requerida en este Teorema; (3) la probabilidad referente al extremo izquierdo se anula por la continuidad por la derecha de las funciones cadlag; (4) la probabilidad del término que controla las oscilaciones en el extremo derecho se anula por la condición *ii*) p. 31 del Teorema. En conclusión, se satisface la condición formulada al comienzo de este punto.

- **Aplicación del Teorema de aproximación:** Se ha justificado que $A_{\sigma(u)}X^n \Rightarrow_n A_{\sigma(u)}X \Rightarrow_u X$ y que $\lim_u \limsup_n P(d(A_{\sigma(u)}(X^n), X^n) \geq \epsilon) = 0$. Esto permite aplicar el Teorema 3.3 p. 18, concluyendo que en las condiciones del Teorema 4.5 p. 31, en $D[0, 1]$ la convergencia de las distribuciones (leyes de probabilidad) finito dimensionales implica $X_n \Rightarrow X$ ($P_n \Rightarrow P$).

5. Convergencia débil en espacio localmente compacto.

Se dice que una **clase o familia $\mathcal{F}$ de funciones es separadora** si $Pf := \int_S f dP \rightarrow \int_S f dQ := Qf$ para todo $f \in \mathcal{F}$ implica que las medidas de probabilidad son equivalentes: i.e., $P(A) = Q(A)$ para todo conjunto A de la σ -álgebra.¹²

En este apartado se relacionarán varios ejercicios propuestos en Billingsley que ilustran una de las enseñanzas principales del texto. Se verá que en un espacio separable, localmente compacto la clase de funciones continuas con soporte compacto (denotada $C_c(S)$) es separadora (determina las medidas de probabilidad). No obstante, una clase separadora no determina necesariamente la convergencia de las medidas de probabilidad en un espacio general. El motivo es que la convergencia requiere una propiedad global que asegure que una parte relevante de la masa probabilidad no se *filtra* fuera de un conjunto compacto. Esto es precisamente lo que evita la propiedad de apretamiento. Si nos restringimos a un espacio de medida que sea separable y localmente compacto, veremos que la familia de funciones $C_c(S)$, además de determinar las medidas de probabilidad, caracterizan la convergencia débil: Si $P_n f \rightarrow P f$ para todo $f \in C_c(S)$ entonces $P_n \Rightarrow P$. El motivo será que, en este contexto, el supuesto $P_n f \rightarrow P f$ fuerza que la sucesión $\{P_n\}$ sea apretada.

La exposición de los siguientes resultados se hace en orden inverso a la forma en que se han trabajado. Los desarrollos se han realizado aplicando estrategias habituales vistas en cursos de

¹²La familia de funciones acotadas y uniformemente continuas forman una clase separadora para un espacio métrico general S (equivalencias de Portmanteau § 3.1 p. 15).

Topología, Integral de Lebesgue y Espacios Normados. Posteriormente se ha revisado que se cumplen los requisitos para aplicarlas en la topología del espacio. Aquí se describen primero las propiedades, lo que permitirá simplificar la justificación de los resultados.

S_{lc} : En lo siguiente S_{lc} denotará un espacio métrico separable y localmente compacto. Las propiedades de un espacio S_{lc} serán empleadas para justificar lo siguiente:

1. **En espacio localmente compacto se tiene:** Dado un compacto $K_0 \subset S_{lc}$ existe otro compacto $K_1 \subset S_{lc}$ tal que $K_0 \subset \text{int}(K_1)$. Por la compacidad local, para cada $x \in K_0$ existe un entorno abierto U_x de x con clausura compacta (Teorema M29.2). La familia de conjuntos $\{U_x : x \in K_0\}$ es un cubrimiento abierto de K_0 , que es compacto y, por tanto, podemos elegir un subcubrimiento finito $\{U_{x_1}, \dots, U_{x_n}\}$. Tomando $K_1 \cup_i \bar{U}_{x_i}$, $i = 1, \dots, n$ se tiene que K_1 es compacto (unión finita de compactos) y $K_0 \subset K_1$.

2. **En espacio métrico se tiene:** Por la normalidad de los espacios métricos (Teorema M32.2), el Lema de Urysohn (Teorema M33.1) permite definir una función continua $\varphi : S_{lc} \rightarrow [0, 1]$ tal que:

$$\begin{cases} \varphi(x) = 1 & \text{si } x \in K_0 \subset \text{int}(K_1), \\ \varphi(x) = 0 & \text{si } x \notin \text{int}(K_1). \end{cases}$$

3. **En espacio separable:** Toda medida de probabilidad es (individualmente) tight (Teorema B1.3). Adicionalmente, este resultado requiere completitud, que no es esencialmente una propiedad topológica del espacio, sino de la métrica. Asumimos, por tanto también que la métrica del espacio es completa. En realidad, como en esta sección no vamos a trabajar con una métrica concreta, esto no será estrictamente un supuesto, ya que todo espacio métrico se puede embeber isométricamente en un espacio métrico completo (Teorema M43.7). La completitud del espacio permite justificar que la clausura de un conjunto totalmente acotado (definido en Teorema B1.3 a partir del conjunto denso del espacio) es compacta (Teorema M45.1).

1. **$C_c(S_{lc})$ es una familia separadora:** Este es un resultado preliminar propuesto en Ejercicio B1.11. En principio podríamos pensar en aproximar funciones uniformemente continuas acotadas por funciones de soporte compacto. Sin embargo esto no es posible si el espacio no es compacto (aquí es localmente compacto); por ejemplo, la función constante $f(x) = 1$ para todo $x \in S_{lc}$ no se puede aproximar de esta forma.

En su lugar, aplicamos la siguiente idea: La regularidad de una medida de probabilidad implica que una medida de probabilidad está completamente determinada por su valor en conjuntos cerrados. Esto se puede reformular de forma que una medida de probabilidad está determinada por funciones uniformemente continuas acotadas (Teoremas B1.1, B1.2). Para ello se aproximan conjuntos cerrados a través de funciones indicatrices. Siguiendo la misma idea, aquí aproximaremos conjuntos compactos a través de funciones $C_c(S_{lc})$. Reformularemos con ello la hipótesis $Pf = Qf$ para toda $f \in C_c(S)$ como $P = Q$ en

compactos. Por la regularidad de toda medida de probabilidad (Valdivia, 1988, Teorema B1.1) esto es extensible a conjuntos de Borel.

