Procesos Empíricos

Alejandro Cholaquidis

Centro de Matemática – Facultad de Ciencias

Universidad de la República - Uruguay

Índice general

1. Fundamentos: integral exterior y convergencia débil **3**
1.1. Introducción y motivación . . . 3
1.2. Probabilidad exterior e integral exterior . . . 7
1.2.1. Integral exterior . . . 7
1.2.2. Fubini exterior . . . 14
1.3. Convergencia débil . . . 18
1.3.1. Redes . . . 19
1.4. Espacios producto . . . 29
1.4.1. Topología y σ-álgebras en productos finitos . . . 29
1.4.2. Tensión y medibilidad en productos finitos . . . 29
1.5. Espacios de funciones acotadas . . . 30
1.6. Convergencias . . . 43

2. Proceso empírico **46**
2.1. Medidas empíricas y clases de Donsker . . . 46
2.2. Desigualdades maximales y números de cubrimiento . . . 48
2.2.1. Desigualdades subgaussianas . . . 57
2.3. Simetrización y medibilidad . . . 59
2.3.1. Simetrización . . . 59
2.4. Teoremas de Glivenko–Cantelli . . . 64
2.4.1. Glivenko-Cantelli por bracketing . . . 64
2.4.2. Glivenko-Cantelli por cubrimientos . . . 64
2.5. Teoremas de Donsker . . . 72
2.5.1. Entropía uniforme . . . 72
2.5.2. Acotamientos (*bracketing*) . . . 76
2.6. Cotas para la entropía y clases VC . . . 77
2.6.1. Clases VC de conjuntos . . . 78
2.6.2. Clases VC de funciones . . . 80
2.7. Números de bracketing . . . 85
2.7.1. Funciones y conjuntos suaves . . . 85

3. Aplicaciones estadísticas **88**
3.1. M-estimadores . . . 88
3.1.1. Tasa de convergencia . . . 91
3.2. Resultados auxiliares . . . 94

Estas notas están muy fuertemente basadas en [5]. Tienen como requisitos probabilidad y medida. Una referencia clásica para los conceptos de probabilidad es [4]. Para teoría de la medida, véase [3]; para procesos en espacios generales, véase [1].

La sección denominada σ-álgebra generada por bolas es complementaria, y se puede omitir en una primera lectura, no es requisito para la lectura de lo que sigue. Lo mismo sucede con la sección Uniformidad y metrización.

Todas las erratas y errores que puedan contener estas notas son de mi responsabilidad. Quedo agradecido al lector que, amablemente, me las haga saber a cholaquidisalejandro@gmail.com

Capítulo 1

Fundamentos: integral exterior y convergencia débil

1.1. Introducción y motivación

La teoría de procesos empíricos estudia el comportamiento de promedios muestrales, no función a función, sino de manera uniforme sobre familias enteras de funciones. Antes de introducir la maquinaria general, conviene ver el ejemplo concreto que originó toda la teoría y que contiene, en miniatura, las dos preguntas centrales.

El caso clásico: la función de distribución empírica

Sea $X_1, \ldots, X_n$ una muestra i.i.d. de una variable aleatoria real X, definidas en un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$, con función de distribución F. La función de distribución empírica es

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty,t]}(X_i), \qquad t \in \mathbb{R}.$$

Para cada t fijo, la ley fuerte de los grandes números da $F_n(t) \to F(t)$ c.s. El teorema de Glivenko–Cantelli refuerza esta afirmación puntual y establece

$$\|F_n - F\|_\infty = \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \to 0 \quad \text{c.s.}$$

Esta es la primera aparición natural de una ley uniforme de grandes números.

La utilidad estadística de esta convergencia es inmediata. Si T es un funcional definido sobre una clase de distribuciones y T es continuo en F respecto de la norma uniforme, entonces

$$\|F_n - F\|_\infty \to 0 \quad \text{implica} \quad T(F_n) \to T(F).$$

Así, el teorema de Glivenko–Cantelli justifica la consistencia de estimadores plug-in. Esta idea es simple pero fundamental: muchos estimadores se obtienen reemplazando la distribución desconocida F por su versión empírica F_n.

Fluctuaciones: el teorema de Donsker

La segunda pregunta natural no es sólo si F_n converge a F, sino cómo fluctúa alrededor de F. Definimos

$$\mathbb{G}_n(t) = \sqrt{n}(F_n(t) - F(t)), \qquad t \in \mathbb{R}.$$

Para cada t fijo, el teorema central del límite da $\mathbb{G}_n(t) \Rightarrow N(0, F(t)(1 - F(t)))$, donde $\Rightarrow$ indica convergencia en distribución (que definiremos más adelante). Pero muchos estadísticos dependen del camino completo $t \mapsto \mathbb{G}_n(t)$, no de un único valor. El teorema de Donsker afirma que, bajo la

topología adecuada, $\sqrt{n}(F_n - F) \Rightarrow \mathbb{U} \circ F$, donde $\mathbb{U}$ es un puente browniano estándar, es decir, un proceso gaussiano centrado con covarianza $\mathrm{Cov}(\mathbb{U}(s), \mathbb{U}(t)) = s \wedge t - st$.

Por ejemplo, bajo $H_0 : F = F_0$, con F_0 continua, el estadístico de Kolmogorov–Smirnov satisface

$$\sqrt{n}\|F_n - F_0\|_\infty \Rightarrow \sup_{0 \le u \le 1} |\mathbb{U}(u)|.$$

De modo similar, el estadístico de Cramér–von Mises tiene límite

$$n \int (F_n - F_0)^2 dF_0 \Rightarrow \int_0^1 \mathbb{U}(u)^2 du.$$

Estos ejemplos muestran por qué no alcanza el TCL para $F_n(t)$ con t fijo: las pruebas globales de bondad de ajuste son funcionales del proceso entero.

Estos dos resultados —una ley uniforme de los grandes números y un teorema central del límite uniforme— son el modelo de toda la teoría que sigue. El objetivo del curso es extenderlos: reemplazar la familia de semirrectas por una clase arbitraria de funciones, y entender en qué espacio viven entonces los objetos límite y qué significa allí converger. Para eso necesitamos primero fijar la notación general.

Notación general y el proceso empírico

Sea ahora $X_1, \dots, X_n$ una muestra i.i.d. de una variable aleatoria X, definidas en un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$, con ley común P, sobre un espacio medible $(\mathcal{X}, \mathcal{B})$. Para una función medible $f : \mathcal{X} \to \mathbb{R}$, escribimos

$$Pf = \int f dP, \qquad P_n f = \frac{1}{n} \sum_{i=1}^n f(X_i),$$

donde $P_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ es la medida empírica. Para una función fija f, la ley fuerte de los grandes números y el teorema central del límite dan, bajo las hipótesis usuales,

$$P_n f \to Pf \quad \text{c.s.}, \qquad \sqrt{n}(P_n - P)f \Rightarrow N(0, P(f - Pf)^2).$$

Estos resultados son puntuales en f. Sin embargo, como acabamos de ver en el caso clásico, en muchas aplicaciones estadísticas no se estudia un único promedio $P_n f$, sino toda una familia de promedios $(P_n f)_{f \in \mathcal{F}}$, donde $\mathcal{F}$ es una clase de funciones $\mathcal{B}$-medibles tales que $P|f| < \infty$ para toda $f \in \mathcal{F}$, de modo que Pf esté definido como número real. Por eso aparecen naturalmente las cantidades

$$\|P_n - P\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |P_n f - Pf|, \qquad \mathbb{G}_n f := \sqrt{n}(P_n - P)f = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{f(X_i) - Pf\}, \quad f \in \mathcal{F}.$$

El objeto $\mathbb{G}_n$ se llama proceso empírico indexado por $\mathcal{F}$; es una función que manda $\omega \in \Omega$ a $\mathbb{G}_n(\omega)$, que actúa sobre las funciones de la clase $\mathcal{F}$, es decir $\mathbb{G}_n(\omega)$ es un elemento en el espacio

$$\ell^\infty(\mathcal{F}) = \{z : \mathcal{F} \to \mathbb{R} : \|z\|_{\mathcal{F}} < \infty\}, \qquad \text{donde} \qquad \|z\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |z(f)|.$$

Para que efectivamente $\|\mathbb{G}_n(\omega)\|_{\mathcal{F}} < \infty$ vamos a tener que pedir una condición de existencia de envolvente para la familia $\mathcal{F}$ que veremos más adelante.

El caso clásico se recupera como caso particular: si $\mathcal{X} = \mathbb{R}$, tomando la clase de semirrectas $\mathcal{F}_0 = \{\mathbb{1}_{(-\infty,t]} : t \in \mathbb{R}\}$ resulta $F_n(t) = P_n \mathbb{1}_{(-\infty,t]}$, de modo que estudiar $F_n - F$ equivale a estudiar $P_n - P$ sobre $\mathcal{F}_0$, y el proceso $\mathbb{G}_n(t)$ de la sección anterior es exactamente $\mathbb{G}_n \mathbb{1}_{(-\infty,t]}$.

Clases de Glivenko–Cantelli y clases de Donsker

Si $f_1, \ldots, f_k \in \mathcal{F}$ satisfacen además $Pf_j^2 < \infty$ para todo j, el TCL multivariado da

$$(\mathbb{G}_n f_1, \ldots, \mathbb{G}_n f_k) \Rightarrow N_k(0, \Sigma), \qquad \Sigma_{ij} = P(f_i - Pf_i)(f_j - Pf_j).$$

Pero esta convergencia finito-dimensional no controla el proceso completo. Cuando la trayectoria $f \mapsto \mathbb{G}_n f$ es acotada, identificamos $\mathbb{G}_n$ con un elemento de $\ell^\infty(\mathcal{F})$. Esta acotación no se deduce de la medibilidad e integrabilidad de cada coordenada y debe verificarse en cada aplicación concreta.

La teoría de procesos empíricos busca condiciones sobre $\mathcal{F}$ que garanticen dos propiedades uniformes, análogas a los dos teoremas del caso clásico. La primera es una ley uniforme de grandes números: $\|P_n - P\|_{\mathcal{F}} \to 0$. Cuando esto ocurre, se dice que $\mathcal{F}$ es una clase P-Glivenko–Cantelli. La segunda es un teorema central del límite uniforme: $\mathbb{G}_n \Rightarrow \mathbb{G}_P$ en $\ell^\infty(\mathcal{F})$, donde $\mathbb{G}_P$ es un proceso gaussiano centrado con covarianza

$$\operatorname{Cov}(\mathbb{G}_P f, \mathbb{G}_P g) = P(fg) - (Pf)(Pg).$$

Cuando esto ocurre, y $\mathbb{G}_n$ se considera como aplicación con valores en $\ell^\infty(\mathcal{F})$, se dice que $\mathcal{F}$ es una clase P-Donsker. Al discutir clases P-Donsker supondremos, como mínimo, que $\mathcal{F} \subset L^2(P)$, de modo que esta covarianza esté bien definida para todo par $f, g \in \mathcal{F}$. El teorema de Glivenko–Cantelli dice entonces que $\mathcal{F}_0$ es P-Glivenko–Cantelli para toda P, y el teorema de Donsker, que $\mathcal{F}_0$ es P-Donsker para toda P.

Motivación estadística: M-estimación

Estas definiciones aparecen de manera natural en M-estimación. Supongamos que

$$\hat{\theta}_n \in \arg\max_{\theta \in \Theta} P_n m_\theta, \qquad \theta_0 \in \arg\max_{\theta \in \Theta} P m_\theta,$$

donde

$$P_n m_\theta = \frac{1}{n} \sum_{i=1}^n m_\theta(X_i).$$

Para deducir que el maximizador empírico $\hat{\theta}_n$ se aproxima al maximizador poblacional θ_0, una condición básica es que el criterio empírico se aproxime uniformemente al poblacional:

$$\sup_{\theta \in \Theta} |P_n m_\theta - P m_\theta| \to 0.$$

Es decir, necesitamos que la clase $\{m_\theta : \theta \in \Theta\}$ sea Glivenko–Cantelli. Esta condición, combinada con una hipótesis de máximo bien separado —para todo $\varepsilon > 0$, $\sup_{d(\theta,\theta_0) \geq \varepsilon} P m_\theta < P m_{\theta_0}$—, implica la consistencia $\hat{\theta}_n \to \theta_0$, como veremos más adelante.

El ejemplo principal es la máxima verosimilitud. Para un modelo dominado $\{p_\theta : \theta \in \Theta\}$ respecto de una medida μ, tomando $m_\theta = \log p_\theta$ resulta

$$\arg\max_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n \log p_\theta(X_i) = \arg\max_{\theta \in \Theta} \prod_{i=1}^n p_\theta(X_i),$$

pues el logaritmo es creciente y dividir por n no altera el argmax; de modo que $\hat{\theta}_n$ es precisamente el estimador de máxima verosimilitud. Conviene normalizar tomando $m_\theta = \log(p_\theta / p_{\theta_0})$, lo que tampoco altera el argmax empírico, ya que difiere del anterior en un término que no depende de θ. El criterio poblacional resulta entonces

$$P m_\theta = -K(p_{\theta_0}, p_\theta) \leq 0, \qquad \text{donde} \qquad K(p, q) = \int p \log \frac{p}{q} \, d\mu$$

es la divergencia de Kullback–Leibler. Así, Pm_θ alcanza su máximo, igual a 0, si y sólo si $p_\theta = p_{\theta_0}$ μ-c.s., y, bajo identificabilidad, esto equivale a $\theta = \theta_0$.[1]

Para estudiar no sólo la consistencia sino también la distribución límite de $\hat{\theta}_n$, es necesario analizar con mayor precisión las fluctuaciones de $P_n - P$ sobre clases de funciones indexadas por entornos de θ_0. Aparecen entonces procesos empíricos locales. Esta es una de las razones por las que la teoría de procesos empíricos es central en M-estimación, estimación semiparamétrica, bootstrap y método delta funcional.

Por qué la convergencia puntual no alcanza

Las definiciones anteriores no son gratuitas: la uniformidad puede fallar incluso cuando toda coordenada converge. Sea P una distribución continua en $\mathbb{R}$, y sea $\mathcal{C}$ la clase de todos los subconjuntos finitos de $\mathbb{R}$. Como P no tiene átomos, para cada $C \in \mathcal{C}$, $PC = P(\mathbb{1}_C) = 0$, y por lo tanto $P_nC \to 0$ c.s. Sin embargo,

$$\sup_{C\in\mathcal{C}} |P_nC - PC| = 1 \quad \text{c.s. para todo } n.$$

En efecto, si $C_n = \{X_1, \ldots, X_n\}$, entonces $C_n \in \mathcal{C}$, $P_nC_n = 1$ y, como P es continua —es decir, sin átomos: $P\{x\} = 0$ para todo x—, $PC_n(\omega) = 0$ para todo ω.

Este ejemplo muestra que la dificultad no está en probar leyes fuertes para cada función individual, sino en controlar la complejidad de la clase completa. Por eso aparecen condiciones de entropía, bracketing y dimensión Vapnik–Chervonenkis.

Un primer problema topológico: el espacio de Skorokhod

En el caso clásico, F_n y $\sqrt{n}(F_n - F)$ son funciones càdlàg[2] sobre la recta; para trabajar sobre un dominio compacto puede reducirse al caso uniforme mediante la transformación cuantil (si F es continua), o identificarse $[-\infty, \infty]$ con $[0, 1]$ mediante el homeomorfismo $\psi(t) = \frac{1}{2} + \frac{1}{\pi}\arctan t$, con $\psi(-\infty) = 0$ y $\psi(+\infty) = 1$, estudiando el proceso reparametrizado $u \mapsto \mathbb{G}_n(\psi^{-1}(u))$, $u \in [0, 1]$. Por eso es natural considerar $\mathcal{D}[0, 1] = \{z : [0, 1] \to \mathbb{R} : z \text{ es càdlàg}\}$.

Una primera tentación sería trabajar en $(\mathcal{D}[0, 1], \|\cdot\|_\infty)$. El problema es que este espacio no es separable. Esa falta de separabilidad puede generar problemas de medibilidad: procesos definidos de manera natural punto a punto pueden no ser variables aleatorias borelianas con valores en ese espacio.

La solución clásica es reemplazar la topología uniforme por la topología de Skorokhod. Sea Λ la clase de biyecciones $\lambda : [0, 1] \to [0, 1]$ estrictamente crecientes y continuas, con $\lambda(0) = 0$ y $\lambda(1) = 1$. Una forma de la métrica J_1 es

$$d(x, y) = \inf_{\lambda\in\Lambda} \{\|\lambda - \mathrm{Id}\,\|_\infty + \|x - y \circ \lambda\|_\infty\}.$$

Esta métrica permite que dos funciones estén cerca aunque sus saltos no ocurran exactamente en el mismo instante, siempre que los tiempos de salto puedan alinearse mediante una pequeña deformación λ. Cabe aclarar que esta métrica genera la topología J_1 pero no es completa; existe una métrica equivalente y completa que exhibe a $\mathcal{D}[0, 1]$ con la topología J_1 como espacio polaco, y que se usa cuando se necesitan propiedades de completitud. Con esta topología, $\mathcal{D}[0, 1]$ es separable, y se obtiene un marco adecuado para la convergencia débil de muchos procesos con trayectorias càdlàg.

Así, Skorokhod resuelve de manera elegante un problema concreto: cómo formular límites débiles para procesos en tiempo real cuyos caminos pueden tener saltos. Pero los procesos empíricos generales

[1] Esta normalización tiene además una ventaja técnica importante. Si se toma $m_\theta = \log(p_\theta/p_{\theta_0})$, entonces, bajo $P = P_{\theta_0}$, la parte positiva de m_θ es siempre integrable, ya que $(\log z)^+ \le z$ y $P_{\theta_0}(p_\theta/p_{\theta_0}) \le 1$. Por lo tanto, Pm_θ está bien definido, aunque eventualmente pueda valer $-\infty$. Además, por la desigualdad de Jensen, $Pm_\theta \le 0$, de modo que $Pm_\theta \in [-\infty, 0]$. En cambio, $P \log p_\theta$ puede no estar definido: como una densidad puede tomar valores tanto mayores como menores que 1, las partes positiva y negativa de $\log p_\theta$ pueden tener ambas esperanza infinita.

[2] Recordar que càdlàg significa continua por la derecha y con límites por la izquierda.

no suelen estar indexados por $t \in [0,1]$, sino por una clase arbitraria de funciones $\mathcal{F}$. El espacio natural pasa a ser $\ell^\infty(\mathcal{F})$. Cuando $\mathcal{F}$ es grande, este espacio también es típicamente no separable. En general no hay una topología de Skorokhod que arregle el problema, porque ya no hay un orden temporal de saltos que podamos deformar.

El problema de medibilidad

El problema de medibilidad aparece de forma muy concreta. Consideremos

$$(\Omega, \mathcal{A}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \lambda), \qquad \mathcal{F}_0 = \{f_t : t \in [0,1]\}, \quad f_t(x) = \mathbb{1}_{\{x \le t\}}.$$

Definimos

$$Z : \Omega \to \ell^\infty(\mathcal{F}_0), \qquad Z(\omega)(f_t) = f_t(\omega) = \mathbb{1}_{\{\omega \le t\}}.$$

Para cada $f_t \in \mathcal{F}_0$ fijo, la coordenada $\omega \mapsto Z(\omega)(f_t)$ es medible. Sin embargo, Z no es Borel-medible como aplicación con valores en $\ell^\infty(\mathcal{F}_0)$.

En efecto, si $\omega \neq \omega'$, entonces $\|Z(\omega) - Z(\omega')\|_{\mathcal{F}_0} = 1$. Sea $A \subset [0,1]$ un conjunto no boreliano y pongamos

$$U_A := \bigcup_{\omega \in A} B_{\ell^\infty(\mathcal{F}_0)}(Z(\omega), 1/2).$$

Entonces U_A es abierto y, por la separación anterior, $Z^{-1}(U_A) = A$.[3] Como $A \notin \mathcal{B}([0,1])$, concluimos que Z no es Borel-medible.

Este ejemplo muestra que la medibilidad de todas las coordenadas no implica la medibilidad del proceso completo. En procesos empíricos ocurre el mismo fenómeno: para cada $f \in \mathcal{F}$, $\omega \mapsto \mathbb{G}_n(\omega)f$ puede ser una variable aleatoria real medible, mientras que $\mathbb{G}_n : \Omega \to \ell^\infty(\mathcal{F})$ puede no ser Borel-medible.

1.2. Probabilidad exterior e integral exterior

Para no excluir estos procesos naturales, la teoría moderna trabaja con probabilidades e integrales exteriores. Estas nociones permiten escribir afirmaciones como $\|P_n - P\|_{\mathcal{F}} \to 0$ en probabilidad exterior.

También permiten formular convergencia débil para procesos no necesariamente medibles. En resumen, los procesos empíricos aparecen porque muchos procedimientos estadísticos dependen de $P_n - P$ simultáneamente sobre una clase de funciones: supremos, argmáximos, cuantiles, pruebas globales, bootstrap y estimadores plug-in. La teoría no sólo debe probar convergencias uniformes; también debe formularlas rigurosamente en espacios donde los procesos naturales pueden no ser medibles. Esta es la razón conceptual por la cual entran la probabilidad exterior y la integral exterior.

1.2.1. Integral exterior

Recordemos que $\overline{\mathbb{R}} := [-\infty, \infty] = \mathbb{R} \cup \{-\infty, +\infty\}$. Consideramos en $\overline{\mathbb{R}}$ su σ-álgebra de Borel $\mathcal{B}(\overline{\mathbb{R}})$, asociada a la topología usual de la recta extendida. Equivalentemente,

$$\mathcal{B}(\overline{\mathbb{R}}) = \{B \cup C : \ B \in \mathcal{B}(\mathbb{R}), \ C \subset \{-\infty, +\infty\}\}.$$

Así, una función $U : \Omega \to \overline{\mathbb{R}}$ es medible si $U^{-1}(B) \in \mathcal{A}$ para todo $B \in \mathcal{B}(\overline{\mathbb{R}})$.

Definición 1.1. Sea $(\Omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad y $T : \Omega \to \overline{\mathbb{R}}$ (no necesariamente medible). Definimos la **esperanza exterior** de T por

$$\mathbb{E}^*(T) = \text{ínf}\left\{\mathbb{E}U : U : \Omega \to \overline{\mathbb{R}} \text{ medible}, U(\omega) \ge T(\omega) \ \ \forall \omega, \ \mathbb{E}U \text{ existe}\right\},$$

[3] Es inmediato que $Z(A) \subset U_A$ por definición de U_A. Si $\omega \in Z^{-1}(U_A)$ existe $\omega' \in A$ tal que $\|Z(\omega') - Z(\omega)\|_{\mathcal{F}_0} < 1/2$, y esto implica $\omega = \omega'$.

donde "$\mathbb{E}U$ existe" significa que

$$\min\{\mathbb{E}(U^+), \mathbb{E}(U^-)\} < \infty \quad (\text{y entonces } \mathbb{E}U = \mathbb{E}(U^+) - \mathbb{E}(U^-) \in [-\infty, \infty]).$$

El conjunto del ínfimo es no vacío porque $U \equiv +\infty$ es medible.

Para $B \subset \Omega$, la **probabilidad exterior** y la **probabilidad interior** se definen como

$$\mathbb{P}^*(B) = \inf\left\{\mathbb{P}(A) : B \subset A, A \in \mathcal{A}\right\}, \qquad \mathbb{P}_*(B) = 1 - \mathbb{P}^*(\Omega \setminus B).$$

De modo análogo, la **esperanza interior** es $\mathbb{E}_*(T) := -\mathbb{E}^*(-T)$.

Observación 1.2

En la definición de $\mathbb{E}^*(T)$ es equivalente exigir $U \geq T$ para todo ω, o bien $U \geq T$ fuera de un conjunto medible de probabilidad cero. En efecto, si $U \geq T$ en N^c, con $N \in \mathcal{A}$ y $\mathbb{P}(N) = 0$, basta reemplazar U por $\widetilde{U} = U\mathbb{1}_{N^c} + \infty\,\mathbb{1}_N$, que es medible, satisface $\widetilde{U} \geq T$ en todo Ω y $\mathbb{E}\widetilde{U} = \mathbb{E}U$. Como T puede no ser medible, hay que interpretar "$U \geq T$ c.s." de esta forma, y no mediante $\mathbb{P}(U < T) = 0$.

Advertencia 1.3

La condición de medibilidad que se pide sobre U en la definición de $\mathbb{E}^*(T)$ implica que $\mathbb{E}^*(T)$ depende de la σ-álgebra subyacente $\mathcal{A}$. Análogamente, la condición de existencia de $\mathbb{E}(U)$ depende de $\mathbb{P}$ (y de $\mathcal{A}$).

El siguiente lema prueba la existencia de un mayorante medible mínimo de una función arbitraria $T : \Omega \to \overline{\mathbb{R}}$. Más precisamente, construye una variable medible T^* que domina a T punto a punto y que es menor o igual, salvo conjuntos nulos, que cualquier otro mayorante medible de T. La prueba trabaja con $\arctan T$ porque arctan es creciente, acotada y biyectiva de $\overline{\mathbb{R}}$ en $[-\pi/2, \pi/2]$; así se preserva el orden, se puede volver a la escala original mediante la tangente extendida y, al estar todo acotado, se puede aplicar el teorema de convergencia acotada.

Lema 1.5. *Para todo $T : \Omega \to \overline{\mathbb{R}}$ existe $T^* : \Omega \to \overline{\mathbb{R}}$ medible tal que:*

(i) $T^* \geq T$ punto a punto;

(ii) si U es medible y $U \geq T$ $\mathbb{P}$-c.s.[4], entonces $T^* \leq U$ $\mathbb{P}$-c.s.;

(iii) si $\mathbb{E}(T^*)$ existe, entonces $\mathbb{E}^*(T) = \mathbb{E}(T^*)$. Más aún, si $\mathbb{E}^*(T) \in \mathbb{R}$, entonces $\mathbb{E}(T^*)$ existe y coincide con $\mathbb{E}^*(T)$.

Demostración. Extendamos la función arcotangente a $\overline{\mathbb{R}}$ definiendo $\phi(\pm\infty) = \pm\pi/2$, donde $\phi(x) = \arctan(x)$ para $x \in \mathbb{R}$, y sea $\psi : [-\pi/2, \pi/2] \to \overline{\mathbb{R}}$ su inversa. Ambas funciones son medibles y estrictamente crecientes.

Pongamos $a := \mathbb{E}^*(\phi(T))$. Como $\phi(T)$ toma valores en $[-\pi/2, \pi/2]$, en la definición de a basta considerar mayorantes medibles con valores en ese intervalo. En efecto, si $W \geq \phi(T)$ es medible, entonces $W \geq -\pi/2$, y $W \wedge (\pi/2)$ sigue siendo un mayorante de $\phi(T)$ cuya esperanza no es mayor que la de W.

Por la definición de a, para cada $m \geq 1$ existe una función medible $\widetilde{V}_m : \Omega \to [-\pi/2, \pi/2]$ tal que $\widetilde{V}_m \geq \phi(T)$ y $\mathbb{E}\widetilde{V}_m \leq a + 1/m$.[5] Definimos $V_m := \min_{1 \leq k \leq m} \widetilde{V}_k$. Entonces (V_m) es decreciente, cada V_m es un mayorante medible de $\phi(T)$, y $a \leq \mathbb{E}V_m \leq \mathbb{E}\widetilde{V}_m \leq a + 1/m$. Por lo tanto, V_m converge

[4] Como en la Observación 1.2: $A \leq B$ c.s. o $A = B$ c.s. significa que la relación vale fuera de un conjunto nulo medible $N \in \mathcal{A}$, aun si alguna de las funciones no es medible. Esta convención se usa en todo el texto.

[5] Nótese que a es finito: como puede restringirse a mayorantes con valores en $[-\pi/2, \pi/2]$, cuyas esperanzas están en ese intervalo, el ínfimo satisface $a \in [-\pi/2, \pi/2]$. Por eso tiene sentido pedir $\mathbb{E}\widetilde{V}_m \leq a + 1/m$: es la caracterización del ínfimo finito.

puntualmente a una función medible $V : \Omega \to [-\pi/2, \pi/2]$, que satisface $V \geq \phi(T)$.[6] Como la sucesión es uniformemente acotada, el teorema de convergencia acotada da

$$\mathbb{E}V = \lim_{m\to\infty} \mathbb{E}V_m = a = \mathbb{E}^*(\phi(T)). \tag{1.1}$$

Definimos ahora $T^* := \psi(V)$. La función T^* es medible y, como $V \geq \phi(T)$ y ψ es creciente,

$$T^* = \psi(V) \geq \psi(\phi(T)) = T \quad \text{punto a punto.}$$

Esto prueba *(i)*.

Sea U medible y tal que $U \geq T$, $\mathbb{P}$-c.s. Modificando U sobre el conjunto nulo excepcional, podemos suponer que $U \geq T$ punto a punto. Entonces $U \wedge T^*$ es un mayorante medible de T, de modo que $\phi(U \wedge T^*)$ es un mayorante medible de $\phi(T)$. Por la definición de la esperanza exterior y por (1.1),

$$\mathbb{E}^*(\phi(T)) \leq \mathbb{E}\Big[\phi(U \wedge T^*)\Big] \leq \mathbb{E}[\phi(T^*)] = \mathbb{E}V = \mathbb{E}^*(\phi(T)).$$

Todas las desigualdades son, por tanto, igualdades.[7] Como $\phi(T^*) - \phi(U \wedge T^*)$ es una variable aleatoria no negativa y acotada, se concluye que $\phi(T^*) = \phi(U \wedge T^*)$, $\mathbb{P}$-c.s. La estricta monotonía de ϕ implica que $T^* = U \wedge T^*$, $\mathbb{P}$-c.s., y por consiguiente $T^* \leq U$, $\mathbb{P}$-c.s. Esto prueba *(ii)*.

Supongamos que $\mathbb{E}T^*$ existe. Como T^* es un mayorante medible de T, la definición de la esperanza exterior da $\mathbb{E}^*T \leq \mathbb{E}T^*$.

Recíprocamente, sea U un mayorante medible de T cuya esperanza exista. Por *(ii)*, $T^* \leq U$, $\mathbb{P}$-c.s., y la monotonía de la esperanza extendida implica $\mathbb{E}T^* \leq \mathbb{E}U$ (esta desigualdad no necesariamente es entre números reales, ambos podrían ser ∞). Tomando ínfimo sobre todos esos mayorantes U, obtenemos $\mathbb{E}T^* \leq \mathbb{E}^*T$. En consecuencia, $\mathbb{E}T^* = \mathbb{E}^*T$. Finalmente, supongamos que $\mathbb{E}^*T \in \mathbb{R}$. Por la definición del ínfimo, existe un mayorante medible $U \geq T$ tal que $\mathbb{E}U < \mathbb{E}^*T + 1$. Necesariamente $\mathbb{E}U \in \mathbb{R}$: no puede ser $+\infty$ por la desigualdad anterior, ni puede ser $-\infty$, pues entonces $\mathbb{E}^*T \leq \mathbb{E}U = -\infty$.

Por *(ii)*, $T^* \leq U$, $\mathbb{P}$-c.s., y por tanto $(T^*)^+ \leq U^+$, $\mathbb{P}$-c.s. Como $\mathbb{E}U \in \mathbb{R}$, se tiene $\mathbb{E}U^+ < \infty$, y en consecuencia $\mathbb{E}(T^*)^+ < \infty$. Por lo tanto, $\mathbb{E}T^*$ existe como valor extendido. Aplicando la identidad anterior, obtenemos $\mathbb{E}T^* = \mathbb{E}^*T \in \mathbb{R}$, lo que prueba *(iii)*. □

Definición 1.6. T^* se llama el **mayorante medible mínimo** de T. Por el punto *(ii)* del Lema anterior, es único salvo modificación en un conjunto de probabilidad $\mathbb{P}$ nula. Definimos además, $T_* := -(-T)^*$.

Advertencia 1.7

Es claro que, como $\mathbb{E}^*(T)$ depende de $\mathcal{A}$ y de $\mathbb{P}$, T^* también depende de $\mathcal{A}$ y de $\mathbb{P}$.

Lema 1.8. *Para todo $B \subset \Omega$ se cumple:*

(i) $\mathbb{P}^*(B) = \mathbb{E}^*(\mathbb{1}_B)$ *y* $\mathbb{P}_*(B) = \mathbb{E}\Big((\mathbb{1}_B)_*\Big)$.

(ii) *Existe $B^* \in \mathcal{A}$ tal que $B \subset B^*$ y $\mathbb{P}(B^*) = \mathbb{P}^*(B)$. Más aún, si $\widetilde{B} \in \mathcal{A}$, $B \subset \widetilde{B}$ y $\mathbb{P}(\widetilde{B}) = \mathbb{P}^*(B)$, entonces $\mathbb{1}_{\widetilde{B}} = (\mathbb{1}_B)^*$ $\mathbb{P}$-c.s. En particular, $(\mathbb{1}_{\widetilde{B}})^* = (\mathbb{1}_B)^*$ $\mathbb{P}$-c.s.*

(iii) $(\mathbb{1}_B)^* + (\mathbb{1}_{\Omega\setminus B})_* = 1$ $\mathbb{P}$-*c.s.*

[6] Una sucesión decreciente y acotada inferiormente (aquí por $-\pi/2$) converge en todo punto, y el límite puntual de funciones medibles es medible.

[7] La primera desigualdad es la definición de $\mathbb{E}^*(\phi(T))$ como ínfimo, aplicada al mayorante $\phi(U \wedge T^*)$, cuya esperanza existe por ser acotado; la segunda usa $U \wedge T^* \leq T^*$, la monotonía de ϕ y la de la esperanza.

Demostración. Sea $B^* := \{(\mathbb{1}_B)^* \geq 1\}$. Entonces $B^* \in \mathcal{A}$ y $B \subset B^*$, pues en B se tiene $(\mathbb{1}_B)^* \geq \mathbb{1}_B = 1$. Por la definición de $\mathbb{P}^*$, $\mathbb{P}^*(B) \leq \mathbb{P}(B^*)$. Como B^* es medible, $\mathbb{P}(B^*) = \mathbb{E}(\mathbb{1}_{B^*})$.

Observemos que $\mathbb{1}_{B^*} \leq (\mathbb{1}_B)^*$ punto a punto: en B^* el miembro izquierdo vale 1, y fuera de B^* vale 0, mientras que el derecho es siempre ≥ 0. Por lo tanto $\mathbb{E}(\mathbb{1}_{B^*}) \leq \mathbb{E}((\mathbb{1}_B)^*)$.

Como $(\mathbb{1}_B)^* \geq \mathbb{1}_B \geq 0$ punto a punto, la esperanza $\mathbb{E}(\mathbb{1}_B)^*$ existe y, por el Lema 1.5*(iii)*, $\mathbb{E}(\mathbb{1}_B)^* = \mathbb{E}^*(\mathbb{1}_B)$. Con esto, $\mathbb{P}^*(B) \leq \mathbb{E}^*(\mathbb{1}_B)$. Veamos que $\mathbb{E}^*(\mathbb{1}_B) \leq \mathbb{P}^*(B)$: si $A \in \mathcal{A}$ y $B \subset A$, entonces $\mathbb{1}_A$ es un mayorante medible de $\mathbb{1}_B$, de donde $\mathbb{E}^*(\mathbb{1}_B) \leq \mathbb{E}\,\mathbb{1}_A = \mathbb{P}(A)$, y se toma ínfimo sobre A. Hemos probado entonces

$$\begin{aligned} \mathbb{P}^*(B) \leq \mathbb{P}(B^*) = \mathbb{E}\mathbb{1}_{B^*} &\leq \mathbb{E}\Big[(\mathbb{1}_B)^*\Big] \\ &= \mathbb{E}^*(\mathbb{1}_B) \leq \mathbb{P}^*(B). \end{aligned} \tag{1.2}$$

Por lo tanto obtenemos directamente la primera identidad de *(i)*, $\mathbb{P}^*(B) = \mathbb{E}^*(\mathbb{1}_B)$, y la existencia afirmada en *(ii)*: $B \subset B^* \in \mathcal{A}$ con $\mathbb{P}(B^*) = \mathbb{P}^*(B)$. Más aún, $\mathbb{1}_{B^*} \leq (\mathbb{1}_B)^*$ y ambas variables tienen la misma esperanza; como una variable no negativa con esperanza nula es nula c.s., aplicado a $(\mathbb{1}_B)^* - \mathbb{1}_{B^*} \geq 0$,

$$(\mathbb{1}_B)^* = \mathbb{1}_{B^*} \qquad \mathbb{P}\text{-c.s.}$$

Sea ahora $\widetilde{B} \in \mathcal{A}$ tal que $B \subset \widetilde{B}$ y $\mathbb{P}(\widetilde{B}) = \mathbb{P}^*(B)$. Como $\mathbb{1}_{\widetilde{B}}$ es un mayorante medible de $\mathbb{1}_B$, la minimalidad da $(\mathbb{1}_B)^* \leq \mathbb{1}_{\widetilde{B}}$, $\mathbb{P}$-c.s.; y las esperanzas coinciden, por (1.2): $\mathbb{E}(\mathbb{1}_B)^* = \mathbb{P}^*(B) = \mathbb{P}(\widetilde{B}) = \mathbb{E}\,\mathbb{1}_{\widetilde{B}}$. El mismo argumento de esperanza nula da $\mathbb{1}_{\widetilde{B}} = (\mathbb{1}_B)^*$, $\mathbb{P}$-c.s. Como $\mathbb{1}_{\widetilde{B}}$ es medible, coincide $\mathbb{P}$-c.s. con su mayorante medible mínimo;[8] por tanto, $(\mathbb{1}_{\widetilde{B}})^* = (\mathbb{1}_B)^*$, $\mathbb{P}$-c.s. Esto prueba *(ii)*.

Probemos ahora *(iii)*; escribimos $B^c := \Omega \setminus B$. Observemos primero que el mayorante medible mínimo se traslada por constantes: para toda $T : \Omega \to \overline{\mathbb{R}}$ y todo $c \in \mathbb{R}$,

$$(T + c)^* = T^* + c \qquad \mathbb{P}\text{-c.s.} \tag{1.3}$$

En efecto, $T^* + c$ es un mayorante medible de $T + c$; y si U es un mayorante medible de $T + c$, entonces $U - c$ lo es de T, de modo que, por la minimalidad, $T^* \leq U - c$, $\mathbb{P}$-c.s., es decir, $T^* + c \leq U$, $\mathbb{P}$-c.s. Como c es finito, ninguna de estas operaciones produce formas indeterminadas.

Con esto, *(iii)* es un cálculo directo a partir de la definición $T_* := -(-T)^*$: como $-\mathbb{1}_{B^c} = \mathbb{1}_B - 1$,

$$(\mathbb{1}_{B^c})_* := -\Big(-\mathbb{1}_{B^c}\Big)^* = -\Big(\mathbb{1}_B - 1\Big)^* = -\Big((\mathbb{1}_B)^* - 1\Big) = 1 - (\mathbb{1}_B)^* \qquad \mathbb{P}\text{-c.s.},$$

donde la tercera igualdad es (1.3) con $c = -1$. Es decir, $(\mathbb{1}_B)^* + (\mathbb{1}_{B^c})_* = 1$, $\mathbb{P}$-c.s. Esto prueba *(iii)*.

Finalmente, intercambiando B y B^c en *(iii)*, $(\mathbb{1}_B)_* = 1 - (\mathbb{1}_{B^c})^*$, $\mathbb{P}$-c.s. Por la definición de probabilidad interior, la primera identidad de *(i)* aplicada a B^c, el Lema 1.5 y la identidad anterior —la linealidad es lícita porque $(\mathbb{1}_{B^c})^*$ es integrable, con esperanza $\mathbb{P}^*(B^c) \leq 1$—,

$$\mathbb{P}_*(B) = 1 - \mathbb{P}^*(B^c) = 1 - \mathbb{E}^*(\mathbb{1}_{B^c}) = 1 - \mathbb{E}(\mathbb{1}_{B^c})^* = \mathbb{E}\Big(1 - (\mathbb{1}_{B^c})^*\Big) = \mathbb{E}\Big((\mathbb{1}_B)_*\Big).$$

Queda probado *(i)*. □

En el lema que sigue adoptamos la siguiente convención de lectura. Las operaciones que involucran a S y T directamente (por ejemplo, $S + T$, $S - T$ o ST) se suponen bien definidas punto a punto en todo Ω, sin formas $\infty - \infty$ ni productos $0 \cdot \infty$ no convenidos; esto garantiza que los mayorantes medibles mínimos del miembro izquierdo, como $(S + T)^*$, tengan sentido. En cambio, las expresiones formadas con S^*, T^*, S_*, T_* (por ejemplo, $S^* + T^*$ o $S^* - T^*$) pueden quedar indefinidas en parte de Ω aun cuando la operación correspondiente sobre S y T esté definida en todas partes. Por ello, cada desigualdad e igualdad del lema se afirma $\mathbb{P}$-c.s. *restringida al conjunto donde todas las expresiones del ítem correspondiente están definidas*; fuera de dicho conjunto no se afirma nada. Cuando S y T son finitas —el único caso que se usa en el resto del capítulo— se tiene $S^* \geq S > -\infty$ y $T^* \geq T > -\infty$

[8] Si h es medible, h es un mayorante medible de sí misma y la parte *(ii)* del Lema 1.5 da $h^* \leq h$ c.s.; junto con $h^* \geq h$ punto a punto, resulta $h^* = h$ c.s.

punto a punto, de modo que las sumas de mayorantes de (i) y (ii) están automáticamente definidas en todo Ω; en (iii), en cambio, todavía puede ocurrir $S^* = T^* = +\infty$, y la desigualdad se afirma únicamente donde $S^* - T^*$ está definida.

Lema 1.9. *Sean $S, T : \Omega \to \overline{\mathbb{R}}$ arbitrarias (no necesariamente medibles), y sean S^*, T^* sus mayorantes medibles mínimos. Entonces,*

(i) $(S+T)^* \leq S^* + T^*$. *Además, si S es medible, entonces $(S+T)^* = S + T^*$; análogamente, si T es medible, $(S+T)^* = S^* + T$.*

(ii) $S^* \leq (S-T)^* + T^*$; *en particular, en el conjunto $\{T^* < +\infty\}$ vale $(S-T)^* \geq S^* - T^*$.*

(iii) $\left|S^* - T^*\right| \leq |S-T|^*$ *en el conjunto donde $S^* - T^*$ está definida.*

(iv) Si S es medible y finita, entonces, definiendo $ST = 0$ en $\{S = 0\}$, $(ST)^* = ST^* \mathbb{1}_{\{S>0\}} + ST_* \mathbb{1}_{\{S<0\}}$.

(v) Si $S, T \geq 0$, entonces, con la convención $0 \cdot \infty = 0$, $(ST)^* \leq S^* T^*$.

(vi) Para todo $c \in \mathbb{R}$, $(\mathbb{1}_{\{T>c\}})^* = \mathbb{1}_{\{T^*>c\}}$, y por tanto $\mathbb{P}^*(T > c) = \mathbb{P}(T^* > c)$.

(vii) $|T|^* = T^* \vee (-T)^* = T^* \vee (-T_*) = |T^*| \vee |T_*|$; *en particular $|T^*| \leq |T|^*$ y $|T_*| \leq |T|^*$.*

(viii) $(S \vee T)^* = S^* \vee T^*$.

(ix) $(S \wedge T)^* \leq S^* \wedge T^*$, *con igualdad si S (o T) es medible.*

Demostración. Recordemos que $S_* \leq S \leq S^*$ y $T_* \leq T \leq T^*$ punto a punto.

Usaremos repetidamente el siguiente *principio de localización*: si $A \in \mathcal{A}$, $v : A \to \overline{\mathbb{R}}$ es medible para la σ-álgebra traza $\mathcal{A}|_A$, y $W \leq v$ punto a punto en A, entonces $W^* \leq v$, $\mathbb{P}$-c.s. en A. En efecto, la función que coincide con v en A y con W^* en A^c es medible y domina a W punto a punto. Por la minimalidad de W^*, este es menor o igual que dicha función $\mathbb{P}$-c.s.; restringiendo a A, obtenemos $W^* \leq v$. Aplicando el mismo argumento a $-W$ se obtiene la versión dual: si $v \leq W$ punto a punto en A, entonces $v \leq W_*$, $\mathbb{P}$-c.s. en A.

(i). Sea A el conjunto medible donde $S^* + T^*$ está definida. Como $S \leq S^*$ y $T \leq T^*$, en A se tiene $S+T \leq S^*+T^*$. Puesto que S^*+T^* es medible en A, el principio de localización da $(S+T)^* \leq S^*+T^*$.

Supongamos ahora que S es medible y sea $B := \{|S| < \infty\}$. En B, la función $S + T^*$ es medible para la σ-álgebra traza y, como $T \leq T^*$, satisface $S+T \leq S+T^*$. Por localización, $(S+T)^* \leq S+T^*$, $\mathbb{P}$-c.s. en B.

Para probar la desigualdad opuesta, pongamos $U := (S+T)^*$. Por definición de mayorante medible mínimo, $U \geq S + T$ punto a punto. Como S es finita en B, podemos restarla y obtener $U - S \geq T$ en B. Además, $U - S$ es medible para la σ-álgebra traza sobre B. Por localización, $T^* \leq U - S$, $\mathbb{P}$-c.s. en B. Sumando S, que es finita en B, resulta $S + T^* \leq U = (S+T)^*$, $\mathbb{P}$-c.s. en B. Por consiguiente, $(S+T)^* = S + T^*$, $\mathbb{P}$-c.s. en B.

En $\{S = +\infty\}$, la buena definición de $S + T$ y de $S + T^*$ implica, respectivamente, $T > -\infty$ y $T^* > -\infty$. Por tanto, $S + T = S + T^* = +\infty$. Como $(S+T)^* \geq S + T$ punto a punto, también $(S+T)^* = +\infty$ en ese conjunto.

En $\{S = -\infty\}$, la buena definición de $S+T^*$ implica $T^* < +\infty$. Como $T \leq T^*$, también $T < +\infty$, y por tanto $S + T = S + T^* = -\infty$. Aplicando el principio de localización a $S + T$, con $v \equiv -\infty$, obtenemos $(S+T)^* = -\infty$, $\mathbb{P}$-c.s. en ese conjunto.

Como $\Omega = B \cup \{S = +\infty\} \cup \{S = -\infty\}$, concluimos que $(S+T)^* = S + T^*$, $\mathbb{P}$-c.s. El caso en que T es medible es simétrico.

(ii). Sea A el conjunto medible donde $(S-T)^*+T^*$ está definida. Debemos probar que $S \leq (S-T)^*+T^*$ punto a punto en A. No podemos limitarnos a escribir $S = (S-T)+T$, pues, cuando T es infinita, la suma del miembro derecho puede adoptar una forma indeterminada.

Si $S = -\infty$, la desigualdad es inmediata. Si $S, T \in \mathbb{R}$, entonces $S = (S-T)+T \leq (S-T)^*+T^*$. Si $S = +\infty$, la buena definición de $S - T$ implica $T < +\infty$, de modo que $S - T = +\infty$ y, como

$(S-T)^* \geq S-T$, $(S-T)^* = +\infty$. Si $S \in \mathbb{R}$ y $T = +\infty$, entonces $T^* = +\infty$. Finalmente, si $S \in \mathbb{R}$ y $T = -\infty$, entonces $S - T = +\infty$ y $(S-T)^* = +\infty$. En estos tres últimos casos, el hecho de que $(S-T)^* + T^*$ esté definida fuerza a que su valor sea $+\infty$. Así, en todos los casos, $S \leq (S-T)^* + T^*$ en A.

Como el miembro derecho es medible en A, el principio de localización da $S^* \leq (S-T)^* + T^*$.

En $\{T^* \in \mathbb{R}\}$ podemos restar T^*, sin producir formas indeterminadas, y obtenemos $(S-T)^* \geq S^* - T^*$. Resta considerar $\{T^* = -\infty\}$. Allí, $T \leq T^* = -\infty$ implica $T = -\infty$. Como $S - T$ está definida, necesariamente $S > -\infty$, de modo que $S - T = +\infty$ y $(S-T)^* = +\infty$. Además, $S^* \geq S > -\infty$, por lo que $S^* - T^* = +\infty$. En consecuencia, $(S-T)^* \geq S^* - T^*$ en todo $\{T^* < +\infty\}$.

(iii). Supongamos primero que $S^*, T^* \in \mathbb{R}$. Aplicando (ii) a (S,T) y luego a (T,S), obtenemos $S^* - T^* \leq (S-T)^*$ y $T^* - S^* \leq (T-S)^*$. Como $S - T \leq |S-T|$ y $T - S \leq |S-T|$, la monotonía del mayorante medible mínimo da $(S-T)^* \leq |S-T|^*$ y $(T-S)^* \leq |S-T|^*$. Por tanto, $S^* - T^* \leq |S-T|^*$ y $T^* - S^* \leq |S-T|^*$, de donde $|S^* - T^*| \leq |S-T|^*$.

Supongamos ahora que $T^* = +\infty$ y $S^* \in \mathbb{R}$. Aplicando (ii) con los papeles de S y T intercambiados, obtenemos $T^* \leq (T-S)^* + S^*$. Como S^* es real, esta desigualdad fuerza $(T-S)^* = +\infty$. Por monotonía, $|S-T|^* \geq (T-S)^* = +\infty$, y la desigualdad buscada es inmediata. El caso $S^* = +\infty$ y $T^* \in \mathbb{R}$ es simétrico.

Si $S^* = -\infty$, entonces $S = -\infty$ punto a punto en ese conjunto. Como $S - T$ está definida, necesariamente $T > -\infty$, y por tanto $|S-T| = +\infty$. De aquí, $|S-T|^* = +\infty$. El caso $T^* = -\infty$ es simétrico. Estos casos incluyen aquellos en que uno de los mayorantes vale $+\infty$ y el otra $-\infty$, y cubren todo el conjunto donde $S^* - T^*$ está definida.

(iv). Supongamos que S es medible y finita, y definamos $R := ST^* \mathbb{1}_{\{S>0\}} + ST_* \mathbb{1}_{\{S<0\}}$. La función R es medible. Como $T_* \leq T \leq T^*$, se tiene $ST \leq R$ punto a punto. En efecto, en $\{S > 0\}$, la desigualdad $T \leq T^*$ da $ST \leq ST^*$; en $\{S < 0\}$, la desigualdad $T_* \leq T$, multiplicada por $S < 0$, da $ST \leq ST_*$; y en $\{S = 0\}$, ambos miembros valen 0. Por la minimalidad de $(ST)^*$, concluimos que $(ST)^* \leq R$.

Para probar la desigualdad opuesta, pongamos $U := (ST)^*$. En $\{S > 0\}$, de $U \geq ST$ se deduce $U/S \geq T$. Como S es finita y estrictamente positiva en ese conjunto, U/S es medible para la σ-álgebra traza. Por localización, $T^* \leq U/S$, y multiplicando por $S > 0$ obtenemos $ST^* \leq U$.

En $\{S < 0\}$, al dividir $U \geq ST$ por $S < 0$, la desigualdad se invierte y obtenemos $U/S \leq T$. La versión dual del principio de localización da $U/S \leq T_*$. Multiplicando nuevamente por $S < 0$, la desigualdad vuelve a invertirse y resulta $ST_* \leq U$.

Finalmente, en $\{S = 0\}$, se tiene $R = 0 = ST \leq U$. Por tanto, $R \leq U = (ST)^*$ en todo Ω. Junto con la desigualdad anterior, esto prueba $(ST)^* = R$.

(v). Como $S^* \geq S \geq 0$ y $T^* \geq T \geq 0$ punto a punto y, con la convención $0 \cdot \infty = 0$, el producto en $[0,\infty]$ es no decreciente en cada factor, se tiene $ST \leq S^*T^*$ punto a punto. El miembro derecho está bien definido y es medible; la minimalidad de $(ST)^*$ da $(ST)^* \leq S^*T^*$, $\mathbb{P}$-c.s.

(vi). Fijemos $c \in \mathbb{R}$. Como $T \leq T^*$, se tiene $\mathbb{1}_{\{T>c\}} \leq \mathbb{1}_{\{T^*>c\}}$ punto a punto, con miembro derecho medible; la minimalidad de $(\mathbb{1}_{\{T>c\}})^*$ da $(\mathbb{1}_{\{T>c\}})^* \leq \mathbb{1}_{\{T^*>c\}}$, $\mathbb{P}$-c.s.

Veamos que $(\mathbb{1}_{\{T>c\}})^* \geq \mathbb{1}_{\{T^*>c\}}$, $\mathbb{P}$-c.s. Sea $A := \{(\mathbb{1}_{\{T>c\}})^* < 1\} \in \mathcal{A}$. En A se tiene $T \leq c$ punto a punto: si fuera $T(\omega) > c$, entonces $(\mathbb{1}_{\{T>c\}})^*(\omega) \geq \mathbb{1}_{\{T>c\}}(\omega) = 1$. Por el principio de localización con $v \equiv c$, $T^* \leq c$, $\mathbb{P}$-c.s. en A; es decir, $\{T^* > c\} \cap A$ tiene probabilidad 0. Como $(\mathbb{1}_{\{T>c\}})^* \geq 0$ punto a punto y en A^c $(\mathbb{1}_{\{T>c\}})^* \geq 1$ se sigue $\mathbb{1}_{\{T^*>c\}} \leq (\mathbb{1}_{\{T>c\}})^*$, $\mathbb{P}$-c.s.

Por tanto, $(\mathbb{1}_{\{T>c\}})^* = \mathbb{1}_{\{T^*>c\}}$, $\mathbb{P}$-c.s.

Por el Lema 1.8(i) $\mathbb{P}^*(T > c) = \mathbb{E}^*(\mathbb{1}_{\{T>c\}})$. Por el Lema 1.5(iii), $\mathbb{E}^*(\mathbb{1}_{\{T>c\}}) = \mathbb{E}\,(\mathbb{1}_{\{T>c\}})^*$. Como $(\mathbb{1}_{\{T>c\}})^* = \mathbb{1}_{\{T^*>c\}}$ $\mathbb{P}$-c.s. y son medibles, podemos tomar esperanza común, y se tiene $\mathbb{E}\,(\mathbb{1}_{\{T>c\}})^* = \mathbb{E}\,\mathbb{1}_{\{T^*>c\}}$. Como $\{T^* > c\}$ es medible, $\mathbb{E}\,\mathbb{1}_{\{T^*>c\}} = \mathbb{P}(T^* > c)$. Finalmente, juntando estas igualdades, probamos que $\mathbb{P}^*(T > c) = \mathbb{P}(T^* > c)$.

Probamos primero (viii), que usaremos en la demostración de (vii).

(viii). Como $S^* \vee T^*$ es medible y domina punto a punto a $S \vee T$, la minimalidad de $(S \vee T)^*$ da $(S \vee T)^* \leq S^* \vee T^*$.

Recíprocamente, $(S \vee T)^*$ domina a $S \vee T$ y, por tanto, domina separadamente a S y a T. Por la minimalidad de S^* y T^*, obtenemos $S^* \leq (S \vee T)^*$ y $T^* \leq (S \vee T)^*$. En consecuencia, $S^* \vee T^* \leq (S \vee T)^*$. Combinando ambas desigualdades, $(S \vee T)^* = S^* \vee T^*$.

(vii). Como $|T| = T \vee (-T)$, el apartado (viii), aplicado a T y $-T$, da $|T|^* = (T \vee (-T))^* = T^* \vee (-T)^*$. Por definición de T_*, se tiene $(-T)^* = -T_*$, y por tanto $|T|^* = T^* \vee (-T_*)$.

Por otra parte, como $T_* \leq T^*$, también $-T^* \leq -T_*$. En consecuencia, $|T^*| \vee |T_*| = T^* \vee (-T^*) \vee T_* \vee (-T_*) = T^* \vee (-T_*)$, pues $T_* \leq T^*$ y $-T^* \leq -T_*$. Por tanto, $|T|^* = |T^*| \vee |T_*|$. En particular, $|T^*| \leq |T|^*$ y $|T_*| \leq |T|^*$.

(ix). Como $S \wedge T \leq S$ y $S \wedge T \leq T$, la monotonía del mayorante medible mínimo implica $(S \wedge T)^* \leq S^*$ y $(S \wedge T)^* \leq T^*$. Por tanto, $(S \wedge T)^* \leq S^* \wedge T^*$.

Supongamos ahora que S es medible y pongamos $U := (S \wedge T)^*$. El conjunto $A := \{U < S\}$ pertenece a $\mathcal{A}$, pues U y S son medibles. En A se tiene necesariamente $U \geq T$. En efecto, si en algún punto de A también se tuviera $U < T$, entonces $U < S \wedge T$, contradiciendo que $U \geq S \wedge T$ punto a punto.

Como U es medible para la σ-álgebra traza sobre A, el principio de localización aplicado a $T \leq U$ da $T^* \leq U$, $\mathbb{P}$-c.s. en A. Por tanto, $S \wedge T^* \leq U$ en A. En A^c, por definición de A, se tiene $S \leq U$, y en consecuencia también $S \wedge T^* \leq U$. Así, $S \wedge T^* \leq U = (S \wedge T)^*$, $\mathbb{P}$-c.s. en Ω.

Como S es medible, $S^* = S$, $\mathbb{P}$-c.s. La desigualdad general ya probada implica $U \leq S^* \wedge T^* = S \wedge T^*$. Combinando ambas desigualdades, obtenemos $(S \wedge T)^* = S \wedge T^* = S^* \wedge T^*$, $\mathbb{P}$-c.s. El caso en que T es medible es simétrico. □

Corolario 1.10 (Desigualdad de Markov exterior)**.** *Sea $T : \Omega \to [0, \infty]$ arbitraria y $c > 0$. Entonces*

$$\mathbb{P}^*(T \geq c) \leq (\mathbb{E}^* T)/c.$$

Demostración. Como $T^* \geq T \geq 0$ punto a punto, la esperanza $\mathbb{E}T^*$ existe (su parte negativa es nula) y, por el Lema 1.5(iii), $\mathbb{E}T^* = \mathbb{E}^* T$. Sea $0 < \varepsilon < c$. Como $\{T \geq c\} \subset \{T > c - \varepsilon\}$, la monotonía de $\mathbb{P}^*$ (si $B_1 \subset B_2$, todo recubrimiento medible de B_2 lo es de B_1, y el ínfimo sólo puede bajar) da $\mathbb{P}^*(T \geq c) \leq \mathbb{P}^*(T > c - \varepsilon)$. Por el Lema 1.9(vi), aplicado con la constante $c - \varepsilon$, $\mathbb{P}^*(T > c - \varepsilon) = \mathbb{P}(T^* > c - \varepsilon)$. Por tanto, $\mathbb{P}^*(T \geq c) \leq \mathbb{P}(T^* > c - \varepsilon)$, $0 < \varepsilon < c$. Haciendo $\epsilon \downarrow 0$, $\mathbb{P}^*(T \geq c) \leq \mathbb{P}(T^* \geq c)$. La variable T^* es medible y no negativa, así que la desigualdad de Markov usual se aplica: de $c\,\mathbb{1}_{\{T^* \geq c\}} \leq T^*$ punto a punto, tomando esperanzas y dividiendo por $c > 0$, $\mathbb{P}(T^* \geq c) \leq (\mathbb{E}T^*)/c$. □

Advertencia 1.11

- La probabilidad exterior puede no ser aditiva. En $([0,1], \mathcal{L}, \lambda)$, sea $V \subset [0,1]$ un conjunto de Vitali: un representante de cada clase $x + \mathbb{Q}$. Entonces $\lambda_*(V) = 0$ y $\lambda^*(V) > 0$.[a] Si $A = V$ y $B = [0,1] \setminus V$, entonces $A \cap B = \varnothing$, $A \cup B = [0,1]$, y, por la dualidad $\mathbb{P}^*(B) = 1 - \lambda_*(V) = 1$,

  $$\mathbb{P}^*(A) + \mathbb{P}^*(B) = \lambda^*(V) + 1 > 1 = \mathbb{P}^*(A \cup B).$$

  Así, la aditividad puede fallar aun para conjuntos disjuntos.