Por las Propiedades 1 y 2 anteriores, podemos tomar los compactos K_0 y K_1 sobre los que definir la función de soporte compacto $\varphi : S_{lc} \rightarrow [0, 1]$.

Por la monotonía de la probabilidad:

$$P(K_0) \leq P\varphi \equiv \int_{S_{lc}} \varphi dP = Q \equiv \int_{S_{lc}} \varphi dQ \leq Q(K_1).$$

Reemplazamos ahora K_1 por la clausura de una sucesión de la clausura de abiertos U_m (compactos) de la forma:

$$K_0 \subset U_{m+1} \subset \overline{U_{m+1}} \subset U_m, \quad \bigcap_m U_m = K_0,$$

Los conjuntos $\overline{U}$ son compactos y podemos aplicar la desigualdad anterior:

$$P(K_0) \leq P\varphi_m = Q\varphi_m \leq Q(\overline{U_m}).$$

Dado que $\overline{U_m} \downarrow K$, tomando $m \rightarrow \infty$ se tiene por la continuidad de Q

$$\lim_{m \rightarrow \infty} Q(\overline{U_m}) = Q(K).$$

Hemos visto por tanto que $P(K_0) \leq Q(K_0)$. De la misma forma se comprueba que $Q(K) \leq P(K_0)$. Así,

$$P(K) = Q(K) \quad \text{para todo compacto } K.$$

2. **Convergencia en $C_c(S_{lc}) \Rightarrow$ apretamiento:** Por la regularidad de toda medida de probabilidad, la probabilidad de cualquier abierto $A \in S_{lc}$ se puede aproximar por compactos:

$$P(A) = \sup\{P(K) : K \subset A, K \text{ compacto}\}, \quad A \text{ abierto.}$$

Podemos así elegir un compacto $K_0 \subset S_{lc}$ que contenga casi toda la probabilidad, i.e.:

$$P(K_0) \geq 1 - \epsilon/2, \quad \text{para cualquier } \epsilon > 0.$$

Por la Propiedad 1 anterior, podemos elegir un compacto K_1 tal que $K_0 \subset \text{int}(K_1)$. De nuevo por la Propiedad 2 definimos a partir de estos compactos la función función continua $\varphi : S \rightarrow [0, 1]$. De esta construcción anterior, tenemos:

$$\int_S \varphi dP \equiv P\varphi \geq \int_{K_0} \varphi dP \equiv P(K_0) \geq 1 - \epsilon/2.$$

Por hipótesis

$$P_n f \rightarrow P f \quad \forall f \in C_c(S_{lc}), \quad \text{y en particular para } \varphi.$$

Existe por tanto N tal que para todo $n \geq N$ se cumple:

$$P_n \varphi \geq P \varphi - \epsilon/2 \geq 1 - \epsilon$$

y por la monotonicidad de la probabilidad (φ se anula fuera de K_1),

$$P_n \varphi \leq P_n(K_1).$$

En este punto se cumple la condición de tightness $P_n(K_1) \geq 1 - \epsilon$ para todo $n \geq N$.

Por la Propiedad 3 anterior, para cada medida de probabilidad P_n podemos obtener un compacto K que satisfaga la condición de tightness para $n = 1, \dots, N - 1$. Para cada P_n existen compactos $K^{(n)} \in S_{lc}$ para los que

$$P_n(K^{(n)}) \geq 1 - \epsilon.$$

Definimos el compacto (unión numerable de compactos)

$$K \equiv K_1 \cup \bigcup_{n=1}^{N-1} K^{(n)},$$

con el que se tiene para $n = 1, \dots, N - 1$:

$$P_n(K) \geq P_n(K^{(n)}) \geq 1 - \epsilon.$$

En conclusión,

$$\inf_n P_n(K) \geq 1 - \epsilon \quad \text{para todo } n,$$

satisfaciéndose así en este contexto la condición de tightness de la familia $\{P_n\}$.

3. **Tightness permite justificar $P_n \Rightarrow P$:** Una vez comprobado que en el espacio S_{lc} la convergencia en $C_c(S_{lc})$ implica tightness en $\{P_n\}$, veremos que esto permite deducir la convergencia a una medida de probabilidad P . Para ello aplicamos la siguiente línea argumental que se describe y justifica en los primeros capítulos de Billingsley.

Por el Teorema de Prohorov la propiedad de tightness comprobada en el punto anterior implica compacidad relativa de las medidas de probabilidad $\{P_n\}$. Esto es, toda subsucesión $\{P_{n_k}\}$ contiene a su vez otra subsucesión $\{P_{n_{k_j}}\}$ convergente a una medida de probabilidad que denotamos Q :

$$P_{n_{k_j}} \Rightarrow Q.$$

Recordamos que la hipótesis inicial es convergencia (de la sucesión completa) en $C_c(S_{lc})$:

$$\int_{S_{lc}} f dP_n \equiv P_n f \Rightarrow \int_{S_{lc}} f dP \equiv P f \quad \text{para toda } f \in C_c(S_{lc}).$$

De manera que la subsucesión $\{P_{n_{k_j}} f\}$ debe converger también a $P f$ y debe ser, por tanto, $Q f = P f$ para toda $f \in C_c(S_{lc})$.

Hemos visto en el paso preliminar 1 que en el espacio S_{lc} , la clase $C_c(S_{lc})$ es separadora (basta que dos medidas de probabilidad coincidan en $C_c(S_{lc})$ para las medidas sean idénticas). En consecuencia, $Q f = P f$ también para toda f acotada y uniformemente continua. Podemos aplicar así la equivalencia de Portmanteau *ii*) § 3.1 y concluir que $P_n \Rightarrow P$.

6. Un Teorema funcional del límite. El Teorema de Donsker.

El Teorema Central del Límite (TCL) establece la convergencia de variables aleatorias en dimensión finita idénticamente distribuidas,¹³ hacia una normal. El Teorema de Donsker es su versión funcional: trata la convergencia de las leyes de probabilidad de trayectorias en espacios de funciones.

El reto de la teoría de convergencia en $D[0, 1]$ es que la convergencia de las distribuciones finito dimensionales no aseguran la convergencia débil en el espacio completo. Como se ha visto, se requiere la propiedad de apretamiento.

El escenario para este teorema son procesos construidos como sumas parciales normalizadas. Estas serán las trayectorias discretas (escalonadas o interpoladas) que se analizan para demostrar la convergencia funcional: primero por las distribuciones finito dimensionales (TCL multivariante) y luego mediante apretamiento en el espacio $D[0, 1]$ introducido en las secciones anteriores.

Se definen las siguientes sumas parciales: Dadas una sucesión de variables aleatorias $z_1, z_2, \dots$ definidas en un espacio de probabilidad $(\Omega, \mathcal{F}, P)$, se definen las sumas parciales $S_n = z_1 + \dots + z_n$ y con ello las siguientes funciones $\tilde{X}^n(\omega) \in C[0, 1]$ y $X^n(\omega) \in D[0, 1]$ que en t toman el valor:

$$\tilde{X}_t^n(\omega) = \frac{1}{\sigma\sqrt{n}}S_{[nt]}(\omega) + (nt - [nt])\frac{1}{\sigma\sqrt{n}}z_{[nt]+1}(\omega), \quad t \in [0, 1]. \quad (21)$$

$$X_t^n(\omega) = \frac{1}{\sigma\sqrt{n}}S_{[nt]}(\omega). \quad (22)$$

Teorema 6.1. *Teorema de Donsker. Sean z_n variables aleatorias independientes e idénticamente distribuidas con media 0 y varianza σ^2 . Para las funciones aleatorias definidas en la expresión (22) se cumple la convergencia $X^n \Rightarrow W$, donde W es un proceso de Wiener.*¹⁴

6.1. Espacio $C[0, 1]$

Por motivos de espacio no reproduciré la demostración dada en Billingsley, que emplea los conceptos y resultados trabajados en las secciones anteriores. En su lugar, analizaremos la siguiente condición,

$$\lim_{\lambda \rightarrow \infty} \limsup_{n \rightarrow \infty} \lambda^2 P \left[\max_{k \leq n} |S_k| \geq \lambda \sigma \sqrt{n} \right] = 0. \quad (23)$$

Suponiendo z_i estacionarias, la condición anterior se emplea en B como equivalente de la siguiente condición de tightness en el espacio $C[0, 1]$ (análoga de la segunda condición en Teorema 4.4 anterior en espacio $D[0, 1]$):

$$\forall \epsilon > 0 \quad \lim_{\delta} \limsup_n P_n(x : w_x(\delta) \geq \epsilon) = 0.$$

¹³Entre otras condiciones, como son: media y varianza finitas o condiciones de Lindeberg/Feller.

¹⁴Un proceso de Wiener se define en el Apéndice § 7.

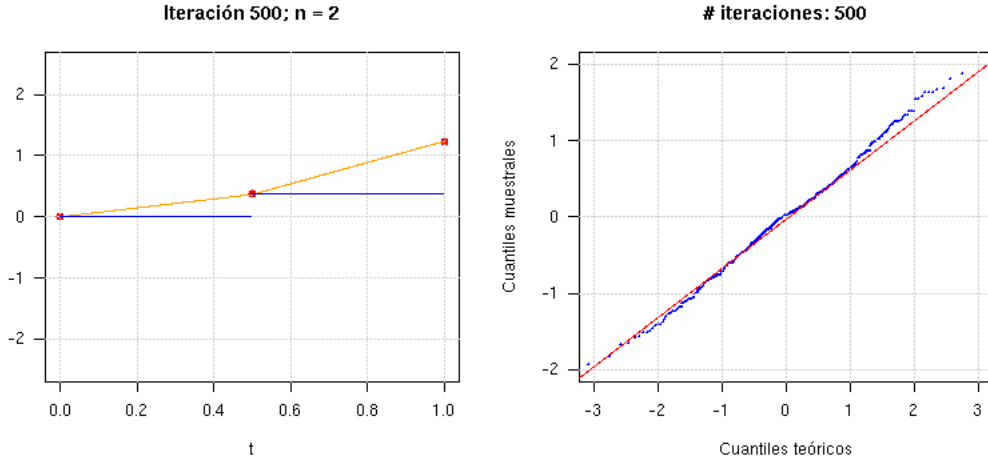


Figura 2: Sumas parciales y gráfico QQ.

Gráfico izquierdo: Procesos simulados con $z_i \sim NID(0, 1)$ (Normal estándar). Naranja: Trayectorias simuladas para el proceso $\tilde{X}^n$ (21). Azul: trayectorias azul representan el proceso X^n (22). Gráfico derecho: A medida que aumenta el número de simulaciones los cuantiles de X_t para un t elegido (en el gráfico se emplea $t = 512$) se aproximan a los cuantiles de la distribución Normal estándar. Se han simulado trayectorias a partir de variables aleatorias Normales y binomiales, observándose en ambos la convergencia. En el gráfico se muestra el caso de variables aleatorias Normales.

El Ejercicio B8.3 propone comprobar que, de hecho, la condición (23) es necesaria para que las sumas parciales $\{X^n\}$ definidas en (22) converjan a un proceso de Wiener W . Para ello el Ejercicio propone emplear la siguiente característica de un proceso de Wiener (se deduce de aplicar la hipótesis $X^n \Rightarrow W$ para obtener la distribución del máximo de las sumas parciales X^n):

$$P \left[\sup_t W_t \leq \alpha \right] = \frac{2}{\sqrt{2\pi}} \int_0^\alpha e^{-u^2/2} du = 2 \left(\Phi(\alpha) - \frac{1}{2} \right) = 2\Phi(\alpha) - 1, \quad \alpha \geq 0.$$

donde Φ es la función de distribución de la Normal estándar N . A partir de esto se sugiere demostrar que el momento de cuarto orden de $\sup_t |W_t|$ es finito y la simetría de W_t .