- La esperanza exterior e interior tampoco son, en general, aditivas. Más precisamente, para variables no necesariamente medibles se tiene (se deja como ejercicio verificarlo)

  $$\mathbb{E}^*(Y + Z) \leq \mathbb{E}^* Y + \mathbb{E}^* Z, \qquad \mathbb{E}_*(Y + Z) \geq \mathbb{E}_* Y + \mathbb{E}_* Z,$$

  siempre que los lados derechos estén bien definidos. La igualdad no tiene por qué valer si las variables no son medibles. Si V es no medible, se puede tomar como ejemplo $Y = \mathbb{1}_V$ y $Z = \mathbb{1}_{V^c}$. Lo que sí se cumple es la homogeneidad positiva: $\mathbb{E}^*(cY) = c\mathbb{E}^* Y$ y $\mathbb{E}_*(cY) = c\mathbb{E}_* Y$ para $c > 0$; para $c = 0$ la identidad vale con la convención $0 \cdot (\pm\infty) = 0$ (en cuyo caso ambos lados son nulos).

[a] Si $K \subset V$ es medible con $\lambda(K) > 0$, los trasladados $K + q$, $q \in \mathbb{Q} \cap [-1, 1]$, son disjuntos dos a dos —los puntos de V están en clases distintas—, tienen todos medida $\lambda(K)$ y están contenidos en $[-1, 2]$, contradiciendo la aditividad numerable; luego $\lambda_*(V) = 0$. Y si fuera $\lambda^*(V) = 0$, los numerables trasladados $V + q$, $q \in \mathbb{Q} \cap [-1, 1]$, serían nulos, pero cubren $[0, 1]$: todo $x \in [0, 1]$ difiere de su representante en un racional de $[-1, 1]$.

1.2.2. Fubini exterior

Sean $(\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$ y $(\Omega_2, \mathcal{A}_2, \mathbb{P}_2)$ espacios de probabilidad. En $\Omega_1 \times \Omega_2$ consideramos la σ-álgebra producto $\mathcal{A}_1 \otimes \mathcal{A}_2$, es decir, la generada por los rectángulos $A \times B$ con $A \in \mathcal{A}_1$, $B \in \mathcal{A}_2$, y la probabilidad producto $\mathbb{P}_1 \otimes \mathbb{P}_2$. Conviene subrayar que $\mathcal{A}_1 \otimes \mathcal{A}_2$ *no está completada* respecto de $\mathbb{P}_1 \otimes \mathbb{P}_2$ (véase la observación que sigue al Lema 1.14). Sea $T : \Omega_1 \times \Omega_2 \to \overline{\mathbb{R}}$ arbitraria. Escribimos $\mathbb{E}$ para la esperanza respecto de $\mathbb{P}_1 \otimes \mathbb{P}_2$.

Definición 1.13 (Notación)**.** Para $\omega_1 \in \Omega_1$ escribimos $T_{\omega_1}(\omega_2) = T(\omega_1, \omega_2)$, y denotamos por $\mathbb{E}_2^* T(\omega_1)$ y $\mathbb{E}_{2*} T(\omega_1)$ la esperanza exterior y la interior de la sección T_{ω_1} en $(\Omega_2, \mathcal{A}_2, \mathbb{P}_2)$, dadas por la Definición 1.1. Así, $\mathbb{E}_2^* T$ y $\mathbb{E}_{2*} T$ son funciones de Ω_1 en $\overline{\mathbb{R}}$, definidas en todo punto y no necesariamente medibles. Análogamente, para $\omega_2 \in \Omega_2$, escribimos $T_{\omega_2}(\omega_1) = T(\omega_1, \omega_2)$ y definimos $\mathbb{E}_1^* T(\omega_2)$ y $\mathbb{E}_{1*} T(\omega_2)$ cambiando los papeles de los índices.

Antes de enunciar el resultado, conviene entender por qué cabe esperar desigualdades y no igualdades. La esperanza exterior $\mathbb{E}^* T$ es el costo de dominar a T mediante un *único* mayorante medible en el espacio producto. En el cálculo iterado $\mathbb{E}_1^* \mathbb{E}_2^* T$, en cambio, se domina cada sección $T(\omega_1, \cdot)$ por separado, y el mayorante elegido para la sección puede depender de ω_1 de forma arbitraria: no se exige que el ensamblado $(\omega_1, \omega_2) \mapsto U_{\omega_1}(\omega_2)$ sea medible en $\mathcal{A}_1 \otimes \mathcal{A}_2$. Dominar sección a sección es, pues, una exigencia más débil que dominar globalmente, y el ínfimo iterado resulta menor o igual que el global: $\mathbb{E}_1^* \mathbb{E}_2^* T \le \mathbb{E}^* T$. Esta es la única desigualdad que requiere trabajo; la cota inferior de la cadena se obtiene por dualidad, aplicando la anterior a $-T$, y la desigualdad central es simplemente "interior $\le$ exterior". En el Teorema de Fubini clásico, la medibilidad conjunta del integrando es precisamente lo que permite convertir estas desigualdades en igualdades.

La prueba usa dos hechos sobre integrales parciales de funciones medibles no negativas en el producto. El primero es un enunciado de medibilidad, y es el que necesita el "para todo ω_1"; el segundo es la igualdad de integrales, el Teorema de Tonelli propiamente dicho.

El siguiente lema se prueba en el apéndice.

Lema 1.14. *Sean $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$, $i = 1, 2$, espacios de probabilidad y $\mathbb{P} := \mathbb{P}_1 \otimes \mathbb{P}_2$ la probabilidad producto sobre $\mathcal{A}_1 \otimes \mathcal{A}_2$, es decir, la única probabilidad sobre $\mathcal{A}_1 \otimes \mathcal{A}_2$ tal que $\mathbb{P}(A \times B) = \mathbb{P}_1(A)\mathbb{P}_2(B)$ para todo $A \in \mathcal{A}_1$, $B \in \mathcal{A}_2$. Sea $g : \Omega_1 \times \Omega_2 \to [0, \infty]$ medible respecto de $\mathcal{A}_1 \otimes \mathcal{A}_2$. Entonces:*

(a) para todo $\omega_1 \in \Omega_1$, la sección $g(\omega_1, \cdot) : \Omega_2 \to [0, \infty]$ es $\mathcal{A}_2$-medible;

(b) la función $\Phi(\omega_1) := \int_{\Omega_2} g(\omega_1, \omega_2)\, d\mathbb{P}_2(\omega_2)$, que por (a) está definida para todo ω_1 con valores en $[0, \infty]$, es $\mathcal{A}_1$-medible;

(c) (Tonelli)

$$\mathbb{E}g = \int_{\Omega_1 \times \Omega_2} g\, d\mathbb{P} = \int_{\Omega_1} \Big(\int_{\Omega_2} g(\omega_1, \omega_2)\, d\mathbb{P}_2(\omega_2) \Big) d\mathbb{P}_1(\omega_1),$$

con ambos miembros en $[0, \infty]$.

Observación 1.15. El Lema 1.14 es puntual en ω_1 porque $\mathcal{A}_1 \otimes \mathcal{A}_2$ no está completada. Si se la reemplaza por su completación respecto de $\mathbb{P}_1 \otimes \mathbb{P}_2$, la parte (a) falla: en $\mathbb{R}^2$ con la σ-álgebra de Lebesgue, si $V \subset \mathbb{R}$ no es medible, el conjunto $V \times \{0\}$ es medible, pues está contenido en el nulo $\mathbb{R} \times \{0\}$, pero su sección en la segunda coordenada 0 es V. Con la completación, (a) y (b) valen sólo para $\mathbb{P}_1$-casi todo ω_1, que es la forma en que se enuncia el Teorema de Tonelli para la medida de Lebesgue en $\mathbb{R}^d$.

Lema 1.16. *Sea $T : \Omega_1 \times \Omega_2 \to \overline{\mathbb{R}}$ arbitraria. Entonces,*

$$\mathbb{E}_* T \le \mathbb{E}_{1*}\mathbb{E}_{2*} T \le \mathbb{E}_1^*\mathbb{E}_2^* T \le \mathbb{E}^* T.$$

Análogamente, intercambiando 1 y 2, $\mathbb{E}_* T \le \mathbb{E}_{2*}\mathbb{E}_{1*} T \le \mathbb{E}_2^*\mathbb{E}_1^* T \le \mathbb{E}^* T$.

Demostración. La prueba tiene tres partes: la desigualdad de la derecha, que es la única que requiere trabajo; la central, que es elemental; y la de la izquierda, que se obtiene de la primera por dualidad.

Si $\mathbb{E}^* T = +\infty$ no hay nada que probar. Supongamos $\mathbb{E}^* T < +\infty$ y sea U un mayorante medible de T[9] cuya esperanza existe y satisface $\mathbb{E}U < +\infty$; en particular $\mathbb{E}U^+ < \infty$. Basta probar $\mathbb{E}_1^*\mathbb{E}_2^* T \le \mathbb{E}U$ para cada tal U y tomar ínfimo; restringir el ínfimo a los U con $\mathbb{E}U < +\infty$ no lo altera cuando $\mathbb{E}^* T < +\infty$.

Fijemos $m \in \mathbb{N}$ y pongamos $V := U \vee (-m)$. Entonces V es medible, $V \ge U \ge T$ en todo punto, $V \ge -m$, y $V \le U^+$ en todo punto.[10] La cota inferior $V \ge -m$ es la clave: elimina toda forma $\infty - \infty$ en lo que sigue.

Por la parte (a) del Lema 1.14 para cada $\omega_1 \in \Omega_1$, la sección $V(\omega_1, \cdot)$ es $\mathcal{A}_2$-medible y su integral respecto de $\mathbb{P}_2$ existe porque $V(\omega_1, \cdot)^- \le m$. Definimos

$$\phi(\omega_1) := \int_{\Omega_2} V(\omega_1, \omega_2)\, d\mathbb{P}_2(\omega_2) = \int_{\Omega_2} (V + m)(\omega_1, \omega_2)\, d\mathbb{P}_2(\omega_2) - m \in [-m, +\infty].$$

La función ϕ es $\mathcal{A}_1$-medible por el Lema 1.14(b) y satisface la identidad de Tonelli[11]

$$\mathbb{E}_1 \phi = \mathbb{E}V. \tag{1.4}$$

En particular

$$-m \le \mathbb{E}_1 \phi = \mathbb{E}V \le \mathbb{E}U^+ < \infty, \tag{1.5}$$

donde $\mathbb{E}V \le \mathbb{E}U^+$ usa $V \le U^+$ y la monotonía de la esperanza extendida.

Veamos que ϕ domina a $\mathbb{E}_2^* T$. Fijemos ω_1. La sección $V(\omega_1, \cdot)$ es $\mathcal{A}_2$-medible, mayor o igual que $T(\omega_1, \cdot)$ en todo punto de Ω_2, y su integral existe y vale $\phi(\omega_1)$. Es decir, es uno de los mayorantes sobre los que se toma el ínfimo que define $\mathbb{E}_2^* T(\omega_1)$. Por tanto,

$$\mathbb{E}_2^* T(\omega_1) \le \phi(\omega_1) \qquad \text{para todo } \omega_1 \in \Omega_1. \tag{1.6}$$

La función $\omega_1 \mapsto \mathbb{E}_2^* T(\omega_1)$ es una función arbitraria sobre Ω_1, en general no medible. Por (1.6), ϕ es un mayorante $\mathcal{A}_1$-medible de ella en todo punto de Ω_1, y su esperanza $\mathbb{E}_1 \phi$ existe y es finita por (1.5). Luego ϕ es admisible en el ínfimo que define $\mathbb{E}_1^*$, y

$$\mathbb{E}_1^*\mathbb{E}_2^* T \le \mathbb{E}_1 \phi = \mathbb{E}V = \mathbb{E}\Big[U \vee (-m)\Big]. \tag{1.7}$$

Se tiene, punto a punto,[12]

$$\Big(U \vee (-m)\Big)^+ = U^+, \qquad \Big(U \vee (-m)\Big)^- = U^- \wedge m \uparrow U^-.$$

Como $\mathbb{E}V$ existe,

$$\mathbb{E}\Big[U \vee (-m)\Big] = \mathbb{E}\Big[(U \vee (-m))^+\Big] - \mathbb{E}\Big[(U \vee (-m))^-\Big] = \mathbb{E}U^+ - \mathbb{E}\Big(U^- \wedge m\Big) \xrightarrow[m\to\infty]{} \mathbb{E}U^+ - \mathbb{E}U^- = \mathbb{E}U,$$

[9] Medible respecto de $\mathcal{A}_1 \otimes \mathcal{A}_2$, sin completar, y con $U \ge T$ en *todo* punto de $\Omega_1 \times \Omega_2$, que es como está definida $\mathbb{E}^*$. Ambas cosas se usan más abajo: la primera para invocar el Lema 1.14, la segunda para que las secciones de U dominen a las de T en todo punto.

[10] Pues $U \le U^+$ y $-m \le 0 \le U^+$.

[11] Por el Lema 1.14(c) aplicado a $V + m \ge 0$, $\mathbb{E}(V + m) = \mathbb{E}_1[\phi + m]$, ambos en $[0, \infty]$. Como m es finito, $\mathbb{E}(V + m) = \mathbb{E}V + m$ (aquí $\mathbb{E}V$ existe porque $V^- \le m$) y $\mathbb{E}_1[\phi + m] = \mathbb{E}_1\phi + m$ (aquí $\mathbb{E}_1\phi$ existe porque $\phi^- \le m$). Restando m se obtiene (1.4), incluido el caso en que ambos miembros valen $+\infty$.

[12] Fijado un punto, sea u el valor de U. Como $-m \le 0$, $\max\{\max(u, -m), 0\} = \max(u, 0) = u^+$. Y $\max\{-\max(u, -m), 0\} = \max\{\min(-u, m), 0\}$: si $u \ge 0$ esto es $0 = u^- \wedge m$; si $u < 0$ es $\min(-u, m) = u^- \wedge m$. La sucesión $u^- \wedge m$ es creciente en m y vale u^- en cuanto $m \ge u^-$; luego $u^- \wedge m \uparrow u^-$.

donde la convergencia es el teorema de convergencia monótona aplicado a $U^- \wedge m \uparrow U^-$,[13] y la resta no produce forma indeterminada porque $\mathbb{E}U^+ < \infty$.[14] Haciendo $m \to \infty$ en (1.7), $\mathbb{E}_1^*\mathbb{E}_2^* T \le \mathbb{E}U$, como queríamos.

Veamos que $\mathbb{E}_{1*}\mathbb{E}_{2*}T \le \mathbb{E}_1^*\mathbb{E}_2^*T$. Para eso probamos primero que $\mathbb{E}_* X \le \mathbb{E}^* X$ para toda X. Sean $L \le X \le W$ con L, W medibles y esperanzas existentes. Por la monotonía de la esperanza extendida, $\mathbb{E}L \le \mathbb{E}W$. Tomando supremo en L e ínfimo en W, $\mathbb{E}_* X \le \mathbb{E}^* X$. Veamos ahora que $\mathbb{E}^*$ es monótona. Si $X \le Y$ en todo punto, todo mayorante medible de Y lo es de X; el conjunto sobre el que se toma el ínfimo para X contiene al de Y, y el ínfimo sólo puede bajar: $\mathbb{E}^* X \le \mathbb{E}^* Y$.

Aplicamos lo primero que probamos en $(\Omega_2, \mathcal{A}_2, \mathbb{P}_2)$ a cada sección $T(\omega_1, \cdot)$, y obtenemos $\mathbb{E}_{2*}T(\omega_1) \le \mathbb{E}_2^*T(\omega_1)$ para todo ω_1. Ahora, en $(\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$: con $X = \mathbb{E}_{2*}T$, $\mathbb{E}_{1*}\mathbb{E}_{2*}T \le \mathbb{E}_1^*\mathbb{E}_{2*}T$. Aplicamos lo segundo que probamos a $\mathbb{E}_{2*}T \le \mathbb{E}_2^*T$, $\mathbb{E}_1^*\mathbb{E}_{2*}T \le \mathbb{E}_1^*\mathbb{E}_2^*T$. Encadenando ambas se obtiene la desigualdad central.

$\mathbb{E}_* T \le \mathbb{E}_{1*}\mathbb{E}_{2*}T$ se obtiene aplicando lo que ya probamos a $-T$ y usando tres identidades de dualidad, todas consecuencia directa de la definición $\mathbb{E}_* X = -\mathbb{E}^*(-X)$:

$$\begin{aligned} \mathbb{E}_* T &= -\mathbb{E}^*(-T), \\ \mathbb{E}_{2*}T(\omega_1) &= -\mathbb{E}_2^*(-T)(\omega_1) \quad \text{para todo } \omega_1, \\ \mathbb{E}_{1*}X &= -\mathbb{E}_1^*(-X) \quad \text{para toda } X \text{ sobre } \Omega_1. \end{aligned}$$

Entonces

$$\mathbb{E}_{1*}\mathbb{E}_{2*}T = -\mathbb{E}_1^*\Big(-\mathbb{E}_{2*}T\Big) = -\mathbb{E}_1^*\Big(\mathbb{E}_2^*(-T)\Big) \ge -\mathbb{E}^*(-T) = \mathbb{E}_* T,$$

donde la primera igualdad es la tercera identidad con $X = \mathbb{E}_{2*}T$, la segunda es la segunda identidad aplicada punto a punto en ω_1, y la desigualdad es la Parte 1 para $-T$ con el signo cambiado.

La cadena con los índices 1 y 2 intercambiados se prueba idénticamente, aplicando el Lema 1.14 y los papeles de Ω_1 y Ω_2 invertidos. □

Advertencia 1.17

En general *no hay* un teorema de Fubini completo para $\mathbb{E}^*$: la igualdad $\mathbb{E}^*(T) = \mathbb{E}_1^*\mathbb{E}_2^*T$ puede fallar si T no es suficientemente regular.

Hay, sin embargo, una clase importante de funciones para las cuales el intercambio es exacto: las que dependen de una sola coordenada. En efecto: si $T = g \circ \pi_1$, cada sección $T(\omega_1, \cdot)$ es constante e igual a $g(\omega_1)$, y la esperanza exterior de una constante es esa constante; luego $\mathbb{E}_2^*T = g$ y $\mathbb{E}_1^*\mathbb{E}_2^*T = \mathbb{E}_1^*g$. La pregunta de Fubini se reduce entonces a si $\mathbb{E}^*(g \circ \pi_1) = \mathbb{E}_1^*g$, es decir, a si la esperanza exterior calculada en el producto —donde hay más mayorantes disponibles, que pueden depender de ω_2— coincide con la calculada en Ω_1. Eso es exactamente la perfección de π_1 del lema siguiente.

Definición 1.18. Sea $\phi : (\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbb{P}}) \to (\Omega, \mathcal{A}, \mathbb{P})$ medible, diremos que ϕ es **perfecta** si, para todo $B \subset \Omega$,

$$\mathbb{P}^*(B) = \tilde{\mathbb{P}}^*\Big(\phi^{-1}(B)\Big)$$

Como esto vale para todo B, $\mathbb{P}_*(B) = \tilde{\mathbb{P}}_*\Big(\phi^{-1}(B)\Big)$.

Lema 1.19. *Sea $\phi : (\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbb{P}}) \to (\Omega, \mathcal{A}, \mathbb{P})$ medible y tal que $\mathbb{P} = \tilde{\mathbb{P}} \circ \phi^{-1}$. Entonces son equivalentes:*

(i) ϕ es perfecta;

(ii) Para toda función acotada $g : \Omega \to \mathbb{R}$, $\mathbb{E}^* g = \tilde{\mathbb{E}}^*(g \circ \phi)$.

[13] Las funciones $U^- \wedge m$ son medibles, no negativas y crecen puntualmente a U^-, así que $\mathbb{E}(U^- \wedge m) \uparrow \mathbb{E}U^- \in [0, \infty]$.

[14] Si $\mathbb{E}U^- = \infty$, entonces $\mathbb{E}U = -\infty$ y el miembro izquierdo tiende a $-\infty$; (1.7) sigue valiendo para cada m y en el límite. En ese caso, de hecho, $\mathbb{E}^*T = -\infty$.

Demostración. Sea $g:\Omega\to\mathbb{R}$ con $|g|\le M$. Como $g^*\ge g\ge -M$; $\mathbb{E}g^*$ existe y, por el Lema 1.5(iii), vale $\mathbb{E}^*g$. Por la fórmula de capas aplicada a $g^*+M\ge 0$ y por el Lema 1.9(vi),

$$\mathbb{E}^*g=\mathbb{E}g^*=-M+\mathbb{E}(g^*+M)=-M+\int_0^{2M}\mathbb{P}(g^*+M>s)\,ds$$
$$=-M+\int_{-M}^{M}\mathbb{P}(g^*>t)\,dt=-M+\int_{-M}^{M}\mathbb{P}^*(g>t)\,dt. \quad (1.8)$$

Es decir, la esperanza exterior de una función acotada queda determinada por las probabilidades exteriores de sus conjuntos de nivel.

(i)⇒(ii). Sea $|g|\le M$; entonces $|g\circ\phi|\le M$. Como $\{g\circ\phi>t\}=\phi^{-1}(\{g>t\})$, la hipótesis da $\widetilde{\mathbb{P}}^*(g\circ\phi>t)=\mathbb{P}^*(g>t)$ para todo t. Aplicando (1.8) en ambos espacios, $\widetilde{\mathbb{E}}^*(g\circ\phi)=\mathbb{E}^*g$.

(ii)⇒(i). Con $g=\mathbb{1}_B$ y $\mathbb{1}_B\circ\phi=\mathbb{1}_{\phi^{-1}(B)}$, el Lema 1.8(i) en ambos espacios da $\mathbb{P}^*(B)=\mathbb{E}^*(\mathbb{1}_B)=\widetilde{\mathbb{E}}^*(\mathbb{1}_{\phi^{-1}(B)})=\widetilde{\mathbb{P}}^*(\phi^{-1}(B))$. □

Lema 1.20 (Las proyecciones coordenadas son perfectas)**.** *Sea $\pi_1:\Omega_1\times\Omega_2\to\Omega_1$ la proyección, definida sobre el espacio producto $(\Omega_1\times\Omega_2,\mathcal{A}_1\otimes\mathcal{A}_2,\mathbb{P}_1\otimes\mathbb{P}_2)$. Entonces:*

(i) para toda $T:\Omega_1\to\overline{\mathbb{R}}$, $(T\circ\pi_1)^*=T^*\circ\pi_1$ $(\mathbb{P}_1\otimes\mathbb{P}_2)$-c.s., donde el mayorante medible mínimo de la izquierda se calcula en el producto y el de la derecha en $(\Omega_1,\mathcal{A}_1,\mathbb{P}_1)$;

(ii) para toda $g:\Omega_1\to\mathbb{R}$ acotada, $\mathbb{E}^*(g\circ\pi_1)=\mathbb{E}_1^*g$; en consecuencia (Lema 1.19), π_1 es perfecta.

Lo mismo vale, intercambiando los índices, para la proyección π_2.

Demostración. (i) La función $T^*\circ\pi_1$ es medible en el producto y domina a $T\circ\pi_1$ punto a punto; por minimalidad, $(T\circ\pi_1)^*\le T^*\circ\pi_1$ c.s.

Recíprocamente, sea $V:=(T\circ\pi_1)^*$. Como $V\ge T\circ\pi_1$ punto a punto y $(T\circ\pi_1)(\omega_1,\omega_2)=T(\omega_1)$, al fijar $\omega_2\in\Omega_2$ la sección $V(\cdot,\omega_2)$ es $\mathcal{A}_1$-medible y domina punto a punto a T (la misma función T para todo ω_2) [15]; por la minimalidad de T^* en Ω_1, se tiene $T^*\le V(\cdot,\omega_2)$, $\mathbb{P}_1$-c.s. El conjunto

$$N:=\{(\omega_1,\omega_2):T^*(\omega_1)>V(\omega_1,\omega_2)\}$$

es medible y, por lo anterior, todas sus ω_2-secciones son $\mathbb{P}_1$-nulas; por Tonelli aplicado a $\mathbb{1}_N$, $(\mathbb{P}_1\otimes\mathbb{P}_2)(N)=0$. Es decir, $T^*\circ\pi_1\le V$ c.s., y queda probada la igualdad.

(ii) Sea $g:\Omega_1\to\mathbb{R}$ acotada. Probamos las dos desigualdades.

$\mathbb{E}^*(g\circ\pi_1)\le\mathbb{E}_1^*g$. Sea $U:\Omega_1\to\overline{\mathbb{R}}$ un mayorante $\mathcal{A}_1$-medible de g cuya esperanza existe. Entonces $U\circ\pi_1$ es $\mathcal{A}_1\otimes\mathcal{A}_2$-medible, domina a $g\circ\pi_1$ en todo punto y, como la ley de π_1 bajo $\mathbb{P}_1\otimes\mathbb{P}_2$ es $\mathbb{P}_1$, $\mathbb{E}[U\circ\pi_1]=\mathbb{E}_1U$. Luego $U\circ\pi_1$ es admisible en el ínfimo que define $\mathbb{E}^*(g\circ\pi_1)$, y $\mathbb{E}^*(g\circ\pi_1)\le\mathbb{E}_1U$. Tomando ínfimo en U se obtiene la desigualdad.

$\mathbb{E}_1^*g\le\mathbb{E}^*(g\circ\pi_1)$. Pongamos $T:=g\circ\pi_1$. Fijado ω_1, la sección $T(\omega_1,\cdot)$ es la constante real $c:=g(\omega_1)$, y $\mathbb{E}_2^*T(\omega_1)=c$: la propia constante es un mayorante medible con integral c, y todo mayorante medible $U\ge c$ tiene $\int U\,d\mathbb{P}_2\ge c$. Así $\mathbb{E}_2^*T=g$ en todo punto de Ω_1, y la desigualdad derecha del Lema 1.16 da

$$\mathbb{E}_1^*g=\mathbb{E}_1^*\mathbb{E}_2^*T\le\mathbb{E}^*T=\mathbb{E}^*(g\circ\pi_1).$$

Como la ley de π_1 bajo $\mathbb{P}_1\otimes\mathbb{P}_2$ es $\mathbb{P}_1$, la implicación (ii)⇒(i) del Lema 1.19 traduce esta identidad en la perfección de π_1. □

En otras palabras: para funciones del producto que dependen de una sola coordenada, la esperanza exterior conjunta y la marginal coinciden, sin ninguna hipótesis de regularidad. Esta es la propiedad que permitirá, en el Capítulo 2, agrandar el espacio muestral —con una muestra fantasma y signos de Rademacher— sin alterar las esperanzas exteriores de los funcionales de la muestra original: el paso clave de la simetrización.

[15] Aquí es donde se usa que π_1 es la proyección: como $(T\circ\pi_1)(\omega_1,\omega_2)=T(\omega_1)$ no depende de ω_2, todas las ω_2-secciones de $T\circ\pi_1$ son la *misma* función T, y la minimalidad en Ω_1 se aplica siempre contra el mismo T^*.

1.3. Convergencia débil

Sea (D,d) un espacio métrico, denotamos

$$C_b(D) = \{f : D \to \mathbb{R} \text{ continua y acotada}\},$$

donde por acotada entendemos $\sup_{x\in D} |f(x)| < \infty$.

Definición 1.22. Sea L una probabilidad de Borel sobre $(D, \mathcal{B}(D))$. Decimos que L es **tensa** si

$$\forall \varepsilon > 0 \ \exists K \subset D \text{ compacto tal que } \mathsf{L}(K) \geq 1-\varepsilon.$$

Si $X : (\Omega, \mathcal{A}, \mathbb{P}) \to D$ es Borel-medible, diremos que X *es tenso* cuando su ley $\mathsf{L}_X := \mathbb{P} \circ X^{-1}$ es tensa.

Definición 1.25. Diremos que una probabilidad de Borel L es **separable** si existe un conjunto numerable $M \subset D$ tal que $\mathsf{L}(\overline{M}) = 1$.

Advertencia 1.26

Una medida puede ser separable sin que el espacio métrico lo sea. Basta considerar en $\mathbb{R}$ la distancia $d(x,y) = 0$ si $x = y$, 1 en otro caso. Dicho espacio no es separable pero la masa de Dirac en cualquier punto es separable. Otro ejemplo menos trivial se obtiene si consideramos $\ell^\infty = \{x = (x_n)_{n\geq 1} : \|x\|_\infty := \sup_n |x_n| < \infty\}$ con la distancia inducida por $\|\cdot\|_\infty$. El espacio ℓ^∞ no es separable (aunque si es completo). Sea $c_0 = \{x : x_n \to 0\}$, que sí es separable (por ejemplo, las sucesiones de soporte finito con coordenadas racionales son densas en c_0).
Sea $(\Omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad y sea $Z : \Omega \to \mathbb{R}$ una variable aleatoria real. Definimos $\xi : \Omega \to \ell^\infty$, $\xi(\omega) = (Z(\omega)/n)_{n\geq 1}$. La aplicación $z \mapsto (z/n)_{n\geq 1}$ de $\mathbb{R}$ en ℓ^∞ es continua; por tanto, ξ es Borel-medible. Además, $\xi(\omega) \in c_0$ para todo ω. De este modo, la ley $\mu := \mathbb{P} \circ \xi^{-1}$ es una probabilidad de Borel en $(\ell^\infty, \mathcal{B}(\ell^\infty))$ tal que $\mu(c_0) = 1$. Concluimos que μ es *separable* (está concentrada en c_0) mientras que el espacio métrico total $(\ell^\infty, \|\cdot\|_\infty)$ no es separable.

Definición 1.27. En un espacio métrico (D,d), una probabilidad L de Borel es **pretensa** si, para todo $\varepsilon > 0$, existe $C \subset D$ totalmente acotado[16] y medible tal que $\mathsf{L}(C) \geq 1-\varepsilon$.

Lema 1.28. *Sea L una probabilidad de Borel en un espacio métrico (D,d). Entonces*

$$\mathsf{L} \textit{ es pretensa} \iff \mathsf{L} \textit{ es separable.}$$

Si además (D,d) es completo*, son equivalentes*

$$\mathsf{L} \textit{ es pretensa} \iff \mathsf{L} \textit{ es tensa} \iff \mathsf{L} \textit{ es separable.}$$

Demostración. Supongamos L pretensa. Para $n \geq 1$ sea $C_n \subset D$ totalmente acotado con $\mathsf{L}(C_n) \geq 1-2^{-n}$. Cada C_n es separable;[17] por tanto $C := \bigcup_n C_n$ es separable y $\mathsf{L}(C) = 1$. Sea $M \subset C$ numerable y denso en C; entonces $\mathsf{L}(\overline{M}) = 1$, es decir, L es separable.

Sea $M = \{x_1, x_2, \dots\} \subset D$ numerable con $\mathsf{L}(\overline{M}) = 1$. Fijado $\varepsilon > 0$, para cada $j \geq 1$ los conjuntos $B(x_m, 1/j)$, $m \geq 1$, cubren $\overline{M}$; por tanto

$$\mathsf{L}\Big(\bigcup_{m\geq 1} B(x_m, 1/j)\Big) = 1.$$

[16] $A \subset D$ es totalmente acotado si para todo $\varepsilon > 0$ existen $x_1, \dots, x_N \in D$ tales que $A \subset \cup_{i=1}^N B(x_i, \varepsilon)$. Nótese que depende de la métrica elegida. La relación entre pretensión, tensión y separabilidad es el Lema 1.28.

[17] Todo conjunto totalmente acotado C es separable: para cada k, C se cubre con finitas bolas de radio $1/k$; eligiendo un punto de C en cada bola que lo corte y uniendo sobre k, se obtiene un subconjunto numerable y denso en C.

Luego existe N_j tal que $G_j := \cup_{m=1}^{N_j} B(x_m, 1/j)$ satisface $\mathsf{L}(G_j) \geq 1 - \varepsilon 2^{-j}$.[18] Sea $C := \bigcap_{j\geq 1} G_j$. Entonces $C \in \mathcal{B}(D)$ y

$$\mathsf{L}(C) \geq 1 - \sum_{j\geq 1} \varepsilon 2^{-j} = 1 - \varepsilon.$$

Además, C es totalmente acotado: dado $\eta > 0$, sea j con $1/j < \eta$; como $C \subset G_j$, C queda cubierto por finitas bolas de radio η. Así, L es pretensa.

Si D es completo, el cierre de un conjunto totalmente acotado es compacto[19]; así, pretensa $\Rightarrow$ tensa. La implicación tensa $\Rightarrow$ pretensa es trivial ("compacto $\Rightarrow$ totalmente acotado"). Junto con lo ya probado, queda la equivalencia de las tres propiedades. □

Definición 1.29. Un par $(A, \leq)$ es un **orden parcial** si, para todo $x, y, z \in A$,

$$x \leq x, \qquad (x \leq y \& y \leq x) \Rightarrow x = y, \qquad (x \leq y \& y \leq z) \Rightarrow x \leq z.$$

1.3.1. Redes

En algunos problemas es necesario indexar las variables por conjuntos más generales que $\mathbb{N}$, como $\mathbb{R}$, $\mathbb{Z}^d$ o familias ordenadas por inclusión. Para tratar todos estos casos con una única notación, introducimos los conjuntos dirigidos y las redes. Las pruebas siguen, en gran medida, el mismo patrón que para sucesiones, pero las redes son necesarias para formular correctamente los argumentos de compacidad en espacios no secuencialmente compactos.

Definición 1.30. Un conjunto parcialmente ordenado $(A, \leq)$ es **dirigido** si para cualesquiera $x, y \in A$ existe $z \in A$ tal que $x \leq z$ y $y \leq z$ (*sucesor común*).

Definición 1.31. Sea $\{(\Omega_\alpha, \mathcal{A}_\alpha, \mathbb{P}_\alpha)\}_{\alpha\in A}$ una familia de espacios de probabilidad indexada por un conjunto dirigido $(A, \leq)$, y sea D un espacio métrico. Una **red** en D es una familia $\{X_\alpha : \Omega_\alpha \to D\}_{\alpha\in A}$ (no exigimos medibilidad). En general denotaremos simplemente X_α sin hacer explícita la mención de A. Análogamente, una red de probabilidades de Borel *sobre* D es una familia $\{\mathsf{L}_\alpha\}_{\alpha\in A}$ de probabilidades de Borel en $(D, \mathcal{B}(D))$: los espacios de probabilidad de partida pueden variar con α, pero todas las aplicaciones toman valores en un mismo espacio métrico D, y es sobre ese espacio común donde se formula la convergencia débil.

Definición 1.32. Dadas una red X_α y una probabilidad de Borel L en D, decimos que X_α **converge débilmente** a L, y denotamos $X_\alpha \Rightarrow \mathsf{L}$, si

$$\mathbb{E}_\alpha^*\big(f(X_\alpha)\big) \longrightarrow \int f d\mathsf{L} \qquad \forall f \in C_b(D),$$

donde $\mathbb{E}_\alpha^*$ denota la esperanza exterior respecto de $\mathbb{P}_\alpha$.

Definición 1.33. Sea (D, d) un espacio métrico, $\{\mathsf{L}_\alpha\}_{\alpha\in A}$ una red de probabilidades de Borel en D, y L una probabilidad de Borel en D. Diremos que L_α **converge débilmente a** L, y denotamos $\mathsf{L}_\alpha \Rightarrow \mathsf{L}$ si

$$\int f d\mathsf{L}_\alpha \longrightarrow \int f d\mathsf{L} \qquad \forall f \in C_b(D).$$

Cuando el límite se describe mediante una aplicación Borel-medible $X : \Omega \to D$, escribimos $X_\alpha \Rightarrow X$ como abreviatura de $X_\alpha \Rightarrow \mathsf{L}_X$, donde L_X es la ley de X (Definición 1.22); en particular, $X_\alpha \Rightarrow c$, con $c \in D$, significa $X_\alpha \Rightarrow \delta_c$.

Teorema 1.34 (Portmanteau). *Sea (D, d) métrico y $X_\alpha : \Omega_\alpha \to D$ una red, no necesariamente medible. Fijada una probabilidad de Borel L en D, son equivalentes:*

[18] Por continuidad desde abajo de L a lo largo de la unión creciente $\bigcup_{m\leq N} B(x_m, 1/j) \uparrow \bigcup_{m\geq 1} B(x_m, 1/j)$, cuya probabilidad es 1.

[19] esto se debe a que en un espacio métrico, compacto es equivalente a completo más totalmente acotado

(i) $X_\alpha \Rightarrow \mathsf{L}$.

(ii) $\liminf_\alpha \mathbb{P}_{\alpha,*}(X_\alpha \in G) \geq \mathsf{L}(G)$, *para todo abierto* $G \subset D$.

(iii) $\limsup_\alpha \mathbb{P}^*_\alpha(X_\alpha \in F) \leq \mathsf{L}(F)$, *para todo cerrado* $F \subset D$.

(iv) $\liminf_\alpha \mathbb{E}_{\alpha,*} f(X_\alpha) \geq \int f d\mathsf{L}$, *para toda* $f : D \to \mathbb{R}$ *l.s.c.* [20] *y acotada por abajo.*

(v) $\limsup_\alpha \mathbb{E}^*_\alpha f(X_\alpha) \leq \int f d\mathsf{L}$, *para toda* $f : D \to \mathbb{R}$ *u.s.c. y acotada por arriba.*

(vi) Para todo boreliano B *con* $\mathsf{L}(\partial B) = 0$, $\lim_\alpha \mathbb{P}_{\alpha,*}(X_\alpha \in B) = \lim_\alpha \mathbb{P}^*_\alpha(X_\alpha \in B) = \mathsf{L}(B)$.

(vii) Si $f \geq 0$ *es acotada y Lipschitz, entonces* $\liminf_\alpha \mathbb{E}_{\alpha,*} f(X_\alpha) \geq \int f d\mathsf{L}$.

Demostración. Probamos las equivalencias $(ii) \Leftrightarrow (iii)$ y $(iv) \Leftrightarrow (v)$, que son cuestión de complementos y de signos; luego el ciclo $(i) \Rightarrow (vii) \Rightarrow (ii) \Rightarrow (iv) \Rightarrow (i)$, que es el contenido real del teorema; y por último $(ii), (iii) \Rightarrow (vi) \Rightarrow (iii)$, que incorpora (vi).

$(ii) \Leftrightarrow (iii)$**.** Si F es cerrado y $G = F^c$, entonces $\{X_\alpha \in F\} = \{X_\alpha \in G\}^c$ y, por definición de probabilidad interior, $\mathbb{P}^*_\alpha(X_\alpha \in F) = 1 - \mathbb{P}_{\alpha,*}(X_\alpha \in G)$; también $\mathsf{L}(F) = 1 - \mathsf{L}(G)$. Como $\limsup_\alpha(1 - z_\alpha) = 1 - \liminf_\alpha z_\alpha$, la desigualdad de (iii) para F es exactamente la de (ii) para G, y G recorre los abiertos cuando F recorre los cerrados.

$(iv) \Leftrightarrow (v)$**.** Una función f es u.s.c. y acotada por arriba si y sólo si $-f$ es l.s.c. y acotada por abajo, y valen $\mathbb{E}_{\alpha,*}(-Y) = -\mathbb{E}^*_\alpha(Y)$ y $\int(-f)\, d\mathsf{L} = -\int f\, d\mathsf{L}$. Cambiando f por $-f$, la desigualdad de (iv) se convierte en la de (v) y recíprocamente.

$(i) \Rightarrow (vii)$**.** Si $f \geq 0$ es acotada y Lipschitz, entonces $-f \in C_b(D)$ y (i) da $\mathbb{E}^*_\alpha(-f(X_\alpha)) \to -\int f\, d\mathsf{L}$. Como $\mathbb{E}_{\alpha,*} f(X_\alpha) = -\mathbb{E}^*_\alpha(-f(X_\alpha))$, resulta $\mathbb{E}_{\alpha,*} f(X_\alpha) \to \int f\, d\mathsf{L}$, y en particular vale (vii).

$(vii) \Rightarrow (ii)$**.** Sea G abierto. Si $G = D$, (ii) es trivial. Si no, G^c es cerrado y no vacío, y para $m \geq 1$ ponemos $f_m(x) := \big(m\, d(x, G^c)\big) \wedge 1$. Como $d(\cdot, G^c)$ es 1-Lipschitz, f_m es Lipschitz y toma valores en $[0,1]$. Además $0 \leq f_m \leq \mathbb{1}_G$ y $f_m \uparrow \mathbb{1}_G$: si $x \notin G$, entonces $d(x, G^c) = 0$ y $f_m(x) = 0$; si $x \in G$, al ser G^c cerrado, $d(x, G^c) > 0$, de modo que $f_m(x)$ crece con m y vale 1 en cuanto $m\, d(x, G^c) \geq 1$. Como $f_m(X_\alpha) \leq \mathbb{1}_G(X_\alpha) = \mathbb{1}_{\{X_\alpha \in G\}}$ punto a punto, la monotonía de $\mathbb{E}_{\alpha,*}$ [21] y (vii) dan, para cada m,

$$\liminf_\alpha \mathbb{P}_{\alpha,*}(X_\alpha \in G) = \liminf_\alpha \mathbb{E}_{\alpha,*} \mathbb{1}_{\{X_\alpha \in G\}} \geq \liminf_\alpha \mathbb{E}_{\alpha,*} f_m(X_\alpha) \geq \int f_m\, d\mathsf{L}.$$

Haciendo $m \to \infty$, el teorema de convergencia monótona da $\int f_m\, d\mathsf{L} \uparrow \int \mathbb{1}_G\, d\mathsf{L} = \mathsf{L}(G)$, que es (ii).

$(ii) \Rightarrow (iv)$**.** Sea $f : D \to \mathbb{R}$ l.s.c. y acotada por abajo. Reducimos primero al caso $f \geq 0$. Si $f \geq -c$ con $c \geq 0$, la función $g := f + c$ es l.s.c. y no negativa, y $\mathbb{E}_{\alpha,*} g(X_\alpha) = \mathbb{E}_{\alpha,*} f(X_\alpha) + c$, pues $Y \mapsto Y + c$ es una biyección entre los minorantes medibles de $f(X_\alpha)$ y los de $g(X_\alpha)$ que traslada la esperanza en c. Como también $\int g\, d\mathsf{L} = \int f\, d\mathsf{L} + c$, la desigualdad de (iv) para g equivale a la de f. Suponemos entonces $f \geq 0$, con lo cual $\int f\, d\mathsf{L} \in [0, +\infty]$.

Aproximamos f por funciones simples cuyos escalones son abiertos. Para $m \geq 1$ y $1 \leq i \leq m^2$, el conjunto $G_{i,m} := \{f > i/m\}$ es abierto por ser f l.s.c., y definimos

$$f_m := \frac{1}{m} \sum_{i=1}^{m^2} \mathbb{1}_{G_{i,m}}.$$

[20] Una función $f : D \to \mathbb{R}$ es *semicontinua inferiormente* (l.s.c.) si $\{f > a\}$ es abierto para todo $a \in \mathbb{R}$, y *semicontinua superiormente* (u.s.c.) si $\{f < a\}$ es abierto para todo $a \in \mathbb{R}$; equivalentemente, si $-f$ es l.s.c. En espacios métricos esto equivale a la formulación puntual $\liminf_{y \to x} f(y) \geq f(x)$ para todo x (resp. $\limsup_{y \to x} f(y) \leq f(x)$). Notemos que f es continua si y sólo si es l.s.c. y u.s.c. a la vez, y que toda f l.s.c. o u.s.c. es Borel-medible, pues los conjuntos $\{f > a\}$ (resp. $\{f < a\}$) generan $\mathcal{B}(\mathbb{R})$; en particular $\int f\, d\mathsf{L}$ está bien definida en (iv) y (v), con valor en $(-\infty, +\infty]$ y $[-\infty, +\infty)$ respectivamente, por estar f acotada por abajo o por arriba.

[21] Si $Y \leq Z$ punto a punto, todo minorante medible de Y lo es de Z; luego $\mathbb{E}_{\alpha,*} Y \leq \mathbb{E}_{\alpha,*} Z$.

Fijado x, el número $m\,f_m(x)$ cuenta los índices $i \in \{1,\dots,m^2\}$ con $i < mf(x)$. Si $f(x) \le m$, es la cantidad de enteros positivos menores que $mf(x)$, que está entre $mf(x)-1$ y $mf(x)$; si $f(x) > m$, cuentan todos y $f_m(x) = m$. En ambos casos $0 \le f_m \le f$, y $f(x) - 1/m \le f_m(x)$ en cuanto $m \ge f(x)$, así que $f_m \to f$ puntualmente.

Por la superaditividad de $\mathbb{E}_{\alpha,*}$ (Advertencia 1.11, iterada a m^2 sumandos) y su homogeneidad positiva[22],

$$\mathbb{E}_{\alpha,*} f_m(X_\alpha) \ge \frac{1}{m}\sum_{i=1}^{m^2} \mathbb{E}_{\alpha,*}\mathbb{1}_{G_{i,m}}(X_\alpha) = \frac{1}{m}\sum_{i=1}^{m^2} \mathbb{P}_{\alpha,*}(X_\alpha \in G_{i,m}).$$

Tomando $\liminf_\alpha$, usando que el $\liminf$ de una suma finita es mayor o igual que la suma de los $\liminf$, y aplicando (ii) a cada abierto $G_{i,m}$,

$$\liminf_\alpha \mathbb{E}_{\alpha,*} f_m(X_\alpha) \ge \frac{1}{m}\sum_{i=1}^{m^2} \mathsf{L}(G_{i,m}) = \int f_m \, d\mathsf{L}.$$

Como $f_m \le f$, la monotonía de $\mathbb{E}_{\alpha,*}$ da $\liminf_\alpha \mathbb{E}_{\alpha,*} f(X_\alpha) \ge \int f_m\, d\mathsf{L}$ para todo m. Finalmente, como $f_m \ge 0$ y $f_m \to f$ puntualmente, el lema de Fatou da

$$\int f\, d\mathsf{L} \le \liminf_m \int f_m\, d\mathsf{L} \le \liminf_\alpha \mathbb{E}_{\alpha,*} f(X_\alpha),$$

que es (iv).

$(iv) \Rightarrow (i)$. Por $(iv) \Leftrightarrow (v)$, de (iv) también se dispone de (v). Si $f \in C_b(D)$, entonces f es a la vez l.s.c. y u.s.c., y acotada, así que (iv), la desigualdad $\mathbb{E}_{\alpha,*} \le \mathbb{E}^*_\alpha$ y (v) dan

$$\int f\, d\mathsf{L} \le \liminf_\alpha \mathbb{E}_{\alpha,*} f(X_\alpha) \le \liminf_\alpha \mathbb{E}^*_\alpha f(X_\alpha) \le \limsup_\alpha \mathbb{E}^*_\alpha f(X_\alpha) \le \int f\, d\mathsf{L}.$$

Luego $\mathbb{E}^*_\alpha f(X_\alpha) \to \int f\, d\mathsf{L}$ para toda $f \in C_b(D)$, que es $X_\alpha \Rightarrow \mathsf{L}$.

$(ii), (iii) \Rightarrow (vi)$. Sea B boreliano con $\mathsf{L}(\partial B) = 0$. Como $\overline{B} = B^\circ \cup \partial B$ y $B^\circ \subset B \subset \overline{B}$, se tiene $\mathsf{L}(B^\circ) = \mathsf{L}(B) = \mathsf{L}(\overline{B})$. Aplicando (ii) al abierto B°, la monotonía de las probabilidades interior y exterior, la desigualdad $\mathbb{P}_{\alpha,*} \le \mathbb{P}^*_\alpha$ y (iii) al cerrado $\overline{B}$,

$$\mathsf{L}(B) = \mathsf{L}(B^\circ) \le \liminf_\alpha \mathbb{P}_{\alpha,*}(X_\alpha \in B) \le \limsup_\alpha \mathbb{P}^*_\alpha(X_\alpha \in B) \le \mathsf{L}(\overline{B}) = \mathsf{L}(B),$$

de modo que ambas probabilidades convergen a $\mathsf{L}(B)$.

$(vi) \Rightarrow (iii)$. Sea F cerrado. Si $F = \varnothing$ la desigualdad es trivial; supongamos $F \ne \varnothing$ y pongamos $F^\delta := \{x : d(x,F) < \delta\}$, que es abierto. Como $d(\cdot, F)$ es continua, $\partial F^\delta \subset \{x : d(x,F) = \delta\}$, y estos últimos conjuntos son disjuntos dos a dos para distintos $\delta > 0$; una probabilidad da masa positiva a lo sumo a numerables conjuntos disjuntos,[23] así que podemos elegir $\delta_n \downarrow 0$ con $\mathsf{L}(\partial F^{\delta_n}) = 0$. Por la monotonía de $\mathbb{P}^*_\alpha$ (pues $F \subset F^{\delta_n}$) y por (vi) aplicado a F^{δ_n},

$$\limsup_\alpha \mathbb{P}^*_\alpha(X_\alpha \in F) \le \lim_\alpha \mathbb{P}^*_\alpha(X_\alpha \in F^{\delta_n}) = \mathsf{L}(F^{\delta_n}) \qquad \text{para todo } n.$$

Como F es cerrado, $\bigcap_n F^{\delta_n} = \{x : d(x,F) = 0\} = F$, y la continuidad de L desde arriba da $\mathsf{L}(F^{\delta_n}) \downarrow \mathsf{L}(F)$, que es (iii).

Con esto, (i), (vii), (ii) y (iv) forman un ciclo de implicaciones; (iii) y (v) se les unen por las dos equivalencias iniciales, y (vi) por las dos últimas flechas. Las siete condiciones son equivalentes. □

Sean (D,d) y (E,e) espacios métricos. Para $g : D \to E$ definimos el conjunto de puntos de discontinuidad

$$D_g := \Big\{x \in D : g \text{ no es continua en } x\Big\}.$$

[22] Para $c > 0$, $Y \mapsto cY$ es una biyección entre los minorantes medibles de Z y los de cZ, y $\mathbb{E}(cY) = c\,\mathbb{E}Y$; luego $\mathbb{E}_{\alpha,*}(cZ) = c\,\mathbb{E}_{\alpha,*}Z$.

[23] Para cada k, a lo sumo k de ellos tienen masa mayor que $1/k$.

Lema 1.35. *Sea $g : D \to E$. Entonces D_g es boreliano. Más precisamente,*

$$D_g = \bigcup_{m\geq 1} \bigcap_{k\geq 1} G_k^{(m)},$$

donde $G_k^{(m)} := \Big\{x : \exists y, z \in B(x, 1/k),\ e(g(y), g(z)) > 1/m\Big\}$, y cada $G_k^{(m)}$ es abierto.

Demostración. Si $x \in G_k^{(m)}$, tome $y, z \in B(x, 1/k)$ con $e(g(y), g(z)) > 1/m$. Para

$$\rho := \min\{1/k - d(x, y), 1/k - d(x, z)\} > 0,$$

todo x' con $d(x', x) < \rho$ satisface $y, z \in B(x', 1/k)$, luego $x' \in G_k^{(m)}$. Así $G_k^{(m)}$ es abierto. La identidad para D_g es la caracterización usual de continuidad por oscilación local: g es continua en x si y sólo si, para todo m, existe k tal que $e(g(y), g(z)) \leq 1/m$ para todos $y, z \in B(x, 1/k)$. □

Teorema 1.36 (Aplicación continua). *Sean D, E espacios métricos, $D_0 \subset D$ y $g : D \to E$ continua en cada punto de D_0. Sea $X_\alpha : \Omega_\alpha \to D$ una red no necesariamente medible y sea $X : \Omega \to D$ Borel-medible que toma valores en D_0. Si $X_\alpha \Rightarrow X$, entonces $g(X)$ es Borel-medible y $g(X_\alpha) \Rightarrow g(X)$.*

Demostración. Vemos primero que $g(X)$ es Borel-medible. Sea $V \subset E$ abierto. La continuidad de g en cada punto de D_0 dice que $g|_{D_0}$ es continua para la topología inducida en D_0, así que $g^{-1}(V) \cap D_0 = G \cap D_0$ para algún abierto $G \subset D$. Como X toma valores en D_0,

$$\{g(X) \in V\} = \{X \in g^{-1}(V) \cap D_0\} = \{X \in G\},$$

que es medible porque X es Borel-medible. Los abiertos generan $\mathcal{B}(E)$, y esto prueba la medibilidad.

Sea ahora $F \subset E$ cerrado. Lo esencial es que la clausura de $g^{-1}(F)$ no incorpora puntos de D_0:

$$\overline{g^{-1}(F)} \cap D_0 \subset g^{-1}(F).$$

En efecto, si $x \in D_0$ es límite de una sucesión $x_n \in g^{-1}(F)$, la continuidad de g en x da $g(x_n) \to g(x)$; como $g(x_n) \in F$ y F es cerrado, $g(x) \in F$. Como X toma valores en D_0, esta inclusión de conjuntos se traduce en la inclusión de eventos

$$\{X \in \overline{g^{-1}(F)}\} \subset \{X \in g^{-1}(F)\} = \{g(X) \in F\},$$

ambos medibles: el primero porque $\overline{g^{-1}(F)}$ es cerrado y el segundo por lo ya probado.

Por otro lado, para cada α vale trivialmente $\{g(X_\alpha) \in F\} = \{X_\alpha \in g^{-1}(F)\} \subset \{X_\alpha \in \overline{g^{-1}(F)}\}$. Juntando esto con la monotonía de la probabilidad exterior, el teorema de Portmanteau aplicado a $X_\alpha \Rightarrow X$ y al cerrado $\overline{g^{-1}(F)}$, y la inclusión de eventos recién obtenida,

$$\limsup_\alpha \mathbb{P}_\alpha^*\Big(g(X_\alpha) \in F\Big) \leq \limsup_\alpha \mathbb{P}_\alpha^*\Big(X_\alpha \in \overline{g^{-1}(F)}\Big) \leq \mathbb{P}\Big(X \in \overline{g^{-1}(F)}\Big) \leq \mathbb{P}\Big(g(X) \in F\Big).$$

Como F es un cerrado arbitrario de E y $g(X)$ es Borel-medible, el teorema de Portmanteau da $g(X_\alpha) \Rightarrow g(X)$. □

Corolario 1.37. *Sean D, E espacios métricos, $g : D \to E$ arbitraria y D_g su conjunto de discontinuidades. Sea $X : \Omega \to D$ Borel-medible y supongamos que $\mathbb{P}(X \in D_g) = 0$. Si $X_\alpha \Rightarrow X$, entonces la aplicación $\omega \mapsto g(X(\omega))$ admite una versión Borel-medible $Y : \Omega \to E$, es decir, $Y = g \circ X$ $\mathbb{P}$-c.s., y además $g(X_\alpha) \Rightarrow Y$.*

Demostración. Sea $D_0 := D \setminus D_g$, que es boreliano por el Lema 1.35. Como $\mathbb{P}(X \in D) = 1$ y $\mathbb{P}(X \in D_g) = 0$, necesariamente $D_0 \neq \varnothing$; fijemos $x_0 \in D_0$ y definamos

$$\widetilde{X}(\omega) := \begin{cases} X(\omega), & X(\omega) \in D_0, \\ x_0, & X(\omega) \notin D_0. \end{cases}$$

Veamos que $\widetilde{X}$ es Borel-medible. Sea

$$A := \{\omega : X(\omega) \in D_0\} = X^{-1}(D_0).$$

Como $D_0 \in \mathcal{B}(D)$ y X es Borel-medible, $A \in \mathcal{A}$. Ahora, para todo $B \in \mathcal{B}(D)$,

$$\widetilde{X}^{-1}(B) = \begin{cases} A \cap X^{-1}(B), & x_0 \notin B, \\ \Big(A \cap X^{-1}(B)\Big) \cup A^c, & x_0 \in B. \end{cases}$$

En ambos casos $\widetilde{X}^{-1}(B) \in \mathcal{A}$, pues $A \in \mathcal{A}$ y $X^{-1}(B) \in \mathcal{A}$. Por lo tanto, $\widetilde{X} : \Omega \to D$ es Borel-medible.

Además, $\widetilde{X}$ toma valores en D_0 y coincide con X salvo en $\{X \notin D_0\} = \{X \in D_g\}$, que tiene probabilidad cero. En particular, $\widetilde{X}$ y X tienen la misma ley, de modo que $X_\alpha \Rightarrow \widetilde{X}$. Como g es continua en todo punto de D_0 y $\widetilde{X}(\Omega) \subset D_0$, por el Teorema 1.36, $g(X_\alpha) \Rightarrow g(\widetilde{X})$. Finalmente, $g(\widetilde{X})$ es Borel-medible y coincide con $g(X)$ en $\{X \in D_0\}$, evento de probabilidad uno. Por tanto, si definimos $Y := g(\widetilde{X})$, entonces Y es una versión Borel-medible de $g(X)$ y $g(X_\alpha) \Rightarrow Y$. □

Definición 1.38. Sea (D, d) un espacio métrico y $\{X_\alpha\}_{\alpha \in A}$ una red de aplicaciones (no necesariamente medibles) con valores en D. Decimos que $\{X_\alpha\}$ es **asintóticamente medible** si

$$\mathbb{E}^*_\alpha\Big(f(X_\alpha)\Big) - \mathbb{E}_{\alpha,*}\Big(f(X_\alpha)\Big) \longrightarrow 0 \qquad \forall f \in C_b(D),$$

donde $\mathbb{E}^*_\alpha$ y $\mathbb{E}_{\alpha,*}$ son la esperanza exterior e interior respectivamente.