- Observamos que

$$2\Phi(\alpha) - 1 = \Phi(\alpha) - \Phi(-\alpha) = P(|N| \leq \alpha),$$

de manera que $\sup_{0 \leq t \leq 1} W_t$ se distribuye como la distribución del valor absoluto de la Normal estándar, $|N|$. Así, se tiene:

$$E \left[\left(\sup_t W_t \right)^4 \right] = E[|N|^4] = E[N^4] = 3 < \infty.$$

Por la simetría del proceso de Wiener, W_t se distribuye como $-W_t$ y con ello $\sup_t (-W_t)$ se distribuye como $\sup_t W_t$.

Denotando $A := \sup_t |W_t|$ $B := \sup_t W_t$:

$$A = \max \left(\sup_t W_t, \sup_t (-W_t) \right) \leq \sup_t W_t + \sup_t (-W_t).$$

Usando la desigualdad $(a + b)^4 \leq 8(a^4 + b^4)$ cuando $a, b \geq 0$,

$$A^4 \leq \left(\sup_t W_t + \sup_t (-W_t) \right)^4 \leq 8 \left(\left(\sup_t W_t \right)^4 + \left(\sup_t (-W_t) \right)^4 \right).$$

Tomando esperanzas y por la igualdad de las distribuciones comentada:

$$\mathbb{E}[A^4] \leq 8(\mathbb{E}[M^4] + \mathbb{E}[M^4]) = 16 \mathbb{E}[M^4] = 16 \cdot 3 = 48 < \infty.$$

- Para el proceso poligonal X^n definido en (22)

$$\sup_{0 \leq t \leq 1} |X_t^n| = \frac{1}{\sigma \sqrt{n}} \max_{0 \leq k \leq n} |S_k|,$$

de manera que (23) se puede escribir en términos de X_t^n :

$$\lim_{\lambda \rightarrow \infty} \limsup_{n \rightarrow \infty} \lambda^2 \mathbb{P} \left(\sup_t |X_t^n| \geq \lambda \right) = 0.$$

Por hipótesis $X^n \Rightarrow W$ en $C[0, 1]$ y por el Teorema de la Aplicación Continua para la aplicación continua $h(x) = \sup_t |x(t)|$:

$$\sup_t |X_t^n| \Rightarrow \sup_t |W_t| = A.$$

En consecuencia,

$$\mathbb{P} \left(\sup_t |X_t^n| \geq \lambda \right) \rightarrow \mathbb{P}(A \geq \lambda).$$

Tomando el límite superior en n y multiplicando por el factor λ^2 que aparece en la condición (23) (veremos que resulta útil):

$$\limsup_{n \rightarrow \infty} \lambda^2 \mathbb{P} \left(\sup_t |X_t^n| \geq \lambda \right) = \lambda^2 \mathbb{P}(A \geq \lambda).$$

Aplicando la versión de la desigualdad de Markov, por la que:

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|^n)}{a^n},$$

se tiene

$$\lambda^2 \mathbb{P}(A \geq \lambda) \leq \lambda^2 \frac{\mathbb{E}[A^4]}{\lambda^4} = \frac{\mathbb{E}[A^4]}{\lambda^2} \rightarrow 0 \quad \text{cuando } \lambda \rightarrow \infty,$$

concluyendo que si $X^n \Rightarrow W$, entonces se debe satisfacer (23).

El factor λ^2 de la condición anterior empleada en B para justificar tightness en el Teorema de Donsker resulta arbitrario o al menos no directamente observable de las condiciones originales de tightness. Este ejercicio ilustra la relevancia de este factor en la condición (23).

6.2. Espacio $D[0, 1]$

6.2.1. Paso 1. Las distribuciones finito dimensionales son convergentes

Se aplica la misma estrategia empleada en § B8 en el espacio de funciones continuas $C[0, 1]$. La demostración de Billingsley detalla especialmente la justificación de la existencia de un proceso de Wiener con las características de las distribuciones finito dimensionales de dicho proceso. Para el espacio de funciones cadlag la justificación es similar. Aquí expondré la demostración suponiendo la existencia de un proceso con las características distribucionales de un proceso de Wiener y desarrollaré algunos aspectos que no se detallan en B.

- **Expresar X^n como suma de incrementos independientes:** Sean $k \geq 1$, $0 \leq t_1 \leq \dots \leq t_k \leq 1$ y denotamos $m_i = \lfloor nt_i \rfloor$ con $m_0 = 0$. Los siguientes incrementos son función directa de las variables aleatorias $\{z\}$.

$$\Delta S_i := S_{m_i} - S_{m_{i-1}} = \sum_{j=m_{i-1}+1}^{m_i} z_j.$$

Reescribimos así el proceso X^n como

$$X_{t_i}^n = \frac{1}{\sigma\sqrt{n}} S_{m_i} = \frac{1}{\sigma\sqrt{n}} \sum_{r=1}^i \Delta S_r,$$

función lineal de sumas disjuntas de incrementos (bloques disjuntos de sumas de z_i). Dado que z_j son i.i.d., las variables aleatorias $\Delta S_1, \dots, \Delta S_k$ son independientes, con

$$\mathbb{E}[\Delta S_i] = 0, \quad \text{Var}(\Delta S_i) = (m_i - m_{i-1}) \sigma^2.$$

El vector de variables aleatorias $(X_{t_1}^n, \dots, X_{t_k}^n)$ se puede expresar, por tanto, de la forma:

$$(X_{t_1}^n, \dots, X_{t_k}^n) = \frac{1}{\sigma\sqrt{n}} (S_{m_1}, \dots, S_{m_k}) = \frac{1}{\sigma\sqrt{n}} (\Delta S_1, \Delta S_1 + \Delta S_2, \dots, \Delta S_1 + \dots + \Delta S_k).$$