Definición 1.39. Para $K \subset D$ y $\delta > 0$, escribimos $K^\delta := \{x \in D : d(x, K) < \delta\}$. Diremos que $\{X_\alpha\}$ es **asintóticamente tensa** si para todo $\varepsilon > 0$ existe un compacto $K \subset D$ tal que

$$\liminf_\alpha \mathbb{P}_{\alpha,*}(X_\alpha \in K^\delta) \geq 1 - \varepsilon, \qquad \forall \delta > 0.$$

Equivalentemente, $\limsup_\alpha \mathbb{P}^*_\alpha(X_\alpha \notin K^\delta) \leq \varepsilon$, $\forall \delta > 0$.

Lema 1.40. *Si $X_\alpha \Rightarrow X$ en D, con X Borel-medible, entonces:*

(a) $\{X_\alpha\}$ es asintóticamente medible.

(b) $\{X_\alpha\}$ es asintóticamente tensa si y sólo si X es tenso.

Demostración. Para (a), sea $f \in C_b(D)$. Como también $-f \in C_b(D)$, la convergencia débil aplicada a f da $\mathbb{E}^*_\alpha f(X_\alpha) \to \mathbb{E} f(X)$, y aplicada a $-f$ da

$$\mathbb{E}_{\alpha,*} f(X_\alpha) = -\mathbb{E}^*_\alpha\Big(-f(X_\alpha)\Big) \to -\mathbb{E}\Big(-f(X)\Big) = \mathbb{E} f(X).$$

Las esperanzas exterior e interior de $f(X_\alpha)$ convergen al mismo límite, así que su diferencia tiende a cero: $\{X_\alpha\}$ es asintóticamente medible.

Para (b) fijamos $\varepsilon \in (0, 1)$; para $\varepsilon \geq 1$ las dos condiciones son triviales. Los compactos K que aparecen abajo tienen probabilidad (interior o exterior) al menos $1 - \varepsilon > 0$, luego son no vacíos y $x \mapsto d(x, K)$ es continua. Usaremos que K^δ es abierto y que su clausura $\overline{K^\delta}$ está contenida en el cerrado $\{x : d(x, K) \leq \delta\}$.

Supongamos primero que $\{X_\alpha\}$ es asintóticamente tensa y tomemos un compacto $K \subset D$ con $\liminf_\alpha \mathbb{P}_{\alpha,*}(X_\alpha \in K^\delta) \geq 1 - \varepsilon$ para todo $\delta > 0$. Como $K^\delta \subset \overline{K^\delta}$ y $\mathbb{P}_{\alpha,*} \leq \mathbb{P}^*_\alpha$, el Teorema 1.34(iii) aplicado al cerrado $\overline{K^\delta}$ da, para todo $\delta > 0$,

$$1 - \varepsilon \leq \liminf_\alpha \mathbb{P}_{\alpha,*}(X_\alpha \in K^\delta) \leq \limsup_\alpha \mathbb{P}^*_\alpha\Big(X_\alpha \in \overline{K^\delta}\Big) \leq \mathbb{P}\Big(X \in \overline{K^\delta}\Big).$$

Los cerrados $\overline{K^{1/k}}$ decrecen con k, contienen a K y están contenidos en $\{x : d(x, K) \leq 1/k\}$; como K es cerrado, $\bigcap_k \overline{K^{1/k}} = \{x : d(x, K) = 0\} = K$. Por continuidad de la probabilidad desde arriba, $\mathbb{P}(X \in K) = \lim_k \mathbb{P}(X \in \overline{K^{1/k}}) \geq 1 - \varepsilon$ y por lo tanto X es tenso.

Recíprocamente, si X es tenso, tomemos un compacto $K \subset D$ con $\mathbb{P}(X \in K) \geq 1 - \varepsilon$. Para cada $\delta > 0$ el abierto K^δ contiene a K, y el Teorema 1.34(ii) da

$$\liminf_{\alpha} \mathbb{P}_{\alpha,*}(X_\alpha \in K^\delta) \geq \mathbb{P}(X \in K^\delta) \geq \mathbb{P}(X \in K) \geq 1 - \varepsilon.$$

El compacto K no depende de δ, de modo que $\{X_\alpha\}$ es asintóticamente tensa. □

El teorema de Prokhorov usa el siguiente resultado, que generaliza la representación de Riesz–Markov y que tomaremos como conocido.[24]

Teorema 1.41 (Daniell–Stone)**.** *Sea (D, d) un espacio métrico y sea $I : C_b(D) \to \mathbb{R}$ un funcional tal que:*

(i) I es lineal;

(ii) $I(f) \geq 0$ si $f \geq 0$;

(iii) $I(1) = 1$;

(iv) si $f_n \in C_b(D)$ y $f_n \downarrow 0$ puntualmente, entonces $I(f_n) \downarrow 0$.

Entonces existe una única probabilidad de Borel L sobre D tal que

$$I(f) = \int_D f d\mathsf{L}, \qquad f \in C_b(D).$$

Definición 1.42. Sea $(A, \leq)$ un conjunto dirigido y $\{X_\alpha\}_{\alpha \in A}$ una red en D. Una **subred** de $\{X_\alpha\}$ es una familia $\{X_{\alpha(\beta)}\}_{\beta \in B}$, donde $(B, \preceq)$ es un conjunto dirigido y $\alpha(\cdot) : B \to A$ satisface:

(a) *monotonía*: para todo $\beta \preceq \beta'$ se tiene que $\alpha(\beta) \leq \alpha(\beta')$;

(b) *cofinalidad*: para todo $\alpha_0 \in A$ existe $\beta_0 \in B$ con $\alpha(\beta_0) \geq \alpha_0$.

Los espacios de probabilidad subyacentes son los correspondientes $(\Omega_{\alpha(\beta)}, \mathcal{A}_{\alpha(\beta)}, \mathbb{P}_{\alpha(\beta)})$, y escribimos $\mathbb{E}^*_{\alpha(\beta)}$, $\mathbb{P}_{\alpha(\beta),*}$, etc.

Advertencia 1.43

En la Definición 1.42 no se pide que $B \subset A$, ni que $\alpha(\cdot)$ sea inyectiva, ni que B sea numerable. De (a) y (b) se obtiene la propiedad que efectivamente se usa en las pruebas, y que llamaremos *finalidad*: para todo $\alpha_0 \in A$ existe $\beta_0 \in B$ tal que $\beta \succeq \beta_0$ implica $\alpha(\beta) \geq \alpha_0$, pues $\alpha(\beta) \geq \alpha(\beta_0) \geq \alpha_0$. Toda sucesión $\{X_n\}_{n \in \mathbb{N}}$ es una red (con $A = \mathbb{N}$), y toda subsucesión es una subred: si $n_1 < n_2 < \cdots$, basta tomar $B = \mathbb{N}$ y $\alpha(j) = n_j$. El recíproco es **falso**: una subred de una sucesión no es, en general, una subsucesión, ni admite extraer una. Por ejemplo, en $D = \{0,1\}^{[0,1]}$ con la topología producto (compacto por Tychonoff), sea $f_n(x)$ el n-ésimo dígito binario de $x \in [0,1]$. Ninguna subsucesión f_{n_j} converge puntualmente: dados $n_1 < n_2 < \cdots$, se elige x con dígitos 0 en las posiciones $n_1, n_3, \ldots$ y 1 en $n_2, n_4, \ldots$, de modo que $f_{n_j}(x)$ oscila. Sin embargo, por compacidad, $\{f_n\}$ sí tiene subredes convergentes.

Teorema 1.44 (Prokhorov). *Sea (D, d) métrico y $X_\alpha : \Omega_\alpha \to D$ una red no necesariamente medible.*

(i) Si $\{X_\alpha\}$ es asintóticamente tensa y asintóticamente medible, entonces existe una subred $X_{\alpha(\beta)}$ y una probabilidad de Borel tensa L en D tales que $X_{\alpha(\beta)} \Rightarrow \mathsf{L}$.

(ii) Si $\{X_n\}$ es una sucesión asintóticamente tensa y asintóticamente medible, entonces existe una subsucesión $X_{n_j} \Rightarrow \mathsf{L}$, para alguna probabilidad de Borel tensa L.

[24] Para el teorema de Daniell–Stone, véase R. M. Dudley, *Real Analysis and Probability*, Cambridge University Press, 2002, §4.5.

Demostración. Para $f \in C_b(D)$, pongamos $I_f := [-\|f\|_\infty, \|f\|_\infty]$. Definimos

$$Y_\alpha := \left(\mathbb{E}^*_\alpha f(X_\alpha)\right)_{f \in C_b(D)} \in \prod_{f \in C_b(D)} I_f.$$

Por Tychonoff,[25] existe una subred $\alpha(\beta)$ tal que, para todo $f \in C_b(D)$, $\mathbb{E}^*_{\alpha(\beta)} f(X_{\alpha(\beta)}) \longrightarrow L(f)$. Como la red es asintóticamente medible,[26] también $\mathbb{E}_{\alpha(\beta),*} f(X_{\alpha(\beta)}) \longrightarrow L(f)$, $f \in C_b(D)$. Probemos que L es lineal y positivo. Para $f, g \in C_b(D)$, usando subaditividad de $\mathbb{E}^*$ y superaditividad de $\mathbb{E}_*$,

$$\begin{aligned} \mathbb{E}_* f(X_{\alpha(\beta)}) + \mathbb{E}_* g(X_{\alpha(\beta)}) &\le \mathbb{E}_*(f+g)(X_{\alpha(\beta)}) \\ &\le \mathbb{E}^*(f+g)(X_{\alpha(\beta)}) \\ &\le \mathbb{E}^* f(X_{\alpha(\beta)}) + \mathbb{E}^* g(X_{\alpha(\beta)}), \end{aligned}$$

donde las esperanzas son respecto de $P_{\alpha(\beta)}$. Pasando al límite,

$$L(f) + L(g) \le L(f+g) \le L(f) + L(g),$$

luego $L(f+g) = L(f) + L(g)$. La homogeneidad para $a \ge 0$ se obtiene pasando al límite en

$$\mathbb{E}^*_{\alpha(\beta)}\left[a f(X_{\alpha(\beta)})\right] = a\, \mathbb{E}^*_{\alpha(\beta)} f(X_{\alpha(\beta)}).$$

Por aditividad, $L(-f) = -L(f)$, y entonces $L(af) = aL(f)$ para todo $a \in \mathbb{R}$. Si $f \ge 0$, entonces $\mathbb{E}_* f(X_{\alpha(\beta)}) \ge 0$, luego $L(f) \ge 0$. Además $L(1) = 1$.

Resta verificar la continuidad monótona. Sea $f_m \in C_b(D)$ con $f_m \downarrow 0$; en particular $0 \le f_m \le f_1$. Fijado $\varepsilon \in (0,1)$, por tensión asintótica de la subred[27] existe compacto $K \subset D$ tal que, para todo $\delta > 0$,

$$\liminf_\beta P_{\alpha(\beta),*}(X_{\alpha(\beta)} \in K^\delta) \ge 1 - \varepsilon.$$

Por el Teorema de Dini (ver Apéndice), $f_m \downarrow 0$ uniformemente en K; tome m tal que $\sup_K f_m \le \varepsilon$. Por continuidad y compacidad, existe $\delta > 0$ tal que $\sup_{K^\delta} f_m \le 2\varepsilon$.[28] Con $B_\beta = \{X_{\alpha(\beta)} \in K^\delta\}$ y $M = \|f_1\|_\infty$, $f_m(X_{\alpha(\beta)}) \le 2\varepsilon + M \mathbb{1}_{B_\beta^c}$,[29] y por tanto $\mathbb{E}^*_{\alpha(\beta)} f_m(X_{\alpha(\beta)}) \le 2\varepsilon + M P^*_{\alpha(\beta)}(B_\beta^c)$. Como $\limsup_\beta P^*_{\alpha(\beta)}(B_\beta^c) \le \varepsilon$, se obtiene $0 \le L(f_m) \le (2+M)\varepsilon$ para algún m. Como $f_m \downarrow 0$ y L es positivo, la sucesión $L(f_m)$ es decreciente y no negativa. Por tanto $L(f_m) \downarrow \ell$ para algún $\ell \ge 0$. Para cada $\varepsilon \in (0,1)$, el índice $m = m(\varepsilon)$ elegido arriba satisface

$$0 \le \ell \le L(f_{m(\varepsilon)}) \le (2+M)\varepsilon.$$

Como $M = \|f_1\|_\infty$ no depende de ε, su arbitrariedad implica $\ell = 0$. Así, $L(f_m) \downarrow 0$.

Por Daniell–Stone, existe una probabilidad de Borel L sobre D tal que

$$L(f) = \int f d\mathsf{L}, \qquad f \in C_b(D).$$

Luego

$$\mathbb{E}^*_{\alpha(\beta)} f(X_{\alpha(\beta)}) \to \int f d\mathsf{L}, \qquad f \in C_b(D),$$

[25] El producto $\prod_f I_f$ de intervalos compactos es compacto por el teorema de Tychonoff; en un espacio topológico compacto, toda red tiene una subred convergente, y la convergencia en la topología producto es exactamente la convergencia coordenada a coordenada, es decir, para cada f por separado.

[26] La diferencia $\mathbb{E}^*_\alpha - \mathbb{E}_{\alpha,*}$ tiende a 0 a lo largo de toda la red y, por lo tanto, a lo largo de cualquier subred.

[27] Las subredes heredan la tensión asintótica: para toda red numérica, $\liminf_\beta z_{\alpha(\beta)} \ge \liminf_\alpha z_\alpha$, de modo que la cota inferior sólo puede mejorar.

[28] Si no, habría puntos x_n con $d(x_n, K) \to 0$ y $f_m(x_n) > 2\varepsilon$; tomando $y_n \in K$ con $d(x_n, y_n) \to 0$ y una subsucesión $y_n \to y \in K$ (compacidad), resulta $x_n \to y$ y, por continuidad de f_m, $f_m(y) \ge 2\varepsilon > \varepsilon \ge \sup_K f_m$, contradicción.

[29] La cota puntual: en B_β, $X_{\alpha(\beta)} \in K^\delta$ y allí $f_m \le 2\varepsilon$; fuera, $f_m \le f_1 \le M$. Luego se usan la monotonía y la subaditividad de $\mathbb{E}^*$, que las constantes salen de la esperanza exterior, y $\mathbb{E}^*_{\alpha(\beta)} \mathbb{1}_{B_\beta^c} = \mathbb{P}^*_{\alpha(\beta)}(B_\beta^c)$, del Lema 1.8(i).

es decir, $X_{\alpha(\beta)} \Rightarrow \mathsf{L}$. Como la subred sigue siendo asintóticamente tensa, el lema anterior implica que L es tensa.[30] Esto concluye la prueba del punto (i).

Para probar (ii) sea, para cada $m \geq 1$, $K_m \subset D$ compacto tal que, para todo $\delta > 0$,

$$\liminf_n P_{n,*}(X_n \in K_m^\delta) \geq 1 - \frac{1}{2m}.$$

Como K_m es compacto métrico, la bola unidad de $C(K_m)$ es separable.[31] Por Tietze, [32] toda $h \in C(K_m)$, $\|h\|_\infty \leq 1$, se extiende a una función de $C_b(D)$ con la misma cota. Por tanto, para cada m, existe una familia numerable $\mathcal{F}_m \subset C_b(D)$, $\|h\|_\infty \leq 1$, cuyas restricciones a K_m son densas en la bola unidad de $C(K_m)$. Sea $\mathcal{F} := \bigcup_{m\geq 1} \mathcal{F}_m$, que es numerable; enumerémosla $\mathcal{F} = \{h_1, h_2, \dots\}$. Como $|\mathbb{E}_n^* h_k(X_n)| \leq 1$ para todo n, k, la sucesión $(\mathbb{E}_n^* h_1(X_n))_n$ tiene una subsucesión convergente, indexada por $n^{(1)} \subset \mathbb{N}$; de ella extraemos $n^{(2)} \subset n^{(1)}$ a lo largo de la cual también converge $\mathbb{E}^* h_2(X_\cdot)$, y así sucesivamente. La subsucesión diagonal $n_j := n_j^{(j)}$ es, salvo finitos términos, una subsucesión de cada $n^{(k)}$; por lo tanto $\mathbb{E}_{n_j}^* h(X_{n_j})$ converge para todo $h \in \mathcal{F}$.

Mostremos que entonces $\mathbb{E}_{n_j}^* f(X_{n_j})$ converge para todo $f \in C_b(D)$. Basta tomar $\|f\|_\infty \leq 1$.[33] Fijados m y $\eta > 0$, sea $h \in \mathcal{F}_m$ tal que $\sup_{x\in K_m} |f(x) - h(x)| \leq \eta$. Por continuidad y compacidad, existe $\delta > 0$ tal que $\sup_{x\in K_m^\delta} |f(x) - h(x)| \leq 2\eta$.[34] Con $B_n = \{X_n \in K_m^\delta\}$,

$$|\mathbb{E}_n^* f(X_n) - \mathbb{E}_n^* h(X_n)| \leq \mathbb{E}_n^* |f - h|(X_n) \leq 2\eta + 2P_n^*(B_n^c).$$

Luego

$$\limsup_j |\mathbb{E}_{n_j}^* f(X_{n_j}) - \mathbb{E}_{n_j}^* h(X_{n_j})| \leq 2\eta + \frac{1}{m}.$$

Como $\eta > 0$, la cota del límite superior implica que, para j suficientemente grande, la diferencia anterior es a lo sumo $3\eta + 1/m$. Por tanto, si j, k son suficientemente grandes,

$$|\mathbb{E}_{n_j}^* f(X_{n_j}) - \mathbb{E}_{n_k}^* f(X_{n_k})| \leq 6\eta + \frac{2}{m} + |\mathbb{E}_{n_j}^* h(X_{n_j}) - \mathbb{E}_{n_k}^* h(X_{n_k})|.$$

El último término tiende a 0, porque $h \in \mathcal{F}$. Haciendo luego $\eta \downarrow 0$ y $m \to \infty$, se concluye que $\mathbb{E}_{n_j}^* f(X_{n_j})$ es de Cauchy. Definimos

$$L(f) := \lim_j \mathbb{E}_{n_j}^* f(X_{n_j}), \qquad f \in C_b(D).$$

Como la subsucesión sigue siendo asintóticamente medible, $\mathbb{E}_{n_j,*} f(X_{n_j}) \longrightarrow L(f)$, $f \in C_b(D)$. Repitiendo el argumento de la parte (i), L es lineal, positivo, $L(1) = 1$ y satisface la continuidad monótona. Por Daniell–Stone existe una probabilidad de Borel L tal que

$$L(f) = \int f d\mathsf{L}, \qquad f \in C_b(D).$$

Por tanto $X_{n_j} \Rightarrow \mathsf{L}$. Como la subsucesión es asintóticamente tensa, L es tensa. □

Definición 1.45. Sea $\mathcal{F} \subset C_b(D)$.

[30] Lema 1.40(b), aplicado a la subred y a la variable canónica $X = \mathrm{id}$ sobre $(D, \mathcal{B}(D), \mathsf{L})$, cuya ley es L.

[31] Para K métrico compacto, $C(K)$ con la norma uniforme es separable: por Stone–Weierstrass, el álgebra generada con coeficientes racionales por las constantes y las funciones $x \mapsto d(x, q)$, con q en un denso numerable de K, es densa en $C(K)$. Todo subconjunto de un métrico separable es separable; en particular, la bola unidad.

[32] Teorema de extensión de Tietze: si A es cerrado en un espacio métrico D y $h : A \to \mathbb{R}$ es continua y acotada, entonces existe $\tilde{h} \in C_b(D)$ tal que $\tilde{h}|_A = h$ y $\|\tilde{h}\|_{\infty,D} = \|h\|_{\infty,A}$.

[33] Para $f \neq 0$, se aplica lo que sigue a $f/\|f\|_\infty$ y se recupera la convergencia multiplicando por $\|f\|_\infty$, gracias a la homogeneidad positiva de $\mathbb{E}^*$; $f = 0$ es trivial.

[34] Mismo argumento de compacidad de la parte (i): puntos x_n con $d(x_n, K_m) \to 0$ y $|f - h|(x_n) > 2\eta$ acumularían en un punto de K_m donde $|f - h| \geq 2\eta > \eta$. La cadena que sigue usa además $\|f - h\|_\infty \leq 2$.

- $\mathcal{F}$ es un **retículo vectorial** si es un subespacio vectorial cerrado bajo partes positivas: si $f \in \mathcal{F}$, entonces $f^+ = f \vee 0 \in \mathcal{F}$ (en particular, $f \vee g = f + (g-f)^+$, $f \wedge g = g - (g-f)^+ \in \mathcal{F}$ para $f, g \in \mathcal{F}$).
- Decimos que $\mathcal{F}$ **separa puntos** de D si para todo par $x \neq y$ en D existe $f \in \mathcal{F}$ tal que $f(x) \neq f(y)$.
- Una **subálgebra** $\mathcal{F} \subset C_b(D)$ es un subespacio vectorial cerrado bajo productos puntuales: si $f, g \in \mathcal{F}$, entonces $fg \in \mathcal{F}$.

El siguiente teorema es clave para probar los lemas que siguen

Teorema 1.46 (Stone–Weierstrass)**.** *Sea (K, d) un espacio métrico compacto y sea $\mathcal{A} \subset C(K)$ una clase de funciones que contiene las funciones constantes y separa puntos de K. Entonces $\mathcal{A}$ es densa en $C(K)$ con la norma uniforme en cualquiera de los dos casos siguientes:*

(i) $\mathcal{A}$ es un retículo vectorial.

(ii) $\mathcal{A}$ es una subálgebra.

Lema 1.47. *Sean L_1 y L_2 medidas de Borel finitas en D.*

(i) Si $\int f \, d\mathsf{L}_1 = \int f \, d\mathsf{L}_2$ para toda $f \in C_b(D)$, entonces $\mathsf{L}_1 = \mathsf{L}_2$.

(ii) Si L_1 y L_2 son probabilidades de Borel tensas en D y

$$\int f \, d\mathsf{L}_1 = \int f \, d\mathsf{L}_2 \qquad \text{para toda } f \in \mathcal{F},$$

donde $\mathcal{F} \subset C_b(D)$ es un retículo vectorial que contiene las funciones constantes y separa puntos de D, entonces $\mathsf{L}_1 = \mathsf{L}_2$.

Demostración. (i) Para todo cerrado no vacío $F \subset D$, las funciones $f_m(x) := (1 - m\, d(x, F))^+$ pertenecen a $C_b(D)$ y satisfacen $0 \leq f_m \leq 1$ y $f_m \downarrow \mathbb{1}_F$. Por convergencia dominada y la hipótesis,

$$\mathsf{L}_1(F) = \lim_m \int f_m \, d\mathsf{L}_1 = \lim_m \int f_m \, d\mathsf{L}_2 = \mathsf{L}_2(F).$$

Como los cerrados forman un π-sistema que contiene a D y genera $\mathcal{B}(D)$, el teorema de Dynkin (ver apéndice) implica $\mathsf{L}_1 = \mathsf{L}_2$[35].

(ii) Por linealidad y (i), basta probar la igualdad de integrales para toda $g \in C_b(D)$ con $0 \leq g \leq 1$.[36] Fijemos una de estas funciones y $\varepsilon \in (0, 1)$. Por tensión, existe un compacto $K \subset D$ tal que

$$\mathsf{L}_1(K^c) + \mathsf{L}_2(K^c) < \varepsilon.$$

Las restricciones de $\mathcal{F}$ a K forman un retículo vectorial que contiene las constantes y separa puntos. Por Stone–Weierstrass, existe $f_0 \in \mathcal{F}$ tal que $\sup_K |f_0 - g| < \varepsilon$. Definiendo $f := (f_0 \vee 0) \wedge 1$, tenemos $f \in \mathcal{F}$, $0 \leq f \leq 1$ y $\sup_K |f - g| < \varepsilon$, pues truncar en $[0, 1]$ no aumenta la distancia a g.

[35] Una familia no vacía $\mathcal{P}$ de conjuntos es un π-sistema si $A \cap B \in \mathcal{P}$ para cualesquiera $A, B \in \mathcal{P}$. La consecuencia del teorema de Dynkin que usamos establece que, si dos medidas finitas coinciden sobre un π-sistema $\mathcal{P}$ que contiene al espacio total D, entonces coinciden sobre toda la σ-álgebra $\sigma(\mathcal{P})$ generada por esa familia.

[36] Para $h \in C_b(D)$, sean $m := \inf_D h$ y $M := \sup_D h$. Si $M > m$, la función $g := (h - m)/(M - m)$ pertenece a $C_b(D)$, satisface $0 \leq g \leq 1$ y permite escribir $h = m + (M - m)g$. Como las medidas son probabilidades, $\int h \, d\mathsf{L}_i = m + (M - m) \int g \, d\mathsf{L}_i$; por tanto, la igualdad de integrales para g implica la igualdad para h. Si $M = m$, h es constante y la igualdad es inmediata.

Como $\int f\,d\mathsf{L}_1 = \int f\,d\mathsf{L}_2$, tenemos

$$\begin{aligned}\left|\int g\,d\mathsf{L}_1 - \int g\,d\mathsf{L}_2\right| &= \left|\int (g-f)\,d\mathsf{L}_1 - \int (g-f)\,d\mathsf{L}_2\right| \\ &\le \left|\int (g-f)\,d\mathsf{L}_1\right| + \left|\int (g-f)\,d\mathsf{L}_2\right| \\ &\le \int |g-f|\,d\mathsf{L}_1 + \int |g-f|\,d\mathsf{L}_2 \\ &= \int_K |g-f|\,d(\mathsf{L}_1+\mathsf{L}_2) + \int_{K^c} |g-f|\,d(\mathsf{L}_1+\mathsf{L}_2) \\ &\le \varepsilon\Big[\mathsf{L}_1(K)+\mathsf{L}_2(K)\Big] + \mathsf{L}_1(K^c) + \mathsf{L}_2(K^c) \\ &\le 2\varepsilon + \mathsf{L}_1(K^c) + \mathsf{L}_2(K^c) < 3\varepsilon.\end{aligned}$$

Aquí usamos que $|g-f|<\varepsilon$ en K, que $|g-f|\le 1$ en D porque $0\le f,g\le 1$, y que $\mathsf{L}_1(K)+\mathsf{L}_2(K)\le 2$ por ser probabilidades. La última desigualdad se sigue de la elección de K. Haciendo $\varepsilon\downarrow 0$, obtenemos la igualdad buscada. □

Lema 1.48. *Sea $X_\alpha:\Omega_\alpha\to D$ una red con valores en un espacio métrico (D,d), y supongamos que X_α es asintóticamente tensa. Sea $\mathcal{F}\subset C_b(D)$ una subálgebra que separa puntos de D. Si*

$$\mathbb{E}^*_\alpha f(X_\alpha) - \mathbb{E}_{\alpha,*} f(X_\alpha)\to 0 \qquad \text{para toda } f\in\mathcal{F},$$

entonces X_α es asintóticamente medible.

Demostración. Para cada $h\in C_b(D)$, definamos $\Delta_\alpha(h) := \mathbb{E}^*_\alpha h(X_\alpha) - \mathbb{E}_{\alpha,*}h(X_\alpha)$. Esta diferencia está bien definida y es finita: si $M:=\|h\|_\infty$, por monotonía tenemos $-M\le \mathbb{E}_{\alpha,*}h(X_\alpha)\le \mathbb{E}^*_\alpha h(X_\alpha)\le M$.

Sea $\tau(t):=(t\vee 0)\wedge 1$. Primero probaremos que $\Delta_\alpha(\tau\circ h)\le\Delta_\alpha(h)$ para toda $h\in C_b(D)$, sin restringir sus valores a $[0,1]$. Fijemos α y pongamos $Y:=h(X_\alpha)$.

Por las definiciones de esperanza exterior e inferior, para cada $\eta>0$ existen funciones medibles y acotadas U,V tales que $U\le Y\le V$, $\mathbb{E}_\alpha U > \mathbb{E}_{\alpha,*}Y-\eta$ y $\mathbb{E}_\alpha V<\mathbb{E}^*_\alpha Y+\eta$. Podemos tomarlas acotadas porque Y es acotada. Como τ es creciente y continua, $\tau(U)\le\tau(Y)\le\tau(V)$, y los extremos son medibles. Por ello, $\mathbb{E}_{\alpha,*}\tau(Y)\ge\mathbb{E}_\alpha\tau(U)$ y $\mathbb{E}^*_\alpha\tau(Y)\le\mathbb{E}_\alpha\tau(V)$. Además, la monotonía y el carácter 1-Lipschitz de τ dan $0\le\tau(V)-\tau(U)\le V-U$. En consecuencia,

$$\Delta_\alpha(\tau\circ h)\le \mathbb{E}_\alpha\tau(V)-\mathbb{E}_\alpha\tau(U)=\mathbb{E}_\alpha[\tau(V)-\tau(U)]\le\mathbb{E}_\alpha[V-U]=\mathbb{E}_\alpha V-\mathbb{E}_\alpha U<\Delta_\alpha(h)+2\eta.$$

Haciendo $\eta\downarrow 0$, obtenemos la desigualdad buscada.

Podemos reemplazar $\mathcal{F}$ por $\mathcal{F}+\mathbb{R}:=\{f+c: f\in\mathcal{F},\ c\in\mathbb{R}\}$. Esta clase contiene las constantes, sigue separando puntos y es una subálgebra: el producto de f_1+c_1 y f_2+c_2 es

$$(f_1f_2+c_1f_2+c_2f_1)+c_1c_2\in\mathcal{F}+\mathbb{R}.$$

Además, $\Delta_\alpha(f+c)=\Delta_\alpha(f)$, por lo que se conserva la hipótesis del lema. En adelante denotamos de nuevo por $\mathcal{F}$ a esta subálgebra ampliada. Además, por reescalamiento, basta probar que $\Delta_\alpha(g)\to 0$ para $g\in C_b(D)$ con $0\le g\le 1$.

Fijemos una de estas funciones y $\varepsilon\in(0,1)$. Por tensión asintótica, existe un compacto $K\subset D$ tal que $\limsup_\alpha \mathbb{P}^*_\alpha(X_\alpha\notin K^\delta)\le\varepsilon$ para todo $\delta>0$. Por Stone–Weierstrass aplicado a las restricciones a K, existe $f_0\in\mathcal{F}$ con $\sup_K|f_0-g|<\varepsilon$. Definamos $f:=\tau\circ f_0$. Entonces $0\le f\le 1$ y $\sup_K|f-g|<\varepsilon$, pues $\tau(g)=g$ y τ es 1-Lipschitz. Además, $\Delta_\alpha(f)\le\Delta_\alpha(f_0)\to 0$. Por continuidad y compacidad, existe $\delta>0$ tal que $|f-g|<\varepsilon$ en K^δ.

Por la definición de esperanza interior (Definición 1.1), $\Delta_\alpha(h)=\mathbb{E}^*_\alpha h(X_\alpha)+\mathbb{E}^*_\alpha[-h(X_\alpha)]$. Aplicando la subaditividad de la esperanza exterior a las identidades $g=f+(g-f)$ y $-g=-f+(f-g)$, y sumando, obtenemos

$$\Delta_\alpha(g)\le\Delta_\alpha(f)+\mathbb{E}^*_\alpha[(g-f)(X_\alpha)]+\mathbb{E}^*_\alpha[(f-g)(X_\alpha)]\le\Delta_\alpha(f)+2\mathbb{E}^*_\alpha|g-f|(X_\alpha).$$

La última desigualdad usa monotonía, pues tanto $g-f$ como $f-g$ están acotadas superiormente por $|g-f|$.

Pongamos $A_\alpha := \{X_\alpha \notin K^\delta\}$. Como $|g-f|<\varepsilon$ en K^δ y $|g-f|\le 1$ en todo D, punto a punto tenemos $|g-f|(X_\alpha) \le \varepsilon \mathbb{1}_{A_\alpha^c} + \mathbb{1}_{A_\alpha} \le \varepsilon + \mathbb{1}_{A_\alpha}$. Por monotonía, invariancia por traslación por constantes y la identidad $\mathbb{E}_\alpha^* \mathbb{1}_A = \mathbb{P}_\alpha^*(A)$,

$$\mathbb{E}_\alpha^* |g-f|(X_\alpha) \le \mathbb{E}_\alpha^*[\varepsilon + \mathbb{1}_{A_\alpha}] = \varepsilon + \mathbb{E}_\alpha^* \mathbb{1}_{A_\alpha} = \varepsilon + \mathbb{P}_\alpha^*(A_\alpha).$$

Sustituyendo, resulta $\Delta_\alpha(g) \le \Delta_\alpha(f) + 2\varepsilon + 2\mathbb{P}_\alpha^*(A_\alpha)$. Como $\Delta_\alpha(f) \to 0$ y $\limsup_\alpha \mathbb{P}_\alpha^*(A_\alpha) \le \varepsilon$, obtenemos $\limsup_\alpha \Delta_\alpha(g) \le 4\varepsilon$. Finalmente, haciendo $\varepsilon \downarrow 0$ y usando $\Delta_\alpha(g) \ge 0$, concluimos que $\Delta_\alpha(g) \to 0$. □

1.4. Espacios producto

1.4.1. Topología y σ-álgebras en productos finitos

Si (D,d) y (E,e) son espacios métricos, usamos en $D \times E$ la métrica

$$d_\times\Big((x,y),(x',y')\Big) := \max\{d(x,x'), e(y,y')\}.$$

Sus bolas abiertas son productos de bolas del mismo radio, y esta métrica genera la topología producto. Las proyecciones coordenadas son continuas. Para productos de un número finito de espacios se procede del mismo modo, tomando el máximo de las distancias coordenadas.

Advertencia 1.50

En $D \times E$ están la σ-álgebra producto $\mathcal{B}(D) \otimes \mathcal{B}(E)$ y la Borel de la topología producto $\mathcal{B}(D \times E)$. Siempre $\mathcal{B}(D) \otimes \mathcal{B}(E) \subset \mathcal{B}(D \times E)$, pero en general pueden diferir. Si D y E son separables, coinciden. En efecto, los rectángulos borelianos son intersecciones de preimágenes por las proyecciones, lo que prueba la inclusión. Si ambos espacios son separables, sus bases numerables dan una base numerable de rectángulos abiertos del producto; todo abierto es entonces una unión numerable de esos rectángulos y pertenece a la σ-álgebra producto.

1.4.2. Tensión y medibilidad en productos finitos

Lema 1.53. *Para una red (o sucesión) (X_α, Y_α) con valores en $\mathcal{X} \times \mathcal{Y}$:*

$$X_\alpha \text{ e } Y_\alpha \text{ son asintóticamente tensas} \iff (X_\alpha, Y_\alpha) \text{ es asintóticamente tenso.}$$

Lema 1.54. *Si X_α y Y_α son asintóticamente tensas, entonces*

$$X_\alpha \text{ e } Y_\alpha \text{ son asintóticamente medibles} \iff (X_\alpha, Y_\alpha) \text{ es asintóticamente medible.}$$

Corolario 1.55. *Sea L una probabilidad de Borel separable en $\mathcal{X} \times \mathcal{Y}$. Si, para toda f, g Lipschitz acotadas y no negativas,*

$$\mathbb{E}_\alpha^*\Big[f(X_\alpha) g(Y_\alpha)\Big] \longrightarrow \int f(x) g(y) d\mathsf{L}(x,y),$$

entonces $(X_\alpha, Y_\alpha) \Rightarrow \mathsf{L}$ en $\mathcal{X} \times \mathcal{Y}$.

Ejemplo 1.56 (Slutsky)**.** Si $X_\alpha \Rightarrow X$ (con X separable) y $Y_\alpha \Rightarrow c$ (constante), entonces $(X_\alpha, Y_\alpha) \Rightarrow (X, c)$.[37] Del teorema de aplicación continua se deducen entonces las conclusiones clásicas, con el cuidado de que cada operación involucrada esté definida y sea continua en el espacio en cuestión:

[37] Esbozo vía el Corolario 1.55: para f, g Lipschitz acotadas no negativas, la subaditividad y la homogeneidad positiva de la esperanza exterior dan

$$|\mathbb{E}_\alpha^*[f(X_\alpha)g(Y_\alpha)] - g(c)\mathbb{E}_\alpha^* f(X_\alpha)| \le \|f\|_\infty \mathbb{E}_\alpha^* |g(Y_\alpha) - g(c)| \to 0.$$

La convergencia a cero se obtiene directamente de $Y_\alpha \Rightarrow c$, aplicada a la función continua y acotada $y \mapsto |g(y) - g(c)|$. Además, $\mathbb{E}_\alpha^* f(X_\alpha) \to \mathbb{E} f(X)$. La ley de (X, c) es separable por serlo la de X, y el corolario da la conclusión.

- si X_α e Y_α toman valores en un espacio vectorial topológico metrizable D, la suma $(x, y) \mapsto x + y$ es continua y $X_\alpha + Y_\alpha \Rightarrow X + c$;
- si Y_α toma valores escalares (en $\mathbb{R}$) y X_α valores en D, el producto por escalares $(a, x) \mapsto ax$ es continuo y $Y_\alpha X_\alpha \Rightarrow cX$;
- si X_α e Y_α toman valores en $\mathbb{R}$ (o, más generalmente, en un álgebra topológica), el producto es continuo y $X_\alpha Y_\alpha \Rightarrow cX$; y si además $c \neq 0$, como la inversión $y \mapsto 1/y$ es continua fuera del origen, $X_\alpha / Y_\alpha \Rightarrow X/c$.

Nótese que en un espacio vectorial topológico general no están definidos ni el producto de dos vectores ni el cociente, de modo que las dos últimas conclusiones no tienen sentido en ese nivel de generalidad.

1.5. Espacios de funciones acotadas

Definición 1.60. Sea T un conjunto arbitrario. Definimos

$$\ell^\infty(T) := \Big\{ z : T \to \mathbb{R} : \sup_{t \in T} |z(t)| < \infty \Big\}, \qquad \|z\|_T := \sup_{t \in T} |z(t)|.$$

Con la distancia uniforme $d(z_1, z_2) = \|z_1 - z_2\|_T$, $\ell^\infty(T)$ es un espacio métrico. Para $T = \mathcal{F}$ es el espacio $\ell^\infty(\mathcal{F})$ de la Sección 1.1.

Observación 1.61. Se deja como ejercicio verificar que $\ell^\infty(T)$ es separable si y sólo si T es finito.

Definición 1.62. Un **proceso estocástico** indexado por T es una familia $\{X(t) : t \in T\}$ de variables aleatorias definidas en un mismo espacio de probabilidad, es decir, cada $X(t) : \Omega \to \mathbb{R}$ es medible.

Para $\omega \in \Omega$, la trayectoria muestral es el mapa $t \mapsto X(t, \omega)$. Si toda trayectoria muestral es acotada, el proceso induce un mapa

$$X : \Omega \longrightarrow \ell^\infty(T), \qquad \omega \longmapsto \Big(t \mapsto X(t, \omega) \Big).$$

A veces las trayectorias poseen propiedades adicionales, como continuidad o medibilidad, y resulta conveniente considerar X como una aplicación con valores en un subespacio $D \subset \ell^\infty(T)$, provisto de la métrica inducida por la norma uniforme. Esta elección no altera la noción de medibilidad, pues $\mathcal{B}(D) = \{D \cap B : B \in \mathcal{B}(\ell^\infty(T))\}$. Por tanto, una aplicación $X : \Omega \to D$ es medible como aplicación con valores en D si y sólo si lo es como aplicación con valores en $\ell^\infty(T)$.

Si, además, D es cerrado en $\ell^\infty(T)$, podemos estudiar la convergencia débil en cualquiera de los dos espacios. Precisamente, si X_n y X toman valores en D, equipado con la distancia uniforme, entonces $X_n \Rightarrow X$ en D si y sólo si $X_n \Rightarrow X$ en $\ell^\infty(T)$.

En efecto, supongamos primero que $X_n \Rightarrow X$ en D. Para verificar la convergencia en $\ell^\infty(T)$, tomemos $g \in C_b(\ell^\infty(T))$. Su restricción $g|_D$ pertenece a $C_b(D)$. Como X_n y X toman valores en D, evaluar g o $g|_D$ sobre ellos da lo mismo. Por tanto, la convergencia en D, aplicada a $g|_D$, implica $\mathbb{E}_n^* g(X_n) \to \mathbb{E} g(X)$.

Recíprocamente, supongamos que $X_n \Rightarrow X$ en $\ell^\infty(T)$ y tomemos $f \in C_b(D)$. Para aplicar esa convergencia necesitamos una función definida en todo $\ell^\infty(T)$. Como D es cerrado, el teorema de extensión de Tietze (ver Apéndice) garantiza que existe $\widetilde{f} \in C_b(\ell^\infty(T))$ que coincide con f en D. En particular, $\widetilde{f}(X_n) = f(X_n)$ y $\widetilde{f}(X) = f(X)$. Aplicando la convergencia del espacio grande a $\widetilde{f}$, obtenemos $\mathbb{E}_n^* f(X_n) \to \mathbb{E} f(X)$. Como esto vale para toda $f \in C_b(D)$, concluimos que $X_n \Rightarrow X$ en D.

La ventaja de trabajar en D es que su estructura adicional puede simplificar considerablemente las verificaciones. El ejemplo siguiente muestra cómo la continuidad de las trayectorias reduce la medibilidad del proceso a la medibilidad de sus coordenadas. En la próxima sección aparecerán otros ejemplos.

Ejemplo 1.63. Sea (T,d) un espacio métrico compacto y sea $\{X(t) : t \in T\}$ un proceso cuyas trayectorias $t \mapsto X(t,\omega)$ son todas continuas, de modo que X toma valores en el subespacio $C(T) \subset \ell^\infty(T)$ de las funciones continuas. Entonces

$$X : \Omega \to C(T) \text{ es Borel-medible} \iff X(t) \text{ es medible para todo } t \in T.$$

La implicación directa se sigue de que cada evaluación $\pi_t : C(T) \to \mathbb{R}$, $\pi_t(z) = z(t)$, es continua, y por lo tanto $X(t) = \pi_t \circ X$ es medible.

Para la recíproca, la clave es que $C(T)$, con la norma uniforme, es separable.[38] Sea $T_0 \subset T$ denso y numerable. Por continuidad de las trayectorias, para todo $z_0 \in C(T)$ vale

$$\|X - z_0\|_T = \sup_{t \in T_0} |X(t) - z_0(t)|,$$

que es un supremo numerable de variables medibles y, por tanto, una variable medible. En consecuencia, la preimagen por X de toda bola (abierta o cerrada) de $C(T)$ es medible. Como $C(T)$ es separable, todo abierto es unión numerable de bolas abiertas, de donde la preimagen de todo abierto es medible, y X resulta Borel-medible.

Para $k \in \mathbb{N}$ y $t_1, \dots, t_k \in T$, denotamos por $\pi_{t_1,\dots,t_k} : \ell^\infty(T) \to \mathbb{R}^k$ la proyección $\pi_{t_1,\dots,t_k}(z) := (z(t_1), \dots, z(t_k))$, y equipamos $\mathbb{R}^k$ con la norma del máximo. Llamamos *funciones cilindro* a los elementos de la clase

$$\mathcal{C} := \Big\{ g \circ \pi_{t_1,\dots,t_k} : \ k \in \mathbb{N},\ t_1, \dots, t_k \in T,\ g \in C_b(\mathbb{R}^k) \Big\}.$$

Lema 1.64. *Se cumplen las siguientes propiedades:*

1. *Cada proyección $\pi_{t_1,\dots,t_k}$ es 1-Lipschitz; en particular, es continua y $\mathcal{C} \subset C_b(\ell^\infty(T))$.*
2. *Dadas $f_1, \dots, f_n \in \mathcal{C}$, existen $m \in \mathbb{N}$, $t_1, \dots, t_m \in T$ y $h_1, \dots, h_n \in C_b(\mathbb{R}^m)$ tales que $f_i = h_i \circ \pi_{t_1,\dots,t_m}$ para $i = 1, \dots, n$.*
3. *$\mathcal{C}$ contiene las constantes y es un subespacio vectorial, una subálgebra y un retículo vectorial de $C_b(\ell^\infty(T))$.*
4. *$\mathcal{C}$ separa los puntos de $\ell^\infty(T)$.*

Demostración. (1) Para $z, x \in \ell^\infty(T)$,

$$\|\pi_{t_1,\dots,t_k}(z) - \pi_{t_1,\dots,t_k}(x)\|_\infty = \max_{1 \le i \le k} |z(t_i) - x(t_i)| \le \|z - x\|_T.$$

Luego $\pi_{t_1,\dots,t_k}$ es continua y, si $g \in C_b(\mathbb{R}^k)$, $g \circ \pi_{t_1,\dots,t_k}$ es continua y acotada por $\sup|g|$.

(2) Escribamos $f_i = g_i \circ \pi_{s_1^i,\dots,s_{k_i}^i}$, con $g_i \in C_b(\mathbb{R}^{k_i})$. Sea $(t_1, \dots, t_m)$, con $m := k_1 + \cdots + k_n$, la concatenación de las listas $(s_1^1, \dots, s_{k_1}^1), \dots, (s_1^n, \dots, s_{k_n}^n)$, y sea $p_i : \mathbb{R}^m \to \mathbb{R}^{k_i}$ la proyección sobre las coordenadas del i-ésimo bloque. Entonces $\pi_{s_1^i,\dots,s_{k_i}^i} = p_i \circ \pi_{t_1,\dots,t_m}$, luego $f_i = h_i \circ \pi_{t_1,\dots,t_m}$ con $h_i := g_i \circ p_i$. Cada h_i es continua, por ser composición de continuas, y acotada, pues $\sup|h_i| \le \sup|g_i|$; es decir, $h_i \in C_b(\mathbb{R}^m)$.

(3) Las constantes están en $\mathcal{C}$: basta tomar $k = 1$ y g constante. Dadas $f_1, f_2 \in \mathcal{C}$ y $a, b \in \mathbb{R}$, por (2) podemos escribir $f_i = h_i \circ \pi_{t_1,\dots,t_m}$ con $h_i \in C_b(\mathbb{R}^m)$. Entonces

$$af_1 + bf_2 = (ah_1 + bh_2) \circ \pi_{t_1,\dots,t_m}, \qquad f_1 f_2 = (h_1 h_2) \circ \pi_{t_1,\dots,t_m},$$

$$f_1 \vee f_2 = (h_1 \vee h_2) \circ \pi_{t_1,\dots,t_m}, \qquad f_1 \wedge f_2 = (h_1 \wedge h_2) \circ \pi_{t_1,\dots,t_m}.$$

[38] Por el teorema de Stone–Weierstrass, o bien directamente: si $T_0 = \{t_1, t_2, \dots\}$ es denso numerable en T, las funciones de la forma $z(t) = \max_i (q_i - r_i\, d(t,t_i)) \vee c$, con parámetros racionales y máximos finitos, forman una familia numerable densa.

Como $C_b(\mathbb{R}^m)$ es cerrado por combinaciones lineales, productos, máximos y mínimos, todas estas funciones pertenecen a $\mathcal{C}$. Junto con (1), esto prueba que $\mathcal{C}$ es un subespacio vectorial, una subálgebra y un retículo vectorial de $C_b(\ell^\infty(T))$.

(4) Si $z_1 \neq z_2$, existe $t \in T$ con $z_1(t) \neq z_2(t)$. La función $f := \arctan \circ \pi_t$ pertenece a $\mathcal{C}$, pues $\arctan \in C_b(\mathbb{R})$, y cumple $f(z_1) \neq f(z_2)$ por ser la arcotangente inyectiva. □

Lema 1.65. *Sea $X_\alpha : \Omega_\alpha \to \ell^\infty(T)$ una red asintóticamente tensa. Entonces X_α es asintóticamente medible si y sólo si, para todo $t \in T$, la marginal real $X_\alpha(t)$ es asintóticamente medible.*

Demostración. ($\Rightarrow$) Sea $t \in T$. Si $\varphi \in C_b(\mathbb{R})$, entonces $\varphi \circ \pi_t \in C_b(\ell^\infty(T))$ por el Lema 1.64(1), luego

$$\mathbb{E}^*_\alpha \varphi(X_\alpha(t)) - \mathbb{E}_{\alpha,*}\varphi(X_\alpha(t)) = \mathbb{E}^*_\alpha(\varphi \circ \pi_t)(X_\alpha) - \mathbb{E}_{\alpha,*}(\varphi \circ \pi_t)(X_\alpha) \longrightarrow 0,$$

y cada marginal $X_\alpha(t)$ es asintóticamente medible.

($\Leftarrow$) Veamos primero que todo vector finito de coordenadas es asintóticamente medible. Dados $t_1, \dots, t_k \in T$, escribamos $\pi := \pi_{t_1,\dots,t_k}$ y $Y_\alpha := \pi(X_\alpha)$.

Y_α es asintóticamente tensa: dado $\varepsilon > 0$, sea $K \subset \ell^\infty(T)$ compacto tal que $\liminf_\alpha \mathbb{P}_{\alpha,*}(X_\alpha \in K^\delta) \geq 1 - \varepsilon$ para todo $\delta > 0$. Por el Lema 1.64(1), π es 1-Lipschitz; luego $\pi(K)$ es compacto y, si $z \in K^\delta$, existe $x \in K$ con $\|z - x\|_T < \delta$, de donde $\|\pi(z) - \pi(x)\|_\infty < \delta$. Así $\pi(K^\delta) \subseteq \pi(K)^\delta$, es decir, $\{X_\alpha \in K^\delta\} \subseteq \{Y_\alpha \in \pi(K)^\delta\}$, y por monotonía de la probabilidad interior,

$$\liminf_\alpha \mathbb{P}_{\alpha,*}\Big(Y_\alpha \in \pi(K)^\delta\Big) \geq \liminf_\alpha \mathbb{P}_{\alpha,*}(X_\alpha \in K^\delta) \geq 1 - \varepsilon \qquad \text{para todo } \delta > 0.$$

Como esto vale para toda lista finita de coordenadas, aplicamos inductivamente el Lema 1.54, agrupando en cada paso las primeras $k-1$ coordenadas y añadiendo la última: ambos factores son asintóticamente tensos y asintóticamente medibles, luego todo vector Y_α de coordenadas es asintóticamente medible.

Ahora, si $f = g \circ \pi_{t_1,\dots,t_k} \in \mathcal{C}$, entonces $f(X_\alpha) = g(Y_\alpha)$ con $Y_\alpha = \pi_{t_1,\dots,t_k}(X_\alpha)$, y por lo anterior

$$\mathbb{E}^*_\alpha f(X_\alpha) - \mathbb{E}_{\alpha,*} f(X_\alpha) = \mathbb{E}^*_\alpha g(Y_\alpha) - \mathbb{E}_{\alpha,*} g(Y_\alpha) \longrightarrow 0 \qquad \text{para toda } f \in \mathcal{C}.$$

Por el Lema 1.64, $\mathcal{C}$ es una subálgebra de $C_b(\ell^\infty(T))$ que contiene las constantes y separa puntos. Como X_α es asintóticamente tensa, el Lema 1.48 extiende la convergencia anterior a toda $f \in C_b(\ell^\infty(T))$, que es precisamente la medibilidad asintótica de X_α. □

Lema 1.66. *Sean L_1 y L_2 probabilidades de Borel* tensas *en $\ell^\infty(T)$. Son equivalentes:*

1. *$\mathsf{L}_1 = \mathsf{L}_2$ en $\mathcal{B}(\ell^\infty(T))$;*
2. *$\mathsf{L}_1 \circ \pi^{-1}_{t_1,\dots,t_k} = \mathsf{L}_2 \circ \pi^{-1}_{t_1,\dots,t_k}$ para todo $k \in \mathbb{N}$ y todo $t_1, \dots, t_k \in T$.*

En particular, si X e Y son aplicaciones Borel-medibles con valores en $\ell^\infty(T)$ y leyes tensas, definidas en espacios de probabilidad posiblemente distintos, entonces tienen la misma ley si y sólo si

$$\Big(X(t_1), \dots, X(t_k)\Big) \overset{d}{=} \Big(Y(t_1), \dots, Y(t_k)\Big)$$

para toda elección finita $t_1, \dots, t_k \in T$.

Demostración. (1) $\Rightarrow$ (2). Por el Lema 1.64(1), las proyecciones $\pi_{t_1,\dots,t_k}$ son continuas y, en particular, Borel-medibles; luego, para todo $B \in \mathcal{B}(\mathbb{R}^k)$, se tiene $\pi^{-1}_{t_1,\dots,t_k}(B) \in \mathcal{B}(\ell^\infty(T))$ y $\mathsf{L}_1\Big(\pi^{-1}_{t_1,\dots,t_k}(B)\Big) = \mathsf{L}_2\Big(\pi^{-1}_{t_1,\dots,t_k}(B)\Big)$.

(2) $\Rightarrow$ (1). Si $f = g \circ \pi_{t_1,\dots,t_k} \in \mathcal{C}$, por cambio de variables y la igualdad de las marginales finito-dimensionales,

$$\int f \, d\mathsf{L}_1 = \int g \, d(\mathsf{L}_1 \circ \pi^{-1}_{t_1,\dots,t_k}) = \int g \, d(\mathsf{L}_2 \circ \pi^{-1}_{t_1,\dots,t_k}) = \int f \, d\mathsf{L}_2.$$

Por el Lema 1.64, $\mathcal{C}$ es una subálgebra y un retículo vectorial de $C_b(\ell^\infty(T))$ que contiene las constantes y separa puntos. Como L_1 y L_2 son tensas, el punto (ii) del Lema 1.47 aplicado en $D = \ell^\infty(T)$ da $\mathsf{L}_1 = \mathsf{L}_2$ en $\mathcal{B}(\ell^\infty(T))$.

La afirmación sobre X e Y es la equivalencia (1) $\Leftrightarrow$ (2) aplicada a $\mathsf{L}_1 := \mathbb{P} \circ X^{-1}$ y $\mathsf{L}_2 := \mathbb{P}' \circ Y^{-1}$: la condición (2) para estas leyes es exactamente la igualdad en distribución de los vectores finito-dimensionales, y no requiere que X e Y estén definidas en un mismo espacio de probabilidad. □

Teorema 1.67. *Sea $X_\alpha : \Omega_\alpha \to \ell^\infty(T)$ una red de aplicaciones, no necesariamente medibles. Entonces X_α converge débilmente hacia alguna probabilidad de Borel tensa en $\ell^\infty(T)$ si y solo si:*

(i) X_α es asintóticamente tensa;

(ii) para toda elección finita $t_1, \dots, t_k \in T$, existe una probabilidad $\mathsf{L}_{t_1,\dots,t_k}$ en $\mathbb{R}^k$ tal que

$$\Big(X_\alpha(t_1), \dots, X_\alpha(t_k)\Big) \Rightarrow \mathsf{L}_{t_1,\dots,t_k}.$$

En tal caso, la ley límite L es la única probabilidad de Borel tensa en $\ell^\infty(T)$ cuyas marginales finito-dimensionales son las leyes de (ii); es decir,

$$\mathsf{L} \circ \pi_{t_1,\dots,t_k}^{-1} = \mathsf{L}_{t_1,\dots,t_k}, \qquad \pi_{t_1,\dots,t_k}(z) := (z(t_1), \dots, z(t_k)).$$

Si, además, $X : \Omega \to \ell^\infty(T)$ es Borel-medible, tiene ley tensa y sus distribuciones finito-dimensionales son las leyes de (ii)*, entonces $\mathcal{L}(X) = \mathsf{L}$ y $X_\alpha \Rightarrow X$.*

Demostración. ($\Rightarrow$). Supongamos que $X_\alpha \Rightarrow \mathsf{L}$, con L una probabilidad de Borel tensa en $\ell^\infty(T)$. Sea X la identidad sobre $(\ell^\infty(T), \mathcal{B}(\ell^\infty(T)), \mathsf{L})$; es Borel-medible y su ley es L, de modo que, con la convención que sigue a la definición de convergencia débil, $X_\alpha \Rightarrow X$. Como X es tenso, el Lema 1.40(b) da (i).

Para (ii), fijemos $t_1, \dots, t_k \in T$ y $g \in C_b(\mathbb{R}^k)$. Por el Lema 1.64(1), $\pi_{t_1,\dots,t_k}$ es continua y $g \circ \pi_{t_1,\dots,t_k} \in C_b(\ell^\infty(T))$. Por la convergencia débil y el cambio de variables,

$$\mathbb{E}_\alpha^* g\Big(\pi_{t_1,\dots,t_k}(X_\alpha)\Big) = \mathbb{E}_\alpha^*(g \circ \pi_{t_1,\dots,t_k})(X_\alpha) \longrightarrow \int g \circ \pi_{t_1,\dots,t_k}\, d\mathsf{L} = \int g\, d\Big(\mathsf{L} \circ \pi_{t_1,\dots,t_k}^{-1}\Big).$$

Como $\pi_{t_1,\dots,t_k}$ es Borel-medible, $\mathsf{L} \circ \pi_{t_1,\dots,t_k}^{-1}$ es una probabilidad de Borel en $\mathbb{R}^k$, y vale (ii) con $\mathsf{L}_{t_1,\dots,t_k} = \mathsf{L} \circ \pi_{t_1,\dots,t_k}^{-1}$. (Este es el caso particular del Teorema 1.36 para la proyección $\pi_{t_1,\dots,t_k}$, continua en todo $\ell^\infty(T)$.)

($\Leftarrow$). Supongamos (i) y (ii). Para aplicar el Teorema 1.44 necesitamos que X_α sea asintóticamente tensa, que es la hipótesis (i), y asintóticamente medible, que probamos ahora. Sea $t \in T$. El caso $k = 1$ de (ii) da $X_\alpha(t) \Rightarrow \mathsf{L}_t$ en $\mathbb{R}$, es decir, $X_\alpha(t) \Rightarrow Z_t$, donde Z_t es la identidad sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathsf{L}_t)$. Como Z_t es Borel-medible, el Lema 1.40(a) implica que $X_\alpha(t)$ es asintóticamente medible. Como esto vale para todo $t \in T$ y X_α es asintóticamente tensa por (i), el Lema 1.65 implica que X_α es asintóticamente medible en $\ell^\infty(T)$.

Sea $X_{\alpha(\beta)}$ una subred de X_α en el sentido de la Definición 1.42. Por la propiedad de finalidad de la Advertencia 1.43, para toda red real (z_α) se tiene $\liminf_\beta z_{\alpha(\beta)} \geq \liminf_\alpha z_\alpha$, y si $z_\alpha \to z$ también $z_{\alpha(\beta)} \to z$. Lo primero muestra que el compacto que da la tensión asintótica de X_α para cada ε sirve también para la subred; lo segundo, aplicado a $\mathbb{E}_\alpha^* f(X_\alpha) - \mathbb{E}_{\alpha,*} f(X_\alpha) \to 0$ para cada $f \in C_b(\ell^\infty(T))$, muestra que la subred es asintóticamente medible. Por el Teorema 1.44(i), aplicado a $X_{\alpha(\beta)}$, esta tiene una subred ulterior que converge débilmente a una probabilidad de Borel tensa. Esa subred ulterior es también una subred de X_α, pues la composición de las funciones que definen dos subredes sucesivas es monótona y cofinal.

Sea Q uno de esos límites, digamos $X_{\alpha(\beta)} \Rightarrow \mathsf{Q}$ para una subred $X_{\alpha(\beta)}$ de X_α. Fijemos $t_1, \dots, t_k \in T$, escribamos $\pi := \pi_{t_1,\dots,t_k}$ y sea $g \in C_b(\mathbb{R}^k)$. Por un lado, el argumento de ($\Rightarrow$) aplicado a la subred da

$$\mathbb{E}_{\alpha(\beta)}^* g\Big(\pi(X_{\alpha(\beta)})\Big) \longrightarrow \int g\, d(\mathsf{Q} \circ \pi^{-1}).$$

Por otro, por (ii) la red real $\mathbb{E}^*_\alpha g(\pi(X_\alpha))$ converge a $\int g\, d\mathsf{L}_{t_1,\dots,t_k}$, y por lo tanto también su subred $\mathbb{E}^*_{\alpha(\beta)} g(\pi(X_{\alpha(\beta)}))$. Por la unicidad de los límites en $\mathbb{R}$,

$$\int g\, d(\mathsf{Q}\circ\pi^{-1}) = \int g\, d\mathsf{L}_{t_1,\dots,t_k} \qquad \text{para toda } g\in C_b(\mathbb{R}^k),$$

y el Lema 1.47(i), aplicado en $D=\mathbb{R}^k$, da

$$\mathsf{Q}\circ\pi^{-1}_{t_1,\dots,t_k} = \mathsf{L}_{t_1,\dots,t_k} \qquad \text{para toda elección finita } t_1,\dots,t_k\in T.$$

Veamos que el límite es único. Si Q_1 y Q_2 son dos límites débiles de subredes como los anteriores, son probabilidades de Borel tensas y, por lo recién probado,

$$\mathsf{Q}_1\circ\pi^{-1}_{t_1,\dots,t_k} = \mathsf{L}_{t_1,\dots,t_k} = \mathsf{Q}_2\circ\pi^{-1}_{t_1,\dots,t_k} \qquad \text{para toda elección finita } t_1,\dots,t_k\in T.$$

Por el Lema 1.66, $\mathsf{Q}_1=\mathsf{Q}_2$. Denotemos por L esta ley común.

Veamos que toda la red converge. Sea $(A,\le)$ el conjunto dirigido que indexa la red. Fijemos $f\in C_b(\ell^\infty(T))$ y sean $a_\alpha := \mathbb{E}^*_\alpha f(X_\alpha)$ y $a:=\int f\, d\mathsf{L}$. Dada una subred $(a_{\alpha(\beta)})_\beta$, la subred correspondiente $(X_{\alpha(\beta)})_\beta$ tiene, por los pasos anteriores, una subred ulterior $(X_{\alpha(\beta(\gamma))})_\gamma$ que converge débilmente a una probabilidad de Borel tensa, que por la unicidad anterior es L. Aplicando la definición de convergencia débil a f,

$$a_{\alpha(\beta(\gamma))} = \mathbb{E}^*_{\alpha(\beta(\gamma))} f(X_{\alpha(\beta(\gamma))}) \longrightarrow \int f\, d\mathsf{L} = a.$$

Así, toda subred de (a_α) tiene una subred ulterior que converge a a. Veamos que esto implica $a_\alpha\to a$. Si no fuera así, existiría $\varepsilon>0$ tal que para todo $\alpha_0\in A$ hay $\alpha\ge\alpha_0$ con $|a_\alpha-a|\ge\varepsilon$; es decir, el conjunto

$$A_\varepsilon := \{\alpha\in A : |a_\alpha - a|\ge\varepsilon\}$$

sería cofinal en A. Con el orden heredado de A, A_ε es dirigido: dados $\alpha_1,\alpha_2\in A_\varepsilon$, tomamos una cota superior común $\alpha_3\in A$ y, por cofinalidad, $\alpha_4\in A_\varepsilon$ con $\alpha_4\ge\alpha_3$. La inclusión $A_\varepsilon\hookrightarrow A$ es monótona, y es cofinal por la cofinalidad de A_ε; por la Definición 1.42, $(a_\alpha)_{\alpha\in A_\varepsilon}$ es una subred de $(a_\alpha)_{\alpha\in A}$, y todos sus términos satisfacen $|a_\alpha-a|\ge\varepsilon$. Ninguna subred ulterior suya puede converger a a, pues todos sus términos satisfacen la misma desigualdad, lo que contradice lo anterior. Por lo tanto,

$$\mathbb{E}^*_\alpha f(X_\alpha)\longrightarrow\int f\, d\mathsf{L},$$

y como $f\in C_b(\ell^\infty(T))$ era arbitraria, $X_\alpha\Rightarrow\mathsf{L}$.

Si L' es otra probabilidad de Borel tensa con $\mathsf{L}'\circ\pi^{-1}_{t_1,\dots,t_k}=\mathsf{L}_{t_1,\dots,t_k}$ para toda elección finita $t_1,\dots,t_k\in T$, entonces L' y L tienen las mismas marginales finito-dimensionales, y el Lema 1.66 da $\mathsf{L}'=\mathsf{L}$.

Finalmente, si X es Borel-medible, tiene ley tensa y sus marginales finito-dimensionales son las $\mathsf{L}_{t_1,\dots,t_k}$, entonces $\mathcal{L}(X)=\mathbb{P}\circ X^{-1}$ es una probabilidad de Borel tensa con esas marginales y, por el paso anterior, $\mathcal{L}(X)=\mathsf{L}$. Como $X_\alpha\Rightarrow\mathsf{L}$, la convención que sigue a la definición de convergencia débil da $X_\alpha\Rightarrow X$. □

La convergencia de las marginales puede establecerse mediante cualquiera de los métodos bien conocidos para probar convergencia débil en espacios euclídeos (TCL multivariado, Cramér-Wold, convergencia de funciones características, Slutsky, etc). La tensión puede darse en forma más concreta, ya sea mediante *aproximación finita* o (esencialmente) con ayuda del teorema de Arzelà–Ascoli. La aproximación finita conduce a la más simple de las dos caracterizaciones, mientras que la segunda es tal vez de mayor interés porque conecta la tensión con la (asintótica, uniforme, equi-) continuidad de las trayectorias $t\mapsto X_\alpha(t)$.

La idea de la *aproximación finita* supone que para todo $\varepsilon,\eta>0$ existe una partición finita

$$T=\bigcup_{i=1}^{k} T_i$$

tal que

$$\limsup_{\alpha} \mathbb{P}^*_\alpha \left(\sup_{1\le i\le k} \sup_{s,t\in T_i} \left| X_\alpha(s) - X_\alpha(t) \right| > \varepsilon \right) < \eta. \tag{1.9}$$

Bajo esta condición, el comportamiento asintótico del proceso puede describirse, con error a lo sumo ε, η, por el comportamiento de las marginales finito–dimensionales $\big(X_\alpha(t_1), \dots, X_\alpha(t_k)\big)$ para puntos fijos $t_i \in T_i$. Si, de este modo, el proceso puede reducirse a un número finito de coordenadas para cualquier par $\varepsilon, \eta > 0$, y las redes de distribuciones marginales son tensas, entonces la red X_α es asintóticamente tensa.

Teorema 1.68. *Una red $X_\alpha : \Omega_\alpha \to \ell^\infty(T)$ es* asintóticamente tensa *si y sólo si*

1. *$X_\alpha(t)$ es asintóticamente tensa en $\mathbb{R}$ para todo $t \in T$;*
2. *para todo $\varepsilon, \eta > 0$ existe una partición finita $T = \bigcup_{i=1}^k T_i$ tal que se verifica (1.9).*

Este resultado es el Teorema 1.5.6 de [5]. En el libro, la necesidad se deduce del criterio de equicontinuidad (Teorema 1.71, más adelante), tomando la partición que se obtiene al disjuntar bolas de una semimétrica[39] que hace a T totalmente acotado, y sólo la suficiencia se demuestra directamente; aquí incluimos, además, una demostración directa de la necesidad.

Demostración. Suponemos $T \neq \varnothing$, pues el otro caso es inmediato, y descartamos siempre los bloques vacíos de las particiones. Para una partición finita $\mathcal{P} = \{T_1, \dots, T_k\}$ de T y $z \in \ell^\infty(T)$, escribimos

$$\operatorname{osc}_{\mathcal{P}}(z) := \max_{i\le k} \sup_{s,t\in T_i} |z(s) - z(t)|,$$

de modo que (1.9) dice $\limsup_\alpha \mathbb{P}^*_\alpha\big(\operatorname{osc}_{\mathcal{P}}(X_\alpha) > \varepsilon\big) < \eta$. Si s, t están en un mismo bloque y $w \in \ell^\infty(T)$, la desigualdad triangular da $|z(s) - z(t)| \le |w(s) - w(t)| + 2\|z - w\|_T$; tomando supremos e intercambiando z y w,

$$\big|\operatorname{osc}_{\mathcal{P}}(z) - \operatorname{osc}_{\mathcal{P}}(w)\big| \le 2\|z - w\|_T, \tag{1.10}$$

y en particular $\operatorname{osc}_{\mathcal{P}}$ es continua en $\ell^\infty(T)$. Por último, usaremos la tensión asintótica en la forma equivalente que figura en la Definición 1.39, con γ en lugar de ε.

Necesidad. Supongamos que X_α es asintóticamente tensa. Por el Lema 1.64(1), cada evaluación π_t es 1-Lipschitz; por lo tanto lleva compactos en compactos y satisface $\pi_t(K^\delta) \subseteq \pi_t(K)^\delta$. Luego $\{X_\alpha(t) \notin \pi_t(K)^\delta\} \subseteq \{X_\alpha \notin K^\delta\}$, y la tensión asintótica de X_α implica la de cada $X_\alpha(t)$.

Fijemos $\varepsilon, \eta > 0$. Por tensión asintótica existe un compacto K tal que

$$\limsup_{\alpha} \mathbb{P}^*_\alpha(X_\alpha \notin K^{\varepsilon/6}) \le \eta/2;$$

agregando la función nula, lo que sólo agranda $K^{\varepsilon/6}$, podemos suponer $K \neq \varnothing$. Por compacidad existen $z_1, \dots, z_N \in K$ con $K \subseteq \bigcup_{j=1}^N B(z_j, \varepsilon/6)$.

Construimos una sola partición de T en cuyos bloques todas las z_j oscilen a lo sumo $\varepsilon/3$. Para $t \in T$ y $j \le N$, sea $\ell_j(t) := \lfloor 3z_j(t)/\varepsilon \rfloor$, el único entero tal que

$$z_j(t) \in \big[\ell_j(t)\varepsilon/3,\ (\ell_j(t) + 1)\varepsilon/3\big).$$

Como $|z_j(t)| \le \|z_j\|_T$, se tiene $|\ell_j(t)| \le 3\|z_j\|_T/\varepsilon + 1$, así que cada ℓ_j toma sólo finitos valores. Decimos que $s, t \in T$ son equivalentes si $\ell_j(s) = \ell_j(t)$ para todo $j \le N$, y llamamos $\mathcal{P}$ a la partición de T en las clases de esta relación. Cada clase queda determinada por el valor de $(\ell_1, \dots, \ell_N)$ en ella, que es uno entre finitos posibles; luego $\mathcal{P}$ es finita. Si s, t están en un mismo bloque, para cada j

[39] Una *semimétrica* (o *pseudométrica*) en T es una función $\rho : T \times T \to [0, \infty)$ tal que, para todos $s, t, u \in T$, $\rho(t, t) = 0$, $\rho(s, t) = \rho(t, s)$ y $\rho(s, t) \le \rho(s, u) + \rho(u, t)$. Es decir, cumple los axiomas de una métrica salvo que $\rho(s, t) = 0$ no implica $s = t$: puntos distintos pueden estar a distancia nula. Las nociones métricas, como la total acotación (nota de la Definición 1.27), se definen igual que para métricas.

los números $z_j(s)$ y $z_j(t)$ están en un mismo intervalo semiabierto de longitud $\varepsilon/3$, de modo que $|z_j(s) - z_j(t)| < \varepsilon/3$. Por lo tanto, $\operatorname{osc}_{\mathcal{P}}(z_j) \le \varepsilon/3$ para todo j. La construcción sólo usa los valores de las z_j: no requiere ninguna estructura en T.

Si $x \in K^{\varepsilon/6}$, existen $z \in K$ y un centro z_j con $\|x - z\|_T < \varepsilon/6$ y $\|z - z_j\|_T < \varepsilon/6$. Luego $\|x - z_j\|_T < \varepsilon/3$ y, por (1.10),

$$\operatorname{osc}_{\mathcal{P}}(x) \le \operatorname{osc}_{\mathcal{P}}(z_j) + 2\|x - z_j\|_T < \varepsilon/3 + 2\varepsilon/3 = \varepsilon.$$

En consecuencia, $\{\operatorname{osc}_{\mathcal{P}}(X_\alpha) > \varepsilon\} \subseteq \{X_\alpha \notin K^{\varepsilon/6}\}$ y, por monotonía de la probabilidad exterior,

$$\limsup_\alpha \mathbb{P}^*_\alpha\Big(\operatorname{osc}_{\mathcal{P}}(X_\alpha) > \varepsilon\Big) \le \eta/2 < \eta,$$

que es (1.9).

Suficiencia. Supongamos que valen ambas condiciones y fijemos $\gamma > 0$. Vamos a construir un compacto K tal que $\limsup_\alpha \mathbb{P}^*_\alpha(X_\alpha \notin K^\delta) < \gamma$ para todo $\delta > 0$.

Por la segunda condición, para cada $m \ge 1$ existe una partición finita $\mathcal{P}_m$ tal que

$$\limsup_\alpha \mathbb{P}^*_\alpha\Big(\operatorname{osc}_{\mathcal{P}_m}(X_\alpha) > 2^{-m}\Big) < \gamma 2^{-m-1}. \tag{1.11}$$

Podemos suponer que cada $\mathcal{P}_{m+1}$ refina a $\mathcal{P}_m$: basta reemplazar $\mathcal{P}_m$ por el refinamiento común de $\mathcal{P}_1, \dots, \mathcal{P}_m$, cuyos bloques son las intersecciones no vacías de un bloque de cada una[40]. Este refinamiento es una partición finita que refina a $\mathcal{P}_m$, y conserva (1.11) porque refinar no aumenta la oscilación: si cada bloque de una partición $\mathcal{Q}$ está contenido en uno de $\mathcal{P}$, entonces $\operatorname{osc}_{\mathcal{Q}}(z) \le \operatorname{osc}_{\mathcal{P}}(z)$.

Elegimos un representante en cada bloque de cada $\mathcal{P}_n$ y, para $z \in \ell^\infty(T)$, definimos $\Pi_n z$ como la función que en cada bloque de $\mathcal{P}_n$ vale constantemente el valor de z en el representante de ese bloque (ver Figura 1.1). Como $\Pi_n z$ sólo toma valores de z, $\|\Pi_n z\|_T \le \|z\|_T$; como cada t está en el mismo bloque de $\mathcal{P}_n$ que su representante, $\|z - \Pi_n z\|_T \le \operatorname{osc}_{\mathcal{P}_n}(z)$. Si $m \ge n$, cada bloque de $\mathcal{P}_m$ está contenido en un bloque de $\mathcal{P}_n$, donde $\Pi_n z$ es constante; si $m \le n$ y s, t están en un mismo bloque de $\mathcal{P}_m$, los representantes de sus bloques en $\mathcal{P}_n$ también están en él, porque $\mathcal{P}_n$ refina a $\mathcal{P}_m$. En resumen,

$$\begin{gathered}\|\Pi_n z\|_T \le \|z\|_T, \qquad \|z - \Pi_n z\|_T \le \operatorname{osc}_{\mathcal{P}_n}(z), \\ \operatorname{osc}_{\mathcal{P}_m}(\Pi_n z) \le \begin{cases} \operatorname{osc}_{\mathcal{P}_m}(z), & m \le n, \\ 0, & m \ge n. \end{cases}\end{gathered} \tag{1.12}$$

Sean $t_1, \dots, t_k$ los representantes de los bloques de $\mathcal{P}_1$. Como todo compacto de $\mathbb{R}$ está contenido en un intervalo $[-c, c]$, cuyo 1-entorno es $(-c-1, c+1)$, la primera condición da un $R > 0$ tal que

$$\limsup_\alpha \mathbb{P}^*_\alpha\Big(|X_\alpha(t_j)| > R\Big) \le \frac{\gamma}{4k} \qquad \text{para todo } j \le k \tag{1.13}$$

(basta tomar el mayor de los radios obtenidos para cada j). Sea

$$K := \Big\{z \in \ell^\infty(T) : \ \|z\|_T \le R + 1/2, \quad \operatorname{osc}_{\mathcal{P}_m}(z) \le 2^{-m} \text{ para todo } m \ge 1\Big\}.$$

Veamos que K es compacto. Como $\ell^\infty(T)$ es completo, basta ver que K es cerrado y totalmente acotado: un subconjunto cerrado de un espacio completo es completo, y un espacio métrico completo y totalmente acotado es compacto.

K es cerrado porque es la intersección de los conjuntos $\{z : \|z\|_T \le R + 1/2\}$ y $\{z : \operatorname{osc}_{\mathcal{P}_m}(z) \le 2^{-m}\}$, $m \ge 1$, que son cerrados por la continuidad de la norma y de las oscilaciones (1.10).

[40] Si $\mathcal{Q}_m$ es el refinamiento común de $\mathcal{P}_1, \dots, \mathcal{P}_m$, sus bloques son los conjuntos no vacíos $A_1 \cap \dots \cap A_m$ con $A_i \in \mathcal{P}_i$. Son disjuntos y cubren T, porque cada t está en exactamente un bloque de cada $\mathcal{P}_i$, y son a lo sumo $|\mathcal{P}_1| \cdots |\mathcal{P}_m|$. Cada bloque de $\mathcal{Q}_{m+1}$ está contenido en uno de $\mathcal{Q}_m$ (el que resulta de omitir A_{m+1}), y cada bloque de $\mathcal{Q}_m$ en uno de $\mathcal{P}_m$ (el A_m); así, $\mathcal{Q}_{m+1}$ refina a $\mathcal{Q}_m$, y $\mathcal{Q}_m$ refina a $\mathcal{P}_m$.

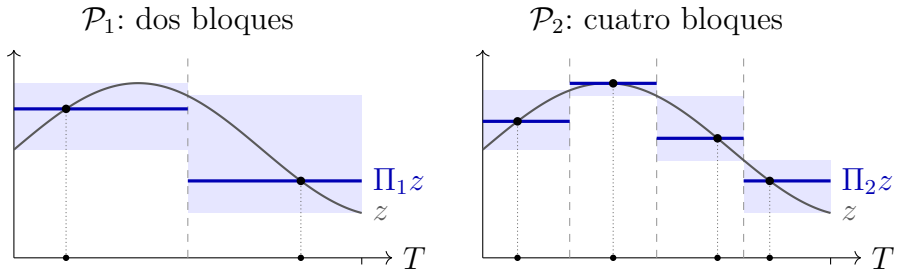


Figura 1.1: La función z (en gris) y su aproximación $\Pi_n z$ (en azul), para una partición $\mathcal{P}_1$ con dos bloques y un refinamiento $\mathcal{P}_2$ con cuatro; por simplicidad, T se dibuja como un intervalo. Los puntos marcan los representantes, y $\Pi_n z$ vale en cada bloque lo que vale z en su representante. La franja sombreada de cada bloque va del mínimo al máximo de z en él: como $\Pi_n z$ queda dentro de la franja, $|z - \Pi_n z|$ no supera su altura, lo que da $\|z - \Pi_n z\|_T \le \operatorname{osc}_{\mathcal{P}_n}(z)$. Al refinar, las franjas se achican y $\Pi_n z$ se acerca a z.

Para ver que es totalmente acotado, fijemos n y llamemos S_n al conjunto de las funciones constantes en los bloques de $\mathcal{P}_n$ con valores en $[-(R+1/2),(R+1/2)]$. Si $\mathcal{P}_n$ tiene p bloques, asociar a cada función de S_n el vector de sus p valores es una isometría de S_n sobre $[-(R+1/2),(R+1/2)]^p$ con la norma del máximo; por lo tanto, S_n es compacto y existen $f_1,\dots,f_q \in S_n$ tales que las bolas $B(f_i, 2^{-n})$ cubren S_n. Sea ahora $z \in K$. Por (1.12), $\Pi_n z \in S_n$, porque es constante en los bloques de $\mathcal{P}_n$ y $\|\Pi_n z\|_T \le \|z\|_T \le R+1/2$; además, $\|z - \Pi_n z\|_T \le \operatorname{osc}_{\mathcal{P}_n}(z) \le 2^{-n}$. Si i es tal que $\|\Pi_n z - f_i\|_T < 2^{-n}$, la desigualdad triangular da

$$\|z - f_i\|_T \le \|z - \Pi_n z\|_T + \|\Pi_n z - f_i\|_T < 2^{-n+1}.$$

Así, $K \subseteq \bigcup_{i=1}^{q} B(f_i, 2^{-n+1})$ y, como n es arbitrario, K se cubre con finitas bolas de radio tan chico como se quiera; es decir, es totalmente acotado. (Si se prefiere que los centros estén en K, basta elegir un punto de K en cada una de esas bolas que corte a K y duplicar el radio.)

Fijemos $\delta > 0$ y n con $2^{-n} < \delta$, y sea $x \in \ell^\infty(T)$ tal que $|x(t_j)| \le R$ para todo $j \le k$ y $\operatorname{osc}_{\mathcal{P}_m}(x) \le 2^{-m}$ para $m = 1,\dots,n$. Veamos que $x \in K^\delta$. Primero acotamos su norma: $\Pi_1 x$ toma los valores $x(t_1),\dots,x(t_k)$, así que $\|\Pi_1 x\|_T \le R$, y el caso $n=1$ de (1.12) da

$$\|x\|_T \le \|\Pi_1 x\|_T + \|x - \Pi_1 x\|_T \le R + \operatorname{osc}_{\mathcal{P}_1}(x) \le R + \tfrac{1}{2}.$$

Por (1.12), entonces, $\|\Pi_n x\|_T \le \|x\|_T \le R + 1/2$, $\operatorname{osc}_{\mathcal{P}_m}(\Pi_n x) \le \operatorname{osc}_{\mathcal{P}_m}(x) \le 2^{-m}$ para $m \le n$ y $\operatorname{osc}_{\mathcal{P}_m}(\Pi_n x) = 0$ para $m \ge n$; es decir, $\Pi_n x \in K$. Además, $\|x - \Pi_n x\|_T \le \operatorname{osc}_{\mathcal{P}_n}(x) \le 2^{-n} < \delta$, de modo que $x \in K^\delta$.

En consecuencia,

$$\{X_\alpha \notin K^\delta\} \subseteq \bigcup_{j=1}^{k} \Big\{|X_\alpha(t_j)| > R\Big\} \cup \bigcup_{m=1}^{n} \Big\{\operatorname{osc}_{\mathcal{P}_m}(X_\alpha) > 2^{-m}\Big\},$$

y la subaditividad de la probabilidad exterior, la desigualdad del límite superior para sumas finitas, (1.13) y (1.11) dan

$$\limsup_\alpha \mathbb{P}^*_\alpha(X_\alpha \notin K^\delta) < k \cdot \frac{\gamma}{4k} + \sum_{m=1}^{n} \gamma 2^{-m-1} < \frac{\gamma}{4} + \frac{\gamma}{2} < \gamma.$$

Como K no depende de δ, X_α es asintóticamente tensa. □

El segundo tipo de caracterización de la tensión asintótica es más profundo: relaciona la tensión en $\ell^\infty(T)$ con una forma asintótica de continuidad uniforme de las trayectorias.

Definición 1.70. Sea ρ una semimétrica en T. Una red $X_\alpha : \Omega_\alpha \to \ell^\infty(T)$ es **asintóticamente uniformemente ρ-equicontinua en probabilidad** si, para todo $\varepsilon, \eta > 0$, existe $\delta > 0$ tal que

$$\limsup_\alpha \mathbb{P}^*_\alpha \left(\sup_{\rho(s,t)<\delta} |X_\alpha(s) - X_\alpha(t)| > \varepsilon \right) < \eta.$$

Equivalentemente,

$$\lim_{\delta \downarrow 0} \limsup_\alpha \mathbb{P}^*_\alpha \left(\sup_{\rho(s,t)<\delta} |X_\alpha(s) - X_\alpha(t)| > \varepsilon \right) = 0, \qquad \varepsilon > 0.$$

Teorema 1.71. *Una red $X_\alpha : \Omega_\alpha \to \ell^\infty(T)$ es asintóticamente tensa si y sólo si:*

1. *para todo $t \in T$, $X_\alpha(t)$ es asintóticamente tensa en $\mathbb{R}$;*
2. *existe una semimétrica ρ en T tal que (T, ρ) es totalmente acotado y X_α es asintóticamente uniformemente ρ-equicontinua en probabilidad.*

Demostración. Si $T = \varnothing$, el espacio $\ell^\infty(T)$ tiene un solo elemento y el resultado es inmediato. Suponemos, en adelante, que $T \neq \varnothing$ y descartamos los bloques vacíos de todas las particiones.

Con la notación $\mathrm{osc}_\mathcal{P}$ de su demostración, usaremos el Teorema 1.68 en esta forma: la tensión asintótica equivale a la tensión asintótica de cada coordenada y a que, para todo $\varepsilon, \eta > 0$, exista una partición finita $\mathcal{P}$ tal que $\limsup_\alpha \mathbb{P}^*_\alpha(\mathrm{osc}_\mathcal{P}(X_\alpha) > \varepsilon) < \eta$.

Necesidad. Supongamos que X_α es asintóticamente tensa. Por el criterio anterior, $X_\alpha(t)$ es asintóticamente tensa en $\mathbb{R}$ para todo $t \in T$. Resta construir la semimétrica.

a) Elección de particiones cada vez más finas. Aplicando el criterio con $\varepsilon = \eta = 2^{-m}$, para cada $m \geq 1$ elegimos una partición finita $\mathcal{Q}_m$ tal que

$$\limsup_\alpha \mathbb{P}^*_\alpha \Big(\mathrm{osc}_{\mathcal{Q}_m}(X_\alpha) > 2^{-m} \Big) < 2^{-m}.$$

Estas particiones no tienen por qué estar encajadas. Para conseguirlo, definimos su refinamiento común acumulado:

$$\mathcal{P}_m := \Big\{ A_1 \cap \cdots \cap A_m : A_j \in \mathcal{Q}_j \ (1 \leq j \leq m), \quad A_1 \cap \cdots \cap A_m \neq \varnothing \Big\}.$$

Es una partición finita: cada punto pertenece a una única de esas intersecciones, y hay a lo sumo $\prod_{j=1}^m |\mathcal{Q}_j|$ intersecciones posibles. Además, $\mathcal{P}_{m+1}$ refina a $\mathcal{P}_m$, pues cada bloque de la primera se obtiene intersectando un bloque de la segunda con un bloque de $\mathcal{Q}_{m+1}$.

Cada bloque de $\mathcal{P}_m$ está contenido en un bloque de $\mathcal{Q}_m$. Por lo tanto, $\mathrm{osc}_{\mathcal{P}_m}(z) \leq \mathrm{osc}_{\mathcal{Q}_m}(z)$ para todo z, y obtenemos

$$\limsup_\alpha \mathbb{P}^*_\alpha \Big(\mathrm{osc}_{\mathcal{P}_m}(X_\alpha) > 2^{-m} \Big) < 2^{-m}. \tag{1.14}$$

b) Construcción de la semimétrica. Para $m \geq 1$ y $s, t \in T$, ponemos

$$d_m(s,t) := \begin{cases} 0, & \text{si } s,t \text{ pertenecen al mismo bloque de } \mathcal{P}_m, \\ 1, & \text{si pertenecen a bloques distintos,} \end{cases} \qquad \rho(s,t) := \sum_{m=1}^\infty 2^{-m} d_m(s,t).$$

La serie converge y $0 \leq \rho(s,t) \leq 1$, porque $0 \leq d_m \leq 1$. También es inmediato que $\rho(t,t) = 0$ y $\rho(s,t) = \rho(t,s)$.

Verifiquemos la desigualdad triangular. Fijados $s, t, u \in T$, si $d_m(s,t) = 0$, entonces $d_m(s,t) \leq d_m(s,u) + d_m(u,t)$ porque el segundo miembro es no negativo. Si $d_m(s,t) = 1$, los puntos s, t están en bloques distintos. En ese caso no pueden cumplirse simultáneamente $d_m(s,u) = 0$ y $d_m(u,t) = 0$, pues ello situaría a los tres puntos en un mismo bloque. Así, al menos uno de esos dos términos vale 1, y nuevamente

$$d_m(s,t) \leq d_m(s,u) + d_m(u,t).$$

Multiplicando por 2^{-m}, sumando de $m=1$ a $m=N$ y pasando al límite cuando $N\to\infty$, resulta $\rho(s,t)\le\rho(s,u)+\rho(u,t)$. Por consiguiente, ρ es una semimétrica. Puede ocurrir $\rho(s,t)=0$ con $s\neq t$: sucede si ninguna de las particiones separa esos puntos, lo cual está permitido para una semimétrica.

c) Relación entre distancias y bloques. Necesitamos las dos implicaciones siguientes:

$$s,t \text{ en un mismo bloque de } \mathcal{P}_m \quad\Longrightarrow\quad \rho(s,t)\le 2^{-m}, \tag{1.15}$$

$$\rho(s,t)<2^{-m} \quad\Longrightarrow\quad s,t \text{ en un mismo bloque de } \mathcal{P}_m. \tag{1.16}$$

Para la primera, si s,t están juntos en $\mathcal{P}_m$, también están juntos en $\mathcal{P}_j$ para todo $j\le m$, porque las particiones son cada vez más finas. Entonces $d_j(s,t)=0$ para $j\le m$ y

$$\rho(s,t)=\sum_{j>m}2^{-j}d_j(s,t)\le\sum_{j>m}2^{-j}=2^{-m}.$$

Para la segunda, si estuvieran en bloques distintos de $\mathcal{P}_m$, tendríamos $d_m(s,t)=1$ y, como todos los sumandos son no negativos, $\rho(s,t)\ge 2^{-m}$, contradiciendo la hipótesis.

d) Total acotación de (T,ρ). Fijemos $r>0$ y elijamos m con $2^{-m}<r$. Escribamos $\mathcal{P}_m=\{A_{m,1},\dots,A_{m,k_m}\}$ y elijamos un representante $t_{m,i}\in A_{m,i}$ en cada bloque. Si $t\in A_{m,i}$, la implicación (1.15) da $\rho(t,t_{m,i})\le 2^{-m}<r$. Por tanto, las k_m bolas abiertas de radio r centradas en esos representantes cubren T. Como $r>0$ era arbitrario, (T,ρ) es totalmente acotado.

e) Equicontinuidad asintótica. Fijemos $\varepsilon,\eta>0$ y elijamos m suficientemente grande para que $2^{-m}<\min\{\varepsilon,\eta\}$. Ponemos $\delta=2^{-m}$. Por (1.16), cualquier par con $\rho(s,t)<\delta$ está contenido en un bloque de $\mathcal{P}_m$. En consecuencia, para toda $z\in\ell^\infty(T)$,

$$\sup_{\rho(s,t)<\delta}|z(s)-z(t)|\le \operatorname{osc}_{\mathcal{P}_m}(z).$$

Esta es una desigualdad determinista, que podemos aplicar a $z=X_\alpha(\omega)$ para cada $\omega\in\Omega_\alpha$. Como $2^{-m}<\varepsilon$, obtenemos las inclusiones de eventos

$$\begin{aligned}\left\{\sup_{\rho(s,t)<\delta}|X_\alpha(s)-X_\alpha(t)|>\varepsilon\right\}&\subseteq\{\operatorname{osc}_{\mathcal{P}_m}(X_\alpha)>\varepsilon\}\\&\subseteq\left\{\operatorname{osc}_{\mathcal{P}_m}(X_\alpha)>2^{-m}\right\}.\end{aligned}$$

La monotonía de la probabilidad exterior y (1.14) implican

$$\begin{aligned}\limsup_\alpha\mathbb{P}_\alpha^*\left(\sup_{\rho(s,t)<\delta}|X_\alpha(s)-X_\alpha(t)|>\varepsilon\right)&\le\limsup_\alpha\mathbb{P}_\alpha^*\left(\operatorname{osc}_{\mathcal{P}_m}(X_\alpha)>2^{-m}\right)\\&<2^{-m}<\eta.\end{aligned}$$

Esto prueba la equicontinuidad asintótica requerida. Obsérvese que las particiones y ρ se construyeron antes de fijar ε,η: la misma semimétrica sirve para todos los pares de tolerancias; sólo cambia δ.

Suficiencia. Supongamos ahora que cada $X_\alpha(t)$ es asintóticamente tensa y que existe una semimétrica ρ con las propiedades del enunciado. Por el Teorema 1.68, basta verificar la condición de oscilación sobre particiones finitas.

Fijemos $\varepsilon,\eta>0$. Por la equicontinuidad asintótica, existe $\delta>0$ tal que

$$\limsup_\alpha\mathbb{P}_\alpha^*\left(\sup_{\rho(s,t)<\delta}|X_\alpha(s)-X_\alpha(t)|>\varepsilon\right)<\eta. \tag{1.17}$$

Como (T,ρ) es totalmente acotado, existen $t_1,\dots,t_k\in T$ cuyas bolas abiertas $B_i:=\{t\in T:\rho(t,t_i)<\delta/3\}$ cubren T. Para convertir esta cubierta en una partición, definimos

$$A_1:=B_1,\qquad A_i:=B_i\setminus\bigcup_{j<i}B_j\quad(2\le i\le k)$$

y descartamos los conjuntos vacíos. Los conjuntos restantes son disjuntos: cada punto se asigna únicamente a la primera bola de la lista que lo contiene. Además, cubren T, porque todo punto pertenece a alguna de las bolas. Denotemos por $\mathcal{P}$ esta partición finita.

Si s, t pertenecen a un mismo bloque A_i, ambos están en B_i. La desigualdad triangular da

$$\rho(s,t) \le \rho(s,t_i) + \rho(t_i,t) < 2\delta/3 < \delta.$$

Por consiguiente, todos los pares que intervienen en $\operatorname{osc}_{\mathcal{P}}(z)$ también intervienen en el supremo sobre $\rho(s,t) < \delta$, y para todo $z \in \ell^\infty(T)$ se tiene

$$\operatorname{osc}_{\mathcal{P}}(z) \le \sup_{\rho(s,t)<\delta} |z(s) - z(t)|.$$

Aplicando esta desigualdad a cada trayectoria, obtenemos

$$\{\operatorname{osc}_{\mathcal{P}}(X_\alpha) > \varepsilon\} \subseteq \left\{ \sup_{\rho(s,t)<\delta} |X_\alpha(s) - X_\alpha(t)| > \varepsilon \right\}.$$

Por monotonía de la probabilidad exterior y (1.17),

$$\limsup_\alpha \mathbb{P}^*_\alpha \left(\operatorname{osc}_{\mathcal{P}}(X_\alpha) > \varepsilon\right) < \eta.$$

Queda verificada la condición de particiones finitas para todo $\varepsilon, \eta > 0$. Junto con la tensión asintótica de las coordenadas, el Teorema 1.68 implica que X_α es asintóticamente tensa en $\ell^\infty(T)$.

En ambas direcciones sólo usamos inclusiones de eventos y monotonía de la probabilidad exterior; no se requiere suponer que esos eventos sean medibles. □

Teorema 1.72. *Sea $X_\alpha : \Omega_\alpha \to \ell^\infty(T)$ una red y sea $X : \Omega \to \ell^\infty(T)$ Borel-medible tal que $X_\alpha \Rightarrow X$.*

(i) Si X_α es asintóticamente tensa, existe una semimétrica ρ en T tal que

- *(a) (T, ρ) es totalmente acotado;*
- *(b) casi todas las trayectorias $t \mapsto X(t,\omega)$ son uniformemente ρ-continuas.*[41]

(ii) Si ρ es una semimétrica en T que satisface (a) *y* (b)*, entonces X_α es asintóticamente uniformemente ρ-equicontinua en probabilidad.*

En (ii) *no se supone que X_α sea asintóticamente tensa; sí se mantiene la hipótesis $X_\alpha \Rightarrow X$.*

Demostración. Si $T = \varnothing$, el resultado es inmediato. Supongamos $T \neq \varnothing$. Para una semimétrica ρ y $\delta > 0$, definimos

$$w_\delta(z) := \sup_{\rho(s,t)<\delta} |z(s) - z(t)|, \qquad z \in \ell^\infty(T).$$

La desigualdad triangular da $|w_\delta(z) - w_\delta(z')| \le 2\|z - z'\|_T$; por tanto, w_δ es continua y $w_\delta(X)$ es medible. Además, $w_\delta(z)$ decrece cuando $\delta \downarrow 0$.

(i) Por el Teorema 1.71, existe una semimétrica ρ que hace a T totalmente acotado y para la cual X_α es asintóticamente uniformemente equicontinua en probabilidad. Para cada $\varepsilon > 0$, el conjunto $\{z : w_\delta(z) > \varepsilon\}$ es abierto. Por Portmanteau (Teorema 1.34(ii)),

$$\mathbb{P}\Big(w_\delta(X) > \varepsilon\Big) \le \liminf_\alpha \mathbb{P}_{\alpha,*}\Big(w_\delta(X_\alpha) > \varepsilon\Big) \le \limsup_\alpha \mathbb{P}^*_\alpha\Big(w_\delta(X_\alpha) > \varepsilon\Big).$$

El último miembro tiende a cero cuando $\delta \downarrow 0$. Como $w_{1/n}(X)$ es no negativa y decreciente, tiene un límite medible $W \ge 0$. Para todo $k, n \ge 1$, $\mathbb{P}(W > 1/k) \le \mathbb{P}(w_{1/n}(X) > 1/k)$; haciendo $n \to \infty$

[41] Una función $z : T \to \mathbb{R}$ es *uniformemente ρ-continua* si para todo $\varepsilon > 0$ existe $\delta > 0$ tal que $|z(s) - z(t)| < \varepsilon$ siempre que $\rho(s,t) < \delta$; equivalentemente, $\sup_{\rho(s,t)<\delta} |z(s) - z(t)| \to 0$ cuando $\delta \downarrow 0$.

y luego tomando la unión numerable en k, resulta $\mathbb{P}(W > 0) = 0$. Así, fuera de un único conjunto medible nulo, $w_\delta(X) \to 0$, que es la continuidad uniforme requerida.

(ii) Por (b), $w_\delta(X) \to 0$ c.s. cuando $\delta \downarrow 0$. Para $\varepsilon > 0$, el conjunto $\{z : w_\delta(z) \geq \varepsilon\}$ es cerrado. Por Portmanteau (Teorema 1.34(iii)),

$$\begin{aligned}\limsup_\alpha \mathbb{P}_\alpha^*\Big(w_\delta(X_\alpha) > \varepsilon\Big) &\leq \limsup_\alpha \mathbb{P}_\alpha^*\Big(w_\delta(X_\alpha) \geq \varepsilon\Big) \\ &\leq \mathbb{P}\Big(w_\delta(X) \geq \varepsilon\Big).\end{aligned}$$

El último miembro tiende a cero por convergencia dominada aplicada a sus indicadores. Esto prueba la equicontinuidad asintótica. Esta implicación usa (b) y $X_\alpha \Rightarrow X$; no necesita adicionalmente la total acotación de (a). □

Equicontinuidad uniforme ρ-asintótica. La equicontinuidad uniforme ρ-asintótica es una condición delicada, pero es precisamente la que permite pasar de la convergencia de marginales finito-dimensionales a la convergencia débil en $\ell^\infty(T)$, una vez controlada la tensión. En aplicaciones concretas suele verificarse mediante propiedades de Markov, desigualdades maximales o argumentos de encadenamiento; en la Parte 2 desarrollaremos estas herramientas para procesos empíricos.

A nivel general, la elección de la semimétrica no es parte de los datos del problema. El conjunto de índices T puede no venir provisto de una semimétrica natural y, aun si la tiene, esa semimétrica puede no ser la adecuada para capturar las oscilaciones del proceso. El criterio anterior muestra que lo relevante es encontrar alguna semimétrica ρ que haga (T, ρ) totalmente acotado y para la cual los incrementos de X_α sean asintóticamente pequeños en probabilidad.

Si X es un posible límite de X_α, sus marginales finito-dimensionales sugieren la ley límite candidata. Para deducir la convergencia débil en $\ell^\infty(T)$ a partir de ellas, hay que verificar además la tensión asintótica de X_α, por ejemplo mediante el Teorema 1.71. La continuidad uniforme de las trayectorias del candidato X, por sí sola, no garantiza esa tensión: la parte (ii) del Teorema 1.72 supone ya que $X_\alpha \Rightarrow X$.

La siguiente proposición se refiere al propio proceso límite X: si alguna semimétrica hace a T totalmente acotado y a casi todas las trayectorias de X uniformemente continuas, entonces la semimétrica intrínseca ρ_0 tiene ambas propiedades. En ese caso, ρ_0 es una candidata natural para verificar la equicontinuidad asintótica de X_α, pero esa verificación sobre la red sigue siendo necesaria al aplicar el criterio de tensión. Que una semimétrica arbitraria no tenga las dos propiedades no descarta la existencia de otra que sí las tenga; para ρ_0, en cambio, la proposición garantiza que las tiene siempre que exista alguna que las cumpla.

Proposición 1.73 (Semimétrica intrínseca)**.** *Sea $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \ell^\infty(T)$ una aplicación Borel-medible. Supongamos que existe una semimétrica ρ en T tal que* (i) *(T, ρ) es totalmente acotado y* (ii) *casi todas las trayectorias $t \mapsto X(t, \omega)$ son uniformemente ρ-continuas. Definamos*

$$\rho_0(s, t) := \mathbb{E}\Big[\arctan|X(s) - X(t)|\Big], \qquad s, t \in T.$$

Entonces ρ_0 es una semimétrica finita en T, el espacio (T, ρ_0) es totalmente acotado y casi todas las trayectorias del propio proceso X son uniformemente ρ_0-continuas. Más precisamente, existe $\Omega_0 \in \mathcal{A}$, con $\mathbb{P}(\Omega_0) = 1$, tal que, para todo $\omega \in \Omega_0$ y todo $\varepsilon > 0$, existe $\delta > 0$ tal que $\rho_0(s, t) < \delta$ implica $|X(s, \omega) - X(t, \omega)| < \varepsilon$ para todos $s, t \in T$. Aquí δ puede depender de ω y de ε, pero no de s, t.

Demostración. Como las evaluaciones $z \mapsto z(t)$ son continuas en $\ell^\infty(T)$, cada $X(t)$ es una variable aleatoria, y como $0 \leq \arctan \leq \pi/2$ en $[0, \infty)$, ρ_0 está bien definida y toma valores en $[0, \pi/2]$. La simetría y la igualdad $\rho_0(t, t) = 0$ son inmediatas. Para la desigualdad triangular usamos que arctan es creciente y subaditiva en $[0, \infty)$. Lo segundo se debe a que es cóncava y se anula en el origen: si $a, b \geq 0$ y $a + b > 0$, la concavidad da $\arctan a \geq \frac{a}{a+b} \arctan(a + b)$ y $\arctan b \geq \frac{b}{a+b} \arctan(a + b)$, y basta sumar. Así,

$$\arctan|X(s) - X(t)| \leq \arctan|X(s) - X(u)| + \arctan|X(u) - X(t)|,$$

y tomando esperanzas se obtiene $\rho_0(s,t) \le \rho_0(s,u) + \rho_0(u,t)$.

Para ver que (T, ρ_0) es totalmente acotado, comparamos ρ_0 con ρ a través del módulo de continuidad. Para $z \in \ell^\infty(T)$ y $\delta > 0$ sea

$$W_\delta(z) := \sup_{\rho(s,t)\le\delta} |z(s) - z(t)| \le 2\|z\|_T.$$

Como $\big|\,|z(s)-z(t)| - |z'(s)-z'(t)|\,\big| \le 2\|z - z'\|_T$ para cada par s, t, la función W_δ es 2-lipschitziana en $\ell^\infty(T)$, de modo que $W_\delta(X)$ es una variable aleatoria aunque el supremo recorra una familia no numerable de pares. Si una trayectoria es uniformemente ρ-continua, su módulo W_δ tiende a cero cuando $\delta \downarrow 0$; por (ii), esto ocurre casi seguramente. Como W_δ crece con δ, basta considerar $\delta = 1/n$, y por convergencia dominada

$$\eta(\delta) := \mathbb{E}\big[\arctan W_\delta(X)\big] \longrightarrow 0 \qquad (\delta \downarrow 0).$$

Si $\rho(s,t) \le \delta$, entonces $|X(s) - X(t)| \le W_\delta(X)$ en todas las trayectorias, y por lo tanto $\rho_0(s,t) \le \eta(\delta)$. Dado $r > 0$, elegimos δ con $\eta(\delta) < r$ y, por (i), puntos $t_1, \dots, t_N$ tales que todo $t \in T$ cumple $\rho(t, t_i) < \delta$ para algún i; entonces las ρ_0-bolas de radio r centradas en los t_i cubren T.

Queda la continuidad uniforme de las trayectorias, que es la parte delicada. Que $\rho_0(s,t)$ sea pequeño sólo dice que $X(s)$ y $X(t)$ están próximos en probabilidad para ese par concreto, mientras que necesitamos un único suceso de probabilidad uno en el que el control valga para todos los pares a la vez. La idea es la siguiente: si f está en el soporte de la ley de X, entonces X está uniformemente cerca de f con probabilidad positiva, así que una oscilación grande de f entre s y t obliga a X a oscilar también con probabilidad positiva, lo que acota $\rho_0(s,t)$ inferiormente. Basta, pues, ver que el soporte tiene probabilidad uno, y para ello necesitamos que la ley de X viva en un conjunto separable.

Sea $\mu := \mathbb{P} \circ X^{-1}$. Para cada $n \ge 1$, disjuntando un cubrimiento finito de T por ρ-bolas de radio $1/n$ y descartando los conjuntos vacíos, obtenemos una partición finita $\{A_{n,1}, \dots, A_{n,k_n}\}$ de T en la que dos puntos de un mismo bloque distan menos de $2/n$. Sea $\mathcal{D}$ la familia, numerable, de las funciones que para algún n son constantes en cada $A_{n,i}$ con valores racionales. Si z es uniformemente ρ-continua y $\varepsilon > 0$, tomamos n tal que $|z(s) - z(t)| < \varepsilon/2$ siempre que $\rho(s,t) < 2/n$, un punto $t_{n,i}$ en cada bloque y racionales q_i con $|q_i - z(t_{n,i})| < \varepsilon/2$; la función que vale q_i en $A_{n,i}$ pertenece a $\mathcal{D}$ y dista a lo sumo ε de z. Por lo tanto, la clausura S de $\mathcal{D}$ en $\ell^\infty(T)$ es un cerrado separable que contiene a todas las funciones uniformemente ρ-continuas, y por (ii), $\mu(S) = 1$.

Sea ahora $(U_j)_{j\ge1}$ una base numerable de la topología relativa de S, por ejemplo las intersecciones con S de las bolas con centro en $\mathcal{D}$ y radio racional; cada U_j es boreliano. El conjunto $N := \bigcup\{U_j : \mu(U_j) = 0\}$ es abierto en S y μ-nulo, por ser unión numerable, así que $F := S \setminus N$ es cerrado en $\ell^\infty(T)$ y $\mu(F) = 1$. Además, si $f \in F$ y $r > 0$, existe j con $f \in U_j \subseteq S \cap B(f, r)$, y $\mu(U_j) > 0$ porque $f \notin N$. Es decir,

$$\mathbb{P}(\|X - f\|_T < r) > 0 \qquad \text{para todo } f \in F \text{ y todo } r > 0. \tag{1.18}$$

Fijemos $f \in F$ y $\varepsilon > 0$, y sea $A := \{\|X - f\|_T < \varepsilon/4\}$, cuya probabilidad p es positiva por (1.18). Si $|f(s) - f(t)| \ge \varepsilon$, en A se tiene

$$|X(s) - X(t)| \ge |f(s) - f(t)| - 2\|X - f\|_T > \varepsilon/2,$$

y en consecuencia

$$\rho_0(s,t) \ge \mathbb{E}\big[\mathbf{1}_A \arctan|X(s) - X(t)|\big] \ge p \arctan(\varepsilon/2) =: \delta.$$

Como $\delta > 0$ no depende de s ni de t, hemos probado que $\rho_0(s,t) < \delta$ implica $|f(s) - f(t)| < \varepsilon$, es decir, que toda $f \in F$ es uniformemente ρ_0-continua. Para terminar, tomamos $\Omega_0 := X^{-1}(F)$, que pertenece a $\mathcal{A}$ porque F es cerrado y X es Borel-medible, y cumple $\mathbb{P}(\Omega_0) = \mu(F) = 1$. Para cada $\omega \in \Omega_0$, la trayectoria $X(\omega)$ está en F y es, por tanto, uniformemente ρ_0-continua. $\square$

En particular, para cada $\omega \in \Omega_0$, $\rho_0(s,t) = 0$ implica $X(s,\omega) = X(t,\omega)$ simultáneamente para todos los pares $s, t \in T$. No hace falta cambiar de versión; si se quiere que *todas* las trayectorias sean uniformemente ρ_0-continuas, basta tomar $\widetilde{X} := X$ en Ω_0 y $\widetilde{X} := 0$ fuera de Ω_0, que es Borel-medible y coincide con X en un conjunto de probabilidad uno.

Otra familia natural de candidatas es

$$\rho_p(s,t) = \left(\mathbb{E}|X(s) - X(t)|^p\right)^{1/(p\vee 1)}, \qquad 0 < p < \infty.$$

Cuando estas cantidades son finitas, ρ_p es una semimétrica: para $p \geq 1$ esto es Minkowski, y para $0 < p < 1$ se usa $|a+b|^p \leq |a|^p + |b|^p$. Estas semimétricas son especialmente útiles cuando el proceso es continuo en media-p. Más precisamente, si para alguna semimétrica ρ que hace totalmente acotado a T se tiene

$$\lim_{\delta \downarrow 0} \sup_{\rho(s,t)<\delta} \mathbb{E}|X(s) - X(t)|^p = 0,$$

entonces los incrementos son uniformemente pequeños en media-p a escala ρ. En particular, para usar directamente ρ_p como semimétrica de equicontinuidad no basta con la continuidad de trayectorias: también hay que verificar que ρ_p sea finita, que (T, ρ_p) sea totalmente acotado y que las trayectorias de X sean uniformemente ρ_p-continuas. En el caso gaussiano, la elección usual es la semimétrica canónica $\rho_2(s,t)$, la raíz del segundo momento de los incrementos (que coincide con su desviación estándar cuando el proceso está centrado), y es la herramienta natural para aplicar métodos de entropía y encadenamiento.

Definición 1.74. Un proceso estocástico $X = \{X(t) : t \in T\}$ se llama **gaussiano** si, para todo $k \in \mathbb{N}$ y todo $t_1, \ldots, t_k \in T$, el vector $(X(t_1), \ldots, X(t_k))$ tiene distribución normal multivariante, posiblemente degenerada, en $\mathbb{R}^k$.

Proposición 1.75. *Sea $X = \{X(t) : t \in T\}$ un proceso gaussiano y sean ρ_p, $0 < p < \infty$, las semimétricas definidas más arriba. Sea $X_\alpha : \Omega_\alpha \to \ell^\infty(T)$ una red de elementos aleatorios. Entonces son equivalentes:*

1. *existe una versión $\widetilde{X}$ de X que es Borel medible como mapa en $\ell^\infty(T)$, cuya ley es tensa, y tal que $X_\alpha \Rightarrow \widetilde{X}$ en $\ell^\infty(T)$;*
2. *para algún $p \in (0,\infty)$, y entonces para todo $p \in (0,\infty)$, se cumplen:*
   - *para todo $k \in \mathbb{N}$ y todo $t_1, \ldots, t_k \in T$, $(X_\alpha(t_1), \ldots, X_\alpha(t_k)) \Rightarrow (X(t_1), \ldots, X(t_k))$;*
   - *X_α es asintóticamente uniformemente ρ_p-equicontinuo en probabilidad;*
   - *(T, ρ_p) es totalmente acotado.*

1.6. Convergencias

Para *redes* de aplicaciones definidas sobre un único espacio de probabilidad fijo $(\Omega, \mathcal{A}, \mathbb{P})$, la convergencia *casi segura* y *en probabilidad* son modos de convergencia estocástica usados con frecuencia, más fuertes que la convergencia débil. En esta sección consideramos sus extensiones a mapas *no medibles*, junto con el concepto de *convergencia casi uniforme*, el cual es equivalente a la *convergencia casi seguramente exterior* para *sucesiones*, pero es más fuerte (y más útil) para redes generales. En lo que sigue (D, d) es un espacio métrico.

Definición 1.79. Sean $X_\alpha, X : \Omega \to D$ aplicaciones arbitrarias.

(i) Decimos que X_α **converge en probabilidad exterior** a X, y escribimos $X_\alpha \xrightarrow{\mathbb{P}^*} X$, si $d(X_\alpha, X)^* \xrightarrow{\mathbb{P}} 0$. Equivalentemente,[42] para todo $\varepsilon > 0$,

$$\mathbb{P}\Big(d(X_\alpha, X)^* > \varepsilon\Big) = \mathbb{P}^*\Big(d(X_\alpha, X) > \varepsilon\Big) \longrightarrow 0,$$

[42] La igualdad $\mathbb{P}(d(X_\alpha, X)^* > \varepsilon) = \mathbb{P}^*(d(X_\alpha, X) > \varepsilon)$ es el Lema 1.9(vi); nótese que, como el evento de la izquierda es medible, la convergencia $d(X_\alpha, X)^* \xrightarrow{\mathbb{P}} 0$ es la usual.

(*ii*) X_α **converge casi uniformemente** a X si, para todo $\varepsilon > 0$, existe un conjunto medible A con $\mathbb{P}(A) \geq 1 - \varepsilon$ tal que $d(X_\alpha, X) \to 0$ *uniformemente* en A. Se denota $X_\alpha \xrightarrow{c.u.} X$.

(*iii*) X_α **converge casi seguramente exterior** a X si $d(X_\alpha, X)^* \to 0$ casi seguramente, para alguna elección de versiones de $d(X_\alpha, X)^*$ (cada versión es un mayorante medible mínimo; en particular, domina a $d(X_\alpha, X)$ en todo punto). Se denota $X_\alpha \xrightarrow{c.s.^*} X$.

(*iv*) X_α **converge casi seguramente** a X si

$$\mathbb{P}_*\Big(\lim_\alpha d(X_\alpha, X) = 0\Big) = 1,$$

y se denota $X_\alpha \xrightarrow{c.s.} X$.

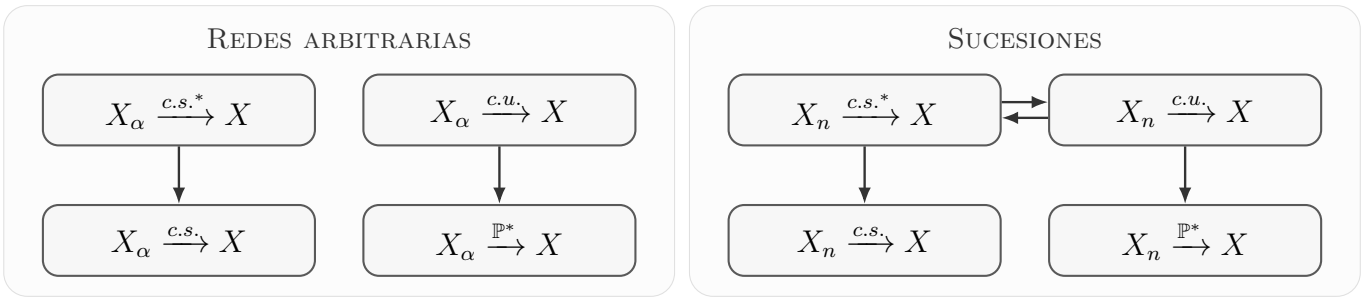


Figura 1.2: Relaciones básicas entre nociones de convergencia para mapas posiblemente no medibles. Sólo se dibujan implicancias válidas. En redes arbitrarias se muestran $c.s.^* \Rightarrow c.s.$ y $c.u. \Rightarrow \mathbb{P}^*$. En sucesiones, $c.s.^*$ y $c.u.$ son equivalentes, además $c.s.^* \Rightarrow c.s.$ y $c.u. \Rightarrow \mathbb{P}^*$. Las flechas ausentes no deben interpretarse automáticamente como no-implicancias; algunas relaciones adicionales pueden deducirse por transitividad.

Los tres primeros conceptos de la lista son los más importantes. El cuarto es delicado: en general no se comporta como uno esperaría. De hecho, para sucesiones no medibles, *incluso la convergencia en todas partes* (que es más fuerte que (*iv*)) no implica ninguna de las otras tres formas de convergencia. Éste es uno de esos fenómenos que recuerdan que la medibilidad, aunque a veces pase desapercibida, no debe tomarse a la ligera.

Para *sucesiones* (*ii*) y (*iii*) son equivalentes, pero para *redes* generales el (*iii*) pierde buena parte de su valor. Primero, está la molestia de que, para redes generales, la convergencia casi seguramente exterior depende de las versiones medibles mínimas que se usen. Segundo, incluso para redes medibles X_α, la convergencia casi seguramente exterior es más bien débil: por ejemplo, *no* implica convergencia en probabilidad.[43] Por estas razones consideramos (*iii*) y (*iv*) sólo para sucesiones. En lo que sigue, usaremos la notación X_α para una red general y X_n para una sucesión, sin mención adicional.

Para sucesiones, las cosas salen en parte como deberían: la convergencia casi seguramente exterior (*iii*) es más fuerte que la convergencia en probabilidad exterior. La equivalencia entre convergencia casi uniforme y convergencia casi seguramente exterior —la parte (iii) del lema siguiente— es el conocido *teorema de Egorov*.

Lema 1.81. *Sea (D, d) un espacio métrico, sea $X : \Omega \to D$ Borel–medible y sean $X_n : \Omega \to D$ aplicaciones arbitrarias. Entonces:*

(i) Si $X_n \xrightarrow{\text{c.s.}^*} X$, entonces $X_n \xrightarrow{\mathbb{P}^*} X$.

(ii) $X_n \xrightarrow{\mathbb{P}^*} X$ si y sólo si toda subsucesión admite una subsucesión ulterior que converge a X c.s.*.

[43] Un ejemplo estándar: en $[0,1]$ con la medida de Lebesgue, sea $X_F = \mathbb{1}_F$, donde F recorre los subconjuntos finitos de $[0,1]$, ordenados por inclusión. Para cada ω fijo se tiene eventualmente $\omega \in F$, de modo que $X_F \to 1$ en todo punto y, al ser cada X_F medible, también c.s.*; sin embargo, $\mathbb{P}(X_F > 1/2) = 0$ para todo F, así que $X_F \to 0$ en probabilidad.