Simplificando la suma telescópica, la distribución del vector será función lineal del siguiente vector de sumas parciales de z_j independientes:

$$Y^n := X_{t_i}^n - X_{t_{i-1}}^n = \frac{1}{\sigma\sqrt{n}} (\Delta S_1, \dots, \Delta S_k).$$

- **Aplicar Teorema Central del Límite (TCL) a cada bloque:** Para cada i

$$Z_n := \frac{1}{\sigma\sqrt{m_i - m_{i-1}}} \Delta S_i \Rightarrow N \quad \text{cuando } n \rightarrow \infty,$$

ya que $m_i - m_{i-1} \rightarrow \infty$ para $t_i > t_{i-1}$ y $\text{Var}(\Delta S_i) = (m_i - m_{i-1}) \sigma^2$. Reescalando para Y_i^n :

$$Y_i^n = \frac{1}{\sigma\sqrt{n}} \Delta S_i = \sqrt{\frac{m_i - m_{i-1}}{n}} \frac{1}{\sigma\sqrt{m_i - m_{i-1}}} \Delta S_i.$$

Observando que $\frac{m_i - m_{i-1}}{n} \rightarrow t_i - t_{i-1}$ cuando $n \rightarrow \infty$ se tiene:

$$Y_i^n \Rightarrow \sqrt{t_i - t_{i-1}} N.$$

- **Distribución conjunta a partir de producto de funciones características:** Sabemos que la función característica de la distribución Normal estándar es $\varphi_N(u) = \exp(-u^2/2)$. Por el TCL anterior, la función característica de Z_n converge a ella: $\varphi_{Z_n}(u) \rightarrow \exp(-u^2/2)$. Reescalando como antes:

$$Y_i^n = \sqrt{\frac{m_i - m_{i-1}}{n}} Z_n$$

y observando de nuevo que $\sqrt{\frac{m_i - m_{i-1}}{n}} \rightarrow \sqrt{t_i - t_{i-1}}$, se tiene el siguiente límite de la función característica de Y_i^n

$$\varphi_{Y_i^n}(u) \rightarrow \exp\left(-\frac{(t_i - t_{i-1})u^2}{2}\right),$$

que es la función característica de $\mathcal{N}(0, t_i - t_{i-1})$.

Dado que los componentes del vector Y^n son sumas parciales independientes, la función característica conjunta de Y^n es el producto:

$$\varphi_{Y^n}(u_1, \dots, u_k) = \prod_{i=1}^k \varphi_{Y_i^n}(u_i).$$

Tomando el límite y usando el resultado anterior para cada componente:

$$\varphi_{Y^n}(u_1, \dots, u_k) \rightarrow \prod_{i=1}^k \exp\left(-\frac{(t_i - t_{i-1})u_i^2}{2}\right),$$

que precisamente caracterizan los incrementos del proceso de Wiener (incrementos independientes de varianza $t_i - t_{i-1}$), es decir,

$$Y^n \Rightarrow (W_{t_1} - W_{t_0}, \dots, W_{t_k} - W_{t_{k-1}}),$$

- **Distribución de los niveles aplicando el Teorema de la Aplicación Continua a los incrementos:** Sea h la aplicación continua $h : \mathbb{R}^k \rightarrow \mathbb{R}^k$,

$$h(y_1, \dots, y_k) = (y_1, y_1 + y_2, \dots, y_1 + \dots + y_k).$$

Así,

$$h(Y^n) = (X_{t_1}^n, \dots, X_{t_k}^n), \quad h(W_{t_1} - W_{t_0}, \dots, W_{t_k} - W_{t_{k-1}}) = (W_{t_1}, \dots, W_{t_k}).$$

Por el Teorema de la Aplicación Continua

$$(X_{t_1}^n, \dots, X_{t_k}^n) \Rightarrow (W_{t_1}, \dots, W_{t_k}).$$

que es precisamente la distribución finito-dimensional del proceso de Wiener (incrementos independientes de varianza $t_i - t_{i-1}$) cuya existencia se justifica en Billingsley.

6.2.2. Paso 2. Las funciones de distribución de $\{X^n\}$ son apretadas

En este punto nos separamos de la demostración dada en B, basada en el Teorema B13.5. Brevemente, este teorema se apoya en el Teorema 4.4 anterior, donde se dan condiciones de tightness a partir del módulo de oscilación w'' definido en (6). Las condiciones de este teorema se pueden comprobar a partir de las desigualdades maximales descritas en § B10.4.

La observación de las trayectorias muestrales de X^n como las mostradas en la Figura 2 sugiere la siguiente conjetura: Sabiendo que las distribuciones de probabilidad del proceso poligonal $\tilde{X}^n$ (interpolación de sumas parciales) descrito en la ecuación (21) convergen en el espacio de funciones continuas, entonces las trayectorias del proceso X^n (sumas parciales escalonadas) definido en (22) deberían estar también asociadas a distribuciones que satisfacen el criterio de tightness en el espacio de funciones cadlag. Al fin y al cabo, están acotadas por las propias trayectorias poligonales convergentes. Esta observación no se justifica de manera inmediata. De hecho, acotación por punto no es un criterio válido para concluir tightness, como muestra el siguiente

■ **Contraejemplo:**

- $X^n \in C[0, 1]$:

$$X^n(t) = 1, \quad 0 \leq t \leq 1.$$

- $Y^n \in D[0, 1]$:

$$Y^n(t) = \begin{cases} 0, & t \in \left[\frac{2k}{n}, \frac{2k+1}{n}\right), \\ 1, & t \in \left[\frac{2k+1}{n}, \frac{2k+2}{n}\right), \end{cases} \quad k = 0, 1, \dots$$

Lo que debe ocurrir es que las oscilaciones en cada intervalo de longitud δ sean con elevada probabilidad pequeñas cuando $\delta \rightarrow 0$. Para ello no es suficiente con acotación puntual, sino uniforme en la métrica de Skorohod definida en el espacio de funciones cadlag.