(iii) $X_n \xrightarrow{\text{c.s.}^*} X$ *si y sólo si* $X_n \xrightarrow{\text{c.u.}} X$.

Lema 1.82. *Sea* (D, d) *un espacio métrico, sea* $X : \Omega \to D$ *Borel–medible y sea* (X_α) *una red de aplicaciones arbitrarias* $\Omega \to D$. *Entonces:*

(i) Definamos $R_\alpha := \sup_{\beta \geq \alpha} d(X_\beta, X)$, *el supremo puntual de la cola, que para redes no numerables puede no ser medible. Entonces* $X_\alpha \xrightarrow{\text{c.u.}} X$ *si y sólo si* $R_\alpha \xrightarrow{\mathbb{P}^*} 0$, *es decir, si y sólo si* $R_\alpha^* \xrightarrow{\mathbb{P}} 0$.

(ii) Si $X_\alpha \xrightarrow{\text{c.u.}} X$, *entonces* $X_\alpha \xrightarrow{\mathbb{P}^*} X$.

En el teorema clásico de representación casi segura —para aplicaciones medibles— las versiones $\widetilde{X_\alpha}$ se piden con la misma ley que las X_α originales. En la versión no medible se conserva más información: se exige

$$\widetilde{\mathbb{E}}^* f(\widetilde{X_\alpha}) = \mathbb{E}_\alpha^* f(X_\alpha) \qquad \text{para toda } f : D \to \mathbb{R} \text{ acotada.}$$

En particular, coinciden las probabilidades exteriores e interiores de $\{X_\alpha \in B\}$ y $\{\widetilde{X_\alpha} \in B\}$ para todo $B \subset D$. Esta igualdad de expectativas exteriores es, en general, más fuerte que la igualdad de las leyes sobre una σ-álgebra prefijada.

El siguiente teorema de representación no medible vale para sucesiones y, más generalmente, para redes indexadas por conjuntos dirigidos no triviales; no vale, en general, para conjuntos dirigidos arbitrarios. Llamaremos *no trivial* a un conjunto dirigido que permite una red de números estrictamente positivos δ_α con $\delta_\alpha \to 0$. En particular, $\mathbb{N}$, con el orden usual, es no trivial.[44]

Teorema 1.91. *Sea* $X_\alpha : \Omega_\alpha \to D$ *una red arbitraria indexada por un conjunto dirigido no trivial, y sea* X_∞ *Borel–medible y separable. Si* $X_\alpha \Rightarrow X_\infty$, *entonces existen un espacio de probabilidad* $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ *y aplicaciones* $\tilde{X}_\alpha : \tilde{\Omega} \to D$, *junto con una aplicación Borel–medible* $\tilde{X}_\infty : \tilde{\Omega} \to D$ *con la misma ley que* X_∞, *tales que:*

(i) $\tilde{X}_\alpha \xrightarrow{\text{c.u.}} \tilde{X}_\infty$;

(ii) $\tilde{\mathbb{E}}^* f(\tilde{X}_\alpha) = \mathbb{E}_\alpha^* f(X_\alpha)$, *para toda* $f : D \to \mathbb{R}$ *acotada y todo* α.

[44] La restricción al conjunto dirigido no trivial es esencial: existen conjuntos dirigidos triviales y, para cada uno de ellos, una red X_α para la cual el teorema falla; véase [5, Sección 1.10].

Capítulo 2

Proceso empírico

2.1. Medidas empíricas y clases de Donsker

Esta parte se ocupa de la convergencia de un tipo particular de aplicación aleatoria: el *proceso empírico*. En todo el capítulo, $X_1, X_2, \ldots$ es una sucesión de elementos aleatorios independientes e idénticamente distribuidos, definidos sobre un mismo espacio de probabilidad, con valores en un espacio medible $(\mathcal{X}, \mathcal{A})$ y ley común P. La *medida empírica* P_n de la muestra $X_1, \ldots, X_n$ es la medida aleatoria discreta dada por

$$P_n(C) = n^{-1}\#\{1 \le i \le n : X_i \in C\}, \qquad C \in \mathcal{A}.$$

Alternativamente (si los puntos son medibles), puede describirse como la medida aleatoria que pone masa $1/n$ en cada observación, $P_n = n^{-1}\sum_{i=1}^n \delta_{X_i}$.

Dada una colección $\mathcal{F}$ de funciones medibles $f : \mathcal{X} \to \mathbb{R}$, la medida empírica induce una aplicación $\mathcal{F} \to \mathbb{R}$ dada por $f \mapsto P_n f$. Usaremos la notación $Qf = \int f dQ$ para una función medible f y una medida signada Q. La versión centrada y reescalada de la aplicación anterior es el *$\mathcal{F}$-proceso empírico* $\mathbb{G}_n$ dado por

$$f \longmapsto \mathbb{G}_n f = \sqrt{n}(P_n - P)f = \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(f(X_i) - Pf\Big).$$

Frecuentemente identificaremos la medida signada

$$\mathbb{G}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n (\delta_{X_i} - P)$$

con el proceso empírico.

Para una función fija f, por la ley de los grandes números y el teorema central del límite se tiene $P_n f \xrightarrow[n\to\infty]{c.s.} Pf$, $\mathbb{G}_n f \Rightarrow \mathcal{N}(0, P(f - Pf)^2)$, siempre que Pf exista y $Pf^2 < \infty$, respectivamente. Veremos cómo hacer estas afirmaciones uniformes en f, al variar sobre una clase $\mathcal{F}$.

La teoría clásica de procesos empíricos considera, como casos especiales, cuando el espacio muestral $\mathcal{X}$ es el intervalo unitario $[0,1]$, la recta $\mathbb{R}$ o $\mathbb{R}^d$, y la colección de índices $\mathcal{F}$ es el conjunto de indicadores de semirrectas izquierdas en $\mathbb{R}$ o de ortantes inferior–izquierdos en $\mathbb{R}^d$. Trataremos medidas y procesos empíricos indexados por estas $\mathcal{F}$ y por muchas otras, incluyendo indicadores de semiespacios, bolas, elipsoides y otros subconjuntos de $\mathbb{R}^d$, clases de funciones suaves $\mathbb{R}^d \to \mathbb{R}$ y funciones monótonas.

Con la notación $\|Q\|_{\mathcal{F}} = \sup\{|Qf| : f \in \mathcal{F}\}$, la versión uniforme de la ley de los grandes números queda $\|P_n - P\|_{\mathcal{F}} \longrightarrow 0$, donde la convergencia es *casi segura exterior*: $\|P_n - P\|^*_{\mathcal{F}} \to 0$ casi seguramente. Una clase $\mathcal{F}$ para la cual esto es cierto se denomina *clase de Glivenko–Cantelli* (o *clase P-Glivenko–Cantelli* para resaltar la dependencia en P). Cuando sólo se tiene $\|P_n - P\|_{\mathcal{F}} \to 0$ en probabilidad exterior, diremos que la clase es Glivenko–Cantelli *débil*; bajo envolvente integrable, el argumento de submartingala inversa de la Sección 2.4 promueve la versión débil a la casi segura.

Para ver al proceso empírico como una aplicación a $\ell^\infty(\mathcal{F})$, necesitamos que sus trayectorias sean acotadas en $\mathcal{F}$. Una condición suficiente es $\sup_{f\in\mathcal{F}}|f(x)-Pf|<\infty$, $x\in\mathcal{X}$. Bajo esta condición, el proceso empírico $\{\mathbb{G}_n f:\ f\in\mathcal{F}\}$ puede verse como una aplicación a $\ell^\infty(\mathcal{F})$.[1]

En consecuencia, tiene sentido investigar condiciones bajo las cuales

$$\mathbb{G}_n=\sqrt{n}(P_n-P)\Rightarrow\mathbb{G},\qquad\text{en }\ell^\infty(\mathcal{F}),\tag{2.1.1}$$

donde el límite $\mathbb{G}$ es un elemento Borel-medible y tenso de $\ell^\infty(\mathcal{F})$. Una clase $\mathcal{F}$ para la cual esto es cierto se denomina *clase de Donsker* (o *P-Donsker*).

La naturaleza del proceso límite $\mathbb{G}$ se deduce de sus marginales. Si $f\in L_2(P)$, el teorema central del límite da $\mathbb{G}_n f\Rightarrow\mathcal{N}\big(0,P(f-Pf)^2\big)$; recíprocamente, la convergencia en distribución de $\mathbb{G}_n f$ obliga a que $f\in L_2(P)$, por el recíproco clásico del teorema central del límite.[2] En particular, si $\mathcal{F}$ es P-Donsker, cada $f\in\mathcal{F}$ pertenece a $L_2(P)$ y los marginales del proceso límite son gaussianos, con la covarianza que se indica a continuación. El TCL multivariado da que, para cualquier conjunto finito $f_1,\dots,f_k\subset\mathcal{F}$, $(\mathbb{G}_n f_1,\dots,\mathbb{G}_n f_k)\Rightarrow\mathcal{N}_k(0,\Sigma)$, donde la matriz $k\times k$ Σ tiene (i,j)-ésimo elemento $\Sigma_{ij}=P(f_i-Pf_i)(f_j-Pf_j)$. Como la convergencia en $\ell^\infty(\mathcal{F})$ implica la convergencia de las marginales, se sigue que el proceso límite $\{\mathbb{G}f:\ f\in\mathcal{F}\}$ debe ser gaussiano de media cero, con función de covarianza

$$\mathbb{E}\Big[\mathbb{G}f_1\mathbb{G}f_2\Big]=P(f_1-Pf_1)(f_2-Pf_2)=Pf_1f_2-(Pf_1)(Pf_2).\tag{2.1.2}$$

Por el Lema de Slutsky, toda clase P-Donsker es Glivenko–Cantelli débil: $\|P_n-P\|_{\mathcal{F}}\to0$ en probabilidad exterior.[3] De hecho, toda clase P-Donsker es P-Glivenko–Cantelli, con convergencia casi segura exterior. Esta conclusión más fuerte no se obtiene sólo con Slutsky: requiere el argumento de submartingala inversa del Lema 2.30, cuya hipótesis de envolvente integrable no forma parte de la definición de clase de Donsker, pero se cumple automáticamente: la propiedad Donsker implica que la envolvente de la clase centrada $\{f-Pf:f\in\mathcal{F}\}$ tiene segundo momento débil, $x^2\,\mathbb{P}^*\big(\sup_{f\in\mathcal{F}}|f-Pf|>x\big)\to0$, y en particular es integrable (véase [5, Lema 2.3.9]); como $\|P_n-P\|_{\mathcal{F}}$ no cambia al centrar la clase, el lema se aplica a la clase centrada. El recíproco no vale en general, aunque el conjunto de clases de Donsker contiene muchos ejemplos de interés. Varias clases son además *universalmente Donsker*, es decir, P-Donsker para toda medida de probabilidad P en el espacio muestral.

Ejemplo 2.1. Sean $X_1,\dots,X_n$ i.i.d. en $\mathbb{R}^d$ y sea $\mathcal{F}$ la colección de todos los indicadores de rectángulos inferiores $\{\mathbb{1}_{(-\infty,t]}:\ t\in\mathbb{R}^d\}$. La medida empírica indexada por $\mathcal{F}$ se identifica con la función de distribución empírica $t\longmapsto P_n\mathbb{1}_{(-\infty,t]}=\frac{1}{n}\sum_{i=1}^n\mathbb{1}\{X_i\le t\}$, donde $X_i\le t$ es coordenada a coordenada. En este caso es natural identificar $f=\mathbb{1}_{(-\infty,t]}$ con $t\in\mathbb{R}^d$ y $\ell^\infty(\mathcal{F})$ con $\ell^\infty(\mathbb{R}^d)$. Por supuesto, en el caso $d=1$ las trayectorias muestrales de la función de distribución empírica están contenidas en un espacio mucho más pequeño (como $\mathcal{D}[-\infty,\infty]$), pero esto no afecta al TCL funcional: si las aplicaciones toman valores en un subespacio con la métrica inducida y el límite es Borel–medible y tenso en ese subespacio, la convergencia débil puede comprobarse equivalentemente en el subespacio o en el espacio ambiente (véase [5, Sección 1.3]).

Se sabe por resultados clásicos (para la recta, debidos a Donsker) que la clase de rectángulos inferiores es de Donsker para cualquier ley subyacente P de $X_1,\dots,X_n$. Los resultados posteriores permiten recuperar este teorema y obtener conclusiones análogas para el proceso empírico indexado por bolas cerradas, rectángulos, semiespacios y elipsoides, así como por muchas colecciones de funciones, siempre bajo las condiciones de medibilidad correspondientes.

[1] En efecto, $|\mathbb{G}_n f|\le n^{-1/2}\sum_{i=1}^n|f(X_i)-Pf|\le\sqrt{n}\,\max_{i\le n}\sup_{f\in\mathcal{F}}|f(X_i)-Pf|$, y cada factor es finito por la condición, evaluada en $x=X_i(\omega)$; lo mismo vale para P_n-P.

[2] Si $Y_1,Y_2,\dots$ son i.i.d. y $n^{-1/2}\sum_{i=1}^n(Y_i-c)$ converge en distribución para alguna constante c, entonces $\mathbb{E}Y_1^2<\infty$ y el límite es normal; véase [5, Sección 2.1].

[3] En detalle: $\|P_n-P\|_{\mathcal{F}}=n^{-1/2}\|\mathbb{G}_n\|_{\mathcal{F}}$. La norma $z\mapsto\|z\|_{\mathcal{F}}$ es continua en $\ell^\infty(\mathcal{F})$, así que $\|\mathbb{G}_n\|_{\mathcal{F}}\Rightarrow\|\mathbb{G}\|_{\mathcal{F}}$ por el teorema de aplicación continua (Teorema 1.36); multiplicando por la sucesión determinista $n^{-1/2}\to0$ (Slutsky), $\|P_n-P\|_{\mathcal{F}}\Rightarrow0$, y la convergencia débil a una constante equivale a la convergencia en probabilidad exterior.

Ejemplo 2.2. Sea $\mathcal{C}$ una colección de conjuntos medibles en $(\mathcal{X}, \mathcal{A})$, y $\mathcal{F}$ igual al conjunto de funciones indicadoras de los conjuntos de $\mathcal{C}$. Esto conduce a la distribución empírica indexada por conjuntos $C \longmapsto P_n(C) = \frac{1}{n}\#\{X_i \in C\}$. En este caso conviene hacer la identificación $C \leftrightarrow \mathbb{1}_C$, tanto en notación como en terminología. Así, que $\mathcal{C}$ sea de Glivenko–Cantelli (débil) o de Donsker significa que lo es la clase $\mathcal{F}$ de sus indicadores, en el sentido definido al comienzo de esta sección; es sólo notación, y se escribe $\|P_n - P\|_{\mathcal{C}}$ en lugar de $\|P_n - P\|_{\mathcal{F}}$.

2.2. Desigualdades maximales y números de cubrimiento

En este capítulo derivamos una clase de desigualdades maximales que pueden usarse para establecer la equicontinuidad asintótica del proceso empírico. Dado que estas desigualdades tienen una aplicabilidad más amplia, dejamos temporalmente el marco específico de los procesos empíricos.

Definición 2.3. Sea $\psi : [0, \infty) \to [0, \infty)$ una función convexa (y por lo tanto continua [4]), no idénticamente nula, con $\psi(0) = 0$[5]. Para una variable aleatoria X, la **norma de Orlicz** $\|X\|_\psi$ se define como

$$\|X\|_\psi := \text{ínf}\left\{C > 0 : \mathbb{E}\,\psi\left(\frac{|X|}{C}\right) \leq 1\right\}.$$

Aquí el ínfimo sobre el conjunto vacío se interpreta como ∞.

Denotamos por L_ψ al conjunto de variables aleatorias X con $\|X\|_\psi < \infty$, identificando, como es usual, variables aleatorias que coinciden casi seguramente. El siguiente lema justifica el nombre de *norma*.

Lema 2.4. *Sea ψ como en la definición anterior. Para todas las variables aleatorias X, Y y todo $\lambda \in \mathbb{R}$ se cumple:*

(i) $\|X\|_\psi = 0$ *si y sólo si* $X = 0$ *c.s.;*

(ii) $\|\lambda X\|_\psi = |\lambda|\,\|X\|_\psi$ *(con la convención* $0 \cdot \infty = 0$*);*

(iii) $\|X + Y\|_\psi \leq \|X\|_\psi + \|Y\|_\psi$.

En particular, L_ψ es un espacio vectorial y $\|\cdot\|_\psi$ es una norma sobre L_ψ.

Demostración. Para cada variable aleatoria X llamemos

$$S_X := \left\{C > 0 : \mathbb{E}\,\psi\left(\frac{|X|}{C}\right) \leq 1\right\},$$

de modo que $\|X\|_\psi = \text{ínf}\, S_X$. La esperanza $\mathbb{E}\,\psi(|X|/C) \in [0, \infty]$ está bien definida porque $\psi \geq 0$ es monótona (y por lo tanto boreliana). Hacemos primero dos observaciones.

Si $C \in S_X$ y $C' \geq C$, entonces $|X|/C' \leq |X|/C$ y, como ψ es no decreciente, $\mathbb{E}\,\psi(|X|/C') \leq \mathbb{E}\,\psi(|X|/C) \leq 1$, es decir, $C' \in S_X$. En consecuencia, *todo $C > \|X\|_\psi$ pertenece a S_X*: por definición de ínfimo existe $C_0 \in S_X$ con $C_0 < C$.

Como $\psi \not\equiv 0$, existe s_0 con $\psi(s_0) > 0$, y necesariamente $s_0 > 0$. Para $t \geq s_0$, escribiendo $s_0 = \frac{s_0}{t}\,t + \left(1 - \frac{s_0}{t}\right)0$ y usando la convexidad y $\psi(0) = 0$,

$$\psi(s_0) \leq \frac{s_0}{t}\,\psi(t), \qquad \text{es decir,} \qquad \psi(t) \geq \frac{\psi(s_0)}{s_0}\,t \xrightarrow[t\to\infty]{} \infty.$$

[4] Sea f convexa en un intervalo abierto I y sea $s(x, y) = \frac{f(y)-f(x)}{y-x}$ la pendiente de la cuerda. Si $u < v < w$ están en I, escribiendo $v = tu + (1-t)w$ con $t = \frac{w-v}{w-u}$, la convexidad da $s(u, v) \leq s(u, w) \leq s(v, w)$. Dado $x_0 \in I$, tomamos $a < x_0 < b$ en I; aplicando lo anterior a los puntos a, x_0, x, b ordenados, resulta $s(a, x_0) \leq s(x_0, x) \leq s(x_0, b)$ para todo $x \in (a, b)$, $x \neq x_0$. Luego $|f(x) - f(x_0)| \leq M|x - x_0|$ con $M = \text{máx}\{|s(a, x_0)|, |s(x_0, b)|\}$, y f es continua en x_0. (En el borde puede fallar: $\mathbf{1}_{\{0\}}$ es convexa en $[0, 1]$ y no es continua en 0.)

[5] por ser convexa, tomar valores en $[0, \infty)$ y cumplir $\Psi(0) = 0$ se sigue que es no decreciente

(i) Si $X = 0$ c.s., entonces $\psi(|X|/C) = \psi(0) = 0$ c.s. para todo $C > 0$, así que $S_X = (0, \infty)$ y $\|X\|_\psi = 0$. Recíprocamente, supongamos $\|X\|_\psi = 0$ y, por el absurdo, que $p := \mathbb{P}(|X| > \delta) > 0$ para algún $\delta > 0$. Todo $C > 0$ está en S_X, y como $\psi \geq 0$ es no decreciente,

$$1 \geq \mathbb{E}\,\psi\left(\frac{|X|}{C}\right) \geq \mathbb{E}\left[\psi\left(\frac{|X|}{C}\right)\mathbf{1}_{\{|X|>\delta\}}\right] \geq \psi\left(\frac{\delta}{C}\right)p \qquad \text{para todo } C > 0.$$

$\psi(\delta/C)\,p \to \infty$ cuando $C \downarrow 0$, lo cual es absurdo. Luego $\mathbb{P}(|X| > \delta) = 0$ para todo $\delta > 0$, y tomando $\delta = 1/n$ con $n \to \infty$ resulta $X = 0$ c.s.

(ii) Homogeneidad. El caso $\lambda = 0$ está cubierto por (i). Si $\lambda \neq 0$, para $C > 0$ se tiene

$$\mathbb{E}\,\psi\left(\frac{|\lambda X|}{C}\right) = \mathbb{E}\,\psi\left(\frac{|X|}{C/|\lambda|}\right),$$

de donde $C \in S_{\lambda X}$ si y sólo si $C/|\lambda| \in S_X$; es decir, $S_{\lambda X} = |\lambda|\,S_X$. Tomando ínfimos, $\|\lambda X\|_\psi = |\lambda|\,\|X\|_\psi$ (si $S_X = \varnothing$, también $S_{\lambda X} = \varnothing$ y ambos lados valen ∞).

(iii) Desigualdad triangular. Si $\|X\|_\psi = \infty$ o $\|Y\|_\psi = \infty$ no hay nada que probar. Si no, sean $C_1 > \|X\|_\psi$ y $C_2 > \|Y\|_\psi$; $C_1 \in S_X$ y $C_2 \in S_Y$. Pongamos $\theta := \frac{C_1}{C_1+C_2} \in (0,1)$. Como $|X+Y| \leq |X| + |Y|$ y ψ es no decreciente y convexa,

$$\psi\left(\frac{|X+Y|}{C_1+C_2}\right) \leq \psi\left(\theta\,\frac{|X|}{C_1} + (1-\theta)\,\frac{|Y|}{C_2}\right) \leq \theta\,\psi\left(\frac{|X|}{C_1}\right) + (1-\theta)\,\psi\left(\frac{|Y|}{C_2}\right).$$

Tomando esperanza,

$$\mathbb{E}\,\psi\left(\frac{|X+Y|}{C_1+C_2}\right) \leq \theta \cdot 1 + (1-\theta)\cdot 1 = 1,$$

o sea $C_1 + C_2 \in S_{X+Y}$ y $\|X+Y\|_\psi \leq C_1 + C_2$. Haciendo $C_1 \downarrow \|X\|_\psi$ y $C_2 \downarrow \|Y\|_\psi$ se obtiene (iii).

Finalmente, por (ii) y (iii), si $X, Y \in L_\psi$ y $\lambda \in \mathbb{R}$, entonces $\lambda X + Y \in L_\psi$, así que L_ψ es un espacio vectorial. Además, si $X = X'$ c.s., entonces $\psi(|X|/C) = \psi(|X'|/C)$ c.s. para todo $C > 0$, luego $S_X = S_{X'}$ y $\|X\|_\psi = \|X'\|_\psi$. Así, $\|\cdot\|_\psi$ está bien definida sobre las clases de equivalencia y, por (i)–(iii), es una norma sobre L_ψ. □

Ejemplo 2.5. ▪ $\psi(x) = x^p$, con $p \geq 1$. En este caso, la norma de Orlicz coincide con $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$.

▪ $\psi_p(x) = e^{x^p} - 1$, $p \geq 1$. Como $x^p \leq \psi_p(x)$ para todo $x \geq 0$, se tiene $\|X\|_p \leq \|X\|_{\psi_p}$.

Observación 2.6. Si $p \leq q$, entonces $\|X\|_{\psi_p} \leq (\log 2)^{1/q-1/p}\,\|X\|_{\psi_q}$. En particular, normas ψ_q con q más grande controlan colas más livianas. También se tiene, para $r \geq 1$, $\|X\|_r \leq Cr\,\|X\|_{\psi_1}$, donde C es una constante universal. Más generalmente, para $\alpha \geq 1$, $\|X\|_r \leq C_\alpha r^{1/\alpha}\|X\|_{\psi_\alpha}$.

Veamos que conocer la norma de Orlicz permite acotar la cola de la distribución. En efecto, si $C > \|X\|_\psi$, entonces, por la desigualdad de Markov,

$$\mathbb{P}(|X| > x) \leq \mathbb{P}\left\{\psi\left(\frac{|X|}{C}\right) \geq \psi\left(\frac{x}{C}\right)\right\} \leq \frac{1}{\psi(x/C)}.$$

Haciendo $C \downarrow \|X\|_\psi$, se obtiene $\mathbb{P}(|X| > x) \leq (\psi(x/\|X\|_\psi))^{-1}$, $x > 0$.[6] En particular, para $\psi_p(x) = e^{x^p} - 1$, se obtiene la cota

$$\mathbb{P}(|X| > x) \leq 2\exp\left\{-\left(\frac{x}{\|X\|_{\psi_p}}\right)^p\right\}.$$

Recíprocamente, una cota exponencial de este tipo implica que la norma ψ_p es finita.

[6] El paso al límite está justificado porque ψ, al ser convexa y finita, es continua en $[0,\infty)$. Si $\psi(x/\|X\|_\psi) = 0$, la cota vale trivialmente con la convención $1/0 = \infty$; si $\|X\|_\psi = 0$, entonces $X = 0$ c.s. y también es trivial.

Advertencia 2.7

Los Lemas 2.8 y 2.10 son ejercicios guiados del práctico 4; sólo demostraremos el Lema 2.9.

Lema 2.8. *Sea X una variable aleatoria tal que, para constantes $K, C > 0$ y algún $p \geq 1$,*

$$\mathbb{P}(|X| > x) \leq K e^{-Cx^p}, \qquad x > 0.$$

Entonces $\|X\|_{\psi_p} \leq \left(\frac{1+K}{C}\right)^{1/p}$.

A continuación consideramos la norma de Orlicz del máximo de un número finito de variables aleatorias. Para las normas L_p, usando $\max_i |X_i|^p \leq \sum_i |X_i|^p$, se obtiene

$$\begin{aligned}
\left\| \max_{1\leq i\leq m} |X_i| \right\|_p &= \left(\mathbb{E}\left[\left(\max_{1\leq i\leq m} |X_i| \right)^p \right] \right)^{1/p} && \text{(definición de } \|\cdot\|_p) \\
&= \left(\mathbb{E} \max_{1\leq i\leq m} |X_i|^p \right)^{1/p} && (t \mapsto t^p \text{ es creciente en } [0,\infty)) \\
&\leq \left(\mathbb{E} \sum_{i=1}^m |X_i|^p \right)^{1/p} && (\max_i a_i \leq \textstyle\sum_i a_i \text{ si } a_i \geq 0) \\
&= \left(\sum_{i=1}^m \mathbb{E}|X_i|^p \right)^{1/p} && \text{(linealidad de la esperanza)} \\
&\leq \left(m \max_{1\leq i\leq m} \mathbb{E}|X_i|^p \right)^{1/p} && \text{(cada sumando es } \leq \text{ el máximo)} \\
&= m^{1/p} \left(\max_{1\leq i\leq m} \|X_i\|_p^p \right)^{1/p} && (\mathbb{E}|X_i|^p = \|X_i\|_p^p) \\
&= m^{1/p} \max_{1\leq i\leq m} \|X_i\|_p && (t \mapsto t^{1/p} \text{ es creciente en } [0,\infty)).
\end{aligned}$$

Esta desigualdad elemental tiene constante 1. Una desigualdad análoga, en general con una constante multiplicativa adicional, es válida para muchas normas de Orlicz. En el caso general, el factor $m^{1/p}$ se reemplaza por $\psi^{-1}(m)$, donde ψ^{-1} denota la inversa generalizada de ψ.

Lema 2.9. *Sea $\psi : [0,\infty) \to [0,\infty)$ convexa, no idénticamente nula, con $\psi(0) = 0$. Supongamos que existe $c > 0$ tal que*

$$\limsup_{x,y\to\infty} \frac{\psi(x)\psi(y)}{\psi(cxy)} < \infty.$$

Entonces existe una constante K_ψ, que depende sólo de ψ y de c, tal que, para cualesquiera variables aleatorias $X_1, \ldots, X_m$, $\|\max_{1\leq i\leq m} |X_i|\|_\psi \leq K_\psi\, \psi^{-1}(m) \max_{1\leq i\leq m} \|X_i\|_\psi$.

Demostración. La idea es la siguiente. Si $\mathbb{E}\,\psi(|X_i|) \leq 1$ para todo i, como ψ es no decreciente se tiene $\psi(\max_i |X_i|) = \max_i \psi(|X_i|) \leq \sum_i \psi(|X_i|)$, y entonces $\mathbb{E}\,\psi(\max_i |X_i|) \leq m$. Para volver a una esperanza a lo sumo 1 hay que dividir el argumento de ψ por un número v que divida su valor por m. La hipótesis sobre el límsup proporciona, para argumentos grandes, una cota de la forma $\psi(u)\psi(v) \leq A\psi(cuv)$. Al tomar $u = t/(cv)$, se obtiene $\psi(t/(cv)) \leq A\psi(t)/\psi(v)$; eligiendo v proporcional a $\psi^{-1}(m)$, se divide por m, salvo constantes. Los valores chicos, donde la hipótesis no dice nada, se controlan con la monotonía de ψ.

Empezamos con algunas propiedades elementales de ψ. Si $\lambda \geq 1$ y $x \geq 0$, escribiendo $x/\lambda = \frac{1}{\lambda}\, x + \left(1 - \frac{1}{\lambda}\right) \cdot 0$, la convexidad y $\psi(0) = 0$ dan

$$\psi\left(\frac{x}{\lambda}\right) \leq \frac{\psi(x)}{\lambda}, \qquad \text{o, equivalentemente,} \qquad \psi(\lambda x) \geq \lambda\, \psi(x). \tag{2.1}$$

En particular $0 \leq \psi(t) \leq t\,\psi(1)$ para $t \in (0,1]$, de modo que ψ es continua en 0; en $(0,\infty)$ lo es por ser convexa y finita. Además $\psi(x) \to \infty$ cuando $x \to \infty$, como se vio en la demostración del

Lema 2.4. Para $t > 0$ definimos $\psi^{-1}(t) = \inf\{x \geq 0 : \psi(x) \geq t\}$. Por lo anterior, ese conjunto es el intervalo cerrado $[\psi^{-1}(t), \infty)$, así que $\psi(\psi^{-1}(t)) \geq t$; en particular $\psi^{-1}(t) > 0$, y $t \mapsto \psi^{-1}(t)$ es no decreciente. Por último, para cada variable aleatoria X y cada $r > 0$,

$$\mathbb{E}\,\psi(|X|/r) \leq 1 \implies \|X\|_\psi \leq r, \qquad r > \|X\|_\psi \implies \mathbb{E}\,\psi(|X|/r) \leq 1. \tag{2.2}$$

La primera implicación es la definición de la norma, y la segunda es la observación del comienzo de la demostración del Lema 2.4.

Pasamos a la hipótesis. Como el lím sup es finito, existen $A \geq 1$ y $x_0 > 0$, que dependen sólo de ψ y de c, tales que

$$\psi(u)\,\psi(v) \leq A\,\psi(cuv) \qquad \text{para todos } u, v \geq x_0$$

(agrandando A si hace falta). Si $v \geq x_0$, $\psi(v) > 0$ y $t \geq c\,x_0\,v$, entonces $u = t/(cv) \geq x_0$, y la desigualdad anterior con $cuv = t$ da

$$\psi\Big(\frac{t}{cv}\Big) \leq A\,\frac{\psi(t)}{\psi(v)}. \tag{2.3}$$

Es decir, dividir el argumento por cv divide el valor de ψ por $\psi(v)/A$, siempre que el argumento resultante no sea demasiado chico.

Vamos ahora a la desigualdad del lema. Si algún $\|X_i\|_\psi$ es infinito no hay nada que probar, porque $\psi^{-1}(m) > 0$ y el lado derecho es infinito. En otro caso, sea $M > \max_i \|X_i\|_\psi$. Por (2.2), $\mathbb{E}\,\psi(|X_i|/M) \leq 1$ para todo i, y vamos a probar que

$$\Big\| \max_{1 \leq i \leq m} |X_i| \Big\|_\psi \leq K_\psi\,\psi^{-1}(m)\,M; \tag{2.4}$$

el lema se obtiene haciendo $M \downarrow \max_i \|X_i\|_\psi$.

Sea $v \geq x_0$ con $\psi(v) > 0$. Para cada i,

$$\psi\Big(\frac{|X_i|}{cvM}\Big) \leq \frac{A}{\psi(v)}\,\psi\Big(\frac{|X_i|}{M}\Big) + \psi(x_0).$$

En efecto, si $|X_i| \geq c\,x_0\,v\,M$, esto es (2.3) con $t = |X_i|/M$; si $|X_i| < c\,x_0\,v\,M$, el argumento de la izquierda es menor que x_0 y basta la monotonía de ψ. Como ψ del máximo es el máximo de los valores de ψ, y un máximo de términos no negativos está acotado por su suma,

$$\psi\Big(\frac{\max_i |X_i|}{cvM}\Big) \leq \frac{A}{\psi(v)} \sum_{i=1}^{m} \psi\Big(\frac{|X_i|}{M}\Big) + \psi(x_0),$$

y tomando esperanza,

$$\mathbb{E}\,\psi\Big(\frac{\max_i |X_i|}{cvM}\Big) \leq \frac{A\,m}{\psi(v)} + \psi(x_0). \tag{2.5}$$

Elegimos v de modo que el primer término sea a lo sumo 1 y que además $v \geq x_0$:

$$v = \lambda_0\,\psi^{-1}(m), \qquad \lambda_0 = \max\Big\{A,\ \frac{x_0}{\psi^{-1}(1)}\Big\} \geq 1.$$

Como $m \geq 1$ y ψ^{-1} es no decreciente, $v \geq x_0\,\psi^{-1}(m)/\psi^{-1}(1) \geq x_0$; el segundo término del máximo está sólo para esto. Por (2.1), $\psi(v) \geq \lambda_0\,\psi\big(\psi^{-1}(m)\big) \geq A\,m$, y en particular $\psi(v) > 0$. Llevando esto a (2.5),

$$\mathbb{E}\,\psi\Big(\frac{\max_i |X_i|}{cvM}\Big) \leq 1 + \psi(x_0).$$

Falta bajar esta cota a 1, y para eso agrandamos el denominador. Aplicando (2.1) punto a punto con $\lambda = 1 + \psi(x_0)$,

$$\mathbb{E}\,\psi\Big(\frac{\max_i |X_i|}{(1+\psi(x_0))\,cvM}\Big) \leq \frac{1}{1+\psi(x_0)}\,\mathbb{E}\,\psi\Big(\frac{\max_i |X_i|}{cvM}\Big) \leq 1.$$

Por (2.2), $\|\max_i |X_i|\|_\psi \le (1+\psi(x_0))\, c\, v\, M$, que es (2.4) con

$$K_\psi = c\left(1+\psi(x_0)\right)\lambda_0 = c\left(1+\psi(x_0)\right)\max\left\{A,\ \frac{x_0}{\psi^{-1}(1)}\right\}.$$

Esta constante depende sólo de ψ y de c, a través de A y x_0. □

El siguiente resultado es una versión útil para variables con colas de tipo Bernstein, que combinan un comportamiento sub-gaussiano para desviaciones pequeñas y sub-exponencial para desviaciones grandes.

Lema 2.10. *Sean $X_1, \dots, X_m$ variables aleatorias tales que, para ciertos $a, b > 0$,*

$$\mathbb{P}(|X_i| > x) \le 2\exp\left\{-\frac{x^2}{a+bx}\right\}, \qquad x > 0, \quad i = 1, \dots, m.$$

Entonces

$$\left\|\max_{1\le i\le m} |X_i|\right\|_{\psi_1} \le K\left(\sqrt{a}\,\sqrt{\log(1+m)} + b\log(1+m)\right),$$

donde K es una constante universal.

Para los propósitos actuales, el valor de las constantes en los lemas anteriores es irrelevante. La conclusión importante es que el inverso de la función ψ determina el tamaño de la norma ψ de un máximo en comparación con las normas ψ de los términos individuales. La norma ψ del máximo crece más lentamente cuanto más rápido crece ψ. Para las funciones ψ_p del Ejemplo 2.5, este crecimiento es de orden logarítmico, porque $\psi_p^{-1}(m) = \left(\log(1+m)\right)^{1/p}$.

El lema anterior es inútil en el caso de un máximo sobre un número infinito de variables. Sin embargo, tal caso puede manejarse mediante aplicaciones repetidas del lema, en un método conocido como *chaining* que veremos en la prueba del próximo teorema. Cada variable aleatoria en el supremo se escribe como una suma de "pequeños eslabones". Para un proceso estocástico $\{X_t : t \in T\}$, el número de eslabones depende de la entropía del conjunto índice respecto de la semimétrica $d(s,t) = \|X_s - X_t\|_\psi$.

La definición general de entropía métrica es la siguiente.

Definición 2.11. Sea (T, d) un espacio semimétrico arbitrario. El **número de cubrimiento** $N(\varepsilon, d)$ es el número mínimo de bolas de radio ε necesarias para cubrir T.

Una colección de puntos se llama ε-**separada** si la distancia entre cada par de puntos distintos es estrictamente mayor que ε. El **número de empaquetamiento** $D(\varepsilon, d)$ es el número máximo de puntos ε-separados en T.

Los correspondientes **números de entropía** son los logaritmos de los números de cubrimiento y de empaquetamiento, respectivamente.

Para los propósitos actuales, pueden usarse tanto números de cubrimiento como de empaquetamiento. En los argumentos, uno puede ser reemplazado por el otro mediante las desigualdades

$$N(\varepsilon, d) \le D(\varepsilon, d) \le N(\varepsilon/2, d).^{7} \tag{2.6}$$

Claramente, los números de cubrimiento y de empaquetamiento aumentan cuando $\varepsilon \downarrow 0$. Por definición, el conjunto semimétrico T está totalmente acotado si y sólo si los números de cubrimiento

[7] Para la segunda desigualdad, sea $\{x_1, \dots, x_m\} \subset T$ un conjunto ε-separado, es decir, con $d(x_i, x_j) > \varepsilon$ para $i \ne j$, y sean $B(y_1, \varepsilon/2), \dots, B(y_n, \varepsilon/2)$ bolas cerradas que cubren T, con $n = N(\varepsilon/2, d)$ (si $N(\varepsilon/2, d) = \infty$ no hay nada que probar). Cada x_i pertenece a alguna de estas bolas, y dos puntos distintos x_i, x_j no pueden estar en una misma bola $B(y_k, \varepsilon/2)$, pues en tal caso $d(x_i, x_j) \le d(x_i, y_k) + d(y_k, x_j) \le \varepsilon/2 + \varepsilon/2 = \varepsilon$, lo que contradice la separación. Por lo tanto, asignar a cada x_i una bola que lo contenga define una función inyectiva de $\{x_1, \dots, x_m\}$ en $\{1, \dots, n\}$, de donde $m \le n$. Tomando el supremo sobre todos los conjuntos ε-separados se obtiene $D(\varepsilon, d) \le N(\varepsilon/2, d)$.

son finitos para todo $\varepsilon > 0$, equivalentemente si lo son los números de empaquetamiento. La cota superior en la siguiente desigualdad maximal depende de la rapidez con que $D(\varepsilon, d)$ crece al decrecer ε, medida mediante un criterio integral.

Para evitar ambigüedades, diremos que un proceso $\{X_t : t \in T\}$ es *separable respecto de* d si existen un conjunto numerable $T' \subset T$ y un conjunto medible nulo N tales que, para todo $\omega \notin N$ y todo $\delta > 0$,

$$\sup_{d(s,t)\le\delta} |X_s(\omega) - X_t(\omega)| = \sup_{\substack{s,t\in T'\\ d(s,t)\le\delta}} |X_s(\omega) - X_t(\omega)|.$$

Así, el mismo conjunto excepcional sirve simultáneamente para todos los radios. Esta es la propiedad de separabilidad usada en los argumentos de máximo y encadenamiento que siguen.

Teorema 2.12. *Sea ψ una función convexa, no nula, con $\psi(0) = 0$, tal que $\limsup_{x,y\to\infty} \frac{\psi(x)\psi(y)}{\psi(cxy)} < \infty$ para alguna constante $c > 0$. Sea $\{X_t : t \in T\}$ un proceso estocástico, indexado por un espacio semimétrico (T, d) y separable respecto de d, tal que*

$$\|X_s - X_t\|_\psi \le L\, d(s,t), \qquad s,t \in T,$$

para alguna constante $L < \infty$. Entonces, para todo $\eta, \delta > 0$,

$$\left\| \sup_{d(s,t)\le\delta} |X_s - X_t| \right\|_\psi \le 16 K_\psi L \left\{ \int_0^\eta \psi^{-1}\Big(D(\varepsilon,d)\Big)\, d\varepsilon + \delta\, \psi^{-1}\Big(D^2(\eta,d)\Big) \right\},$$

donde K_ψ es la constante del Lema 2.9.

Corolario 2.13. *En las condiciones del Teorema 2.12,*

$$\left\| \sup_{s,t\in T} |X_s - X_t| \right\|_\psi \le 32 K_\psi L \int_0^{\operatorname{diam} T} \psi^{-1}\Big(D(\varepsilon,d)\Big)\, d\varepsilon,$$

donde $\operatorname{diam} T = \sup_{s,t\in T} d(s,t)$*, y el lado derecho se interpreta como $+\infty$ si la integral diverge.*

Demostración del Teorema 2.12. Primero probamos la desigualdad para el máximo sobre un conjunto finito de índices, con una constante que no depende del conjunto, y después pasamos al caso general por convergencia monótona y separabilidad.

Sea $\varepsilon_j = \eta 2^{-j}$ para $j \ge 0$. Podemos suponer que $\int_0^\eta \psi^{-1}\big(D(\varepsilon,d)\big)\, d\varepsilon < \infty$, ya que en otro caso no hay nada que probar. Como $D(\varepsilon, d)$ es no creciente en ε, esto implica $D(\varepsilon, d) < \infty$ para todo $\varepsilon > 0$. La hipótesis sobre el lím sup interviene sólo a través del Lema 2.9, que usaremos en la forma siguiente: si $\|Y_i\|_\psi \le r$ para $i = 1, \dots, m$, entonces

$$\left\| \max_{1\le i\le m} |Y_i| \right\|_\psi \le K_\psi\, r\, \psi^{-1}(m),$$

con K_ψ la constante del lema. Como ψ^{-1} es no decreciente, la desigualdad sigue valiendo si se reemplaza m por cualquier cota superior.

Sea $F \subset T$ finito. Vamos a mostrar que

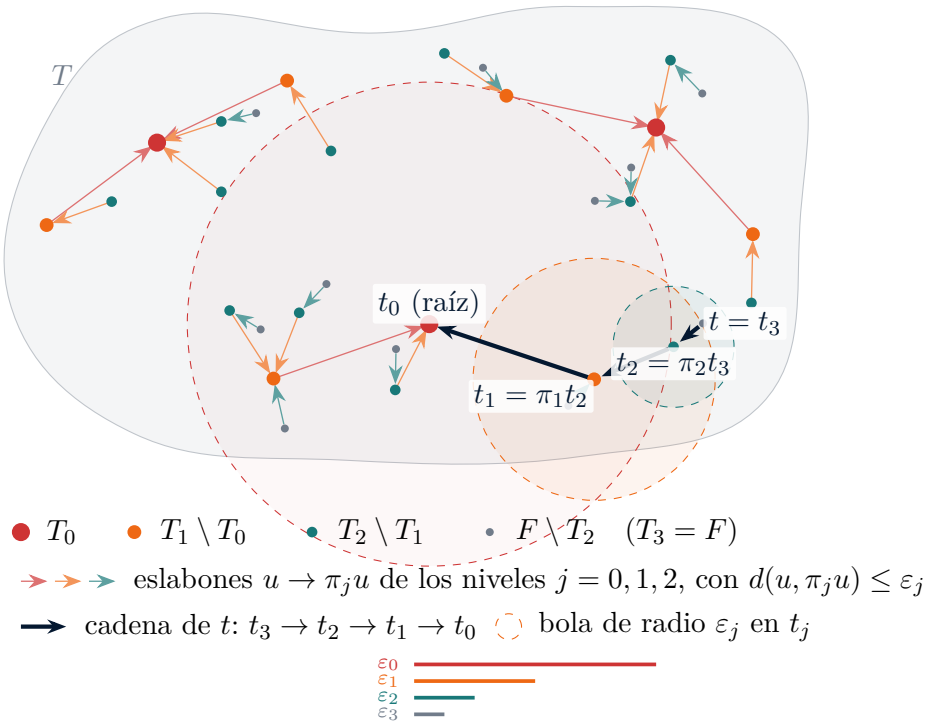
$$\left\| \max_{\substack{s,t\in F\\ d(s,t)\le\delta}} |X_s - X_t| \right\|_\psi \le 16 K_\psi L \int_0^\eta \psi^{-1}\Big(D(\varepsilon,d)\Big)\, d\varepsilon + K_\psi L\, \delta\, \psi^{-1}\Big(D^2(\eta,d)\Big). \tag{2.7}$$

Veamos que podemos suponer que d no se anula entre puntos distintos de F. Si $d(u,v) = 0$, entonces $\|X_u - X_v\|_\psi \le L\, d(u,v) = 0$, es decir, $X_u = X_v$ c.s. Además, por la desigualdad triangular, $d(u,w) = d(v,w)$ para todo $w \in T$.

La relación $s \sim t \iff d(s,t) = 0$ es de equivalencia en F. Sea $F^* \subset F$ un conjunto con un representante de cada clase, y sea $\pi(s) \in F^*$ el representante de s. Por lo anterior, $N = \bigcup_{s\in F}\{X_s \ne X_{\pi(s)}\}$ es una unión finita de conjuntos nulos, y $d(\pi(s),\pi(t)) = d(s,t)$ para todo $s,t \in F$. Así, si

$s, t \in F$ y $d(s,t) \le \delta$, el par $\pi(s), \pi(t)$ también satisface $d(\pi(s), \pi(t)) \le \delta$, y fuera de N se tiene $|X_s - X_t| = |X_{\pi(s)} - X_{\pi(t)}|$. Como además $F^* \subset F$, los máximos de $|X_s - X_t|$ sobre los pares a distancia a lo sumo δ en F y en F^* coinciden fuera de N, y en particular tienen la misma norma $\|\cdot\|_\psi$. Como el lado derecho de (2.7) no depende de F, basta probar la desigualdad para F^*, donde d no se anula entre puntos distintos.

Sea $T_0 \subset F$ un conjunto ε_0-separado maximal. Una vez construido T_j, que en particular es ε_{j+1}-separado, le agregamos puntos de F mientras siga siéndolo. Obtenemos así un conjunto $T_{j+1} \supset T_j$, maximal entre los subconjuntos ε_{j+1}-separados de F.



Como T_j es un subconjunto ε_j-separado de T, se tiene $|T_j| \le D(\varepsilon_j, d)$. Por maximalidad, todo punto de F está a distancia a lo sumo ε_j de algún punto de T_j. En particular, para cada $u \in T_{j+1}$ podemos elegir $\pi_j u \in T_j$ con $d(u, \pi_j u) \le \varepsilon_j$. Por la misma razón, si ε_J es menor que la mínima distancia entre puntos distintos de F, entonces $T_J = F$: un punto de $F \setminus T_J$ estaría a distancia a lo sumo ε_J de un punto distinto de T_J.

Fijemos un tal J. Cada $t \in F = T_J$ determina una cadena $t_J = t$, $t_j = \pi_j t_{j+1}$ para $j = J-1, \dots, 0$, que termina en un punto $t_0 \in T_0$ al que llamamos la raíz de t. La suma telescópica

$$X_t - X_{t_0} = \sum_{j=0}^{J-1} \Big(X_{t_{j+1}} - X_{t_j} \Big)$$

escribe $X_t - X_{t_0}$ como una suma de eslabones de la forma $X_u - X_{\pi_j u}$, con $u \in T_{j+1}$, uno por cada nivel j. Por lo tanto

$$\max_{t \in F} |X_t - X_{t_0}| \le \sum_{j=0}^{J-1} \max_{u \in T_{j+1}} |X_u - X_{\pi_j u}|,$$

donde en cada término t_0 es la raíz de t. El máximo del nivel j se toma sobre a lo sumo $|T_{j+1}| \le D(\varepsilon_{j+1}, d)$ variables, todas con $\|X_u - X_{\pi_j u}\|_\psi \le L\varepsilon_j$. Por el Lema 2.9,

$$\Big\| \max_{u \in T_{j+1}} |X_u - X_{\pi_j u}| \Big\|_\psi \le K_\psi L \varepsilon_j \, \psi^{-1}\Big(D(\varepsilon_{j+1}, d) \Big).$$

Para comparar con la integral usamos que $\psi^{-1}\Big(D(\varepsilon, d)\Big)$ es no creciente en ε, porque $D(\varepsilon, d)$ es no creciente en ε y ψ^{-1} es no decreciente. En el intervalo $[\varepsilon_{j+2}, \varepsilon_{j+1}]$, a la izquierda de ε_{j+1}, vale entonces

$\psi^{-1}\big(D(\varepsilon,d)\big) \geq \psi^{-1}\big(D(\varepsilon_{j+1},d)\big)$. Como ese intervalo tiene longitud $\varepsilon_{j+1}-\varepsilon_{j+2} = \varepsilon_j/4$,

$$\frac{\varepsilon_j}{4}\,\psi^{-1}\big(D(\varepsilon_{j+1},d)\big) = \int_{\varepsilon_{j+2}}^{\varepsilon_{j+1}} \psi^{-1}\big(D(\varepsilon_{j+1},d)\big)\,d\varepsilon \leq \int_{\varepsilon_{j+2}}^{\varepsilon_{j+1}} \psi^{-1}\big(D(\varepsilon,d)\big)\,d\varepsilon. \tag{2.8}$$

Juntando todo,

$$\begin{aligned}
\Big\|\max_{t\in F}|X_t - X_{t_0}|\Big\|_\psi &\leq \Big\|\sum_{j=0}^{J-1}\max_{u\in T_{j+1}}|X_u - X_{\pi_j u}|\Big\|_\psi && \text{(monotonía de } \|\cdot\|_\psi\text{)}\\
&\leq \sum_{j=0}^{J-1}\Big\|\max_{u\in T_{j+1}}|X_u - X_{\pi_j u}|\Big\|_\psi && \text{(desigualdad triangular)}\\
&\leq K_\psi L\sum_{j=0}^{J-1}\varepsilon_j\,\psi^{-1}\big(D(\varepsilon_{j+1},d)\big) && \text{(Lema 2.9)}\\
&= 4K_\psi L\sum_{j=0}^{J-1}\frac{\varepsilon_j}{4}\,\psi^{-1}\big(D(\varepsilon_{j+1},d)\big)\\
&\leq 4K_\psi L\sum_{j=0}^{J-1}\int_{\varepsilon_{j+2}}^{\varepsilon_{j+1}}\psi^{-1}\big(D(\varepsilon,d)\big)\,d\varepsilon && \text{(por (2.8))}\\
&= 4K_\psi L\int_{\varepsilon_{J+1}}^{\varepsilon_1}\psi^{-1}\big(D(\varepsilon,d)\big)\,d\varepsilon && \text{(intervalos contiguos)}\\
&\leq 4K_\psi L\int_0^{\eta}\psi^{-1}\big(D(\varepsilon,d)\big)\,d\varepsilon && \text{(integrando} \geq 0 \text{ y } \varepsilon_1 < \eta\text{)}.
\end{aligned}$$

Sean ahora $s,t\in F$ con $d(s,t)\leq\delta$. Entonces

$$|X_s - X_t| \leq |X_s - X_{s_0}| + |X_{s_0} - X_{t_0}| + |X_{t_0} - X_t| \leq 2\max_{t\in F}|X_t - X_{t_0}| + |X_{s_0} - X_{t_0}|. \tag{2.9}$$

Las raíces forman a lo sumo $|T_0|^2 \leq D^2(\eta,d)$ pares, pero su distancia no está controlada por δ. Sólo sabemos que $d(s_0,t_0)\leq\delta+4\eta$, y usar esta cota introduciría un término $\eta\,\psi^{-1}\big(D^2(\eta,d)\big)$, que en general no es comparable con la integral. Para evitarlo, sea

$$P = \{(s_0,t_0):\ s,t\in F,\ d(s,t)\leq\delta\}\subset T_0\times T_0$$

el conjunto de los pares de raíces de pares cercanos, y para cada $(a,b)\in P$ elijamos $s^{ab},t^{ab}\in F$ con $d(s^{ab},t^{ab})\leq\delta$ y raíces a y b, respectivamente. Como a es la raíz de s^{ab} y b la de t^{ab},

$$|X_a - X_b| \leq |X_a - X_{s^{ab}}| + |X_{s^{ab}} - X_{t^{ab}}| + |X_{t^{ab}} - X_b| \leq 2\max_{t\in F}|X_t - X_{t_0}| + \max_{(a,b)\in P}|X_{s^{ab}} - X_{t^{ab}}|,$$

Como $|P|\leq|T_0|^2\leq D^2(\eta,d)$ y cada $X_{s^{ab}} - X_{t^{ab}}$ tiene norma a lo sumo $L\delta$, el Lema 2.9 da

$$\Big\|\max_{(a,b)\in P}|X_{s^{ab}} - X_{t^{ab}}|\Big\|_\psi \leq K_\psi L\,\delta\,\psi^{-1}\big(D^2(\eta,d)\big).$$

Si $s,t\in F$ y $d(s,t)\leq\delta$, entonces $(s_0,t_0)\in P$, y combinando con (2.9) resulta

$$\begin{aligned}
\Big\|\max_{\substack{s,t\in F\\ d(s,t)\leq\delta}}|X_s - X_t|\Big\|_\psi &\leq 4\Big\|\max_{t\in F}|X_t - X_{t_0}|\Big\|_\psi + \Big\|\max_{(a,b)\in P}|X_{s^{ab}} - X_{t^{ab}}|\Big\|_\psi\\
&\leq 16K_\psi L\int_0^{\eta}\psi^{-1}\big(D(\varepsilon,d)\big)\,d\varepsilon + K_\psi L\,\delta\,\psi^{-1}\big(D^2(\eta,d)\big),
\end{aligned}$$

que es (2.7).

Pasamos a un conjunto numerable $T' \subset T$. Sean F_k conjuntos finitos con $F_k \uparrow T'$, y sean

$$Z_k = \max_{\substack{s,t\in F_k \\ d(s,t)\le\delta}} |X_s - X_t| \;\uparrow\; Z = \sup_{\substack{s,t\in T' \\ d(s,t)\le\delta}} |X_s - X_t|,$$

donde el máximo sobre un conjunto vacío es 0. Por (2.7), $\|Z_k\|_\psi$ está acotada, para todo k, por el lado derecho de (2.7). Si β es mayor que ese lado derecho, resulta $\mathbb{E}\,\psi(Z_k/\beta) \le 1$ para todo k (por (2.2)). Como ψ es continua y no decreciente, $\psi(Z_k/\beta) \uparrow \psi(Z/\beta)$, con la convención $\psi(\infty) = \infty$. Por convergencia monótona, $\mathbb{E}\,\psi(Z/\beta) \le 1$, es decir, $\|Z\|_\psi \le \beta$. Como β es cualquier número mayor que el lado derecho de (2.7), la desigualdad vale también con T' en lugar de F.

Finalmente usamos la separabilidad. Sean $T' \subset T$ numerable y N nulo como en la definición de proceso separable respecto de d. Fuera de N, el supremo de $|X_s - X_t|$ sobre los pares $s, t \in T$ con $d(s,t) \le \delta$ coincide con el supremo sobre los pares de T', cuya norma acabamos de acotar. Luego (2.7) vale también con el supremo sobre todo T, y como

$$\begin{aligned} &16K_\psi L \int_0^\eta \psi^{-1}\Big(D(\varepsilon,d)\Big)\,d\varepsilon + K_\psi L\,\delta\,\psi^{-1}\Big(D^2(\eta,d)\Big) \\ &\qquad \le 16K_\psi L\Big\{\int_0^\eta \psi^{-1}\Big(D(\varepsilon,d)\Big)\,d\varepsilon + \delta\,\psi^{-1}\Big(D^2(\eta,d)\Big)\Big\}, \end{aligned}$$

se obtiene el teorema. □

Demostración del Corolario 2.13. Si $\operatorname{diam} T = \infty$, entonces $D(\varepsilon,d) \ge 2$ para todo $\varepsilon > 0$, la integral diverge y no hay nada que probar. Supongamos $\operatorname{diam} T < \infty$, sea $\eta > \operatorname{diam} T$ y tomemos $\delta = \eta$ en el Teorema 2.12. Como todos los pares cumplen $d(s,t) \le \operatorname{diam} T < \delta$, el supremo del teorema es el supremo sobre todos los $s, t \in T$. Además $D(\eta,d) = 1$, porque no hay dos puntos de T a distancia mayor que η. Luego

$$\Big\|\sup_{s,t\in T} |X_s - X_t|\Big\|_\psi \le 16K_\psi L\Big\{\int_0^\eta \psi^{-1}\Big(D(\varepsilon,d)\Big)\,d\varepsilon + \eta\,\psi^{-1}(1)\Big\} \le 32K_\psi L \int_0^\eta \psi^{-1}\Big(D(\varepsilon,d)\Big)\,d\varepsilon,$$

donde usamos que $D(\varepsilon,d) \ge 1$ y que ψ^{-1} es no decreciente, de modo que el integrando es al menos $\psi^{-1}(1)$. Como $D(\varepsilon,d) = 1$ para $\varepsilon \ge \operatorname{diam} T$, la última integral es igual a

$$\int_0^{\operatorname{diam} T} \psi^{-1}\Big(D(\varepsilon,d)\Big)\,d\varepsilon + (\eta - \operatorname{diam} T)\,\psi^{-1}(1).$$

Haciendo $\eta \downarrow \operatorname{diam} T$ se obtiene el corolario. □

Aunque el teorema proporciona una cota para el módulo de continuidad del proceso, también se necesitará una cota para el máximo del proceso. En efecto, para cualquier $t_0 \in T$,[8]

$$\Big\|\sup_{t\in T} |X_t|\Big\|_\psi \le \|X_{t_0}\|_\psi + 32K_\psi L \int_0^{\operatorname{diam} T} \psi^{-1}\Big(D(\varepsilon,d)\Big)\,d\varepsilon.$$

No obstante, es natural enunciar la desigualdad maximal en términos de incrementos. La cota para los incrementos muestra que el módulo de continuidad del proceso tiende a cero en norma ψ si $\int_0^\eta \psi^{-1}\Big(D(\varepsilon,d)\Big)\,d\varepsilon < \infty$ para algún $\eta > 0$. En efecto, en el Teorema 2.12 el lado derecho puede hacerse arbitrariamente pequeño eligiendo primero η pequeño y luego δ pequeño.[9] De esta desigualdad puede deducirse la continuidad de casi todas las trayectorias.

[8] La desigualdad se obtiene de $|X_t| \le |X_{t_0}| + \sup_{s,u} |X_s - X_u|$, aplicando la monotonía y la desigualdad triangular de la norma de Orlicz, y luego el Corolario 2.13 al término del supremo.