El objetivo en este punto no es determinar si la conjetura anterior –el proceso X^n es convergente (en particular tight) en el espacio de funciones continuas debido a que los procesos poligonales $\tilde{X}^n$ lo son en el espacio de funciones continuas– es cierta o no. Sabemos que es así como se demuestra en § B14. Lo que se tratará de hacer aquí es fundamentar esta conjetura sin recurrir a las herramientas desarrolladas específicamente para el espacio de funciones cadlag (modulos de oscilación y teoremas) y empleadas en la demostración expuesta en B.

El planteamiento propuesto para ello se basa en justificar:

1. $d(X^n, \tilde{X}^n) \rightarrow 0$ en probabilidad, con la distancia de Skorohod d .
2. Lo anterior implica que X^n preserva la propiedad de tightness de $\tilde{X}^n$.

Paso 2.1: $d(X^n, \tilde{X}^n) \rightarrow 0$ en probabilidad: Retomamos la definición de la distancia de Skorohod d° dada en (3):

$$d^\circ(x, y) = \inf_{\lambda \in \Lambda} \{ \|\lambda\|^\circ \vee \|x - y \circ \lambda\| \},$$

con $\|x\| = \sup_t |x(t)|$ y λ es un homeomorfismo estrictamente creciente de $[0, 1]$ a $[0, 1]$. Tomando la identidad $\lambda(t) = t$ se tiene:

$$d^\circ(x, y) \leq \|x - y\|.$$

Dado que d° es el ínfimo, es suficiente probar:

$$\|X^n - \tilde{X}^n\| \xrightarrow{p} 0.$$

Para t en cualquier subintervalo de una partición de $[0, 1]$, i.e. $t \in [i/n, (i+1)/n)$, se tiene:

$$X_t^n = \frac{1}{\sigma\sqrt{n}} S_i, \quad \tilde{X}_t^n = \frac{1}{\sigma\sqrt{n}} S_i + \frac{nt - i}{\sigma\sqrt{n}} z_{i+1}.$$

Así, en cada subintervalo

$$|X_t^n - \tilde{X}_t^n| = \frac{nt - i}{\sigma\sqrt{n}} |z_{i+1}| \leq \frac{1}{\sigma\sqrt{n}} |z_{i+1}|.$$

Tomando el supremo sobre $t \in [0, 1]$:

$$\|X^n - \tilde{X}^n\| \leq \frac{1}{\sigma\sqrt{n}} \max_{1 \leq j \leq n} |z_j|.$$

Se trata entonces de ver que $\frac{1}{\sigma\sqrt{n}} \max_{1 \leq j \leq n} |z_j|$ converge a 0 en probabilidad.

De forma similar a como se hace en el primer lema de Borel-Cantelli, vemos aquí que por la subaditividad de la probabilidad (denotando $A_j = \{|z_j| \geq \sigma\sqrt{n}\epsilon\}$ para $\epsilon > 0$):

$$P\left(\bigcup_{j=1}^n A_j\right) \leq \sum_{j=1}^n P(A_j) = \sum_{j=1}^n P(|z_j| \geq \sigma\sqrt{n}\epsilon) = n P(|z_1| \geq \sigma\sqrt{n}\epsilon),$$

donde se ha empleado que $\{z_j\}$ son idénticamente distribuidas.

La convergencia en probabilidad que se requiere en este paso es así equivalente a comprobar que:

$$nP(|z_1| \geq \sigma\sqrt{n}\epsilon) \rightarrow 0. \quad (24)$$

Motivado por el interés en hacer uso de una estrategia habitual que consiste en relacionar probabilidades con esperanzas a través de funciones indicatrices, proponemos analizar si se satisface la siguiente desigualdad:

$$P(|z_1| \geq x) \leq \frac{\mathbb{E}[z_1^2 \mathbf{1}\{|z_1| \geq x\}]}{x^2},$$

que está muy próxima a la desigualdad de Markov aplicada a la siguiente variable aleatoria auxiliar:

$$Y_x(\omega) := z_1(\omega)^2 \mathbf{1}\{|z_1(\omega)| \geq x\} \geq 0.$$

Observamos que se cumple

$$\{\omega \in \Omega : |z_1(\omega)| \geq x\} \subseteq \{\omega \in \Omega : Y_x(\omega) \geq x^2\}.$$

Por la monotonicidad de la probabilidad:

$$P(|z_1| \geq x) \leq P(Y_x \geq x^2).$$

Aplicando la desigualdad de Markov a Y_x para una cota inferior x^2 :

$$P(|z_1| \geq x) \leq P(Y_x \geq x^2) \leq \frac{E[Y_x]}{x^2} = \frac{E[\xi_1^2 \mathbf{1}\{|z_1| \geq x\}]}{x^2} \xrightarrow{x \rightarrow \infty} 0,$$

donde la esperanza tiende a cero cuando $x \rightarrow \infty$ por ser z_1^2 integrable (Recordamos que la esperanza de las colas de una variable aleatoria integrable tiende a cero.)

Tomando ahora $x = \sigma\sqrt{n}\epsilon$, se tiene:

$$nP(|\xi_1| \geq \sigma\sqrt{n}\epsilon) = \frac{1}{\sigma^2\epsilon^2} (\sigma^2 n \epsilon^2) P(|z_1| \geq \sigma\sqrt{n}\epsilon) = \frac{1}{\sigma^2\epsilon^2} x^2 P(|z_1| \geq x) \xrightarrow{n \rightarrow \infty} 0.$$

Como resultado,

$$P\left(\max_{1 \leq j \leq n} |z_j| \geq \sigma\sqrt{n}\epsilon\right) \rightarrow 0; \quad \text{i.e.} \quad \frac{1}{\sqrt{n}} \max_{1 \leq j \leq n} |z_j| \xrightarrow{P} 0.$$

En conclusión, se ha comprobado que:

$$d^\circ(X^n, \tilde{X}^n) \leq \|X^n - \tilde{X}^n\| \leq \frac{1}{\sigma\sqrt{n}} \max_{1 \leq j \leq n} |z_j| \xrightarrow{P} 0.$$

Paso 2.2: Si $d^\circ(X^n, \tilde{X}^n) \rightarrow 0$ entonces X^n preserva la propiedad de tightness de $\tilde{X}^n$: Dado $\epsilon > 0$ arbitrario podemos encontrar un conjunto compacto en D cuya probabilidad es al menos $1 - \epsilon$.