[9] Primero η: la convergencia de la integral hace pequeña su cola $\int_0^\eta$. Con η fijo, $\psi^{-1}(D^2(\eta,d))$ es un número finito (la integrabilidad fuerza $D(\varepsilon,d) < \infty$ para todo $\varepsilon > 0$), y entonces δ pequeño hace pequeño el segundo término.

2.2.1. Desigualdades subgaussianas

Una variable normal estándar tiene colas del orden $x^{-1}e^{-x^2/2}$ y satisface $\mathbb{P}(|X| > x) \le 2e^{-x^2/2}$, $x > 0$. Por cálculo directo, su norma ψ_2 (Ejemplo 2.5) es $\sqrt{8/3}$.[10] En esta subsección estudiaremos variables y procesos con cotas de cola similares.

La desigualdad de Hoeffding[11] da una cota subgaussiana para sumas de variables independientes, centradas y acotadas. Usaremos el siguiente caso particular.

Lema 2.14. *Sean $a_1, \dots, a_n$ constantes y sean $\varepsilon_1, \dots, \varepsilon_n$ variables de Rademacher independientes, es decir, $\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = -1) = \frac{1}{2}$. Entonces, para todo $x > 0$,*

$$\mathbb{P}\left(\left|\sum_{i=1}^n \varepsilon_i a_i\right| > x\right) \le 2\exp\left\{-\frac{x^2}{2\|a\|^2}\right\},$$

donde $\|a\|$ denota la norma euclidiana de $a = (a_1, \dots, a_n)$. En consecuencia, por el Lema 2.8 $\|\sum_{i=1}^n \varepsilon_i a_i\|_{\psi_2} \le \sqrt{6}\,\|a\|$.

Definición 2.15. Un proceso estocástico $\{X_t : t \in T\}$ se dice **subgaussiano** respecto de una semimétrica d en T si, para todo $s, t \in T$ y todo $x > 0$,

$$\mathbb{P}(|X_s - X_t| > x) \le 2\exp\left\{-\frac{x^2}{2d^2(s,t)}\right\},$$

con la convención de que, si $d(s,t) = 0$, entonces $X_s = X_t$ casi seguramente. Las constantes 2 y 1/2 no tienen importancia especial.

Ejemplo 2.16.
- Todo proceso gaussiano centrado es subgaussiano para $d(s,t) = \sigma(X_s - X_t)$.
- Otro ejemplo es el proceso de Rademacher $X_a = \sum_{i=1}^n a_i\varepsilon_i$, $a \in \mathbb{R}^n$, donde $\varepsilon_1, \dots, \varepsilon_n$ son variables de Rademacher independientes. Por el Lema 2.14, este proceso es subgaussiano para la distancia euclidiana $d(a,b) = \|a - b\|$.

Observación 2.17. Los procesos subgaussianos satisfacen la cota de incrementos $\|X_s - X_t\|_{\psi_2} \le \sqrt{6}\,d(s,t)$: aplicando el Lema 2.8 con $X = X_s - X_t$, $p = 2$, $K = 2$ y $C = (2d^2(s,t))^{-1}$, se obtiene $\|X_s - X_t\|_{\psi_2} \le \left(\frac{1+2}{(2d^2(s,t))^{-1}}\right)^{1/2} = \sqrt{6}\,d(s,t)$.

Como $\psi_2^{-1}(m) = \sqrt{\log(1+m)}$, la desigualdad maximal general implica, para procesos subgaussianos, una cota en términos de una integral de entropía. Debido a las propiedades elementales del logaritmo, el enunciado puede simplificarse ligeramente.

Corolario 2.18 (Desigualdad maximal subgaussiana o cota de Dudley)**.** *Sea $\{X_t : t \in T\}$ un proceso subgaussiano separable respecto de la semimétrica d. Entonces, para todo $\delta > 0$,*

$$\mathbb{E}\sup_{d(s,t)\le\delta} |X_s - X_t| \le K\int_0^\delta \sqrt{\log D(\varepsilon, d)}\,d\varepsilon,$$

donde K es una constante universal. En particular, para todo $t_0 \in T$,

$$\mathbb{E}\sup_{t\in T} |X_t| \le \mathbb{E}|X_{t_0}| + K\int_0^\infty \sqrt{\log D(\varepsilon, d)}\,d\varepsilon. \tag{2.10}$$

[10] Para $X \sim N(0,1)$ y $C^2 > 2$, la integral gaussiana da $\mathbb{E}e^{X^2/C^2} = (1 - 2/C^2)^{-1/2}$; igualando $\mathbb{E}\psi_2(X/C) = 1$, es decir $(1 - 2/C^2)^{-1/2} = 2$, queda $C^2 = 8/3$.

[11] Si $Y_1, \dots, Y_n$ son independientes, con $\mathbb{E}\,Y_i = 0$ y $\alpha_i \le Y_i \le \beta_i$ c.s., entonces, para todo $x > 0$, $\mathbb{P}\big(\big|\sum_{i=1}^n Y_i\big| \ge x\big) \le 2\exp\big\{-2x^2/\sum_{i=1}^n(\beta_i - \alpha_i)^2\big\}$.

Demostración. Empezamos descartando algunos casos. Si la integral del lado derecho es infinita, no hay nada que probar, así que la suponemos finita. Si $\operatorname{diam}(T) = 0$, todos los pares están a distancia nula y, por la hipótesis subgaussiana, $X_s = X_t$ c.s. para cada par fijo. Por la separabilidad, el supremo del lado izquierdo coincide c.s. con un supremo sobre un conjunto numerable de pares, y es por lo tanto nulo c.s.; como el lado derecho es no negativo, la desigualdad se cumple. Suponemos entonces $\operatorname{diam}(T) > 0$. También podemos suponer $\delta \le \operatorname{diam}(T)$: si $\delta > \operatorname{diam}(T)$, todos los pares cumplen $d(s,t) \le \operatorname{diam}(T)$, así que el supremo del lado izquierdo es el mismo que con $\operatorname{diam}(T)$ en lugar de δ, mientras que la integral hasta δ es mayor o igual que la integral hasta $\operatorname{diam}(T)$, porque el integrando es no negativo.

La idea es aplicar el Teorema 2.12 con $\psi = \psi_2$ y con $L = \sqrt{6}$ y $\eta = \delta$. Como $\psi_2^{-1}(m) = \sqrt{\log(1+m)}$, obtenemos

$$\Big\| \sup_{d(s,t)\le\delta} |X_s - X_t| \Big\|_{\psi_2} \le 16\sqrt{6}\, K_{\psi_2} \Big\{ \int_0^\delta \sqrt{\log\Big(1 + D(\varepsilon, d)\Big)}\, d\varepsilon + \delta \sqrt{\log\Big(1 + D^2(\delta, d)\Big)} \Big\}. \tag{2.11}$$

El segundo término de la llave se acota por el primero. Por un lado, $1 + D^2 \le (1+D)^2$ para $D \ge 0$, así que $\log(1+D^2) \le 2\log(1+D)$ y

$$\delta \sqrt{\log\Big(1 + D^2(\delta, d)\Big)} \le \sqrt{2}\, \delta \sqrt{\log\Big(1 + D(\delta, d)\Big)}.$$

Por otro lado, $D(\varepsilon, d)$ es no creciente en ε, de modo que el integrando de (2.11) es, en todo el intervalo $(0, \delta]$, al menos su valor en δ. Integrando,

$$\delta \sqrt{\log\Big(1 + D(\delta, d)\Big)} \le \int_0^\delta \sqrt{\log\Big(1 + D(\varepsilon, d)\Big)}\, d\varepsilon.$$

Juntando las dos desigualdades, el segundo término es a lo sumo $\sqrt{2}$ veces el primero, y

$$\Big\| \sup_{d(s,t)\le\delta} |X_s - X_t| \Big\|_{\psi_2} \le 16\sqrt{6}\,(1+\sqrt{2})\, K_{\psi_2} \int_0^\delta \sqrt{\log\Big(1 + D(\varepsilon, d)\Big)}\, d\varepsilon.$$

Para pasar de $\log(1+D)$ a $\log D$, observemos que si $0 < \varepsilon < \operatorname{diam}(T)$, hay dos puntos de T a distancia mayor que ε, y entonces $D(\varepsilon, d) \ge 2$. Para $D \ge 2$ vale $1 + D \le D^2$, es decir, $\log(1+D) \le 2\log D$. Como $\delta \le \operatorname{diam}(T)$, esto vale para todo $\varepsilon \in (0, \delta)$, y como $\sqrt{6} \cdot \sqrt{2} = 2\sqrt{3}$,

$$\Big\| \sup_{d(s,t)\le\delta} |X_s - X_t| \Big\|_{\psi_2} \le 32\sqrt{3}\,(1+\sqrt{2})\, K_{\psi_2} \int_0^\delta \sqrt{\log D(\varepsilon, d)}\, d\varepsilon. \tag{2.12}$$

Falta pasar de la norma ψ_2 a la esperanza. Para todo $u \ge 0$ vale $u \le \psi_2(u) + 1$: si $u \ge 1$, $u \le u^2 \le e^{u^2} - 1$, y si $u < 1$, el lado derecho es al menos 1. Dada una variable aleatoria Z y $r > \|Z\|_{\psi_2}$, aplicando esto a $u = |Z|/r$ y usando (2.2),

$$\frac{\mathbb{E}|Z|}{r} \le \mathbb{E}\, \psi_2\Big(\frac{|Z|}{r}\Big) + 1 \le 2.$$

Haciendo $r \downarrow \|Z\|_{\psi_2}$ se obtiene $\mathbb{E}|Z| \le 2\|Z\|_{\psi_2}$. Aplicado al supremo, cuya medibilidad está garantizada por la separabilidad, (2.12) da la primera afirmación del corolario con

$$K = 64\sqrt{3}\,(1+\sqrt{2})\, K_{\psi_2},$$

donde K_{ψ_2} es la constante del Lema 2.9 para ψ_2.

Para probar (2.10), fijamos $t_0 \in T$. Para todo $t \in T$,

$$|X_t| \le |X_{t_0}| + |X_t - X_{t_0}| \le |X_{t_0}| + \sup_{s,t'\in T} |X_s - X_{t'}|,$$

y tomando supremo y esperanza, $\mathbb{E}\sup_t |X_t| \le \mathbb{E}|X_{t_0}| + \mathbb{E}\sup_{s,t'} |X_s - X_{t'}|$. Si $\operatorname{diam}(T) = \infty$, entonces $D(\varepsilon, d) \ge 2$ para todo $\varepsilon > 0$, la integral hasta ∞ diverge y no hay nada que probar. Si $\operatorname{diam}(T) < \infty$, aplicamos la primera parte con $\delta = \operatorname{diam}(T)$, que abarca todos los pares; y como $D(\varepsilon, d) = 1$, es decir $\log D(\varepsilon, d) = 0$, para $\varepsilon \ge \operatorname{diam}(T)$, la integral hasta $\operatorname{diam}(T)$ coincide con la integral hasta ∞. Esto es (2.10). □

2.3. Simetrización y medibilidad

2.3.1. Simetrización

Uno de los enfoques principales para probar teoremas de Glivenko–Cantelli y de Donsker consiste en comparar el proceso empírico con su versión simetrizada. Si $\varepsilon_1, \dots, \varepsilon_n$ son Rademacher i.i.d., independientes de $X_1, \dots, X_n$, escribimos

$$(P_n - P)f = \frac{1}{n}\sum_{i=1}^{n}\{f(X_i) - Pf\}, \qquad P_n^\circ f = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i f(X_i).$$

Ambos procesos están centrados, pues $\mathbb{E}(\varepsilon_i f(X_i) \mid X_i) = 0$. La ventaja de P_n° es que, condicionalmente en $X_1, \dots, X_n$, es un proceso de Rademacher y por tanto subgaussiano para la semimétrica empírica; así pueden aplicarse las desigualdades maximales subgaussianas de la Sección 2.2, en particular el Corolario 2.18.

Como los supremos $\|P_n - P\|_{\mathcal{F}}$ pueden no ser medibles, el resultado se formula con esperanza exterior. Cuando aparecen esperanzas exteriores suponemos que $X_1, \dots, X_n$ son las proyecciones coordenadas en $(\mathcal{X}^n, \mathcal{A}^n, P^n)$, y que las variables auxiliares independientes viven en factores producto adicionales. Esta convención permite usar la perfección de las proyecciones coordenadas y las desigualdades iteradas para esperanzas exteriores. El teorema de Fubini ordinario sólo se usará más adelante, cuando una hipótesis de P-medibilidad garantice la medibilidad del supremo correspondiente.

Lema 2.19. *Sea $(\Omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad y sea $\tau : \Omega \to \Omega$ una biyección tal que τ y τ^{-1} son medibles y $\mathbb{P} \circ \tau^{-1} = \mathbb{P}$. Entonces, para toda $T : \Omega \to \overline{\mathbb{R}}$,*

$$(T \circ \tau)^* = T^* \circ \tau \qquad \mathbb{P}\text{-c.s.}$$

En particular, si $\mathbb{E}\, T^*$ existe (por ejemplo, si $T \geq 0$), entonces $\mathbb{E}^*(T \circ \tau) = \mathbb{E}^* T$.

Demostración. Como $T^* \circ \tau$ es medible y domina a $T \circ \tau$, el Lema 1.5(ii) da $(T \circ \tau)^* \leq T^* \circ \tau$ c.s. Recíprocamente, $(T \circ \tau)^* \circ \tau^{-1}$ es medible y domina a $T \circ \tau \circ \tau^{-1} = T$, así que, otra vez por el Lema 1.5(ii), el conjunto medible $N = \{T^* > (T \circ \tau)^* \circ \tau^{-1}\}$ es nulo. Como $\{T^* \circ \tau > (T \circ \tau)^*\} = \tau^{-1}(N)$[12] y $\mathbb{P}(\tau^{-1}(N)) = \mathbb{P}(N) = 0$, también $T^* \circ \tau \leq (T \circ \tau)^*$ c.s. Para la última afirmación, el cambio de variables da $\mathbb{E}[T^* \circ \tau] = \mathbb{E}\, T^*$, que existe por hipótesis; como $(T \circ \tau)^* = T^* \circ \tau$ c.s., también existe $\mathbb{E}[(T \circ \tau)^*]$, y el Lema 1.5(iii), aplicado a $T \circ \tau$ y a T, da

$$\mathbb{E}^*(T \circ \tau) = \mathbb{E}\Big[(T \circ \tau)^*\Big] = \mathbb{E}\Big[T^* \circ \tau\Big] = \mathbb{E}\, T^* = \mathbb{E}^* T.$$

□

Lema 2.20. *Sea $\Phi : \mathbb{R} \to \mathbb{R}$ convexa, no decreciente y finita, con la convención $\Phi(+\infty) := \lim_{t\to\infty} \Phi(t)$, y sea $\mathcal{F}$ una clase de funciones medibles con $P|f| < \infty$ para todo $f \in \mathcal{F}$. Entonces*

$$\mathbb{E}^* \Phi\Big(\|P_n - P\|_{\mathcal{F}}\Big) \leq \mathbb{E}^* \Phi\Big(2\|P_n^\circ\|_{\mathcal{F}}\Big).$$

Demostración. Trabajamos en el espacio producto $\mathcal{X}^n \times \mathcal{X}^n \times \{-1, 1\}^n$ con la probabilidad $P^n \otimes P^n \otimes R_n$, donde R_n es la uniforme sobre $\{-1, 1\}^n$, y llamamos $X = (X_1, \dots, X_n)$, $Y = (Y_1, \dots, Y_n)$ y $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$ a las tres proyecciones coordenadas. Así, Y es una copia independiente de la muestra y ε es un vector de signos de Rademacher independiente de ambas. Escribimos $\mathbb{E}^*_X$, $\mathbb{E}^*_{X,Y}$, etc., para indicar respecto de qué bloques de coordenadas se calcula la esperanza exterior. Agregar factores no altera la esperanza exterior de una función no negativa que depende sólo de algunos bloques: por

[12] Pues $\tau(\omega) \in N$ si y sólo si $T^*(\tau(\omega)) > (T \circ \tau)^*\big(\tau^{-1}(\tau(\omega))\big) = (T \circ \tau)^*(\omega)$.

ejemplo, $\mathbb{E}^*_{X,Y}\, g(X) = \mathbb{E}^*_X\, g(X)$ para toda $g : \mathcal{X}^n \to [0,\infty]$.[13] Usaremos este hecho varias veces sin mencionarlo.

Todos los argumentos de Φ que aparecen son no negativos, así que toman valores en $[0,\infty]$; los supremos sobre $\mathcal{F}$ pueden valer $+\infty$, y en ese caso se usa la convención $\Phi(+\infty) = \lim_{t\to\infty} \Phi(t)$ del enunciado, cuyo límite existe porque Φ es no decreciente. Así extendida, Φ es no decreciente y continua en $[0,\infty]$,[14] y la desigualdad de convexidad

$$\Phi\Big(\frac{s+t}{2}\Big) \le \frac{1}{2}\,\Phi(s) + \frac{1}{2}\,\Phi(t) \tag{2.13}$$

sigue valiendo para $s,t \in [0,\infty]$.[15] Una consecuencia de la monotonía y de la continuidad que usaremos más adelante es que el supremo conmuta con Φ: para cualquier familia $(a_f)_{f\in\mathcal{F}}$ en $[0,\infty]$,[16]

$$\sup_{f\in\mathcal{F}} \Phi(a_f) = \Phi\Big(\sup_{f\in\mathcal{F}} a_f\Big). \tag{2.14}$$

Podemos suponer además que $\Phi \ge 0$ en $[0,\infty]$. En efecto, si T está acotada inferiormente y c es una constante real, el Lema 1.9(i), con el sumando medible c, da $(T+c)^* = T^* + c$, y el Lema 1.5(iii) da $\mathbb{E}^*(T+c) = \mathbb{E}^*T + c$. Las funciones a las que se aplica $\mathbb{E}^*$ en el enunciado son de la forma $\Phi(Z)$ con $Z \ge 0$, y por lo tanto están acotadas inferiormente por $\Phi(0)$. Así, reemplazar Φ por $\Phi - \Phi(0)$, que sigue siendo convexa y no decreciente, resta la misma constante $\Phi(0)$ a ambos lados de la desigualdad; y $\Phi - \Phi(0) \ge 0$ en $[0,\infty]$ por la monotonía. Con esta reducción, todas las esperanzas exteriores que siguen están bien definidas y toman valores en $[0,\infty]$.

Comenzamos reemplazando Pf por el promedio de f sobre la muestra fantasma Y. Fijemos $x = (x_1,\dots,x_n) \in \mathcal{X}^n$ y, para $f \in \mathcal{F}$ e $y \in \mathcal{X}^n$, sean

$$h_f(y) = \frac{1}{n}\sum_{i=1}^{n}\{f(x_i) - f(y_i)\}, \qquad H(x,y) = \sup_{f\in\mathcal{F}} |h_f(y)|.$$

Cada h_f es medible en y, pero $H(x,\cdot)$ puede no serlo, porque es un supremo sobre una clase posiblemente no numerable. Como $P|f| < \infty$, la variable $h_f(Y)$ es integrable y

$$\mathbb{E}_Y h_f(Y) = \frac{1}{n}\sum_{i=1}^{n} f(x_i) - Pf, \tag{2.15}$$

porque cada Y_i tiene ley P y los $f(x_i)$ son constantes.[17] Entonces

$$\begin{aligned}\Phi\Big(\Big|\frac{1}{n}\sum_{i=1}^{n} f(x_i) - Pf\Big|\Big) &= \Phi\Big(|\mathbb{E}_Y h_f(Y)|\Big) \le \Phi\Big(\mathbb{E}_Y|h_f(Y)|\Big)\\ &\le \mathbb{E}_Y\,\Phi\Big(|h_f(Y)|\Big) \le \mathbb{E}^*_Y\,\Phi\Big(H(x,Y)\Big).\end{aligned} \tag{2.16}$$

[13] Es una consecuencia del Lema 1.20(i). Si π es la proyección sobre el bloque, $(g\circ\pi)^* = g^*\circ\pi$ c.s. Como $g \ge 0$, existen las esperanzas de $(g\circ\pi)^*$ y de g^*, y por el Lema 1.5(iii)

$$\mathbb{E}^*(g\circ\pi) = \mathbb{E}\big[(g\circ\pi)^*\big] = \mathbb{E}\big[g^*\circ\pi\big] = \mathbb{E}g^* = \mathbb{E}^*g,$$

donde la penúltima igualdad usa que la ley de π es la marginal correspondiente. Usamos la parte (i) del lema y no la (ii) porque ésta está enunciada sólo para funciones acotadas.

[14] En $[0,\infty)$ es continua porque una función convexa y finita en $\mathbb{R}$ lo es; en $+\infty$ la continuidad es la definición de $\Phi(+\infty)$.

[15] Si $s,t < \infty$ es la convexidad usual. Si $t = +\infty$ y $s < \infty$, basta hacer $t' \to \infty$ en $\Phi((s+t')/2) \le \frac{1}{2}\Phi(s) + \frac{1}{2}\Phi(t')$ y usar la continuidad en $+\infty$. Si $s = t = +\infty$ no hay nada que probar.

[16] Sea $a = \sup_f a_f$. La desigualdad $\le$ vale porque $a_f \le a$ para todo f. Para la otra, tomamos f_k con $a_{f_k} \to a$; por la continuidad de Φ en $[0,\infty]$, $\Phi(a_{f_k}) \to \Phi(a)$, y cada $\Phi(a_{f_k})$ es a lo sumo el supremo de la izquierda.

[17] Éste es el único lugar de la prueba donde se usa la hipótesis $P|f| < \infty$: gracias a ella, la diferencia $\frac{1}{n}\sum_i f(x_i) - Pf$ se escribe como una esperanza respecto de la muestra fantasma.

La primera desigualdad usa $|\mathbb{E}_Y h_f(Y)| \le \mathbb{E}_Y|h_f(Y)|$ y la monotonía de Φ; la segunda es la desigualdad de Jensen para la función convexa Φ y la variable integrable $|h_f(Y)|$.[18] La tercera se debe a que $|h_f| \le H(x,\cdot)$ punto a punto, de modo que $\Phi(|h_f|) \le \Phi(H(x,\cdot))$, y a que la esperanza de una función medible no supera la esperanza exterior de cualquier función mayor.[19] El lado derecho de (2.16) no depende de f, así que tomando supremo en $f \in \mathcal{F}$ y usando (2.14) obtenemos, para todo $x \in \mathcal{X}^n$,

$$\Phi\Big(\Big\|\frac{1}{n}\sum_{i=1}^{n} f(x_i) - Pf\Big\|_{\mathcal{F}}\Big) \le \mathbb{E}_Y^*\,\Phi\Big(H(x,Y)\Big).$$

El lado izquierdo es $\Phi(\|P_n - P\|_{\mathcal{F}})$ evaluado en la muestra $X = x$. Como la desigualdad vale para todo x, la monotonía de la esperanza exterior y el Lema 1.16 en $\mathcal{X}^n \times \mathcal{X}^n$ dan[20]

$$\mathbb{E}^*\,\Phi\Big(\|P_n - P\|_{\mathcal{F}}\Big) \le \mathbb{E}_X^*\,\mathbb{E}_Y^*\,\Phi\Big(H(X,Y)\Big) \le \mathbb{E}_{X,Y}^*\,\Phi\Big(H(X,Y)\Big). \tag{2.17}$$

Introducimos ahora los signos. Para $e = (e_1,\dots,e_n) \in \{-1,1\}^n$ sea

$$H_e(x,y) = \Big\|\frac{1}{n}\sum_{i=1}^{n} e_i\{f(x_i) - f(y_i)\}\Big\|_{\mathcal{F}},$$

de modo que $H = H_{\mathbf{1}}$, con $\mathbf{1} = (1,\dots,1)$. Aquí e es un vector fijo; los signos aleatorios ε aparecen recién al final.

Cambiar el signo del i-ésimo sumando equivale a intercambiar x_i con y_i. Para $i = 1,\dots,n$ sea σ_i la aplicación de $\mathcal{X}^n \times \mathcal{X}^n$ en sí mismo que intercambia x_i con y_i y deja fijas las demás coordenadas, y sea τ_e la composición de las σ_i con $e_i = -1$ (el orden no importa, y $\tau_{\mathbf{1}}$ es la identidad). Si $e_i = -1$, tras el intercambio el i-ésimo sumando de $H \circ \tau_e$ es $f(y_i) - f(x_i) = e_i\{f(x_i) - f(y_i)\}$, y si $e_i = 1$ no cambia. Como esto vale para cada f,

$$H_e = H \circ \tau_e \qquad \text{punto a punto en } \mathcal{X}^n \times \mathcal{X}^n. \tag{2.18}$$

De la ley $Q = P^n \otimes P^n$ de (X,Y) sólo se necesita que cada intercambio la preserve:

$$Q \circ \sigma_i^{-1} = Q, \qquad i = 1,\dots,n, \tag{2.19}$$

es decir, que la ley de (X,Y) no cambie al intercambiar X_i con Y_i. Para $Q = P^n \otimes P^n$ esto vale: $Q \circ \sigma_i^{-1}$ y Q coinciden sobre los rectángulos $A_1 \times \cdots \times A_n \times B_1 \times \cdots \times B_n$, porque intercambiar A_i con B_i no altera $\prod_j P(A_j)P(B_j)$, y como los rectángulos forman un π-sistema que contiene al espacio total y genera $\mathcal{A}^n \otimes \mathcal{A}^n$, el teorema de Dynkin (ver apéndice) da (2.19).

Cada σ_i es una biyección bimedible, porque es una permutación de coordenadas y $\sigma_i \circ \sigma_i$ es la identidad. Por (2.19), cada τ_e es una biyección bimedible que preserva Q. Aplicando Φ a (2.18) se obtiene $\Phi(H_e) = \Phi(H) \circ \tau_e$ punto a punto, y el Lema 2.19, aplicado en $(\mathcal{X}^n \times \mathcal{X}^n, \mathcal{A}^n \otimes \mathcal{A}^n, Q)$ con $\tau = \tau_e$ y $T = \Phi(H) \ge 0$, da[21]

$$\mathbb{E}_{X,Y}^*\,\Phi\Big(H_e(X,Y)\Big) = \mathbb{E}_{X,Y}^*\,\Big(\Phi(H)\circ\tau_e\Big)(X,Y) = \mathbb{E}_{X,Y}^*\,\Phi\Big(H(X,Y)\Big) \qquad \text{para todo } e \in \{-1,1\}^n.$$

El lado derecho de (2.17) coincide entonces con el promedio de estas cantidades sobre los 2^n valores de e, y

$$\mathbb{E}^*\,\Phi\Big(\|P_n - P\|_{\mathcal{F}}\Big) \le \frac{1}{2^n}\sum_{e\in\{-1,1\}^n} \mathbb{E}_{X,Y}^*\,\Phi\Big(H_e(X,Y)\Big). \tag{2.20}$$

[18] La esperanza $\mathbb{E}_Y\Phi(|h_f(Y)|)$ existe en $[0,\infty]$ porque $\Phi \ge 0$ en $[0,\infty)$, y Jensen vale aunque sea infinita.

[19] Por el Lema 1.5, $\mathbb{E}^*U = \mathbb{E}U$ si U es medible, y $\mathbb{E}^*$ es monótona (véase la demostración del Lema 1.16).

[20] El Lema 1.16, con Ω_1 la coordenada X y Ω_2 la coordenada Y, da $\mathbb{E}_1^*\mathbb{E}_2^* T \le \mathbb{E}^* T$ para cualquier T. Es la dirección que siempre vale; la inversa sería un Fubini exterior, que puede fallar si T no es conjuntamente medible.

[21] El lema se aplica a la función $\Phi(H)$ como un todo. Como vale para funciones arbitrarias, no hace falta relacionar $\Phi(H)^*$ con H^*.

Falta separar las contribuciones de X y de Y. Fijemos e y sea

$$A_e(x) = \left\| \frac{1}{n} \sum_{i=1}^{n} e_i f(x_i) \right\|_{\mathcal{F}}, \qquad x \in \mathcal{X}^n.$$

Para cada f, la desigualdad triangular da

$$\left| \frac{1}{n} \sum_{i=1}^{n} e_i \{f(x_i) - f(y_i)\} \right| \le \left| \frac{1}{n} \sum_{i=1}^{n} e_i f(x_i) \right| + \left| \frac{1}{n} \sum_{i=1}^{n} e_i f(y_i) \right| \le A_e(x) + A_e(y),$$

y como la cota no depende de f, $H_e(x,y) \le A_e(x) + A_e(y)$. Escribiendo $A_e(x) + A_e(y) = \frac{1}{2}\{2A_e(x)\} + \frac{1}{2}\{2A_e(y)\}$, la monotonía de Φ y la convexidad (2.13) dan

$$\Phi\Big(H_e(x,y)\Big) \le \frac{1}{2}\,\Phi\Big(2A_e(x)\Big) + \frac{1}{2}\,\Phi\Big(2A_e(y)\Big).$$

Tomando esperanza exterior conjunta y usando la monotonía, la subaditividad y la homogeneidad positiva de $\mathbb{E}^*$,[22]

$$\mathbb{E}^*_{X,Y}\,\Phi\Big(H_e(X,Y)\Big) \le \frac{1}{2}\,\mathbb{E}^*_{X,Y}\,\Phi\Big(2A_e(X)\Big) + \frac{1}{2}\,\mathbb{E}^*_{X,Y}\,\Phi\Big(2A_e(Y)\Big).$$

Cada término de la derecha depende de un solo bloque de coordenadas, y los dos son iguales: ambos son la esperanza exterior, respecto de P^n, de la misma función $x \mapsto \Phi(2A_e(x))$, porque X e Y tienen la misma ley.[23] Por lo tanto,

$$\mathbb{E}^*_{X,Y}\,\Phi\Big(H_e(X,Y)\Big) \le \mathbb{E}^*_{X}\,\Phi\Big(2A_e(X)\Big) \qquad \text{para todo } e \in \{-1,1\}^n. \tag{2.21}$$

Sustituyendo (2.21) en (2.20),

$$\mathbb{E}^*\,\Phi\Big(\|P_n - P\|_{\mathcal{F}}\Big) \le \frac{1}{2^n} \sum_{e \in \{-1,1\}^n} \mathbb{E}^*_X\,\Phi\Big(2A_e(X)\Big) = \mathbb{E}_\varepsilon\,\mathbb{E}^*_X\,\Phi\Big(2A_\varepsilon(X)\Big),$$

porque ε es uniforme en $\{-1,1\}^n$.[24] Por último, el Lema 1.16, ahora en el producto $\{-1,1\}^n \times \mathcal{X}^n$ con ε en el primer factor, da $\mathbb{E}_\varepsilon \mathbb{E}^*_X \le \mathbb{E}^*_{X,\varepsilon}$; y como $A_\varepsilon(X) = \|P^\circ_n\|_{\mathcal{F}}$,

$$\mathbb{E}^*\,\Phi\Big(\|P_n - P\|_{\mathcal{F}}\Big) \le \mathbb{E}^*_{X,\varepsilon}\,\Phi\Big(2A_\varepsilon(X)\Big) = \mathbb{E}^*\,\Phi\Big(2\|P^\circ_n\|_{\mathcal{F}}\Big).$$

Ésta es la desigualdad para $\Phi - \Phi(0)$, y por la reducción del comienzo vale también para Φ. □

Lema de simetrización y medibilidad. El lema de simetrización es válido para cualquier clase de funciones medibles $\mathcal{F}$ para la cual las cantidades anteriores estén definidas. En las pruebas de los teoremas de Glivenko–Cantelli y de Donsker se aplicará no sólo a la clase original, sino también a clases construidas a partir de ella, por ejemplo clases de diferencias pequeñas como $\mathcal{F}_\delta = \{f - g : f, g \in \mathcal{F},\ \|f - g\|_{P,2} < \delta\}$. Luego se aplica una desigualdad maximal al lado derecho del lema, condicionalmente en $X_1, \ldots, X_n$. Para justificar ese paso se quiere calcular la esperanza exterior conjunta del proceso simetrizado mediante Fubini, como una esperanza iterada en los signos y en la muestra. Pero Fubini no vale en general para esperanzas exteriores. Por eso se impone una condición de medibilidad que haga medible el supremo simetrizado.

[22] Por el Lema 1.9(i), $(S+T)^* \le S^* + T^*$, y por el Lema 1.5(iii), $\mathbb{E}^*(S+T) \le \mathbb{E}^*S + \mathbb{E}^*T$ si $S, T \ge 0$. La homogeneidad positiva figura en la Advertencia 1.11.

[23] Por la observación del primer párrafo, $\mathbb{E}^*_{X,Y}\Phi(2A_e(X)) = \mathbb{E}^*_{P^n}\Phi(2A_e)$, y lo mismo con Y en lugar de X, puesto que Y es la proyección sobre otra copia de $(\mathcal{X}^n, \mathcal{A}^n, P^n)$. No se usa ninguna medibilidad de A_e.

[24] En el factor finito $\{-1,1\}^n$, con la σ-álgebra de todas las partes, toda función es medible; por eso $\mathbb{E}^*_\varepsilon = \mathbb{E}_\varepsilon$, y el promedio sobre e es la esperanza en ε de la función $e \mapsto \mathbb{E}^*_X\Phi(2A_e(X))$.

Escribimos $\overline{\mathcal{A}^n}$ para la completación de $\mathcal{A}^n$ respecto de P^n y $\overline{P^n}$ para la extensión de P^n a ella (Definición 3.11 del apéndice) y, como en la prueba del Lema 2.20, R_n para la distribución uniforme sobre $\{-1,1\}^n$. Como el vector de signos toma sólo finitos valores, todos con probabilidad positiva, la medibilidad conjunta de $(x_1,\dots,x_n,\varepsilon_1,\dots,\varepsilon_n)\longmapsto\|\sum_{i=1}^n \varepsilon_i f(x_i)\|_{\mathcal{F}}$ respecto de la completación de $\mathcal{A}^n\otimes 2^{\{-1,1\}^n}$ para $P^n\otimes R_n$ (donde $2^{\{-1,1\}^n}$ es la σ-álgebra de todas las partes de $\{-1,1\}^n$) equivale a la $\overline{\mathcal{A}^n}$-medibilidad, para todo $e=(e_1,\dots,e_n)\in\{-1,1\}^n$, de

$$(x_1,\dots,x_n)\longmapsto\left\|\sum_{i=1}^n e_i f(x_i)\right\|_{\mathcal{F}}=\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n e_i f(x_i)\right|. \tag{2.3.2}$$

En efecto, (2.3.2) es la sección en e de la función anterior, y la equivalencia es el Lema 3.13(c) del apéndice, con $(\Omega,\mathcal{C},\mu)=(\mathcal{X}^n,\mathcal{A}^n,P^n)$, $E=\{-1,1\}^n$ y $Q=R_n$. Para esta aplicación concreta de Fubini basta considerar vectores de signos. La Definición 2.3.3 de [5], que adoptamos a continuación, exige la misma medibilidad para todo vector real y es, por tanto, ligeramente más fuerte.

Definición 2.21. Sea $\mathcal{F}$ una clase de funciones medibles $f:\mathcal{X}\to\mathbb{R}$ en un espacio de probabilidad $(\mathcal{X},\mathcal{A},P)$. Diremos que $\mathcal{F}$ es una **clase P-medible** si, para todo $n\geq 1$ y todo $e=(e_1,\dots,e_n)\in\mathbb{R}^n$, la aplicación $(x_1,\dots,x_n)\longmapsto\sup_{f\in\mathcal{F}}|\sum_{i=1}^n e_i f(x_i)|$ es medible respecto de la completación de $(\mathcal{X}^n,\mathcal{A}^n,P^n)$.

Fubini para el proceso de Rademacher. Sea $\mathcal{F}$ una clase P-medible; de hecho, basta que (2.3.2) sea $\overline{\mathcal{A}^n}$-medible para todo $e\in\{-1,1\}^n$. Denotemos por $h_{n,e}$ la aplicación (2.3.2) y pongamos $S_n(x,e):=h_{n,e}(x)$ para $(x,e)\in\mathcal{X}^n\times\{-1,1\}^n$, de modo que $S_n(X,\varepsilon)=\left\|\sum_{i=1}^n\varepsilon_i f(X_i)\right\|_{\mathcal{F}}$. Entonces la función $x\mapsto\mathbb{E}_\varepsilon S_n(x,\varepsilon)$ es $\overline{\mathcal{A}^n}$-medible y

$$\mathbb{E}^*_{X,\varepsilon}\left\|\sum_{i=1}^n\varepsilon_i f(X_i)\right\|_{\mathcal{F}}=\mathbb{E}_X\mathbb{E}_\varepsilon\left\|\sum_{i=1}^n\varepsilon_i f(X_i)\right\|_{\mathcal{F}}, \tag{FR}$$

donde $\mathbb{E}_X$ es la integral respecto de $\overline{P^n}$.

Demostración. Por lo visto antes de la Definición 2.21, S_n es medible respecto de la completación de $\mathcal{A}^n\otimes 2^{\{-1,1\}^n}$ para $P^n\otimes R_n$; denotemos por $\overline{P^n\otimes R_n}$ la extensión de $P^n\otimes R_n$ a esa completación. Para cada $x\in\mathcal{X}^n$ fijo, la esperanza en los signos es la suma finita

$$\mathbb{E}_\varepsilon S_n(x,\varepsilon)=\frac{1}{2^n}\sum_{e\in\{-1,1\}^n}h_{n,e}(x),$$

que es $\overline{\mathcal{A}^n}$-medible porque cada sumando lo es. El Lema 3.12(iii), aplicado en el espacio producto, y el Lema 3.13(c) dan

$$\mathbb{E}^*_{X,\varepsilon}S_n(X,\varepsilon)=\int S_n\,d\overline{P^n\otimes R_n}=\int_{\mathcal{X}^n}\frac{1}{2^n}\sum_{e\in\{-1,1\}^n}h_{n,e}\,d\overline{P^n}=\mathbb{E}_X\mathbb{E}_\varepsilon S_n(X,\varepsilon).$$

□

Por el Lema 3.12(iii), ahora en $(\mathcal{X}^n,\mathcal{A}^n,P^n)$, la esperanza $\mathbb{E}_X$ de (FR) coincide también con la esperanza exterior $\mathbb{E}^*_X$. Sin la hipótesis de medibilidad, el Lema 1.16, con $\mathbb{E}^*_\varepsilon=\mathbb{E}_\varepsilon$ en el factor finito, sólo da $\mathbb{E}^*_X\mathbb{E}_\varepsilon S_n(X,\varepsilon)\leq\mathbb{E}^*_{X,\varepsilon}S_n(X,\varepsilon)$: lo que aporta la P-medibilidad es la igualdad. Como $S_n\geq 0$, (FR) vale aun cuando ambos miembros son $+\infty$. Si $\mathcal{F}$ tiene una envolvente F con $P^*F<\infty$ (Definición 2.25), ambos son finitos: $S_n(x,e)\leq\sum_{i=1}^n F^*(x_i)$, y el lado derecho es un mayorante medible de S_n con esperanza $nPF^*=nP^*F$ (Lema 1.5(iii)).

2.4. Teoremas de Glivenko–Cantelli

2.4.1. Glivenko-Cantelli por bracketing

Definición 2.22. Sean $l, u : \mathcal{X} \to \mathbb{R}$ medibles, con $l \le u$. El **bracket** $[l, u]$ es la clase de funciones f tales que $l \le f \le u$. Si $\|\cdot\|$ es una seminorma, un **ε-bracket** es un bracket con $\|u - l\| < \varepsilon$. El **número de bracketing** $N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|)$ es el mínimo número de ε-brackets necesarios para cubrir $\mathcal{F}$, y $\log N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|)$ es la entropía por bracketing. Los extremos l, u no tienen por qué pertenecer a $\mathcal{F}$, pero se toman medibles y de norma finita.

Teorema 2.23. *Sea $\mathcal{F}$ una clase de funciones medibles tal que $N_{[]}(\varepsilon, \mathcal{F}, L_1(P)) < \infty$ para todo $\varepsilon > 0$. Entonces $\mathcal{F}$ es P-Glivenko–Cantelli, es decir, $\|P_n - P\|^*_{\mathcal{F}} \to 0$ c.s. Si $\|P_n - P\|_{\mathcal{F}}$ es medible, puede omitirse el asterisco.*

Demostración. Fijemos $\varepsilon > 0$ y tomemos brackets $[l_i, u_i]$, $i = 1, \dots, m$, que cubren $\mathcal{F}$ y satisfacen $P(u_i - l_i) < \varepsilon$. Si $f \in [l_i, u_i]$, entonces $(P_n - P)f \le (P_n - P)u_i + P(u_i - f) \le (P_n - P)u_i + \varepsilon$, pues $0 \le u_i - f \le u_i - l_i$.[25] Análogamente, $(P_n - P)f \ge (P_n - P)l_i - P(f - l_i) \ge (P_n - P)l_i - \varepsilon$. Por tanto, para todo $f \in \mathcal{F}$, $|(P_n - P)f| \le \max_{1 \le i \le m} \max\{|(P_n - P)u_i|,\ |(P_n - P)l_i|\} + \varepsilon$. El máximo del lado derecho es medible y, por la ley fuerte aplicada al conjunto finito de extremos $\{l_i, u_i : i = 1, \dots, m\}$, converge c.s. a 0.[26] Así, $\limsup_{n\to\infty} \|P_n - P\|^*_{\mathcal{F}} \le \varepsilon$ c.s. Como $\varepsilon > 0$ es arbitrario, tomando una sucesión $\varepsilon_k \downarrow 0$ se obtiene $\|P_n - P\|^*_{\mathcal{F}} \to 0$ c.s. No se necesita la hipótesis de P-medibilidad usada en los argumentos por simetrización: aquí sólo intervienen aproximaciones finitas y la ley fuerte. □

Ejemplo 2.24. Sea $\mathcal{F} = \{\mathbb{1}_{(-\infty, c]} : c \in \mathbb{R}\}$. Esta clase tiene números de bracketing finitos en $L_1(P)$ para toda distribución P en $\mathbb{R}$; no hace falta suponer que P tenga densidad. Dado $\varepsilon > 0$, elijamos una grilla $-\infty = t_0 < t_1 < \cdots < t_m = \infty$ tal que $P((t_i, t_{i+1})) < \varepsilon$ para todo i[27]. Entonces los brackets $\left[\mathbb{1}_{(-\infty, t_i]},\ \mathbb{1}_{(-\infty, t_{i+1})}\right]$, $i = 0, \dots, m-1$, cubren $\mathcal{F}$ y tienen tamaño $\left\|\mathbb{1}_{(-\infty, t_{i+1})} - \mathbb{1}_{(-\infty, t_i]}\right\|_{P,1} = P((t_i, t_{i+1})) < \varepsilon$. Luego $N_{[]}(\varepsilon, \mathcal{F}, L_1(P)) < \infty$, y por el Teorema 2.23, $\mathcal{F}$ es P-Glivenko–Cantelli. Si P tiene soporte finito, el resultado también vale: de hecho sólo hay finitas funciones distintas P-c.s., de modo que se puede cubrir con brackets exactos.

2.4.2. Glivenko-Cantelli por cubrimientos

Definición 2.25. Una **envolvente** de una clase $\mathcal{F}$ es una función $F : \mathcal{X} \to [0, \infty]$ tal que $|f(x)| \le F(x)$ para todo $x \in \mathcal{X}$ y todo $f \in \mathcal{F}$. La envolvente mínima es $F(x) = \sup_{f \in \mathcal{F}} |f(x)|$. Usualmente se supone que es finita en todo punto; cuando se integra una envolvente no necesariamente medible, se usa integral exterior, por ejemplo $P^*F < \infty$.

Definición 2.26 (Notación)**.** Para clases de funciones usamos la siguiente notación para los números de cubrimiento de la Definición 2.11. Si $\|\cdot\|$ es una seminorma, $N(\varepsilon, \mathcal{F}, \|\cdot\|)$ es el número de cubrimiento de $\mathcal{F}$ para la semimétrica $\|f - g\|$, con la salvedad de que los centros de las bolas pueden no pertenecer a $\mathcal{F}$ (véase el párrafo siguiente). Si Q es una medida de probabilidad y $r \ge 1$, escribimos $\|f\|_{Q,r} = (Q|f|^r)^{1/r}$ y $N(\varepsilon, \mathcal{F}, L_r(Q))$ para el caso $\|\cdot\| = \|\cdot\|_{Q,r}$; por ejemplo, $N(\varepsilon, \mathcal{F}, L_1(P_n))$ corresponde a la semimétrica $P_n|f - g|$.

Los centros de las bolas no están obligados a pertenecer a $\mathcal{F}$, aunque deben tener norma finita. Cuando un argumento requiera centros en la clase, se usará la siguiente *internalización* elemental:

[25] En detalle: $P_n f \le P_n u_i$ por monotonía de P_n (los extremos son medibles e integrables, parte de la definición de bracket), de donde $(P_n - P)f \le (P_n - P)u_i + P(u_i - f)$, y $P(u_i - f) \le P(u_i - l_i) < \varepsilon$.

[26] Por eso la minimalidad del mayorante da $\|P_n - P\|^*_{\mathcal{F}} \le \max_i\{\cdots\} + \varepsilon$ c.s. para cada n; la unión en n de los conjuntos excepcionales sigue siendo nula.

[27] Esto no exige continuidad de P. Con $K = \lfloor 1/\varepsilon \rfloor$ y los cortes cuantílicos $t_i := \inf\{t : F(t) \ge i/(K+1)\}$, $i = 1, \dots, K$, donde F es la función de distribución de P, la masa *abierta* entre cortes consecutivos es $P((t_i, t_{i+1})) \le 1/(K+1) < \varepsilon$: por la continuidad a derecha de F, $F(t_i) \ge i/(K+1)$, mientras que $F(t_{i+1}^-) \le (i+1)/(K+1)$. Los átomos, en particular los de masa $\ge \varepsilon$, quedan absorbidos en los puntos de corte; si algunos t_i coinciden, se descartan las repeticiones.

dado un cubrimiento por bolas de radio ε con centros arbitrarios, eligiendo en cada bola que corta a $\mathcal{F}$ un elemento de la intersección se obtiene, por la desigualdad triangular, un cubrimiento con centros en $\mathcal{F}$, de radio 2ε y con a lo sumo la misma cantidad de bolas.

Definición 2.27. Sea $(\mathcal{G}_n)_{n\geq 1}$ una filtración decreciente, es decir $\mathcal{G}_{n+1} \subset \mathcal{G}_n$. Una sucesión integrable (Y_n) es una **martingala inversa** si Y_n es $\mathcal{G}_n$-medible y $\mathbb{E}(Y_n \mid \mathcal{G}_{n+1}) = Y_{n+1}$ c.s. Es una **submartingala inversa** si $\mathbb{E}(Y_n \mid \mathcal{G}_{n+1}) \geq Y_{n+1}$ c.s., y una **supermartingala inversa** si la desigualdad es la opuesta.

Para pasar a una submartingala usual se indexa con los enteros negativos: para $t \in \{\ldots, -3, -2, -1\}$ sean $M_t = Y_{-t}$, $\mathcal{F}_t = \mathcal{G}_{-t}$. Si $t < s$ entonces $-t > -s$ y $\mathcal{F}_t = \mathcal{G}_{-t} \subset \mathcal{G}_{-s} = \mathcal{F}_s$.

Además, M_t es $\mathcal{F}_t$-medible, y para $t \leq -2$, con $n = -t - 1$, $\mathbb{E}(M_{t+1} \mid \mathcal{F}_t) = \mathbb{E}(Y_n \mid \mathcal{G}_{n+1}) \geq Y_{n+1} = M_t$. $(M_t, \mathcal{F}_t)_{t\leq -1}$ es una submartingala usual indexada por los enteros negativos, y $n \to \infty$ corresponde a $t \to -\infty$.

Teorema 2.28 (Convergencia de submartingalas inversas)**.** *Sea $(\mathcal{G}_n)_{n\geq 1}$ una sucesión decreciente de sub-σ-álgebras de un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$, y sea $\mathcal{G}_\infty := \bigcap_n \mathcal{G}_n$. Sean Y_n variables integrables, con Y_n $\mathcal{G}_n$-medible y $\mathbb{E}(Y_n \mid \mathcal{G}_{n+1}) \geq Y_{n+1}$ c.s. para todo n. Si $\inf_n \mathbb{E}Y_n > -\infty$, existe una variable integrable y $\mathcal{G}_\infty$-medible Y_∞ tal que $Y_n \to Y_\infty$ c.s. y en L^1.*

Teorema 2.29 (Ley 0–1 de Hewitt–Savage)**.** *Si P es una probabilidad sobre $(\mathcal{X}, \mathcal{A})$, todo $A \in \mathcal{A}^\infty$ invariante bajo las permutaciones de $\mathbb{N}$ que mueven sólo un número finito de coordenadas cumple $P^\infty(A) \in \{0, 1\}$.*

Sean $X_1, X_2, \ldots$ i.i.d. con $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$, tomadas como las coordenadas de $\{-1, 1\}^{\mathbb{N}}$, sea $S_n = X_1 + \cdots + X_n$ y consideremos $A = \{S_n = 0 \text{ para infinitos } n\}$. El evento A es invariante por permutaciones finitas: si π permuta sólo las primeras m coordenadas, S_n no cambia para $n \geq m$, y como A depende sólo de los valores de S_n para n grande, A no cambia. La ley 0–1 de Hewitt–Savage se aplica y da $\mathbb{P}(A) \in \{0, 1\}$; de hecho $\mathbb{P}(A) = 1$, por la recurrencia del paseo simple.

A no es un evento cola. Si A estuviera en $\sigma(X_2, X_3, \ldots)$, cambiar sólo la primera coordenada no podría sacar a un punto de A. Pero $\omega = (1, -1, 1, -1, \ldots)$ cumple $S_{2k}(\omega) = 0$ para todo k, así que $\omega \in A$, mientras que $\omega' = (-1, -1, 1, -1, 1, \ldots)$, que difiere de ω sólo en la primera coordenada, cumple $S_n(\omega') = S_n(\omega) - 2 \in \{-2, -1\}$ para todo $n \geq 1$, de modo que $\omega' \notin A$. Por lo tanto, A no pertenece a la σ-álgebra cola, y la ley 0–1 de Kolmogorov no se aplica directamente.

Lema 2.30. *Sea $\mathcal{F}$ una clase de funciones medibles con envolvente F tal que $P^*F < \infty$. Trabajamos en el espacio producto $(\mathcal{X}^\infty, \mathcal{A}^\infty, P^\infty)$, donde P_n es la medida empírica de las n primeras proyecciones coordenadas. Para cada n, sea $\Sigma_n \subset \mathcal{A}^\infty$ la σ-álgebra generada por todas las funciones $\mathcal{A}^\infty$-medibles $h : \mathcal{X}^\infty \to \mathbb{R}$ simétricas en sus primeros n argumentos. Entonces $\Sigma_{n+1} \subset \Sigma_n$ y*

$$\mathbb{E}\Big(\|P_n - P\|_{\mathcal{F}}^* \mid \Sigma_{n+1}\Big) \;\geq\; \|P_{n+1} - P\|_{\mathcal{F}}^* \qquad P^\infty\text{-c.s.}$$

Además, existen versiones de los mayorantes medibles mínimos $\|P_n - P\|_{\mathcal{F}}^*$ simétricas en sus primeros n argumentos, y por tanto Σ_n-medibles. Cualesquiera de tales versiones forman una submartingala inversa respecto de $(\Sigma_n)_{n\geq 1}$ y convergen casi seguramente a una constante.

Demostración. Denotamos por $X_i \colon \mathcal{X}^\infty \to \mathcal{X}$ la i-ésima proyección coordenada. Bajo P^∞, las X_i son i.i.d. con ley P, y $P_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$. Salvo que se indique otra cosa, "medible" para funciones definidas sobre $\mathcal{X}^\infty$ significa $\mathcal{A}^\infty$-medible, y las esperanzas, las afirmaciones casi seguras y los mayorantes medibles mínimos de esas funciones se entienden respecto de P^∞. Escribimos

$$\Delta_n := \|P_n - P\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |P_n f - Pf|,$$

que es una función de $\mathcal{X}^\infty$ en $[0, \infty]$, no necesariamente medible.

Para $m \geq 1$, identificamos cada permutación $\pi \in \mathfrak{S}_m$ con la transformación de $\mathcal{X}^\infty$ que permuta las primeras m coordenadas y deja fijas las restantes:

$$\pi(x_1, x_2, \ldots) := (x_{\pi(1)}, \ldots, x_{\pi(m)}, x_{m+1}, x_{m+2}, \ldots).$$

Estas transformaciones forman un grupo bajo composición. En particular, para cada $\sigma \in \mathfrak{S}_m$ fija, la aplicación $\pi \mapsto \pi \circ \sigma$ es una biyección de $\mathfrak{S}_m$ en sí mismo. Cada π es bimedible, porque las coordenadas de $\pi(x)$ son coordenadas de x. Además preserva P^∞, porque $P^\infty \circ \pi^{-1}$ y P^∞ coinciden sobre los rectángulos medibles, que forman una clase cerrada por intersecciones finitas que genera $\mathcal{A}^\infty$. Por lo tanto, para toda función medible $g \geq 0$,

$$\int g \circ \pi \, dP^\infty = \int g \, dP^\infty. \tag{1}$$

Una función h sobre $\mathcal{X}^\infty$ es simétrica en sus primeras m coordenadas si $h \circ \pi = h$ para toda $\pi \in \mathfrak{S}_m$.

Comenzamos describiendo Σ_m mediante conjuntos. Sea $\mathcal{S}_m$ la clase de los $A \in \mathcal{A}^\infty$ tales que $\pi^{-1}(A) = A$ para toda $\pi \in \mathfrak{S}_m$; veamos que $\Sigma_m = \mathcal{S}_m$. La clase $\mathcal{S}_m$ es una σ-álgebra, porque las preimágenes conmutan con complementos y con uniones numerables. Si h es una de las funciones que generan Σ_m y $B \subset \mathbb{R}$ es boreliano, entonces $h^{-1}(B) \in \mathcal{A}^\infty$ y, para toda $\pi \in \mathfrak{S}_m$,

$$\pi^{-1}\Big(h^{-1}(B)\Big) = (h \circ \pi)^{-1}(B) = h^{-1}(B),$$

así que $h^{-1}(B) \in \mathcal{S}_m$. Como Σ_m es la menor σ-álgebra que contiene a todos estos conjuntos, $\Sigma_m \subset \mathcal{S}_m$. Recíprocamente, si $A \in \mathcal{S}_m$, la función $\mathbf{1}_A$ es medible y $\mathbf{1}_A \circ \pi = \mathbf{1}_{\pi^{-1}(A)} = \mathbf{1}_A$ para toda $\pi \in \mathfrak{S}_m$. Es decir, $\mathbf{1}_A$ es una de las funciones generadoras, y por lo tanto $A = \{\mathbf{1}_A = 1\} \in \Sigma_m$.

De esta caracterización se siguen tres hechos que usaremos. (a) $\Sigma_{m+1} \subset \Sigma_m$. En efecto, toda función medible simétrica en las primeras $m+1$ coordenadas también es simétrica en las primeras m: basta considerar las permutaciones que dejan fija la coordenada $m+1$. Por lo tanto, las funciones que generan Σ_{m+1} están incluidas entre las que generan Σ_m, lo que prueba la inclusión. (b) Toda función medible $h\colon \mathcal{X}^\infty \to [-\infty, \infty]$ simétrica en sus primeras m coordenadas es Σ_m-medible. Si h es real, esto es la definición de Σ_m. El caso general, que necesitamos porque los mayorantes medibles mínimos pueden tomar el valor $+\infty$, sale de que para cada $t \in \mathbb{R}$ el conjunto $\{h > t\}$ está en $\mathcal{A}^\infty$ y cumple $\pi^{-1}(\{h > t\}) = \{h \circ \pi > t\} = \{h > t\}$, así que pertenece a $\mathcal{S}_m = \Sigma_m$. (c) Si $A \in \Sigma_m$ y $\pi \in \mathfrak{S}_m$, entonces $\mathbf{1}_A \circ \pi = \mathbf{1}_A$.

Construimos ahora versiones simétricas de los mayorantes medibles mínimos. Primero, Δ_n es simétrica en sus primeras n coordenadas. Si $\pi \in \mathfrak{S}_n$ y $f \in \mathcal{F}$, como $j \mapsto \pi(j)$ es una biyección de $\{1, \ldots, n\}$,

$$P_n f(\pi(x)) = \frac{1}{n} \sum_{j=1}^{n} f(x_{\pi(j)}) = \frac{1}{n} \sum_{j=1}^{n} f(x_j) = P_n f(x),$$

y tomando supremo en f resulta $\Delta_n \circ \pi = \Delta_n$. Sea $\widetilde{\Delta}_n$ un mayorante medible mínimo de Δ_n, que existe por el Lema 1.5, y definamos

$$Y_n := \min_{\pi \in \mathfrak{S}_n} \widetilde{\Delta}_n \circ \pi.$$

La función Y_n es medible, por ser el mínimo de una cantidad finita de funciones medibles. Es simétrica en sus primeras n coordenadas, porque si $\sigma \in \mathfrak{S}_n$,

$$Y_n \circ \sigma = \min_{\pi \in \mathfrak{S}_n} \widetilde{\Delta}_n \circ (\pi \circ \sigma) = \min_{\rho \in \mathfrak{S}_n} \widetilde{\Delta}_n \circ \rho = Y_n,$$

donde la segunda igualdad se debe a que $\rho = \pi \circ \sigma$ recorre todo $\mathfrak{S}_n$ cuando π lo hace, de modo que se minimiza sobre la misma familia de funciones. Domina a Δ_n en todo punto: como $\widetilde{\Delta}_n \geq \Delta_n$, para cada π se tiene $\widetilde{\Delta}_n \circ \pi \geq \Delta_n \circ \pi = \Delta_n$. Por último, tomando $\pi = \mathrm{id}$ se obtiene $Y_n \leq \widetilde{\Delta}_n$. Como Y_n es un mayorante medible de Δ_n, la minimalidad de $\widetilde{\Delta}_n$ da $\widetilde{\Delta}_n \leq Y_n$ c.s., y así

$$Y_n = \widetilde{\Delta}_n \qquad \text{c.s.}$$

En consecuencia, Y_n es también un mayorante medible mínimo de Δ_n, es decir, una versión de $\|P_n - P\|_{\mathcal{F}}^*$: si U es medible y $U \geq \Delta_n$, entonces $Y_n = \widetilde{\Delta}_n \leq U$ c.s. Además, por (b), Y_n es Σ_n-medible.