Fijamos un compacto $K \subset D$ y definimos el entorno cerrado de K :

$$K^\epsilon := \{x \in D : d(x, K) \leq \epsilon\}, \quad \text{con } d(x, K) = \inf_{k \in K} d(x, k).$$

Se tiene:

$$\{\omega \in \Omega : \tilde{X}^n(\omega) \in K\} \cap \{\omega \in \Omega : d(X^n(\omega), \tilde{X}^n(\omega)) \leq \epsilon\} \subset \{\omega \in \Omega : X^n(\omega) \in K^\epsilon\}.$$

Simplificando la notación de conjuntos como en otras ocasiones, por ejemplo: $\{\omega \in \Omega : \tilde{X}^n(\omega) \in K\} \equiv \{\tilde{X}^n(\omega) \in K\}$ y tomando probabilidades:

$$P(X^n \in K^\epsilon) \geq P(\tilde{X}^n \in K) - P(d(X^n, \tilde{X}^n) > \epsilon).$$

Equivalentemente:

$$P(X^n \notin K^\epsilon) \leq P(\tilde{X}^n \notin K) + P(d(X^n, \tilde{X}^n) > \epsilon). \quad (25)$$

Como $\{\tilde{X}^n\}$ es tight, para todo $\eta > 0$ existe un compacto $K \subset D$ tal que

$$\sup_n P(\tilde{X}^n \notin K) < \eta.$$

Utilizando el resultado del Paso 1 anterior, por el que $d(X^n, \tilde{X}^n) \rightarrow 0$ en probabilidad, para el mismo $\epsilon > 0$ se tiene:

$$P(d(X^n, \tilde{X}^n) > \epsilon) \rightarrow 0,$$

y con ello, $\limsup_n P(d(X^n, \tilde{X}^n) > \epsilon) = 0$.

Tomando límite superior en n en la desigualdad (25) anterior:

$$\limsup_n P(X^n \notin K^\epsilon) \leq \sup_n P(\tilde{X}^n \notin K) + \limsup_n P(d(X^n, \tilde{X}^n) > \epsilon) < \eta.$$

Así, dado que lo anterior se cumple para $\eta > 0$ arbitrario, se concluye que dado $\epsilon > 0$ existe un compacto K tal que

$$\limsup_n P(X^n \notin K^\epsilon) \text{ es tan pequeño como queramos.}$$

Por último, tomando la clausura $\overline{K^\epsilon}$ (el conjunto K^ϵ es cerrado, no necesariamente compacto).

Como K es totalmente acotado por ser compacto, K^ϵ también es totalmente acotado (si K está cubierto por bolas de diámetro δ , entonces K^ϵ está cubierto por bolas de diámetro $\delta + \epsilon$).

Por el Teorema M45.1 un conjunto cerrado y totalmente acotado es compacto si el espacio métrico es completo. Dado que la clausura de un conjunto es cerrada y K^ϵ es totalmente acotado. Como el espacio D es completo con la distancia d° , concluimos que $\overline{K^\epsilon}$ es un conjunto compacto. Hemos encontrado así un conjunto compacto con el que $\{X^n\}$ es tight, es decir, dado $\eta > 0$:

$$\limsup_n P(X^n \notin \overline{K^\epsilon}) < \eta.$$

Lo anterior es intuitivamente coherente con la estabilidad de la convergencia comentada en § 3.2; estabilidad de la propiedad de tightness ante perturbaciones pequeñas/controladas.

6.2.3. Paso 3. Aplicación del Teorema de Prohorov

En espacios de dimensión infinita, la convergencia débil de las distribuciones finito dimensionales no implica la convergencia débil de la sucesión completa (Ejemplo B2.5). Los conjuntos finito-dimensionales de $C[0, 1]$ forman un sistema π (cerrado por intersecciones finitas) y forman además un clase separadora (Ejemplo B1.3). Estas propiedades junto con el teorema de la aplicación continua permite demostrar que si una familia de medidas de probabilidad definidas sobre el espacio $(C[0, 1], \mathcal{C})$ (donde $\mathcal{C}$ es la σ -álgebra generada por los abiertos de $C[0, 1]$) es relativamente compacta, entonces la convergencia débil de las distribuciones finito dimensionales sí implica convergencia débil de la sucesión completa (Ejemplo B5.1).

En los pasos 1 y 2 anteriores se ha comprobado que en el contexto del Teorema de Donsker 6.1 las distribuciones finito dimensionales de las sumas parciales definidas en (22) convergen débilmente y además son tight. Dado que de acuerdo con el Teorema de Prohorov tightness implica compacidad relativa, se concluye que las sumas parciales $\{X^n\}$ definidas en (22) convergen a un proceso cuyas distribuciones finito dimensionales coinciden con las de un proceso de Wiener.

7. Conclusiones

En este trabajo se ha presentado un tema clásico en Probabilidad que extiende las ideas del Teorema de Arzelà-Ascoli estudiado en el grado a espacios de medidas de probabilidad. El trabajo integra conceptos y resultados estudiados en dos áreas diferentes estudiadas en el grado: por un lado Topología y por otro Espacios Normados e Integral de Lebesgue.

Se ha presentado el concepto de apretamiento (tightness) y su papel en el teorema funcional del límite de Donsker. La relevancia de esta propiedad resulta clara en el estudio de la convergencia débil. Se ha visto que una forma de caracterizar la convergencia débil es por medio del comportamiento de integrales sobre la familia de funciones continuas y acotadas. Pero esta convergencia sólo tiene sentido cuando la masa de probabilidad no se extiende al infinito sin ser evaluada por dichas funciones. La propiedad de apretamiento se refiere precisamente a esta característica global, que permite recoger en un conjunto compacto al menos casi toda la probabilidad $(1 - \epsilon)$, no sólo localmente ni limitada a los puntos donde las funciones no se anulan.

Se ha propuesto una justificación del teorema de Donsker en $D[0, 1]$ alternativa a la mostrada en el texto de referencia de Billingsley. Resulta interesante porque muestra de forma más clara el papel de la distancia de Skorohod, que era una de las cuestiones que inicialmente se quería explorar y motivaban este trabajo. En el paso 1 se requiere una expresión de la distancia (no sólo su existencia) y en el paso 2 se emplea que esa distancia convierte el espacio D en completo. En la demostración Billingsley se podría pensar con que es suficiente la existencia de una métrica completa y se podría pensar que es interesante proponer una definición o cálculo alternativo. De la discusión de este teorema se puede concluir que no es especialmente relevante proponer medidas alternativas (que motivaban también este trabajo). La distancia de Skorohod da una expresión simple para paso 1, que se reduce de hecho a la distancia del supremo tomando λ la identidad.

Por un lado, tomando como referencia varios ejercicios propuestos en el texto de referencia se han trabajado los aspectos fundamentales que conducen al teorema de Donsker. Esto me ha permitido consolidar los conceptos y resultados, lo que me permitirá continuar con este estudio: (1) Condiciones alternativas de tightness, como por ejemplo basadas en tiempos óptimos de parada ([Aldous, 1989](#)); (2) Aplicaciones modernas de procesos empíricos ([van der Vaart y Wellner, 1996](#)).

Apéndice.

Principio de Cavalieri: Según el principio de Cavalieri, un volumen se puede calcular sumando las áreas de sus secciones transversales horizontales. En el texto se emplea una idea similar que se conoce como la *representación por capas* de la integral de una función no negativa.

Dada una función medible $g : S \rightarrow [0, \infty)$ y una medida de probabilidad P definida en la

correspondiente álgebra, se satisface la identidad:

$$\int_S g(x) P(dx) = \int_0^\infty P(\{x \in S : g(x) > t\}) dt.$$

Los conjuntos $\{x \in S : g(x) > t\}$ se denominan *conjuntos de supernivel* y en el texto se denotan de forma abreviada como $\{g > t\}$.

Esta idea la hemos encontrado en varias ocasiones durante el grado, por ejemplo para expresar la esperanza de una variable aleatoria no negativa X con función de densidad f como (Ross, 2010, p. 610; Vélez Ibarrola, 2019, Capítulo 11):

$$E[X] = \int_0^\infty P\{X > x\} dx,$$

que a su vez se puede emplear para obtener la desigualdad de Markov (Billingsley, 1995, pp. 79-80).

Proceso de Wiener: Un proceso de Wiener es una medida de probabilidad en el espacio de funciones $C[0, 1]$. Billingsley muestra su equivalencia como un proceso estocástico $W = \{W(t) : 0 \leq t \leq 1\}$ cuyas trayectorias son continuas y además:

- $W(0) = 0$ casi seguro.
- Para cada conjunto $0 \leq t_0 < t_1 < \dots < t_k \leq 1$ los incrementos $W(t_1) - W(t_0)$, $W(t_2) - W(t_1)$, ..., $W(t_k) - W(t_{k-1})$ son independientes.
- Cada incremento se distribuye como la distribución Normal con media 0 y varianza igual a la longitud del intervalo, i.e.: $W(t) - W(s) \sim N(0, t - s)$, $0 \leq s < t \leq 1$.

Así, para cada $t \in [0, 1]$ se tiene $W(t) \sim N(0, t)$ con matriz de covarianzas $E[W(s)W(t)] = \min\{s, t\}$.

Referencias

- David Aldous. Weak convergence and the general theory of processes. Monograph, Institute of Mathematical Statistics, Hayward, CA, 1989. URL <http://www.stat.berkeley.edu/~aldous/Papers/weak-gtp.pdf>. Monograph (available as IMS Lecture Notes / technical report).
- Robert B. Ash. *Real Analysis and Probability*. Academic Press, San Diego, CA, 2nd edition, 2000. ISBN 0-12-064481-8.
- Patrick Billingsley. *Probability and Measure*. John Wiley & Sons, New York, 3rd edition, 1995. ISBN 0-471-00710-2.
- Patrick Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, New York, 2nd edition, 1999. ISBN 978-0471191144.

- A. García y M. J. Muñoz. *Espacios de Hilbert y Análisis de Fourier: Los primeros pasos*. Editorial Sanz y Torres, Madrid, 2a edition, 2014. ISBN 978-84-15550-62-4.
- Mikhail Lifshits y Vladislav Vysotsky. On the completion of Skorokhod space. *Electronic Communications in Probability*, 25(none):1 – 10, 2020. doi:10.1214/20-ECP346.
- James R. Munkres. *Topology*. Prentice Hall, Upper Saddle River, NJ, 2nd edition, 2000. ISBN 0-13-468951-8.
- Yu. V. Prohorov. Convergence of random processes and limit theorems in probability theory. *Theory of Probability & Its Applications*, 1(2):157–214, 1956. doi:10.1137/1101016.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2025. URL <https://www.R-project.org/>.
- Sheldon M. Ross. *Introduction to Probability Models*. Academic Press, 10th edition, 2010. ISBN 978-0-12-375686-2.
- A. V. Skorohod. Limit theorems for stochastic processes. *Theory of Probability and Its Applications*, 1(3):261–290, 1956.
- Manuel Valdivia. *Análisis Matemático V*. Editorial UNED, Madrid, 1a edition, 1988. ISBN 978-84-36223-32-3.
- A. W. van der Vaart y Jon A. Wellner. *Weak Convergence and Empirical Processes*. Springer, 1996. ISBN 0-387-94640-3. URL <https://link.springer.com/book/10.1007/978-3-031-29040-4>.
- R Vélez Ibarrola. *Cálculo de Probabilidades 2*. Editorial UNED, Madrid, 2019. ISBN 978-84-362-7529-2.
- R. Vélez Ibarrola y T. Prieto Rumeau. *Procesos Estocásticos*. Editorial UNED, Madrid, 2013. ISBN 978-84-362-6700-6.
- A. Vera y P. Alegría. *Un Curso de Análisis Funcional*. AVL, 1997. ISBN 978-84-921102-8-5.