Las variables Y_n son integrables. Sea F^* un mayorante medible mínimo de F respecto de P (Lema 1.5), de modo que $PF^* = P^*F < \infty$. Si $f \in \mathcal{F}$, entonces f es medible y $|f| \leq F \leq F^*$, y por lo tanto

$$|P_n f| \leq P_n|f| \leq P_n F^*, \qquad |Pf| \leq P|f| \leq PF^* = P^*F.$$

Luego $|P_n f - Pf| \leq P_n F^* + P^*F$ para toda $f \in \mathcal{F}$, y tomando supremo,

$$\Delta_n \leq P_n F^* + P^*F.$$

El lado derecho es medible, porque cada $F^*(X_i) = F^* \circ X_i$ lo es. Como cada X_i tiene ley P,

$$\mathbb{E}\Big(P_n F^* + P^*F\Big) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}F^*(X_i) + P^*F = PF^* + P^*F = 2P^*F.$$

Por la minimalidad de Y_n, $0 \leq Y_n \leq P_n F^* + P^*F$ c.s., y así $\mathbb{E}Y_n \leq 2P^*F < \infty$.

Para calcular esperanzas condicionales respecto de Σ_m usaremos el siguiente hecho. Si $H\colon \mathcal{X}^\infty \to [0,\infty]$ es medible e integrable, su promedio sobre las permutaciones

$$T_m H := \frac{1}{m!}\sum_{\pi \in \mathfrak{S}_m} H \circ \pi$$

es una versión de $\mathbb{E}(H \mid \Sigma_m)$. En efecto, por (1) cada $H \circ \pi$ es integrable con $\mathbb{E}(H \circ \pi) = \mathbb{E}H$, así que $T_m H$ es integrable. Además $T_m H$ es medible y simétrica en sus primeras m coordenadas: si $\sigma \in \mathfrak{S}_m$,

$$(T_m H) \circ \sigma = \frac{1}{m!}\sum_{\pi \in \mathfrak{S}_m} H \circ (\pi \circ \sigma) = \frac{1}{m!}\sum_{\rho \in \mathfrak{S}_m} H \circ \rho = T_m H,$$

otra vez porque $\pi \mapsto \pi \circ \sigma$ es una biyección de $\mathfrak{S}_m$. Por (b), $T_m H$ es Σ_m-medible. Por último, si $A \in \Sigma_m$ y $\pi \in \mathfrak{S}_m$,

$$\int_A H \circ \pi \, dP^\infty = \int \underbrace{\mathbf{1}_A(x)}_{=\mathbf{1}_A(\pi(x))} H(\pi(x))\, dP^\infty(x) = \int \mathbf{1}_A(\pi(x))\, H(\pi(x))\, dP^\infty(x)$$

$$= \int (\mathbf{1}_A H) \circ \pi \, dP^\infty = \int \mathbf{1}_A(x)\, H(x)\, dP^\infty(x) = \int_A H \, dP^\infty.$$

donde la primera igualdad usa (c), y la segunda es (1). Promediando sobre $\pi \in \mathfrak{S}_m$,

$$\int_A T_m H \, dP^\infty = \int_A H \, dP^\infty \qquad \text{para todo } A \in \Sigma_m.$$

Así, $T_m H$ es Σ_m-medible, integrable y tiene las mismas integrales que H sobre los conjuntos de Σ_m. Por la definición de esperanza condicional, $T_m H = \mathbb{E}(H \mid \Sigma_m)$ c.s.

Ahora relacionamos Δ_{n+1} con Δ_n mediante este promedio, con $m = n+1$. Fijemos $f \in \mathcal{F}$ y miremos $P_n f$ como función sobre $\mathcal{X}^\infty$. Para $\pi \in \mathfrak{S}_{n+1}$ se tiene $(P_n f)(\pi(x)) = n^{-1}\sum_{j=1}^n f(x_{\pi(j)})$, y por lo tanto

$$\frac{1}{(n+1)!}\sum_{\pi \in \mathfrak{S}_{n+1}} (P_n f)(\pi(x)) = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{1}{(n+1)!}\sum_{\pi \in \mathfrak{S}_{n+1}} f(x_{\pi(j)})\right).$$

Para $j \leq n$ y $k \leq n+1$ fijos hay exactamente $n!$ permutaciones $\pi \in \mathfrak{S}_{n+1}$ con $\pi(j) = k$: las que mandan los n índices restantes biyectivamente sobre los n valores restantes. Agrupando según el valor de $\pi(j)$ se obtiene, para cada $j \leq n$,

$$\frac{1}{(n+1)!}\sum_{\pi \in \mathfrak{S}_{n+1}} f(x_{\pi(j)}) = \frac{n!}{(n+1)!}\sum_{k=1}^{n+1} f(x_k) = P_{n+1} f(x),$$

y promediando en j,

$$P_{n+1}f = \frac{1}{(n+1)!}\sum_{\pi\in\mathfrak{S}_{n+1}} (P_n f)\circ\pi.$$

Como Pf es una constante, coincide con su propio promedio, y restando,

$$(P_{n+1}-P)f = \frac{1}{(n+1)!}\sum_{\pi\in\mathfrak{S}_{n+1}} \Big((P_n-P)f\Big)\circ\pi.$$

Por la desigualdad triangular, y como $|(P_n-P)f|\le\Delta_n$ en todo punto,

$$\Big|(P_{n+1}-P)f\Big| \le \frac{1}{(n+1)!}\sum_{\pi\in\mathfrak{S}_{n+1}} \Delta_n\circ\pi.$$

El lado derecho no depende de f. Tomando supremo sobre $f\in\mathcal{F}$ y usando que $\Delta_n\le Y_n$ en todo punto,

$$\Delta_{n+1} \le \frac{1}{(n+1)!}\sum_{\pi\in\mathfrak{S}_{n+1}} \Delta_n\circ\pi \le \frac{1}{(n+1)!}\sum_{\pi\in\mathfrak{S}_{n+1}} Y_n\circ\pi = T_{n+1}Y_n.$$

La función $T_{n+1}Y_n$ es medible, así que es un mayorante medible de Δ_{n+1}, y la minimalidad de Y_{n+1} da $Y_{n+1}\le T_{n+1}Y_n$ c.s. Por otro lado, Y_n es medible, no negativa e integrable, así que el hecho anterior se aplica con $H=Y_n$ y $m=n+1$ y da $T_{n+1}Y_n=\mathbb{E}(Y_n\mid\Sigma_{n+1})$ c.s. Combinando ambas,

$$Y_{n+1}\le\mathbb{E}(Y_n\mid\Sigma_{n+1}) \qquad \text{c.s.},$$

que es la desigualdad del enunciado.

Pasamos a la convergencia, que obtendremos del Teorema 2.28 con $\mathcal{G}_n=\Sigma_n$. Verifiquemos sus hipótesis. Por (a), $(\Sigma_n)_{n\ge1}$ es una sucesión decreciente de sub-σ-álgebras de $\mathcal{A}^\infty$. Cada Y_n es integrable y Σ_n-medible, y acabamos de probar que $\mathbb{E}(Y_n\mid\Sigma_{n+1})\ge Y_{n+1}$ c.s.; es decir, $(Y_n,\Sigma_n)_{n\ge1}$ es una submartingala inversa. Por último, $\inf_n\mathbb{E}Y_n\ge0>-\infty$, porque $Y_n\ge0$. [28] El teorema da entonces una variable integrable Y_∞, medible respecto de

$$\Sigma_\infty := \bigcap_{n\ge1}\Sigma_n,$$

tal que $Y_n\to Y_\infty$ c.s. y en L^1.

Veamos que Σ_∞ es trivial. Sea $A\in\Sigma_\infty$ y sea π una permutación de $\mathbb{N}$ que mueve sólo un número finito de coordenadas. Entonces existe m tal que π sólo mueve coordenadas entre 1 y m, es decir, $\pi\in\mathfrak{S}_m$, y como $A\in\Sigma_m=\mathcal{S}_m$, se tiene $\pi^{-1}(A)=A$. Así, A es un conjunto de $\mathcal{A}^\infty$ invariante bajo las permutaciones que mueven sólo un número finito de coordenadas, y el Teorema 2.29 da $P^\infty(A)\in\{0,1\}$.

De aquí se sigue que Y_∞ es constante c.s. En efecto, la constante $\mathbb{E}Y_\infty$ es Σ_∞-medible e integrable, y tiene las mismas integrales que Y_∞ sobre los conjuntos de Σ_∞: si $A\in\Sigma_\infty$ y $P^\infty(A)=0$, ambas integrales sobre A son nulas, y si $P^\infty(A)=1$, ambas valen $\mathbb{E}Y_\infty$. Por lo tanto $\mathbb{E}(Y_\infty\mid\Sigma_\infty)=\mathbb{E}Y_\infty$ c.s. Pero Y_∞ es Σ_∞-medible, así que también $\mathbb{E}(Y_\infty\mid\Sigma_\infty)=Y_\infty$ c.s., y en consecuencia

$$Y_\infty = c := \mathbb{E}Y_\infty \qquad \text{c.s.}$$

La constante c es finita y no negativa, y $Y_n\to c$ c.s. Además, por la convergencia en L^1, $|\mathbb{E}Y_n-c|\le\mathbb{E}|Y_n-Y_\infty|\to0$, de modo que $c=\lim_n\mathbb{E}Y_n=\lim_n\mathbb{E}^*\|P_n-P\|_{\mathcal{F}}$.

Finalmente, el resultado no depende de las versiones elegidas. Dos mayorantes medibles mínimos de una misma función coinciden c.s., porque la minimalidad de cada uno da que es menor o igual

[28] Las Y_n pueden valer $+\infty$ en un conjunto nulo. Si se quieren variables reales, basta reemplazar cada Y_n por $Y_n\mathbf{1}_{\{Y_n<\infty\}}$, que es Σ_n-medible porque $\{Y_n<\infty\}\in\Sigma_n$, y coincide c.s. con Y_n porque Y_n es integrable. Estas variables cumplen las mismas hipótesis, y como coinciden con las Y_n fuera de un conjunto nulo, la conclusión vale también para las Y_n.

que el otro c.s. Sea (Y'_n) otra sucesión de versiones de los $\|P_n - P\|^*_{\mathcal{F}}$ simétricas en sus primeras n coordenadas. Por (b), cada Y'_n es Σ_n-medible, y $Y'_n = Y_n$ c.s.; en particular, Y'_n es integrable. Como variables integrables iguales c.s. tienen la misma esperanza condicional,

$$\mathbb{E}(Y'_n \mid \Sigma_{n+1}) = \mathbb{E}(Y_n \mid \Sigma_{n+1}) \geq Y_{n+1} = Y'_{n+1} \qquad \text{c.s.},$$

así que (Y'_n, Σ_n) también es una submartingala inversa. Además, el conjunto $\bigcup_n \{Y'_n \neq Y_n\}$ es nulo, por ser unión numerable de conjuntos nulos, y fuera de él las dos sucesiones coinciden, de modo que $Y'_n \to c$ c.s. El mismo argumento muestra que la desigualdad del enunciado vale para cualquier elección de versiones de los mayorantes, sean o no simétricas. □

Teorema 2.31. *Sea $\mathcal{F}$ una clase P-medible de funciones medibles, con envolvente F tal que $P^*F < \infty$. Para $M > 0$, sea $\mathcal{F}_M = \{f\mathbb{1}_{\{F \leq M\}} : f \in \mathcal{F}\}$. Entonces son equivalentes:*

(i) $\|P_n - P\|^*_{\mathcal{F}} \to 0$ *c.s.;*

(ii) para todo $\varepsilon > 0$ *y todo* $M > 0$, $\frac{1}{n} \log N(\varepsilon, \mathcal{F}_M, L_1(P_n)) \xrightarrow{\mathbb{P}^*} 0$.

Cuando se cumplen, ambas convergencias valen también en media exterior. En particular, es suficiente que, para todo $\varepsilon > 0$, $\log N(\varepsilon, \mathcal{F}, L_1(P_n)) = o^*_{\mathbb{P}}(n)$.

Demostración de la suficiencia. Supongamos (ii). Probaremos primero la convergencia en media exterior, $\mathbb{E}^*\|P_n - P\|_{\mathcal{F}} \to 0$, y de ella deduciremos la convergencia casi segura mediante el Lema 2.30.

Podemos suponer que la envolvente F es medible. En efecto, sea $H := F^*$ su mayorante medible mínimo (Lema 1.5), que es una envolvente medible de $\mathcal{F}$ con $PH = P^*F < \infty$.[29] La afirmación (i) sólo depende de la clase $\mathcal{F}$, no de la envolvente, así que si (ii) vale también con H en lugar de F, es decir, para la clase $\mathcal{H}_M := \{f\mathbb{1}_{\{H \leq M\}} : f \in \mathcal{F}\}$, podemos hacer el resto de la prueba con H. La ventaja es que, al ser H medible, $\mathcal{H}_M$ resulta una clase P-medible de funciones medibles, que es lo que se necesita más abajo para aplicar (FR) a la clase truncada.[30] Veamos ambas cosas.

Como $H \geq F$, si $H(x) \leq M$ entonces $F(x) \leq M$; es decir, $\{H \leq M\} \subset \{F \leq M\}$, y por lo tanto $f\mathbb{1}_{\{H \leq M\}} = (f\mathbb{1}_{\{F \leq M\}})\mathbb{1}_{\{H \leq M\}}$: los elementos de $\mathcal{H}_M$ son los de $\mathcal{F}_M$ multiplicados por el indicador $\mathbb{1}_{\{H \leq M\}}$. Ahora bien, multiplicar todos los elementos de una clase por un mismo indicador $\mathbb{1}_A$ no aumenta sus números de cubrimiento en $L_1(P_n)$. En efecto, $|g\mathbb{1}_A - c\mathbb{1}_A| = |g - c|\mathbb{1}_A \leq |g - c|$ punto a punto, de donde $P_n|g\mathbb{1}_A - c\mathbb{1}_A| \leq P_n|g - c|$; por lo tanto, si las bolas de radio ε con centros $c_1, \ldots, c_N$ cubren la clase, las bolas del mismo radio con centros $c_1\mathbb{1}_A, \ldots, c_N\mathbb{1}_A$ cubren la clase multiplicada. Con $A = \{H \leq M\}$ resulta $N(\varepsilon, \mathcal{H}_M, L_1(P_n)) \leq N(\varepsilon, \mathcal{F}_M, L_1(P_n))$, y la condición de (ii) pasa de $\mathcal{F}_M$ a $\mathcal{H}_M$.[31]

Cada $f\mathbb{1}_{\{H \leq M\}}$ es medible porque f y H lo son. Para la P-medibilidad, fijemos n y $e \in \mathbb{R}^n$, y para $I \subset \{1, \ldots, n\}$ sea $A_I := \{x \in \mathcal{X}^n : \ H(x_i) \leq M \text{ si y sólo si } i \in I\}$. Los 2^n conjuntos A_I son medibles, por serlo H, y forman una partición de $\mathcal{X}^n$: cada x está en el de $I = \{i : H(x_i) \leq M\}$. Si $x \in A_I$, entonces $\mathbb{1}_{\{H(x_i) \leq M\}} = \mathbb{1}_{\{i \in I\}}$ para cada i, y por lo tanto

$$\sup_{f \in \mathcal{F}} \left| \sum_{i=1}^n e_i f(x_i) \mathbb{1}_{\{H(x_i) \leq M\}} \right| = \sup_{f \in \mathcal{F}} \left| \sum_{i \in I} e_i f(x_i) \right|.$$

El lado derecho es la aplicación de la Definición 2.21 para $\mathcal{F}$ con el vector que anula las coordenadas de e fuera de I, y es por lo tanto medible respecto de la completación de P^n. Como el lado izquierdo

[29] Por el Lema 1.5(i), $H \geq F \geq |f|$ para toda $f \in \mathcal{F}$, así que H es una envolvente; por el Lema 1.5(iii), que se aplica porque $H \geq 0$, $PH = P^*F$. En particular, $H < \infty$ P-c.s.

[30] En otras palabras, se prueba la cadena (ii) para $\mathcal{F}_M \Rightarrow$ (ii) para $\mathcal{H}_M \Rightarrow$ (i), donde la segunda implicación es el resto de la demostración hecho con H en lugar de F; las demás hipótesis del teorema valen para H, porque $\mathcal{F}$ es la misma clase y $PH < \infty$. Con F no medible, en cambio, el conjunto $\{F \leq M\}$ puede no ser medible, y entonces $f\mathbb{1}_{\{F \leq M\}}$ tampoco.

[31] Los nuevos centros tienen norma finita, pues $P_n|c_j\mathbb{1}_A| \leq P_n|c_j|$, y el argumento no usa que A sea medible. La condición pasa porque, para todo $\eta > 0$, $\{n^{-1} \log N(\varepsilon, \mathcal{H}_M, L_1(P_n)) > \eta\} \subset \{n^{-1} \log N(\varepsilon, \mathcal{F}_M, L_1(P_n)) > \eta\}$, y la probabilidad exterior es monótona.

coincide con una función de ese tipo sobre cada una de las piezas medibles A_I, también es medible respecto de la completación;[32] esto es la P-medibilidad de $\mathcal{H}_M$. En lo que sigue suponemos, entonces, que F es medible; en particular $PF < \infty$, cada $\mathcal{F}_M$ es una clase P-medible de funciones medibles, y $|g| \le M$ para toda $g \in \mathcal{F}_M$, pues $|f\mathbb{1}_{\{F\le M\}}| \le F\mathbb{1}_{\{F\le M\}} \le M$.

Por el Lema 2.20 con $\Phi(x) = x$ (su hipótesis se cumple porque $P|f| \le P^*F < \infty$) y por (FR), multiplicada por $1/n$,[33]

$$\mathbb{E}^*\|P_n - P\|_{\mathcal{F}} \le 2\,\mathbb{E}^*\|P_n^\circ\|_{\mathcal{F}} = 2\,\mathbb{E}_X\mathbb{E}_\varepsilon\|P_n^\circ\|_{\mathcal{F}}.$$

Fijemos $M > 0$. Cada $f \in \mathcal{F}$ se escribe como $f = f\mathbb{1}_{\{F\le M\}} + f\mathbb{1}_{\{F>M\}}$, donde el primer sumando está en $\mathcal{F}_M$ y el segundo está acotado en valor absoluto por $F\mathbb{1}_{\{F>M\}}$. Como $|\varepsilon_i| = 1$, para cualquier valor de la muestra y de los signos,

$$\begin{aligned}|P_n^\circ f| &\le \left|P_n^\circ(f\mathbb{1}_{\{F\le M\}})\right| + \frac{1}{n}\sum_{i=1}^n |f(X_i)|\mathbb{1}_{\{F(X_i)>M\}}\\ &\le \|P_n^\circ\|_{\mathcal{F}_M} + P_n(F\mathbb{1}_{\{F>M\}}).\end{aligned}$$

El lado derecho no depende de f; tomando supremo en $f \in \mathcal{F}$,

$$\|P_n^\circ\|_{\mathcal{F}} \le \|P_n^\circ\|_{\mathcal{F}_M} + P_n(F\mathbb{1}_{\{F>M\}}).$$

Promediando en los signos e integrando en la muestra,[34]

$$\begin{aligned}\mathbb{E}_X\mathbb{E}_\varepsilon\|P_n^\circ\|_{\mathcal{F}} &\le \mathbb{E}_X\mathbb{E}_\varepsilon\|P_n^\circ\|_{\mathcal{F}_M} + \mathbb{E}\,P_n(F\mathbb{1}_{\{F>M\}})\\ &= \mathbb{E}_X\mathbb{E}_\varepsilon\|P_n^\circ\|_{\mathcal{F}_M} + P(F\mathbb{1}_{\{F>M\}}),\end{aligned}$$

y junto con la desigualdad de simetrización,

$$\mathbb{E}^*\|P_n - P\|_{\mathcal{F}} \le 2\,\mathbb{E}_X\mathbb{E}_\varepsilon\|P_n^\circ\|_{\mathcal{F}_M} + 2P(F\mathbb{1}_{\{F>M\}}). \tag{2.22}$$

El último término tiende a 0 cuando $M \to \infty$, por convergencia dominada.[35] Basta probar entonces que $\mathbb{E}_X\mathbb{E}_\varepsilon\|P_n^\circ\|_{\mathcal{F}_M} \to 0$ para cada M fijo.

Fijemos $\delta > 0$ y una realización de la muestra $X_1, \ldots, X_n$. Por la internalización que sigue a la Definición 2.26, existe $\mathcal{G} \subset \mathcal{F}_M$ con $|\mathcal{G}| \le N(\delta/2, \mathcal{F}_M, L_1(P_n))$ tal que cada $f \in \mathcal{F}_M$ está a distancia $P_n|f - g_f| \le \delta$ de algún $g_f \in \mathcal{G}$.[36] Para cualquier valor de los signos, como $|\varepsilon_i| = 1$,

$$|P_n^\circ f| \le |P_n^\circ g_f| + \frac{1}{n}\sum_{i=1}^n |f(X_i) - g_f(X_i)| \le \max_{g\in\mathcal{G}} |P_n^\circ g| + \delta,$$

y tomando supremo en $f \in \mathcal{F}_M$ y promediando en los signos,

$$\mathbb{E}_\varepsilon\|P_n^\circ\|_{\mathcal{F}_M} \le \mathbb{E}_\varepsilon \max_{g\in\mathcal{G}} |P_n^\circ g| + \delta.$$

[32] Si φ coincide con φ_I sobre cada A_I, entonces $\{\varphi \in B\} = \bigcup_I \big(A_I \cap \{\varphi_I \in B\}\big)$ para todo boreliano B, que es una unión finita de conjuntos de la completación.

[33] La esperanza exterior de $\|P_n^\circ\|_{\mathcal{F}}$ es la conjunta en la muestra y los signos, como en el Lema 2.20. La igualdad es (FR), que se aplica porque $\mathcal{F}$ es P-medible, dividida por n gracias a la homogeneidad positiva de $\mathbb{E}^*$ (Advertencia 1.11).

[34] Las funciones $\mathbb{E}_\varepsilon\|P_n^\circ\|_{\mathcal{F}}$ y $\mathbb{E}_\varepsilon\|P_n^\circ\|_{\mathcal{F}_M}$ de la muestra son medibles respecto de la completación por el resultado de Fubini para el proceso de Rademacher (FR), aplicado a las clases P-medibles $\mathcal{F}$ y $\mathcal{F}_M$, y $P_n(F\mathbb{1}_{\{F>M\}})$ es medible porque F lo es; por eso se puede integrar término a término. La igualdad se debe a que $\mathbb{E}\,P_n(F\mathbb{1}_{\{F>M\}}) = n^{-1}\sum_{i=1}^n \mathbb{E}\big[F(X_i)\mathbb{1}_{\{F(X_i)>M\}}\big]$ y cada X_i tiene ley P.

[35] Se tiene $0 \le F\mathbb{1}_{\{F>M\}} \le F$, con $PF < \infty$, y $F(x)\mathbb{1}_{\{F(x)>M\}} = 0$ en cuanto $M \ge F(x)$, de modo que $F\mathbb{1}_{\{F>M\}} \to 0$ en todo punto donde $F < \infty$, es decir, P-c.s.

[36] Se aplica a un cubrimiento mínimo de radio $\delta/2$: los centros pasan a estar en $\mathcal{F}_M$ y el radio se duplica. (Si ese número de cubrimiento fuera infinito, la cota para $\mathbb{E}_\varepsilon\|P_n^\circ\|_{\mathcal{F}_M}$ que se obtiene abajo sería trivial.)

Para acotar el máximo, denotemos por $\|\cdot\|_{\psi_2|X}$ la norma ψ_2 en el espacio de los signos $(\{-1,1\}^n, R_n)$, con la muestra fija. Entonces

$$\begin{aligned}\mathbb{E}_\varepsilon \max_{g\in\mathcal{G}} |P_n^\circ g| &\le 2\left\|\max_{g\in\mathcal{G}} |P_n^\circ g|\right\|_{\psi_2|X}\\ &\le 2K_{\psi_2}\sqrt{1+\log|\mathcal{G}|}\,\max_{g\in\mathcal{G}}\|P_n^\circ g\|_{\psi_2|X}\\ &\le 2\sqrt{6}\,K_{\psi_2}\sqrt{1+\log|\mathcal{G}|}\,\max_{g\in\mathcal{G}}\Big(\frac{P_n g^2}{n}\Big)^{1/2}\\ &\le 2\sqrt{6}\,K_{\psi_2} M\sqrt{\frac{1+\log|\mathcal{G}|}{n}}.\end{aligned}$$

La primera desigualdad es $\mathbb{E}|Z| \le 2\|Z\|_{\psi_2}$, que se probó en la demostración del Corolario 2.18. La segunda es el Lema 2.9 con $\psi = \psi_2$, aplicado a las $|\mathcal{G}|$ variables $P_n^\circ g$, junto con $\psi_2^{-1}(m) = \sqrt{\log(1+m)} \le \sqrt{1+\log m}$.[37] La tercera es el Lema 2.14 con $a_i = g(X_i)/n$, cuya norma euclídea es $\Big(\sum_{i=1}^n g(X_i)^2/n^2\Big)^{1/2} = (P_n g^2/n)^{1/2}$. La última usa que $P_n g^2 \le M^2$, pues $|g| \le M$ para $g \in \mathcal{G}$; es para esto que la red se eligió dentro de $\mathcal{F}_M$. Como $|\mathcal{G}| \le N(\delta/2, \mathcal{F}_M, L_1(P_n))$, concluimos que, para toda realización de la muestra, $\mathbb{E}_\varepsilon\|P_n^\circ\|_{\mathcal{F}_M} \le \delta + W_n$, donde

$$W_n := 2\sqrt{6}\,K_{\psi_2} M\sqrt{\frac{1+\log N(\delta/2, \mathcal{F}_M, L_1(P_n))}{n}}.$$

Integremos ahora respecto de la muestra. Ya vimos que $\mathbb{E}_\varepsilon\|P_n^\circ\|_{\mathcal{F}_M}$ es una función medible de la muestra, y además es a lo sumo M, porque $|P_n^\circ g| \le P_n|g| \le M$ para $g \in \mathcal{F}_M$. En cambio, W_n puede no ser medible, porque el número de cubrimiento, como función de la muestra, no tiene por qué serlo (por eso la hipótesis (ii) está formulada en probabilidad exterior). Sea W_n^* su mayorante medible mínimo. Como $W_n \le W_n^*$, la cota anterior da $\mathbb{E}_\varepsilon\|P_n^\circ\|_{\mathcal{F}_M} \le \delta + W_n^*$, y junto con la cota por M,

$$\mathbb{E}_\varepsilon\|P_n^\circ\|_{\mathcal{F}_M} \le \delta + (W_n^* \wedge M).$$

Por otra parte, como $W_n^2 = 24K_{\psi_2}^2 M^2\big(n^{-1} + n^{-1}\log N(\delta/2, \mathcal{F}_M, L_1(P_n))\big)$, la hipótesis (ii), con $\delta/2$ en lugar de ε, da $W_n \xrightarrow{\mathbb{P}^*} 0$;[38] es decir, $W_n^* \to 0$ en probabilidad, ya que $\mathbb{P}(W_n^* > \eta) = \mathbb{P}^*(W_n > \eta)$ por el Lema 1.9(vi). La variable $W_n^* \wedge M$ es medible, está acotada por M y tiende a 0 en probabilidad, así que su esperanza tiende a 0 por convergencia dominada.[39] Integrando en la muestra la cota anterior,

$$\mathbb{E}_X\mathbb{E}_\varepsilon\|P_n^\circ\|_{\mathcal{F}_M} \le \delta + \mathbb{E}(W_n^* \wedge M) \xrightarrow[n\to\infty]{} \delta,$$

y como el lado izquierdo no depende de δ, que es arbitrario, $\mathbb{E}_X\mathbb{E}_\varepsilon\|P_n^\circ\|_{\mathcal{F}_M} \to 0$.

Volviendo a (2.22), para todo $M > 0$,

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{E}^*\|P_n - P\|_{\mathcal{F}} &\le 2\lim_{n\to\infty} \mathbb{E}_X\mathbb{E}_\varepsilon\|P_n^\circ\|_{\mathcal{F}_M} + 2P(F\mathbb{1}_{\{F>M\}})\\ &= 2P(F\mathbb{1}_{\{F>M\}}),\end{aligned}$$

y haciendo $M \to \infty$ se obtiene $\mathbb{E}^*\|P_n - P\|_{\mathcal{F}} \to 0$, que es la convergencia en media exterior de (i). Por la desigualdad de Markov exterior (Corolario 1.10), $\mathbb{P}^*(\|P_n - P\|_{\mathcal{F}} \ge \eta) \le \mathbb{E}^*\|P_n - P\|_{\mathcal{F}}/\eta \to 0$ para todo $\eta > 0$; es decir, $\|P_n - P\|_{\mathcal{F}} \xrightarrow{\mathbb{P}^*} 0$, lo que equivale a $\|P_n - P\|_{\mathcal{F}}^* \to 0$ en probabilidad. Por otra parte, por el Lema 2.30, las versiones de $\|P_n - P\|_{\mathcal{F}}^*$ simétricas en sus primeros n argumentos

[37] Para $m \ge 1$, $\log(1+m) \le \log(2m) = \log 2 + \log m \le 1 + \log m$.

[38] Sean $c := 24K_{\psi_2}^2 M^2$ y $\eta > 0$. Si $n \ge 2c/\eta^2$, entonces $1/n \le \eta^2/(2c)$ y $\{W_n > \eta\} = \{n^{-1}\log N(\delta/2, \mathcal{F}_M, L_1(P_n)) > \eta^2/c - 1/n\} \subset \{n^{-1}\log N(\delta/2, \mathcal{F}_M, L_1(P_n)) > \eta^2/(2c)\}$, cuya probabilidad exterior tiende a 0 por (ii).

[39] Directamente: para todo $\eta > 0$, $\mathbb{E}(W_n^* \wedge M) \le \eta + M\,\mathbb{P}(W_n^* > \eta) \to \eta$.

convergen casi seguramente a una constante. Como la convergencia casi segura implica la convergencia en probabilidad, esa constante es 0,[40] y por lo tanto $\|P_n - P\|_{\mathcal{F}}^* \to 0$ c.s.

Por último, la condición suficiente del final del enunciado implica (ii), ya que, como vimos, multiplicar por un indicador no aumenta los números de cubrimiento, y por lo tanto $N(\varepsilon, \mathcal{F}_M, L_1(P_n)) \le N(\varepsilon, \mathcal{F}, L_1(P_n))$ para todo $\varepsilon > 0$ y todo $M > 0$.[41] □

La necesidad de la condición de entropía, que completa la equivalencia del Teorema 2.31, no se demuestra aquí. La prueba de [5] usa la desimetrización, desigualdades para multiplicadores gaussianos y la desigualdad de Sudakov; además obtiene la convergencia de la entropía en media exterior.

Proposición 2.32. *Sean $\mathcal{F}, \mathcal{F}_1, \mathcal{F}_2$ clases P-Glivenko–Cantelli. Entonces:*

1. $\{a_1 f_1 + a_2 f_2 : f_i \in \mathcal{F}_i,\ |a_i| \le 1\}$ *es P-Glivenko–Cantelli;*
2. $\mathcal{F}_1 \cup \mathcal{F}_2$ *es P-Glivenko–Cantelli;*
3. *la clase de todas las funciones que son simultáneamente límite puntual y límite en $L_1(P)$ de una sucesión en $\mathcal{F}$ es P-Glivenko–Cantelli.*

Demostración. Para (1), si $\mathcal{H} = \{a_1 f_1 + a_2 f_2 : f_i \in \mathcal{F}_i,\ |a_i| \le 1\}$, entonces $\|P_n - P\|_{\mathcal{H}}^* \le \|P_n - P\|_{\mathcal{F}_1}^* + \|P_n - P\|_{\mathcal{F}_2}^* \to 0$ c.s.[42] Para (2),

$$\|P_n - P\|_{\mathcal{F}_1 \cup \mathcal{F}_2}^* \le \max\{\|P_n - P\|_{\mathcal{F}_1}^*, \|P_n - P\|_{\mathcal{F}_2}^*\} \to 0 \quad \text{c.s.}$$

Para (3), sea $\overline{\mathcal{F}}$ la clase descripta. Si $g \in \overline{\mathcal{F}}$, existe $f_m \in \mathcal{F}$ tal que $f_m \to g$ puntualmente y $P|f_m - g| \to 0$. Para cada muestra fija, $P_n|f_m - g| \to 0$, pues P_n sólo promedia finitos puntos. Luego, dado $\eta > 0$, para cada g puede elegirse m con $P_n|f_m - g| + P|f_m - g| < \eta$, y $|(P_n - P)g| \le \|P_n - P\|_{\mathcal{F}} + P_n|f_m - g| + P|f_m - g| \le \|P_n - P\|_{\mathcal{F}} + \eta$. Tomando supremo en g y luego $\eta \downarrow 0$, $\|P_n - P\|_{\overline{\mathcal{F}}} \le \|P_n - P\|_{\mathcal{F}}$. Por mayorantes medibles mínimos, $\|P_n - P\|_{\overline{\mathcal{F}}}^* \le \|P_n - P\|_{\mathcal{F}}^*$ c.s., y el resultado sigue. □

2.5. Teoremas de Donsker

2.5.1. Entropía uniforme

En esta sección presentamos el teorema de Donsker basado en *entropía uniforme*. Su prueba combina simetrización, una desigualdad maximal subgaussiana condicionada en la muestra y el teorema de Glivenko–Cantelli por entropía aleatoria.

Recordemos que $N(\varepsilon, \mathcal{F}, \|\cdot\|)$ es el número de cubrimiento para una seminorma cualquiera, con centros no necesariamente en $\mathcal{F}$ (Definición 2.26). Si la seminorma es la de $L_r(Q)$, entonces $N(\varepsilon, \mathcal{F}, L_r(Q)) \le N_{[]}(2\varepsilon, \mathcal{F}, L_r(Q))$, pues cada bracket de tamaño 2ε queda contenido en una bola de radio ε, centrada en el punto medio de sus extremos.

La *condición de entropía uniforme* se formula mediante la integral truncada que definimos a continuación.

Definición 2.34. Sea $\mathcal{F}$ una clase de funciones medibles con envolvente F. Para $\delta > 0$, definimos la integral de entropía uniforme truncada por

$$J(\delta, \mathcal{F}, L_2) := \int_0^\delta \sup_Q \sqrt{\log N\Big(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\Big)}\, d\varepsilon,$$

[40] Esas versiones coinciden c.s. con $\|P_n - P\|_{\mathcal{F}}^*$, así que tienden a 0 en probabilidad; como también tienden a la constante c.s., y por lo tanto en probabilidad, la unicidad c.s. del límite en probabilidad da que la constante es 0.

[41] Aquí el indicador es $\mathbb{1}_{\{F \le M\}}$, con F la envolvente del enunciado; aquel argumento no usa medibilidad. La condición pasa a $\mathcal{F}_M$ por la monotonía de la probabilidad exterior, como antes.

[42] Punto a punto, $|(P_n - P)(a_1 f_1 + a_2 f_2)| \le |(P_n - P)f_1| + |(P_n - P)f_2|$ para $|a_i| \le 1$; tomando supremos, $\|P_n - P\|_{\mathcal{H}} \le \|P_n - P\|_{\mathcal{F}_1} + \|P_n - P\|_{\mathcal{F}_2}$, y la regla $(S+T)^* \le S^* + T^*$ del Lema 1.9(i) pasa la desigualdad a los mayorantes. En (2) se usa análogamente $\|\cdot\|_{\mathcal{F}_1 \cup \mathcal{F}_2} = \|\cdot\|_{\mathcal{F}_1} \vee \|\cdot\|_{\mathcal{F}_2}$ y la regla $(S \vee T)^* = S^* \vee T^*$, Lema 1.9(viii).

donde el supremo se toma sobre todas las medidas de probabilidad finitamente discretas Q —es decir, probabilidades que ponen masa en una cantidad finita de puntos— tales que $0 < \|F\|_{Q,2} < \infty$. Si el conjunto de tales Q es vacío, se interpreta el supremo como 0. La condición de entropía uniforme es

$$J(\infty, \mathcal{F}, L_2) < \infty. \tag{UE}$$

Teorema 2.35. *Sea $\mathcal{F}$ una clase de funciones medibles con envolvente F tal que $P^*F^2 < \infty$. Supongamos que se cumple (UE) y que, para todo $\delta > 0$, las clases $\mathcal{F}_\delta = \{f-g : f, g \in \mathcal{F},\ \|f-g\|_{P,2} < \delta\}$, $\mathcal{F}_\infty^2 = \{(f-g)^2 : f, g \in \mathcal{F}\}$ son P-medibles. Entonces $\mathcal{F}$ es P-Donsker.*

Demostración. Escribamos $\rho(f,g) = \|f-g\|_{P,2}$ y $\mathcal{F}_\infty = \{f-g : f,g \in \mathcal{F}\}$; $\mathcal{F}_\delta$ es la clase del enunciado. Por el criterio de convergencia débil en $\ell^\infty(\mathcal{F})$, alcanza probar: (i) convergencia finito-dimensional, (ii) total acotación de $(\mathcal{F}, \rho)$, (iii) equicontinuidad asintótica uniforme.[43] La convergencia finito-dimensional se obtiene del TCL multivariado, pues $|f| \le F$ y $P^*F^2 < \infty$.[44]

Probamos la equicontinuidad. Sea $\delta_n \downarrow 0$. Entonces $\|\mathbb{G}_n\|_{\mathcal{F}_{\delta_n}} = \sup_{\rho(f,g)<\delta_n} |\mathbb{G}_n f - \mathbb{G}_n g|$. Fijado $x > 0$, por Markov exterior y simetrización,[45]

$$\mathbb{P}^*\{\|\mathbb{G}_n\|_{\mathcal{F}_{\delta_n}} > x\} \le \frac{1}{x}\mathbb{E}^*\|\mathbb{G}_n\|_{\mathcal{F}_{\delta_n}} \le \frac{2}{x}\mathbb{E}^*\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i h(X_i)\right\|_{\mathcal{F}_{\delta_n}}. \tag{1}$$

Como $\mathcal{F}_{\delta_n}$ es P-medible, (FR), aplicada a la clase $\mathcal{F}_{\delta_n}$ y multiplicada por $n^{-1/2}$, da

$$\mathbb{E}^*\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i h(X_i)\right\|_{\mathcal{F}_{\delta_n}} = \mathbb{E}_X\mathbb{E}_\varepsilon\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i h(X_i)\right\|_{\mathcal{F}_{\delta_n}}. \tag{2}$$

Condicionando en $X_1, \dots, X_n$, ponemos $\|h\|_{n,2} := \|h\|_{P_n,2} = (P_n h^2)^{1/2}$, $\theta_n^2 = \sup_{h\in\mathcal{F}_{\delta_n}} P_n h^2$. Para $h, k \in \mathcal{F}_{\delta_n}$, $\left\|\frac{1}{\sqrt{n}}\sum_i \varepsilon_i (h-k)(X_i)\right\|_{\psi_2|X} \lesssim \|h-k\|_{n,2}$, luego el proceso de Rademacher es subgaussiano para $d_n(h,k) = \|h-k\|_{n,2}$.[46]

Separabilidad condicional. Para $X = (X_1, \dots, X_n)$ fijo, el proceso depende de h sólo a través del vector $(h(X_1), \dots, h(X_n)) \in \mathbb{R}^n$; el cociente de $\mathcal{F}_{\delta_n}$ por distancia d_n nula se identifica así con un subconjunto de $\mathbb{R}^n$ y es, por tanto, separable. Eligiendo representantes de un conjunto numerable d_n-denso, la desigualdad $|Z_h - Z_k| \le n^{-1/2}\sum_i |h-k|(X_i) \le \sqrt{n}\, d_n(h,k)$, válida para cada realización de los signos, da que el supremo global de los incrementos coincide con el supremo sobre esos representantes. Como aquí el Corolario 2.18 se aplica con $\delta = \operatorname{diam}$, esa igualdad global es lo único que se necesita.

Como $0 \in \mathcal{F}_{\delta_n}$ (tómese $f = g$), el supremo del proceso está dominado por el supremo de los incrementos: $\sup_h |Z_h| = \sup_h |Z_h - Z_0| \le \sup_{h,k} |Z_h - Z_k|$. Además, el diámetro de $(\mathcal{F}_{\delta_n}, d_n)$ es a lo sumo $2\theta_n$, y por (2.6), $\int_0^{2\theta_n} \sqrt{\log D(u, d_n)}\, du \le 2\int_0^{\theta_n} \sqrt{\log N(v, d_n)}\, dv$.[47] Por la desigualdad maximal subgaussiana (Corolario 2.18 con $\delta = \operatorname{diam}$, y K su constante),

$$\mathbb{E}_\varepsilon\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i h(X_i)\right\|_{\mathcal{F}_{\delta_n}} \le 2K\int_0^{\theta_n} \sqrt{\log N(u, \mathcal{F}_{\delta_n}, L_2(P_n))}\, du. \tag{3}$$

[43] El criterio es el Teorema 1.67 combinado con el de equicontinuidad (Teorema 1.71): la tensión asintótica equivale a la tensión de las coordenadas (que aquí da el TCL) más la total acotación de $(\mathcal{F}, \rho)$ y la equicontinuidad asintótica uniforme respecto de ρ.

[44] Cada f es medible con $Pf^2 = P^*f^2 \le P^*F^2 < \infty$, así que el TCL multivariado se aplica a los vectores $(f_1, \dots, f_k)$ centrados.

[45] Corolario 1.10 y Lema 2.20 con $\Phi(x) = x$, aplicado a la clase $\mathcal{F}_{\delta_n}$. En efecto, para $h = f - g \in \mathcal{F}_{\delta_n}$, $P|h| \le 2P^*F \le 2(P^*F^2)^{1/2} < \infty$; la segunda desigualdad es Cauchy–Schwarz aplicada al mayorante F^*, usando $P^*F = PF^*$ y $(F^2)^* = (F^*)^2$ c.s. (por la minimalidad del Lema 1.5, inciso (ii)). Además, $\sqrt{n}(P_n - P)h = \mathbb{G}_n h$ y el factor $\sqrt{n}$ convierte P_n° en $n^{-1/2}\sum_i \varepsilon_i h(X_i)$.

[46] Lema 2.14 condicional, con $a_i = n^{-1/2}(h-k)(X_i)$: $\|a\|^2 = n^{-1}\sum_i (h-k)^2(X_i) = \|h-k\|_{n,2}^2$, y la constante es $\sqrt{6}$.

[47] El diámetro: $\|h-k\|_{n,2} \le \|h\|_{n,2} + \|k\|_{n,2} \le 2\theta_n$, por Minkowski. La integral: con $u = 2v$, $\int_0^{2\theta_n} \sqrt{\log D(u, d_n)}\, du = 2\int_0^{\theta_n} \sqrt{\log D(2v, d_n)}\, dv \le 2\int_0^{\theta_n} \sqrt{\log N(v, d_n)}\, dv$, pues $D(2v, d_n) \le N(v, d_n)$ por (2.6). Obsérvese que el Corolario 2.18 está enunciado con números de empaquetamiento D; reemplazar D por N al mismo radio iría en la dirección incorrecta ($N \le D$), y es el cambio de escala $u = 2v$ el que permite pasar a números de cubrimiento.

Además, para toda probabilidad Q,

$$N(u, \mathcal{F}_\infty, L_2(Q)) \le N^2(u/2, \mathcal{F}, L_2(Q)), \tag{4}$$

pues si $g_1, \dots, g_N$ cubre $\mathcal{F}$ con radio $u/2$, entonces $\{g_j - g_k : 1 \le j, k \le N\}$ cubre $\mathcal{F}_\infty$ con radio u: $\|(f - f') - (g_j - g_k)\|_{Q,2} \le \|f - g_j\|_{Q,2} + \|f' - g_k\|_{Q,2} < u$. Como $\mathcal{F}_{\delta_n} \subset \mathcal{F}_\infty$, de (3)–(4),

$$\begin{aligned} \mathbb{E}_\varepsilon \left\| \frac{1}{\sqrt{n}} \sum_i \varepsilon_i h(X_i) \right\|_{\mathcal{F}_{\delta_n}} &\le 2\sqrt{2}\, K \int_0^{\theta_n} \sqrt{\log N(u/2, \mathcal{F}, L_2(P_n))}\, du \\ &= 4\sqrt{2}\, K \|F\|_{n,2} \int_0^{\theta_n/(2\|F\|_{n,2})} \sqrt{\log N(\varepsilon \|F\|_{n,2}, \mathcal{F}, L_2(P_n))}\, d\varepsilon \\ &\le 4\sqrt{2}\, K \|F\|_{n,2}\, J\left(\frac{\theta_n}{\|F\|_{n,2}}, \mathcal{F}, L_2 \right), \end{aligned} \tag{5}$$

donde usamos que P_n es finitamente discreta.[48] Si $\|F\|_{n,2} = 0$, entonces $h(X_i) = 0$ para todo $h \in \mathcal{F}_{\delta_n}$ e i, y (5) es trivial. Reemplazando F por $F \vee 1$, suponemos $\|F\|_{n,2} \ge 1$.[49] Además,

$$\mathbb{E}^* \|F\|_{n,2}^2 = \mathbb{E}^* P_n F^2 \le P^* F^2 < \infty. \tag{6}$$

Mostramos ahora que $\theta_n \to 0$ en probabilidad exterior.[50] Si $h \in \mathcal{F}_{\delta_n}$, entonces $Ph^2 < \delta_n^2$, y por tanto

$$\theta_n^2 = \sup_{h \in \mathcal{F}_{\delta_n}} P_n h^2 \le \delta_n^2 + \|P_n - P\|^*_{\mathcal{F}_\infty^2}. \tag{7}$$

Basta probar que $\mathcal{F}_\infty^2$ es P-Glivenko–Cantelli. Esta clase tiene envolvente $(2F)^2$, integrable exteriormente, y es P-medible por hipótesis. Para $h, k \in \mathcal{F}_\infty$, $|h^2 - k^2| = |h - k||h + k| \le |h - k|\, 4F$, pues $|h| \le 2F$, $|k| \le 2F$. Luego, para toda probabilidad discreta Q,

$$Q|h^2 - k^2| \le Q(|h - k|\, 4F) \le \|h - k\|_{Q,2} \|4F\|_{Q,2}. \tag{8}$$

En particular, si $h, k \in \mathcal{F}_\infty$ y $\|h - k\|_{Q,2} < \varepsilon \|F\|_{Q,2}$, entonces $Q|h^2 - k^2| < 4\varepsilon \|F\|_{Q,2}^2 = \varepsilon \|2F\|_{Q,2}^2$. Los centros de un cubrimiento de $\mathcal{F}_\infty$ pueden ser externos y no satisfacer $|k| \le 2F$; por eso internalizamos primero (véase el párrafo que sigue a la Definición 2.26): un cubrimiento de radio $\frac{\varepsilon}{2}\|F\|_{Q,2}$ en $L_2(Q)$ da centros $k_j \in \mathcal{F}_\infty$ que cubren con radio $\varepsilon \|F\|_{Q,2}$. Para estos centros vale la desigualdad anterior, y por lo tanto

$$N\Big(\varepsilon \|2F\|_{Q,2}^2, \mathcal{F}_\infty^2, L_1(Q)\Big) \le N\left(\frac{\varepsilon}{2} \|F\|_{Q,2}, \mathcal{F}_\infty, L_2(Q) \right). \tag{9}$$

Usando otra vez (4),

$$N\left(\frac{\varepsilon}{2} \|F\|_{Q,2}, \mathcal{F}_\infty, L_2(Q) \right) \le N^2\left(\frac{\varepsilon}{4} \|F\|_{Q,2}, \mathcal{F}, L_2(Q) \right). \tag{10}$$

Por la condición de entropía uniforme, el lado derecho de (10) está acotado por una constante finita que depende de ε, pero no de Q.[51] Aplicado con $Q = P_n$, obtenemos

$$\log N\Big(\varepsilon \|2F\|_{n,2}^2, \mathcal{F}_\infty^2, L_1(P_n)\Big) = O_{\mathbb{P}^*}(1) = o_{\mathbb{P}^*}(n). \tag{11}$$

[48] El primer paso usa (4): $\sqrt{\log N^2(u/2, \mathcal{F}, L_2(P_n))} = \sqrt{2}\, \sqrt{\log N(u/2, \mathcal{F}, L_2(P_n))}$. El segundo es el cambio de variable $u = 2\varepsilon \|F\|_{n,2}$, que aporta el factor $2\|F\|_{n,2}$; luego se agranda el límite superior de $\theta_n/(2\|F\|_{n,2})$ a $\theta_n/\|F\|_{n,2}$ y se reemplaza el integrando por su supremo sobre las Q finitamente discretas, entre las cuales está P_n cuando $0 < \|F\|_{n,2} < \infty$.

[49] El reemplazo conserva las hipótesis: $F \vee 1$ sigue siendo envolvente, $P^*(F \vee 1)^2 \le P^* F^2 + 1 < \infty$ y, como $\|F \vee 1\|_{Q,2} \ge \|F\|_{Q,2}$, los números de cubrimiento con radio $\varepsilon \|F \vee 1\|_{Q,2}$ no superan a los de radio $\varepsilon \|F\|_{Q,2}$; por lo tanto la integral de entropía uniforme no aumenta.

[50] En (7): para cada h, $P_n h^2 = Ph^2 + (P_n - P)h^2 < \delta_n^2 + \|P_n - P\|_{\mathcal{F}_\infty^2}$; el paso al mayorante $\|\cdot\|^*$ vale c.s. y da una cota medible, uniforme en h.

[51] La finitud del supremo en Q para cada ε fijo sale de (UE): la integral finita fuerza que el integrando $\sup_Q \sqrt{\log N(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q))}$ sea finito para casi todo ε y, por monotonía en ε, para todo $\varepsilon > 0$.

Además, por (6), $\|2F\|_{n,2}^2 = O_{\mathbb{P}^*}(1)$. Fijados $\delta, \eta > 0$, podemos elegir $C < \infty$ de modo que $\mathbb{P}^*(\|2F\|_{n,2}^2 > C) < \eta$ para todo n suficientemente grande. En el evento complementario, tomando $\varepsilon = \delta/C$ y usando la monotonía del número de cubrimiento en el radio, $N(\delta, \mathcal{F}_\infty^2, L_1(P_n)) \le N\Big(\varepsilon\|2F\|_{n,2}^2, \mathcal{F}_\infty^2, L_1(P_n)\Big)$. Por tanto, (11) implica que $\log N(\delta, \mathcal{F}_\infty^2, L_1(P_n)) = o_{\mathbb{P}^*}(n)$ para cada $\delta > 0$, que es la condición suficiente del final del enunciado del Teorema 2.31. Como $(2F)^2$ es una envolvente integrable de $\mathcal{F}_\infty^2$, ese teorema da

$$\|P_n - P\|_{\mathcal{F}_\infty^2}^* \xrightarrow{\mathbb{P}^*} 0. \tag{12}$$

De (7) se sigue

$$\theta_n \xrightarrow{\mathbb{P}^*} 0. \tag{13}$$

Como $J(\infty, \mathcal{F}, L_2) < \infty$, la función $\eta \mapsto J(\eta, \mathcal{F}, L_2)$ es acotada y satisface $J(\eta, \mathcal{F}, L_2) \downarrow 0$ cuando $\eta \downarrow 0$. Por (13) y $\|F\|_{n,2} \ge 1$,

$$J\left(\frac{\theta_n}{\|F\|_{n,2}}, \mathcal{F}, L_2\right) \xrightarrow{\mathbb{P}^*} 0. \tag{14}$$

Además, está acotada por $J(\infty, \mathcal{F}, L_2)$. Por Cauchy–Schwarz y (6),

$$\mathbb{E}_X^*\left[\|F\|_{n,2} J\left(\frac{\theta_n}{\|F\|_{n,2}}, \mathcal{F}, L_2\right)\right] \le (\mathbb{E}^*\|F\|_{n,2}^2)^{1/2} \left(\mathbb{E}^* J^2\left(\frac{\theta_n}{\|F\|_{n,2}}, \mathcal{F}, L_2\right)\right)^{1/2} \longrightarrow 0, \tag{15}$$

por convergencia dominada exterior. Juntando (1), (2), (5) y (15),[52]

$$\|\mathbb{G}_n\|_{\mathcal{F}_{\delta_n}} \xrightarrow{\mathbb{P}^*} 0 \qquad (\delta_n \downarrow 0). \tag{16}$$

Esto es la equicontinuidad asintótica uniforme.

Falta la total acotación de $(\mathcal{F}, \rho)$. Por (12), para cada m existe una probabilidad discreta Q_m tal que

$$\sup_{h \in \mathcal{F}_\infty} |(Q_m - P)h^2| \le m^{-2}. \tag{17}$$

En efecto, elegimos $n = n(m)$ y una realización de la medida empírica para la cual el supremo en (12) es menor que m^{-2} y, simultáneamente, $P_n F^2 < \infty$; esto último ocurre casi seguramente, pues $P^* F^2 < \infty$.[53] Fijado $\varepsilon > 0$, elegimos m con $m^{-1} \le \varepsilon$. Por la condición de entropía uniforme,[54]

$$N(\varepsilon/2, \mathcal{F}, L_2(Q_m)) < \infty. \tag{18}$$

Internalizando un cubrimiento de radio $\varepsilon/2$ (véase el párrafo que sigue a la Definición 2.26), obtenemos $f_1, \ldots, f_N \in \mathcal{F}$ tales que, para todo $f \in \mathcal{F}$, existe j con $\|f - f_j\|_{Q_m,2} \le \varepsilon$. Como $f - f_j \in \mathcal{F}_\infty$, por (17),

$$\|f - f_j\|_{P,2}^2 = P(f - f_j)^2 \le Q_m(f - f_j)^2 + |(P - Q_m)(f - f_j)^2| \le \varepsilon^2 + m^{-2} \le 2\varepsilon^2. \tag{19}$$

Entonces $f_1, \ldots, f_N$ es una $\sqrt{2}\,\varepsilon$-red de $\mathcal{F}$ en $L_2(P)$. Como $\varepsilon > 0$ era arbitrario, $(\mathcal{F}, \rho)$ es totalmente acotada.

Tenemos convergencia finito-dimensional, equicontinuidad asintótica uniforme y total acotación de $(\mathcal{F}, \rho)$. Por el criterio de convergencia débil en $\ell^\infty(\mathcal{F})$, $\mathbb{G}_n \Rightarrow \mathbb{G}_P$ en $\ell^\infty(\mathcal{F})$, donde $\mathbb{G}_P$ es el puente browniano P-indexado por $\mathcal{F}$. Luego $\mathcal{F}$ es P-Donsker. □

[52] El Cauchy–Schwarz exterior para factores no negativos se justifica con $(AB)^* \le A^* B^*$ c.s. (Lema 1.9(v)) seguido del Cauchy–Schwarz clásico para A^*, B^*. El segundo factor tiende a 0: $J^2(\theta_n/\|F\|_{n,2}) \le J^2(\infty)$ está acotada y tiende a 0 en probabilidad exterior por (14), y se razona como en la convergencia dominada del Teorema 2.31.

[53] Sea $(P_n F^2)^*$ un mayorante medible mínimo de $P_n F^2$. Como $\mathbb{E}(P_n F^2)^* = \mathbb{E}^* P_n F^2 \le P^* F^2 < \infty$, se tiene $(P_n F^2)^* < \infty$ c.s. y, fuera de un conjunto nulo, también $P_n F^2 < \infty$. Por (12), para n grande el evento medible $\{\|P_n - P\|_{\mathcal{F}_\infty^2}^* \le m^{-2}\}$ tiene probabilidad positiva; su intersección con el evento de probabilidad 1 recién descripto es no vacía, y cualquier ω en ella da la medida finitamente discreta $Q_m = P_n(\omega)$ requerida (nótese que $\|P_n - P\|_{\mathcal{F}_\infty^2} \le \|P_n - P\|_{\mathcal{F}_\infty^2}^*$ en todo punto).

[54] La integral finita implica que el integrando es finito para casi todo radio; por la monotonía de $\varepsilon \mapsto N(\varepsilon, \mathcal{F}, L_2(Q_m))$, los números de cubrimiento son entonces finitos para *todo* radio positivo: si fueran infinitos en algún $\varepsilon_0 > 0$, lo serían en todo $(0, \varepsilon_0]$ y la integral divergiría.

Ejemplo 2.36. Sea $\mathcal{F}$ el conjunto de todas las funciones indicadoras $1_{(-\infty,t]}$ en $\mathbb{R}$. Para cualquier medida de probabilidad Q y $\varepsilon \leq 1$ se cumple

$$N\big(\varepsilon, \mathcal{F}, L_2(Q)\big) \leq N_{[]}\big(\varepsilon^2, \mathcal{F}, L_1(Q)\big) \leq \frac{2}{\varepsilon^2}.$$

[55] Como la envolvente es $F \equiv 1$, se tiene $\|F\|_{Q,2} = 1$ para toda Q, y la cota anterior es uniforme en Q; como $\int_0^1 \sqrt{\log(2/\varepsilon^2)}\, d\varepsilon < \infty$, la condición de entropía uniforme (UE) se cumple y $\mathcal{F}$ es Donsker para toda P. Los números de cubrimiento de la clase de rectángulos inferiores $(-\infty, t]$ en dimensiones superiores satisfacen una cota similar, pero con una potencia mayor de $(1/\varepsilon)$. Por lo tanto, la clase de todos los rectángulos inferiores $(-\infty, t]$ en $\mathbb{R}^k$ es Donsker para cualquier dimensión.

Más adelante en este capítulo se muestra que estas clases de conjuntos son ejemplos de la amplia colección de clases VC, las cuales son Donsker siempre que sean adecuadamente medibles.

2.5.2. Acotamientos (*bracketing*)

El segundo teorema central del límite empírico usa entropía por acotamientos (*bracketing*) en lugar de entropía uniforme. Su versión más simple usa brackets en $L_2(P)$ y la condición

$$\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_2(P))}\, d\varepsilon < \infty. \tag{2.23}$$

Teorema 2.37. *Sea $\mathcal{F}$ una clase de funciones medibles. Si se cumple (2.23), entonces $\mathcal{F}$ es P-Donsker.*

A diferencia de la condición de entropía uniforme, la integral de bracketing involucra sólo a la medida verdadera P. Esta ventaja se compensa con que los números de bracketing pueden ser mayores que los números de cubrimiento; por eso las condiciones suficientes basadas en entropía uniforme y en bracketing no son comparables (véase [5, Cap. 2.5]).

El resultado anterior se prueba mediante una versión algo más fuerte. A partir de una familia finita de $L_2(P)$-brackets puede construirse una envolvente cuadrado-integrable; por tanto, la existencia de tal envolvente está implícita en (2.23). Sin embargo, una envolvente en $L_2(P)$ no es necesaria para que una clase sea Donsker. Lo que sí debe poseer la envolvente es un *segundo momento débil*, en el sentido $P^*(F > x) = o(x^{-2}), \qquad x \to \infty$. De modo análogo, la norma $L_2(P)$ usada para medir el tamaño de los brackets puede reemplazarse por una cantidad más débil, lo que reduce los números de bracketing y facilita la convergencia de la integral.

Definimos $\|f\|_{P,2,\infty} := \sup_{x>0} \left(x^2 P(|f| > x)\right)^{1/2}$. Esta cantidad no es una norma, porque no satisface necesariamente la desigualdad triangular; pero existe una norma equivalente a ella salvo una constante 2, lo cual no será relevante aquí. Además, por Markov,

$$\|f\|_{P,2,\infty} \leq \|f\|_{P,2}.^{56}$$

Por tanto, los números de bracketing relativos a $L_{2,\infty}(P)$ son menores o iguales que los relativos a $L_2(P)$.

La prueba de los teoremas de bracketing se basa en un argumento de encadenamiento. A diferencia de la prueba por entropía uniforme, aquí el encadenamiento se aplica a los sumandos originales, sin una simetrización inicial. Como los sumandos no están uniformemente acotados del modo adecuado,

[55] Para la primera desigualdad: los brackets de semirrectas pueden tomarse con $0 \leq u - l \leq 1$, y entonces, para f, g en un mismo bracket, $Q(f-g)^2 \leq Q(u-l)^2 \leq Q(u-l) < \varepsilon^2$; cada bracket cabe así en una bola de $L_2(Q)$ de radio ε centrada en cualquiera de sus elementos. Para la segunda: la construcción cuantílica detallada en la nota del Ejemplo 2.24, aplicada con ε^2 en lugar de ε (es decir, con $K = \lfloor \varepsilon^{-2} \rfloor$ cortes), da una grilla con masa abierta $< \varepsilon^2$ entre cortes consecutivos y, por tanto, a lo sumo $K + 1 \leq 2/\varepsilon^2$ brackets, para $0 < \varepsilon \leq 1$.

[56] $x^2 P(|f| > x) \leq \mathbb{E}\big(f^2 \mathbb{1}_{\{|f|>x\}}\big) \leq Pf^2$, y se toma supremo en x.

no se usa Hoeffding ni la desigualdad maximal subgaussiana. En su lugar se usa la desigualdad de Bernstein (véase [5, Sección 2.2]): para toda función f cuadrado-integrable y uniformemente acotada,

$$\mathbb{P}\{|\mathbb{G}_n f| > x\} \le 2\exp\left\{-\frac{1}{2}\,\frac{x^2}{P(f-Pf)^2 + \|f-Pf\|_\infty x/(3\sqrt{n})}\right\}, \qquad x>0.$$

En particular, como $P(f-Pf)^2 \le Pf^2$ y $\|f-Pf\|_\infty \le 2\|f\|_\infty$, puede usarse la versión más simple con el denominador $Pf^2 + 2\|f\|_\infty x/(3\sqrt{n})$. En consecuencia, por la desigualdad maximal finita correspondiente,[57] si $\mathcal{H}$ es finita y $|\mathcal{H}| \ge 2$, entonces

$$\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{H}} \lesssim \max_{h\in\mathcal{H}} \frac{\|h\|_\infty}{\sqrt{n}} \log|\mathcal{H}| + \max_{h\in\mathcal{H}} \|h\|_{P,2}\sqrt{\log|\mathcal{H}|}. \tag{2.24}$$

El encadenamiento se organiza para que los dos términos de (2.24) sean comparables. Esto ocurre cuando $\|h\|_\infty \asymp \frac{\sqrt{n}\,\|h\|_{P,2}}{\sqrt{\log|\mathcal{H}|}}$, y por eso, en la prueba, la desigualdad se aplica después de truncar los incrementos a un nivel de este orden.

Teorema 2.38. *Sea $\mathcal{F}$ una clase de funciones medibles tal que*

$$\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_{2,\infty}(P))}\,d\varepsilon + \int_0^\infty \sqrt{\log N(\varepsilon, \mathcal{F}, L_2(P))}\,d\varepsilon < \infty.$$

Si, además, una envolvente F de $\mathcal{F}$ posee segundo momento débil, es decir, $P^*(F>x) = o(x^{-2})$, $x\to\infty$, entonces $\mathcal{F}$ es P-Donsker.

Demostración del Teorema 2.37 a partir del Teorema 2.38. Verificamos las dos hipótesis. Primero, por lo observado antes del Teorema 2.38 y al comienzo de la Subsección 2.5.1,

$$N_{[]}(\varepsilon, \mathcal{F}, L_{2,\infty}(P)) \le N_{[]}(\varepsilon, \mathcal{F}, L_2(P)),$$
$$N(\varepsilon, \mathcal{F}, L_2(P)) \le N_{[]}(2\varepsilon, \mathcal{F}, L_2(P)).$$

Así, la condición (2.23) implica la finitud de las dos integrales de entropía del enunciado.

Segundo, la envolvente: tomando finitos 1-brackets $[l_i, u_i]$, $i=1,\dots,m$, que cubren $\mathcal{F}$, la función $F := \max_{1\le i\le m}(|l_i| \vee |u_i|)$ es una envolvente de $\mathcal{F}$ y pertenece a $L_2(P)$, por ser un máximo finito de funciones de $L_2(P)$. Toda envolvente de cuadrado integrable tiene segundo momento débil: $x^2\,\mathbb{P}(F>x) \le \mathbb{E}\big(F^2 \mathbb{1}_{\{F>x\}}\big) \to 0$ cuando $x\to\infty$, por convergencia dominada. Aplicando el Teorema 2.38, $\mathcal{F}$ es P-Donsker. □

2.6. Cotas para la entropía y clases VC

En la sección anterior vimos que el proceso empírico converge débilmente para clases $\mathcal{F}$ que satisfacen la condición de entropía uniforme (UE). En particular, si para algún $\delta>0$ y alguna constante K se cumple, para $0<\varepsilon<1$,

$$\sup_Q \log N\big(\varepsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\big) \le K\left(\frac{1}{\varepsilon}\right)^{2-\delta}, \tag{2.25}$$

entonces la integral en (UE) es finita: cerca de 0, el integrando queda acotado por una constante por $\varepsilon^{-1+\delta/2}$, que es integrable; para ε grande, la entropía es nula, pues $|f|\le F$.[58] Por el teorema de Donsker por entropía uniforme, $\mathcal{F}$ es P-Donsker para toda probabilidad P tal que $P^*F^2<\infty$, siempre que se cumplan las condiciones de medibilidad requeridas.

[57] Es el Lema 2.10 con $a = 2\max_h Ph^2$ y $b = \frac{4}{3\sqrt{n}}\max_h\|h\|_\infty$, seguido de $\mathbb{E}|Z| \le C\|Z\|_{\psi_1}$ y $\log(1+m)\le 2\log m$ para $m\ge 2$.

[58] Para $\varepsilon>1$, la bola de radio $\varepsilon\|F\|_{Q,2}$ centrada en 0 contiene a toda la clase ($\|f\|_{Q,2}\le\|F\|_{Q,2}$), así que $N=1$ y $\log N = 0$; y $\varepsilon^{-1+\delta/2}$ es integrable en $(0,1)$ porque el exponente es >-1.

Muchas clases satisfacen incluso la condición más fuerte

$$\sup_Q N\Big(\varepsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\Big) \le K\left(\frac{1}{\varepsilon}\right)^V, \qquad 0<\varepsilon<1, \tag{2.26}$$

para algún $V<\infty$. En efecto, (2.26) implica $\sup_Q \log N\Big(\varepsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\Big) \le \log K + V\log(1/\varepsilon)$, y por tanto (2.25) para cualquier $\delta\in(0,2)$, ya que el logaritmo crece más lentamente que cualquier potencia; en particular, implica (UE). En esta sección mostraremos que cotas de este tipo se obtienen para clases definidas por condiciones combinatorias. Para clases de conjuntos, estas condiciones fueron introducidas por Vapnik y Červonenkis; de allí proviene el nombre de *clases VC*. Más adelante se definen clases VC de funciones a partir de clases VC de conjuntos, y se estudian operaciones que preservan estas propiedades de entropía.

2.6.1. Clases VC de conjuntos

Definición 2.39 (Clases VC de conjuntos)**.** Sea $\mathcal{C}$ una colección de subconjuntos de un conjunto $\mathcal{X}$. Dado un conjunto finito $\{x_1,\dots,x_n\}\subset\mathcal{X}$, diremos que $\mathcal{C}$ **selecciona** un subconjunto de $\{x_1,\dots,x_n\}$ si este puede escribirse como $C\cap\{x_1,\dots,x_n\}$ para algún $C\in\mathcal{C}$. Diremos que $\mathcal{C}$ **fragmenta** (*shatters*) a $\{x_1,\dots,x_n\}$ si selecciona todos sus 2^n subconjuntos.

Definimos $\Delta_n(\mathcal{C},x_1,\dots,x_n) := \#\{C\cap\{x_1,\dots,x_n\} : C\in\mathcal{C}\}$. La **dimensión VC** de $\mathcal{C}$ es

$$V(\mathcal{C}) := \sup\Big\{n : \max_{\{x_1,\dots,x_n\}\subset\mathcal{X}} \Delta_n(\mathcal{C},x_1,\dots,x_n) = 2^n\Big\}.$$

Es decir, $V(\mathcal{C})$ es el mayor tamaño de un conjunto finito que puede ser fragmentado por $\mathcal{C}$. [59] Si existen conjuntos fragmentados de tamaño arbitrariamente grande, ponemos $V(\mathcal{C})=\infty$. Una colección de conjuntos medibles $\mathcal{C}$ se llama **clase VC** si $V(\mathcal{C})<\infty$.

Con esta convención, si $\mathcal{C}$ es VC, entonces ningún conjunto de tamaño $n>V(\mathcal{C})$ puede ser fragmentado por $\mathcal{C}$.

El resultado principal de esta sección es que los números de cubrimiento de una clase VC crecen a lo sumo polinómicamente en $1/\varepsilon$; la consecuencia Donsker se extrae después del Teorema 2.46. Las cotas de cubrimiento que siguen están enunciadas en términos de la dimensión $V(\mathcal{C})$; en [5] aparecen formuladas con el índice $V(\mathcal{C})+1$, y los exponentes difieren en consecuencia.

Ejemplo 2.40. La colección de semirrectas $(-\infty,c]$ en $\mathbb{R}$ fragmenta cualquier conjunto de un punto, pero no fragmenta ningún conjunto de dos puntos: si $x_1<x_2$, no puede seleccionar sólo $\{x_2\}$. Por tanto, su dimensión VC es 1. La colección de intervalos $(a,b]$ en $\mathbb{R}$ fragmenta todo conjunto de dos puntos, pero no puede fragmentar ningún conjunto de tres puntos ordenados, pues no puede seleccionar simultáneamente el menor y el mayor sin seleccionar el punto intermedio. Por tanto, su dimensión VC es 2. En $\mathbb{R}^d$, las clases análogas de ortantes inferiores $(-\infty,c]$ y de rectángulos $(a,b]$ tienen dimensiones VC d y $2d$, respectivamente.[60]

El siguiente resultado combinatorio, conocido como *lema de Sauer*, es la base de las cotas polinomiales de entropía para clases VC. El ejemplo que sigue ilustra la diferencia entre los subconjuntos seleccionados y los fragmentados, que son los que se comparan en su enunciado.

Ejemplo 2.41. Sea $S=\{1,2,3\}$ y $\mathcal{C}=\{\varnothing,S\}$. Sobre todo S, la clase sólo selecciona $\varnothing\cap S=\varnothing$, $S\cap S=S$, luego $\Delta_3(\mathcal{C},1,2,3)=2$. En cambio, para decidir si $A\subset S$ está fragmentado se mira $\{C\cap A : C\in\mathcal{C}\}$, por lo que los singletons están fragmentados. Por tanto, los subconjuntos fragmentados son $\varnothing,\{1\},\{2\},\{3\}$.

[59] Aquí $V(\mathcal{C})$ denota la dimensión VC. En parte de la literatura se llama *índice VC* al entero $V(\mathcal{C})+1$.

[60] Para los ortantes en $\mathbb{R}^d$: los d vértices $e_1,\dots,e_d$ de la base canónica se fragmentan eligiendo c coordenada a coordenada; pero entre $d+1$ puntos siempre hay uno dominado, coordenada a coordenada, por el máximo de los restantes (cada coordenada del máximo la aporta alguno de los otros d), y ese punto no puede excluirse mientras se seleccionan los demás. Para los rectángulos, $V=2d$: la prueba usa separadamente dos extremos por coordenada y puede consultarse en [5, Sección 2.6].

Lema 2.42 (Sauer). *Sean $x_1, \dots, x_n \in \mathcal{X}$. Entonces $\Delta_n(\mathcal{C}, x_1, \dots, x_n)$ está acotado por el número de subconjuntos de $S = \{x_1, \dots, x_n\}$ que son fragmentados por $\mathcal{C}$. En particular, si $V(\mathcal{C}) < \infty$, entonces*

$$\Delta_n(\mathcal{C}, x_1, \dots, x_n) \leq \sum_{j=0}^{V(\mathcal{C})} \binom{n}{j}. \tag{2.27}$$

Demostración. Escribamos $S = \{x_1, \dots, x_n\}$. Reemplazando cada $C \in \mathcal{C}$ por su traza $C \cap S$, podemos suponer que $\mathcal{C} \subset 2^S$. Esto no cambia los subconjuntos seleccionados de S, ni los subconjuntos $A \subset S$ fragmentados, pues $(C \cap S) \cap A = C \cap A$. Así, $\Delta_n(\mathcal{C}, x_1, \dots, x_n) = \#\mathcal{C}$.

Digamos que $\mathcal{C} \subset 2^S$ es *hereditaria* si $B \subset C \in \mathcal{C}$ implica $B \in \mathcal{C}$. Si $\mathcal{C}$ es hereditaria, entonces todo $A \in \mathcal{C}$ es fragmentado por $\mathcal{C}$: en efecto, para cada $B \subset A$, se tiene $B \in \mathcal{C}$ y $B \cap A = B$. Luego, en el caso hereditario, el número de conjuntos fragmentados es al menos $\#\mathcal{C}$.

Reducimos el caso general al hereditario. Fijado $i \in \{1, \dots, n\}$, definimos

$$T_i(C) := \begin{cases} C \setminus \{x_i\}, & x_i \in C \text{ y } C \setminus \{x_i\} \notin \mathcal{C}, \\ C, & \text{en otro caso.} \end{cases}$$

Sea $T_i(\mathcal{C}) = \{T_i(C) : C \in \mathcal{C}\}$. La aplicación $C \mapsto T_i(C)$ es inyectiva. En efecto, si $A \subset S$ no contiene x_i, los únicos posibles preimágenes de A son A y $A \cup \{x_i\}$. Pero si $A \in \mathcal{C}$, entonces $A \cup \{x_i\}$ no se transforma en A; y si $A \notin \mathcal{C}$, entonces A no es preimagen. Por tanto cada imagen tiene a lo sumo una preimagen, y $\#T_i(\mathcal{C}) = \#\mathcal{C}$.

Además, todo $A \subset S$ fragmentado por $T_i(\mathcal{C})$ es fragmentado por $\mathcal{C}$. Si $x_i \notin A$, entonces $T_i(C) \cap A = C \cap A$ para todo C, y no hay nada que probar. Supongamos $x_i \in A$. Sea $B \subset A$ con $x_i \notin B$. Como A es fragmentado por $T_i(\mathcal{C})$, existe $C \in \mathcal{C}$ tal que $T_i(C) \cap A = B \cup \{x_i\}$. Entonces $x_i \in T_i(C)$, luego $T_i(C) = C$. Por la definición de T_i, esto implica $C \setminus \{x_i\} \in \mathcal{C}$. Por tanto $C \cap A = B \cup \{x_i\}$, $\quad (C \setminus \{x_i\}) \cap A = B$. Así, para cada $B \subset A$ que no contiene x_i, la clase $\mathcal{C}$ selecciona tanto B como $B \cup \{x_i\}$ sobre A. Luego selecciona todos los subconjuntos de A, y A es fragmentado por $\mathcal{C}$.

Aplicamos sucesivamente $T_1, \dots, T_n$, y repetimos el ciclo mientras la colección cambie. En cada paso el cardinal se conserva. Además, si $T_i(\mathcal{C}) \neq \mathcal{C}$, entonces $\sum_{D \in T_i(\mathcal{C})} |D| < \sum_{C \in \mathcal{C}} |C|$, pues algún x_i fue borrado y ningún elemento es agregado. Como esta suma es un entero no negativo, el procedimiento termina. Sea $\mathcal{D}$ la colección final. Entonces $\#\mathcal{D} = \#\mathcal{C}$, y, por el párrafo anterior aplicado en cada paso, $\#\{A \subset S : A \text{ es fragmentado por } \mathcal{D}\} \leq \#\{A \subset S : A \text{ es fragmentado por } \mathcal{C}\}$. Además, $\mathcal{D}$ es hereditaria: si $D \in \mathcal{D}$ y $x_i \in D$, pero $D \setminus \{x_i\} \notin \mathcal{D}$, entonces T_i todavía cambiaría la colección final, contradicción. Iterando el borrado de puntos, todo $B \subset D$ pertenece a $\mathcal{D}$.[61]

Como $\mathcal{D}$ es hereditaria, fragmenta al menos $\#\mathcal{D}$ conjuntos. Por tanto

$$\#\mathcal{C} = \#\mathcal{D} \leq \#\{A \subset S : A \text{ es fragmentado por } \mathcal{D}\} \leq \#\{A \subset S : A \text{ es fragmentado por } \mathcal{C}\}.$$

Esto prueba la primera afirmación.

Finalmente, si $V(\mathcal{C}) < \infty$, ningún subconjunto de S de tamaño mayor que $V(\mathcal{C})$ puede ser fragmentado. Luego los subconjuntos fragmentados de S están entre los subconjuntos de tamaños $0, \dots, V(\mathcal{C})$, y por la primera parte, $\Delta_n(\mathcal{C}, x_1, \dots, x_n) \leq \sum_{j=0}^{V(\mathcal{C})} \binom{n}{j}$. □

Así, aunque por definición una clase VC sólo excluye la posibilidad de seleccionar los 2^n subconjuntos cuando $n > V(\mathcal{C})$, el Lema 2.42 muestra que en realidad sólo puede seleccionar un número polinomial en n. Esta es la idea combinatoria detrás de las cotas uniformes de entropía para clases VC, y se registra en el corolario siguiente.

Corolario 2.43. *Si $\mathcal{C}$ es una clase VC de conjuntos con dimensión VC $V(\mathcal{C}) < \infty$, entonces* $\max_{x_1,\dots,x_n} \Delta_n(\mathcal{C}, x_1, \dots, x_n) \leq \sum_{j=0}^{V(\mathcal{C})} \binom{n}{j}$. *En particular, para $1 \leq V(\mathcal{C}) \leq n$,*

$$\max_{x_1,\dots,x_n} \Delta_n(\mathcal{C}, x_1, \dots, x_n) \leq \left(\frac{en}{V(\mathcal{C})}\right)^{V(\mathcal{C})},$$

[61] Con precisión: la estabilidad de $\mathcal{D}$ bajo cada T_i da $D \setminus \{x_i\} \in \mathcal{D}$ para todo $D \in \mathcal{D}$ y todo $x_i \in D$; quitando de a uno los puntos de $D \setminus B$ se llega a $B \in \mathcal{D}$.

de modo que el crecimiento es polinómico de orden a lo sumo $O(n^{V(\mathcal{C})})$.

Demostración. La primera desigualdad es (2.27), tomando máximo sobre $x_1, \dots, x_n$. La última es la cota combinatoria usual para sumas binomiales parciales.[62] □

Teorema 2.44. *Existe una constante universal* K *tal que, para toda clase VC de conjuntos* $\mathcal{C}$*, toda probabilidad* Q*, todo* $r \geq 1$ *y todo* $0 < \varepsilon < 1$*,*

$$N(\varepsilon, \mathcal{C}, L_r(Q)) \leq K\, V(\mathcal{C})(4e)^{V(\mathcal{C})} \left(\frac{1}{\varepsilon}\right)^{rV(\mathcal{C})}.$$

Aquí $\mathcal{C}$ *se identifica con la clase de indicadores* $\{\mathbb{1}_C : C \in \mathcal{C}\}$*, y* $\|\mathbb{1}_C - \mathbb{1}_D\|_{Q,r} = Q(C \triangle D)^{1/r}$*. La fórmula se entiende para* $V(\mathcal{C}) \geq 1$*. Si* $V(\mathcal{C}) = 0$ *y la clase es no vacía, todas sus trazas son idénticas y el número de cubrimiento es* 1*, de modo que el caso límite es trivial.*

La demostración se apoya en el Lema 2.42 y no se incluye aquí; véase [5, Sección 2.6].

2.6.2. Clases VC de funciones

Definición 2.45. El **subgráfico** de una función $f : \mathcal{X} \to \mathbb{R}$ es $C_f := \{(x,t) \in \mathcal{X} \times \mathbb{R} : t < f(x)\}$. Una clase $\mathcal{F}$ de funciones medibles se llama *clase VC de subgráficos*, o simplemente **clase VC de funciones**, si la colección $\{C_f : f \in \mathcal{F}\}$ es una clase VC de conjuntos en $\mathcal{X} \times \mathbb{R}$. Denotamos por $V(\mathcal{F})$ la dimensión VC de esta colección de subgráficos.

Al igual que para conjuntos, los números de cubrimiento de las clases VC de funciones crecen polinómicamente.

Teorema 2.46. *Sea* $\mathcal{F}$ *una clase VC-subgráfico de funciones medibles con envolvente medible* F*, y sea* $r \geq 1$*. Entonces, para toda probabilidad* Q *con* $0 < \|F\|_{Q,r} < \infty$ *y todo* $0 < \varepsilon < 1$*,*

$$N\Big(\varepsilon \|F\|_{Q,r}, \mathcal{F}, L_r(Q)\Big) \leq K\, V(\mathcal{F})(16e)^{V(\mathcal{F})} \left(\frac{1}{\varepsilon}\right)^{rV(\mathcal{F})},$$

donde K *es una constante universal. La fórmula se usa para* $V(\mathcal{F}) \geq 1$*; si* $V(\mathcal{F}) = 0$ *y la clase es no vacía, contiene una sola función distinta y el número de cubrimiento es* 1*.*

Proposición 2.47 (Cota elemental con constante relajada)**.** *Bajo las hipótesis del teorema anterior y suponiendo* $V(\mathcal{F}) \geq 1$*,*

$$N\Big(\varepsilon \|F\|_{Q,r}, \mathcal{F}, L_r(Q)\Big) \leq K\, V(\mathcal{F})(16e)^{V(\mathcal{F})} \left(\frac{2}{\varepsilon}\right)^{rV(\mathcal{F})}.$$

Demostración. Sea $\mathcal{C} = \{C_f : f \in \mathcal{F}\}$ la clase de subgráficos de la Definición 2.45. Para $f, g \in \mathcal{F}$, por Fubini,

$$(Q \times \lambda)(C_f \triangle C_g) = \int \lambda\{t : \mathbb{1}_{\{t<f(x)\}} \neq \mathbb{1}_{\{t<g(x)\}}\}\, dQ(x) = Q|f - g|. \tag{1}$$

Además, $C_f \triangle C_g \subset S := \{(x,t) : |t| \leq F(x)\}$.

Primero sea $r = 1$. Si $QF > 0$, definimos una probabilidad en $\mathcal{X} \times \mathbb{R}$ por $P(A) = \frac{(Q\times\lambda)(A \cap S)}{2QF}$. Como $C_f \triangle C_g \subset S$, de (1)

$$P(C_f \triangle C_g) = \frac{Q|f - g|}{2QF}. \tag{2}$$

Los centros de un cubrimiento de $\mathcal{C}$ pueden ser conjuntos externos a la clase; internalizando un cubrimiento de radio $\varepsilon/2$ en $L_1(P)$ (véase el párrafo que sigue a la Definición 2.26) se obtienen

[62] Con $V = V(\mathcal{C})$ y $1 \leq V \leq n$: como $(V/n)^{j-V} \geq 1$ para $j \leq V$, $\sum_{j\leq V} \binom{n}{j} \leq (n/V)^V \sum_{j\leq n} \binom{n}{j}(V/n)^j = (n/V)^V(1 + V/n)^n \leq (n/V)^V e^V$, usando el binomio de Newton y $1 + x \leq e^x$.

subgráficos $C_{f_j} \in \mathcal{C}$ tales que, para todo $f \in \mathcal{F}$, hay un índice j con $P(C_f \triangle C_{f_j}) < \varepsilon$, y (2) da $Q|f - f_j| < 2\varepsilon QF$. Por tanto,

$$N(2\varepsilon QF, \mathcal{F}, L_1(Q)) \leq N\left(\frac{\varepsilon}{2}, \mathcal{C}, L_1(P)\right). \tag{3}$$

Aplicando el Teorema 2.44 a $\mathcal{C}$, en el radio $\varepsilon/2$,

$$N(2\varepsilon QF, \mathcal{F}, L_1(Q)) \leq K\, V(\mathcal{F})(4e)^{V(\mathcal{F})} \left(\frac{2}{\varepsilon}\right)^{V(\mathcal{F})}. \tag{4}$$

Reescalando el radio, y aumentando la constante de la base, obtenemos el caso $r = 1$:

$$N(\varepsilon QF, \mathcal{F}, L_1(Q)) \leq K\, V(\mathcal{F})(16e)^{V(\mathcal{F})} \left(\frac{1}{\varepsilon}\right)^{V(\mathcal{F})}. \tag{5}$$

Sea ahora $r > 1$. Como $0 < \|F\|_{Q,r} < \infty$, también $0 < QF^{r-1} < \infty$. Definimos $\frac{dR}{dQ} := \frac{F^{r-1}}{QF^{r-1}}$. Usando $|f - g| \leq 2F$,

$$Q|f - g|^r \leq Q\{|f - g|(2F)^{r-1}\} = 2^{r-1}(QF^{r-1})R|f - g|. \tag{6}$$

Luego

$$\|f - g\|_{Q,r} \leq 2\,(QF^{r-1})^{1/r}\|f - g\|_{R,1}^{1/r}. \tag{7}$$

Además,

$$(QF^{r-1})^{1/r}(RF)^{1/r} = (QF^r)^{1/r} = \|F\|_{Q,r}. \tag{8}$$

Así, si $f, g \in \mathcal{F}$ y $\|f - g\|_{R,1} \leq \varepsilon^r RF$, entonces $\|f - g\|_{Q,r} \leq 2\varepsilon\|F\|_{Q,r}$; nótese que la cota $|f - g| \leq 2F$ usada en (6) requiere que ambas funciones estén dominadas por la envolvente, así que este paso vale para redes con centros *en la clase*. Por consiguiente, toda red de radio $\varepsilon^r RF$ en $L_1(R)$ con centros en $\mathcal{F}$ es una red de radio $2\varepsilon\|F\|_{Q,r}$ en $L_r(Q)$, y

$$N(2\varepsilon\|F\|_{Q,r}, \mathcal{F}, L_r(Q)) \leq N_{\text{int}}(\varepsilon^r RF, \mathcal{F}, L_1(R)), \tag{9}$$

donde N_{int} denota el número de cubrimiento con centros en la clase. Ahora bien, el cubrimiento construido en el caso $r = 1$ tiene centros $f_j \in \mathcal{F}$ (la internalización se hizo en (3)), de modo que la cota (5) vale también para N_{int} sin costo adicional. Aplicándola con la probabilidad R y radio ε^r,

$$N_{\text{int}}(\varepsilon^r RF, \mathcal{F}, L_1(R)) \leq K\, V(\mathcal{F})(16e)^{V(\mathcal{F})} \left(\frac{1}{\varepsilon}\right)^{rV(\mathcal{F})}. \tag{10}$$

Combinando (9) y (10), y sustituyendo ε por $\varepsilon/2$, obtenemos

$$N(\varepsilon\|F\|_{Q,r}, \mathcal{F}, L_r(Q)) \leq K\, V(\mathcal{F})(16e)^{V(\mathcal{F})} \left(\frac{2}{\varepsilon}\right)^{rV(\mathcal{F})},$$

como se afirmó. □

El Teorema 2.46 muestra que toda clase VC de funciones satisface la condición de entropía uniforme con holgura. Por tanto, si la clase es suficientemente medible y $P^*F^2 < \infty$, el Teorema 2.35 implica que $\mathcal{F}$ es P-Donsker. Para clases de conjuntos, la envolvente es $F \equiv 1$, de modo que las clases VC de conjuntos son P-Donsker para toda probabilidad P, bajo las condiciones de medibilidad correspondientes.

Lema 2.48 (Espacios vectoriales finito-dimensionales)**.** *Sea $\mathcal{F}$ un espacio vectorial finito-dimensional de funciones medibles $f : \mathcal{X} \to \mathbb{R}$. Entonces $\mathcal{F}$ es una clase VC-subgráfico y $V(\mathcal{F}) \leq \dim(\mathcal{F}) + 1$.*

Demostración. Sea $d = \dim(\mathcal{F})$ y tomemos $n = d + 2$ puntos $(x_1, t_1), \ldots, (x_n, t_n) \in \mathcal{X} \times \mathbb{R}$. Consideremos los vectores $v_f := (f(x_1) - t_1, \ldots, f(x_n) - t_n) \in \mathbb{R}^n$, $f \in \mathcal{F}$. Como $\{(f(x_1), \ldots, f(x_n)) : f \in \mathcal{F}\}$ está contenido en un subespacio de dimensión $\leq d$, el conjunto de vectores v_f está contenido en un subespacio lineal de dimensión $\leq d+1 = n-1$.[63] Por tanto, existe $a = (a_1, \ldots, a_n) \neq 0$ ortogonal a todos los v_f, es decir, $\sum_{i=1}^{n} a_i\{f(x_i) - t_i\} = 0, \qquad f \in \mathcal{F}$. Multiplicando a por -1 si hace falta, podemos suponer que algún $a_i > 0$. Entonces, para todo $f \in \mathcal{F}$, $\sum_{a_i>0} a_i\{f(x_i) - t_i\} = \sum_{a_i<0}(-a_i)\{f(x_i) - t_i\}$. El conjunto de puntos con $a_i > 0$ no puede ser seleccionado por los subgráficos C_f de la Definición 2.45. En efecto, si existiera f tal que $(x_i, t_i) \in C_f \iff a_i > 0$, entonces el lado izquierdo de la igualdad anterior sería estrictamente positivo, mientras que el lado derecho sería no positivo, una contradicción.[64] Así, ningún conjunto de $d + 2$ puntos puede ser fragmentado por los subgráficos de $\mathcal{F}$. Luego $V(\mathcal{F}) \leq d + 1$. □

Lema 2.49 (Traslaciones de una función monótona)**.** *Sea $\psi : \mathbb{R} \to \mathbb{R}$ una función monótona fija y sea $\mathcal{F} = \{\psi_h : \psi_h(x) = \psi(x - h),\ h \in \mathbb{R}\}$. Entonces $\mathcal{F}$ es VC-subgráfico con dimensión VC a lo sumo 1. Si ψ no es constante, entonces $V(\mathcal{F}) = 1$.*

Demostración. Para cada $h \in \mathbb{R}$, sea $G_h := C_{\psi_h} = \{(x, t) \in \mathbb{R}^2 : t < \psi(x - h)\}$ el subgráfico de ψ_h (Definición 2.45). Supongamos primero que ψ es no decreciente. Si $h_1 \geq h_2$, entonces $x - h_1 \leq x - h_2$ para todo x, y por monotonía $\psi(x - h_1) \leq \psi(x - h_2)$. Por tanto, $h_1 \geq h_2 \Longrightarrow G_{h_1} \subseteq G_{h_2}$. Es decir, los subgráficos forman una familia totalmente ordenada por inclusión. Si ψ es no creciente, se obtiene la misma conclusión con el orden de los h invertido.

Una familia totalmente ordenada por inclusión no puede fragmentar dos puntos.[65] En efecto, si $A \subset B$, las trazas de A y B sobre dos puntos no pueden realizar simultáneamente las dos configuraciones singleton opuestas $\{z_1\}$ y $\{z_2\}$. Por lo tanto, los subgráficos de $\mathcal{F}$ no fragmentan ningún conjunto de dos puntos, y $V(\mathcal{F}) \leq 1$.

Si ψ no es constante, existe un punto que sí puede ser seleccionado y no seleccionado por distintos subgráficos. En efecto, como ψ es monótona y no constante, existen u_1, u_2 tales que, por ejemplo, $\psi(u_1) < \psi(u_2)$. Elegimos t_0 con $\psi(u_1) < t_0 < \psi(u_2)$ y fijamos cualquier $x_0 \in \mathbb{R}$. Tomando $h_1 = x_0 - u_2$, $h_2 = x_0 - u_1$, se tiene $t_0 < \psi(x_0 - h_1) = \psi(u_2)$, $t_0 > \psi(x_0 - h_2) = \psi(u_1)$. Así, $z_0 = (x_0, t_0)$ pertenece a G_{h_1} pero no pertenece a G_{h_2}. Luego el conjunto unitario $\{z_0\}$ es fragmentado. Por tanto, si ψ no es constante, $V(\mathcal{F}) = 1$. □

Lema 2.50 (Permanencia de clases VC de conjuntos)**.** *Sean $\mathcal{C}$ y $\mathcal{D}$ clases VC de subconjuntos de $\mathcal{X}$. Sean $\phi : \mathcal{X} \to \mathcal{Y}$ y $\psi : \mathcal{Z} \to \mathcal{X}$ aplicaciones fijas. Entonces:*

1. *$\mathcal{C}^c := \{C^c : C \in \mathcal{C}\}$ es VC;*
2. *$\mathcal{C} \cap \mathcal{D} := \{C \cap D : C \in \mathcal{C},\ D \in \mathcal{D}\}$ es VC;*
3. *$\mathcal{C} \cup \mathcal{D} := \{C \cup D : C \in \mathcal{C},\ D \in \mathcal{D}\}$ es VC;*
4. *$\phi(\mathcal{C}) := \{\phi(C) : C \in \mathcal{C}\}$ es VC si ϕ es inyectiva;*
5. *$\psi^{-1}(\mathcal{C}) := \{\psi^{-1}(C) : C \in \mathcal{C}\}$ es VC;*
6. *el cierre puntual de $\mathcal{C}$, visto como clase de indicadoras, es VC;*
7. *si $\mathcal{C}$ y $\mathcal{D}$ son VC en $\mathcal{X}$ y $\mathcal{Y}$, respectivamente, entonces $\mathcal{C} \times \mathcal{D} := \{C \times D : C \in \mathcal{C},\ D \in \mathcal{D}\}$ es VC en $\mathcal{X} \times \mathcal{Y}$.*

[63] Los vectores $(f(x_1), \ldots, f(x_n))$ recorren la imagen de $\mathcal{F}$ por la aplicación lineal de evaluación, un subespacio W con $\dim W \leq d$; entonces $v_f \in W + \mathbb{R}\,t$, con $t = (t_1, \ldots, t_n)$, de dimensión $\leq d + 1 < n$, y el complemento ortogonal de ese subespacio contiene algún $a \neq 0$.

[64] Si $a_i > 0$, la pertenencia $(x_i, t_i) \in C_f$ da $f(x_i) - t_i > 0$ y el sumando es estrictamente positivo (y hay al menos un $a_i > 0$); si $a_i < 0$, la no pertenencia da $f(x_i) - t_i \leq 0$ y el sumando derecho es ≤ 0. Nótese que basta exhibir *un* subconjunto no seleccionable para descartar la fragmentación.

[65] Para seleccionar $\{z_1\}$ hace falta un conjunto que contenga z_1 pero no z_2; para $\{z_2\}$, uno que contenga z_2 pero no z_1. Pero dos conjuntos comparables no pueden cumplir esto a la vez: el menor está contenido en el mayor.

Demostración. Para (1), una clase selecciona un subconjunto de $\{x_1, \dots, x_n\}$ si y sólo si la clase de complementos selecciona el complemento de ese subconjunto. Por tanto, fragmentar es invariante por complementos.

Para (2), por el Lema 2.42, de n puntos la clase $\mathcal{C}$ puede seleccionar a lo sumo $O(n^{V(\mathcal{C})})$ subconjuntos, y $\mathcal{D}$ a lo sumo $O(n^{V(\mathcal{D})})$. Luego la clase de intersecciones puede seleccionar a lo sumo $O(n^{V(\mathcal{C})+V(\mathcal{D})})$ subconjuntos.[66] Para n grande esto es estrictamente menor que 2^n; por tanto no puede fragmentar conjuntos de tamaño arbitrariamente grande. Así, $\mathcal{C} \cap \mathcal{D}$ es VC.

La afirmación (3) sigue de (1) y (2), ya que $C \cup D = (C^c \cap D^c)^c$.

Para (4), si $\phi(\mathcal{C})$ fragmenta $\{y_1, \dots, y_n\}$, entonces, como ϕ es inyectiva, existen $x_1, \dots, x_n$ tales que $\phi(x_i) = y_i$, y $\mathcal{C}$ fragmenta $\{x_1, \dots, x_n\}$. Luego $\phi(\mathcal{C})$ no puede tener dimensión VC mayor que $\mathcal{C}$.[67]

Para (5), si $\psi^{-1}(\mathcal{C})$ fragmenta $\{z_1, \dots, z_n\}$, entonces los puntos $\psi(z_i)$ deben ser distintos; de lo contrario no se podría seleccionar uno de z_i, z_j sin seleccionar el otro. Además, $\mathcal{C}$ fragmenta $\{\psi(z_1), \dots, \psi(z_n)\}$. Por tanto $\psi^{-1}(\mathcal{C})$ es VC.

Para (6), sea $\overline{C}$ un límite puntual de elementos de $\mathcal{C}$. Fijados $x_1, \dots, x_n$, para cada i se tiene $\mathbb{1}_{C_\alpha}(x_i) \to \mathbb{1}_{\overline{C}}(x_i)$; como los valores son 0 o 1, para α suficientemente grande coinciden simultáneamente en los n puntos.[68] Así, $\overline{C}$ selecciona sobre $\{x_1, \dots, x_n\}$ un subconjunto que ya seleccionaba algún elemento de $\mathcal{C}$. El cierre puntual no aumenta las trazas finitas y, por tanto, conserva la propiedad VC.

Finalmente, para (7), observamos que $C \times D = (C \times \mathcal{Y}) \cap (\mathcal{X} \times D)$. Las clases $\{C \times \mathcal{Y} : C \in \mathcal{C}\}$ y $\{\mathcal{X} \times D : D \in \mathcal{D}\}$ son VC por (5), aplicando las proyecciones. Luego $\mathcal{C} \times \mathcal{D}$ es VC por (2). □

Lema 2.51 (Permanencia de clases VC-subgráfico)**.** *Sean $\mathcal{F}$ y $\mathcal{G}$ clases VC-subgráfico de funciones sobre $\mathcal{X}$. Sean $a : \mathcal{X} \to \mathbb{R}$ y $\psi : \mathcal{Z} \to \mathcal{X}$ funciones medibles fijas, y $\varphi : \mathbb{R} \to \mathbb{R}$ una función fija. Entonces:*

1. $\mathcal{F} \wedge \mathcal{G} := \{f \wedge g : f \in \mathcal{F},\ g \in \mathcal{G}\}$ *es VC-subgráfico;*
2. $\mathcal{F} \vee \mathcal{G} := \{f \vee g : f \in \mathcal{F},\ g \in \mathcal{G}\}$ *es VC-subgráfico;*
3. $\{\mathcal{F} > 0\} := \{\{x : f(x) > 0\} : f \in \mathcal{F}\}$ *es VC;*
4. $-\mathcal{F} := \{-f : f \in \mathcal{F}\}$ *es VC-subgráfico;*
5. $\mathcal{F} + a := \{f + a : f \in \mathcal{F}\}$ *es VC-subgráfico;*
6. $\mathcal{F} \cdot a := \{fa : f \in \mathcal{F}\}$ *es VC-subgráfico;*
7. $\mathcal{F} \circ \psi := \{f \circ \psi : f \in \mathcal{F}\}$ *es VC-subgráfico;*
8. $\varphi \circ \mathcal{F} := \{\varphi \circ f : f \in \mathcal{F}\}$ *es VC-subgráfico si φ es monótona.*

Demostración. Llamamos subgráfico abierto de f al subgráfico C_f de la Definición 2.45, y subgráfico cerrado a $\overline{C}_f := \{(x, t) : t \le f(x)\}$. Primero observamos que las clases de subgráficos abiertos y cerrados tienen la misma dimensión VC. En efecto, si los cerrados fragmentan $(x_i, t_i)_{i \le m}$, para cada $S \subset \{1, \dots, m\}$ elegimos f_S tal que $t_i \le f_S(x_i)$ si y sólo si $i \in S$. Como hay sólo finitos pares (S, i) con $i \notin S$, existe $0 < \eta < \min_{S, i \notin S}\{t_i - f_S(x_i)\}$. Entonces los subgráficos abiertos de las mismas funciones fragmentan $(x_i, t_i - \eta)_{i \le m}$. La desigualdad recíproca se obtiene desplazando los niveles en el sentido opuesto.

Para (1) y (2), $C_{f \wedge g} = C_f \cap C_g$, $C_{f \vee g} = C_f \cup C_g$. Luego el resultado sigue del Lema 2.50.

Para (3), si $\iota : \mathcal{X} \to \mathcal{X} \times \mathbb{R}$ está dada por $\iota(x) = (x, 0)$, entonces $\{x : f(x) > 0\} = \iota^{-1}(C_f)$. Por permanencia bajo preimágenes, la clase $\{\mathcal{F} > 0\}$ es VC.

[66] Porque $(C \cap D) \cap S = (C \cap S) \cap (D \cap S)$: cada traza de $\mathcal{C} \cap \mathcal{D}$ sobre S está determinada por un par de trazas, de modo que $\Delta_n(\mathcal{C} \cap \mathcal{D}) \le \Delta_n(\mathcal{C})\, \Delta_n(\mathcal{D})$; por Sauer, esto es polinómico en n y, para n grande, $< 2^n$.

[67] Si algún y_i no estuviera en la imagen de ϕ, ningún $\phi(C)$ lo contendría y el conjunto no podría fragmentarse (no se selecciona el total); así, $y_i = \phi(x_i)$ para ciertos x_i, y la inyectividad da $\phi(C) \cap \{y_1, \dots, y_n\} = \phi(C \cap \{x_1, \dots, x_n\})$: las trazas se corresponden biunívocamente.

[68] Una red de ceros y unos que converge es eventualmente constante; para cada i hay un índice a partir del cual coinciden, y por la propiedad de conjunto dirigido hay un índice común a los n puntos.

Para (4), los subgráficos de $-f$ se obtienen, mediante la transformación fija $(x,t)\mapsto(x,-t)$, como complementos de subgráficos cerrados de f. Por la observación precedente y la permanencia bajo complementos y preimágenes, $-\mathcal{F}$ es VC-subgráfico.[69]

Para (5), $C_{f+a}=\{(x,t): t-a(x)<f(x)\}$, que es la preimagen de C_f por la aplicación fija $(x,t)\mapsto(x,t-a(x))$. Luego $\mathcal{F}+a$ es VC-subgráfico.

Para (6), descomponemos según el signo de a: $C_{fa}=\{t<fa,\ a>0\}\cup\{t<fa,\ a<0\}\cup\{t<0,\ a=0\}$. En $\{a>0\}$, la condición $t<fa(x)$ equivale a $t/a(x)<f(x)$, una preimagen de C_f. En $\{a<0\}$ equivale a $t/a(x)>f(x)$, una preimagen de un supergráfico de f, que también es VC por el argumento de (4). El tercer término es fijo. Por uniones finitas e intersecciones con conjuntos fijos, $\mathcal{F}\cdot a$ es VC-subgráfico.[70]

Para (7), $C_{f\circ\psi}=\{(z,t): t<f(\psi(z))\}$ es la preimagen de C_f por la aplicación fija $(z,t)\mapsto(\psi(z),t)$. Luego $\mathcal{F}\circ\psi$ es VC-subgráfico.

Finalmente probamos (8). Supongamos primero que φ es no decreciente. Para cada s, el conjunto $A_s:=\{u: s<\varphi(u)\}$ es un intervalo superior: (a_s,∞), $[a_s,\infty)$, $\varnothing$ o $\mathbb{R}$. Supongamos que los subgráficos de $\varphi\circ\mathcal{F}$ fragmentan $(x_i,s_i)_{i\le m}$. Ninguno de los conjuntos A_{s_i} puede ser $\varnothing$ ni $\mathbb{R}$, pues en esos casos la pertenencia de la coordenada i sería constante.[71] Por tanto, para cada i, existe a_i tal que la condición de pertenencia es $a_i<f(x_i)$ o $a_i\le f(x_i)$.

Para cada $S\subset\{1,\dots,m\}$, elegimos $f_S\in\mathcal{F}$ que realiza la traza S. Sea I_c el conjunto de coordenadas en las que aparece la condición cerrada $a_i\le f(x_i)$. Para cada $i\in I_c$, las cantidades $a_i-f_S(x_i),\qquad S\subset\{1,\dots,m\},\quad i\notin S$, son estrictamente positivas. Como son finitas en número, podemos elegir $\eta_i>0$ menor que todas ellas. Definimos

$$b_i:=\begin{cases}a_i-\eta_i, & i\in I_c,\\ a_i, & i\notin I_c.\end{cases}$$

Entonces, para todo S y todo i, $b_i<f_S(x_i)\iff i\in S$. Así, los subgráficos abiertos de $\mathcal{F}$ fragmentan $(x_i,b_i)_{i\le m}$. En consecuencia, la dimensión VC-subgráfico de $\varphi\circ\mathcal{F}$ no excede la de $\mathcal{F}$.

Si φ es no creciente, la función $\widetilde{\varphi}(u):=\varphi(-u)$ es no decreciente y $\varphi\circ f=\widetilde{\varphi}\circ(-f)$. El resultado se sigue del caso anterior y de (4), porque $-\mathcal{F}$ es VC-subgráfico. Esto prueba (8). □

Cierre del círculo: semiespacios, bolas y elipsoides. Los lemas de permanencia permiten verificar la propiedad VC sin volver a la definición combinatoria, y con ella la propiedad Donsker. Sea $\mathcal{G}:=\{x\mapsto\theta^\top x+c:\ \theta\in\mathbb{R}^d,\ c\in\mathbb{R}\}$, un espacio vectorial de funciones sobre $\mathbb{R}^d$ de dimensión $d+1$. Por el Lema 2.48, $\mathcal{G}$ es VC-subgráfico, y por el Lema 2.51(3) la colección de semiespacios abiertos

$$\mathcal{H}=\{\mathcal{G}>0\}=\Big\{\{x\in\mathbb{R}^d:\theta^\top x+c>0\}:\theta\in\mathbb{R}^d,\ c\in\mathbb{R}\Big\}$$

es una clase VC de conjuntos. Su envolvente es $F\equiv1$, de modo que $P^*F^2=1<\infty$ para toda P; el Teorema 2.44 da la cota (2.26), que implica (UE), y el Teorema 2.35 concluye que $\mathcal{H}$ es P-Donsker para toda probabilidad P en $\mathbb{R}^d$, bajo las condiciones de medibilidad correspondientes. Agrandando el espacio vectorial se cubren más clases con el mismo argumento: con $\mathcal{G}_0=\operatorname{span}\{1,x_1,\dots,x_d,\|x\|^2\}$, de dimensión $d+2$, la clase $\{\mathcal{G}_0>0\}$ contiene todas las bolas abiertas, pues $\{x:\|x-a\|<r\}=\{x:r^2-\|a\|^2+2a^\top x-\|x\|^2>0\}$; y con el espacio de todos los polinomios de grado a lo sumo 2, de dimensión $(d+1)(d+2)/2$, se obtienen los elipsoides. En todos los casos la envolvente es 1 y la conclusión Donsker es la misma.

[69] En fórmulas: $C_{-f}=\{(x,t):t<-f(x)\}=\{(x,t):-t>f(x)\}$, la preimagen por la aplicación fija $(x,t)\mapsto(x,-t)$ del complemento del subgráfico cerrado $\overline{C}_f$; se combinan las partes (1) y (5) del Lema 2.50 con la observación inicial.

[70] Formalmente, la clase $\{C_{fa}:f\in\mathcal{F}\}$ está contenida en $\{(A\cap E_+)\cup(B\cap E_-)\cup K:A\in\mathcal{A},\,B\in\mathcal{B}\}$, con $\mathcal{A},\mathcal{B}$ clases VC (preimágenes por aplicaciones fijas de subgráficos y supergráficos), $E_\pm=\{a\gtrless0\}\times\mathbb{R}$ y K fijos; una clase de un solo conjunto es VC, y se aplican las partes (2) y (3) del Lema 2.50. La permanencia (5) de la dimensión VC bajo preimágenes es puramente conjuntista: no requiere inyectividad ni medibilidad. La medibilidad de a y ψ se exige, en cambio, para que las funciones transformadas $(f+a,\ fa,\ f\circ\psi)$ sigan siendo medibles, como pide la Definición 2.45; la φ monótona de (8) es automáticamente boreliana.

[71] La pertenencia $(x_i,s_i)\in C_{\varphi\circ f}$ es $s_i<\varphi(f(x_i))$, es decir $f(x_i)\in A_{s_i}$. Si $A_{s_i}=\varnothing$, el punto no pertenece a ningún subgráfico y no se selecciona el total; si $A_{s_i}=\mathbb{R}$, pertenece a todos y no puede excluirse. En ambos casos falla la fragmentación.

2.7. Números de bracketing

Mientras que la teoría VC controla muchas clases mediante argumentos combinatorios, las cotas de bracketing para clases suaves provienen de la teoría de aproximación. En algunos casos estas cotas son uniformes respecto de la medida subyacente.

2.7.1. Funciones y conjuntos suaves

Para $\alpha > 0$, sea $m := \max\{j \in \mathbb{N}_0 : j < \alpha\} = \lceil \alpha \rceil - 1$. Así, si $0 < \alpha \leq 1$, entonces $m = 0$. Sea $\mathcal{X} \subset \mathbb{R}^d$ acotado. Para un multiíndice $k = (k_1, \ldots, k_d)$ escribimos $|k| = k_1 + \cdots + k_d$ y $D^k = \frac{\partial^{|k|}}{\partial x_1^{k_1} \cdots \partial x_d^{k_d}}$. Para funciones $f : \mathcal{X} \to \mathbb{R}$ suficientemente regulares, definimos

$$\|f\|_\alpha = \max_{|k| \leq m} \sup_{x \in \operatorname{int}(\mathcal{X})} |D^k f(x)| + \max_{|k|=m} \sup_{\substack{x,y \in \operatorname{int}(\mathcal{X}) \\ x \neq y}} \frac{|D^k f(x) - D^k f(y)|}{\|x - y\|^{\alpha - m}}.$$

Sea $C_M^\alpha(\mathcal{X})$ la clase de funciones continuas $f : \overline{\mathcal{X}} \to \mathbb{R}$ para las cuales la norma anterior está bien definida y satisface $\|f\|_\alpha \leq M$. La notación $\|\cdot\|_\infty$ queda reservada para la norma supremo sobre $\overline{\mathcal{X}}$.

Las cotas clásicas de entropía para $C_1^\alpha(\mathcal{X})$ en norma supremo dan inmediatamente cotas de bracketing en $L_r(Q)$, uniformes en Q.

Teorema 2.52 (Entropía supremo de clases Hölder (Teorema 2.7.1 en [5]))**.** *Sea $\mathcal{X}$ un subconjunto convexo y acotado de $\mathbb{R}^d$, con interior no vacío. Existe una constante K, que depende sólo de α y d, tal que*

$$\log N(\varepsilon, C_1^\alpha(\mathcal{X}), \|\cdot\|_\infty) \leq K\, \lambda(\mathcal{X}^1) \left(\frac{1}{\varepsilon}\right)^{d/\alpha}, \qquad \varepsilon > 0,$$

donde $\mathcal{X}^1 := \{x \in \mathbb{R}^d : \operatorname{dist}(x, \mathcal{X}) < 1\}$.

Corolario 2.53 (Bracketing de clases Hölder)**.** *Sea $\mathcal{X} \subset \mathbb{R}^d$ convexo y acotado, con interior no vacío. Existe una constante K, que depende sólo de α, d y $\operatorname{diam}(\mathcal{X})$, tal que, para todo $r \geq 1$, todo $\varepsilon > 0$ y toda probabilidad Q sobre $\mathcal{X}$,*

$$\log N_{[]}(\varepsilon, C_1^\alpha(\mathcal{X}), L_r(Q)) \leq K \left(\frac{1}{\varepsilon}\right)^{d/\alpha}.$$

Demostración. Sea $p = N(\eta/2, C_1^\alpha(\mathcal{X}), \|\cdot\|_\infty)$. Internalizando el cubrimiento (véase el párrafo que sigue a la Definición 2.26), obtenemos $f_1, \ldots, f_p \in C_1^\alpha(\mathcal{X})$ tales que todo $f \in C_1^\alpha(\mathcal{X})$ satisface $\|f - f_i\|_\infty \leq \eta$ para algún i; en particular, los f_i son medibles y los brackets que siguen son admisibles. Entonces $f_i - \eta \leq f \leq f_i + \eta$, de modo que $[f_i - \eta, f_i + \eta]$ cubre a f. Como Q es probabilidad, $\|(f_i + \eta) - (f_i - \eta)\|_{Q,r} = 2\eta$. Por tanto,

$$N_{[]}(2\eta, C_1^\alpha(\mathcal{X}), L_r(Q)) \leq N\left(\frac{\eta}{2}, C_1^\alpha(\mathcal{X}), \|\cdot\|_\infty\right).$$

Sea K_1 la constante del Teorema 2.52. Aplicándolo en el radio $\eta/2 = \varepsilon/4$, con $\eta = \varepsilon/2$,

$$\log N_{[]}(\varepsilon, C_1^\alpha(\mathcal{X}), L_r(Q)) \leq 4^{d/\alpha} K_1\, \lambda(\mathcal{X}^1) \left(\frac{1}{\varepsilon}\right)^{d/\alpha}.$$

Si $D = \operatorname{diam}(\mathcal{X})$ y $x_0 \in \mathcal{X}$, entonces $\mathcal{X}^1 \subset B(x_0, D+1)$; por tanto $\lambda(\mathcal{X}^1) \leq v_d (D+1)^d$.[72] Por lo tanto vale la cota anunciada con $K = 4^{d/\alpha} v_d (D+1)^d K_1$, que depende sólo de α, d y D. □

[72] v_d es el volumen de la bola unitaria de $\mathbb{R}^d$. La inclusión: si $y \in \mathcal{X}^1$, hay $x \in \mathcal{X}$ con $\|y - x\| < 1$, y $\|y - x_0\| \leq \|y - x\| + \|x - x_0\| < 1 + D$.

Consecuencia. El Corolario 2.53, junto con el teorema de Donsker por bracketing, implica que $C_1^\alpha([0,1]^d)$ es universalmente Donsker si $\alpha > d/2$. En efecto, para $r = 2$ la integral de bracketing queda dominada cerca de 0 por $\int_0^1 \varepsilon^{-d/(2\alpha)}\, d\varepsilon$, que es finita exactamente cuando $d/(2\alpha) < 1$.[73] Por ejemplo, en $[0,1]$ basta una condición Hölder uniforme de orden $> 1/2$; en $[0,1]^2$ basta cualquier suavidad $\alpha > 1$.

Subgráficos suaves. La suavidad de las fronteras no basta por sí sola para obtener una clase Donsker de conjuntos: también se requiere una condición sobre la medida subyacente. Si la masa de probabilidad se concentra de forma errática cerca de las fronteras, la suavidad geométrica ayuda poco. El siguiente resultado da una cota de bracketing para subgráficos suaves bajo una hipótesis de densidad acotada.

Corolario 2.54 (Subgráficos Hölder)**.** *Sea $d \geq 2$ y sea $\mathcal{C}_{\alpha,d}$ la colección de subgráficos en $[0,1]^d$ de funciones $f \in C_1^\alpha([0,1]^{d-1})$, es decir,*

$$\mathcal{C}_{\alpha,d} := \Big\{ \{(x,t) \in [0,1]^{d-1} \times [0,1] : t \leq f(x)\} : f \in C_1^\alpha([0,1]^{d-1}) \Big\}.$$

Si Q es una probabilidad en $\mathbb{R}^d$ con densidad de Lebesgue acotada q, entonces, para todo $r \geq 1$ y todo $\varepsilon > 0$,

$$\log N_{[]}(\varepsilon, \mathcal{C}_{\alpha,d}, L_r(Q)) \leq K \, \|q\|_\infty^{(d-1)/\alpha} \left(\frac{1}{\varepsilon}\right)^{(d-1)r/\alpha},$$

donde K depende sólo de α y d. Aquí $\mathcal{C}_{\alpha,d}$ se identifica con la clase de indicadoras $\{\mathbb{1}_C : C \in \mathcal{C}_{\alpha,d}\}$.

Demostración. Sea $p = N(\eta/2, C_1^\alpha([0,1]^{d-1}), \|\cdot\|_\infty)$. Internalizando el cubrimiento (véase el párrafo que sigue a la Definición 2.26), obtenemos $f_1, \ldots, f_p \in C_1^\alpha([0,1]^{d-1})$ que forman una η-red en norma supremo; en particular, los f_i son continuos y los conjuntos que siguen son medibles. Sea K_1 la constante del Teorema 2.52 en dimensión $d-1$. Aplicándolo en el radio $\eta/2$, $\log p \leq K_0 \eta^{-(d-1)/\alpha}$, con $K_0 = 2^{(d-1)/\alpha} K_1 \, \lambda\Big(([0,1]^{d-1})^1\Big)$. Para cada i, definimos los conjuntos inferior y superior dentro del mismo cilindro que aparece en la definición de $\mathcal{C}_{\alpha,d}$:

$$C_i^- := \{(x,t) \in [0,1]^{d-1} \times [0,1] : t \leq f_i(x) - \eta\},$$
$$C_i^+ := \{(x,t) \in [0,1]^{d-1} \times [0,1] : t \leq f_i(x) + \eta\}.$$

Si $S_f := \{(x,t) \in [0,1]^{d-1} \times [0,1] : t \leq f(x)\}$ es el elemento de $\mathcal{C}_{\alpha,d}$ asociado a f y $\|f - f_i\|_\infty \leq \eta$, entonces $C_i^- \subset S_f \subset C_i^+$. Además, $C_i^+ \setminus C_i^- \subset \{(x,t) : x \in [0,1]^{d-1},\ f_i(x) - \eta < t \leq f_i(x) + \eta\}$. Como Q tiene densidad acotada q, $Q(C_i^+ \setminus C_i^-) \leq \int_{[0,1]^{d-1}} \int_{f_i(x)-\eta}^{f_i(x)+\eta} q(x,t)\, dt\, dx \leq 2\eta \|q\|_\infty$. Por tanto, el tamaño $L_r(Q)$ del bracket de indicadoras $[\mathbb{1}_{C_i^-}, \mathbb{1}_{C_i^+}]$ es a lo sumo $\|\mathbb{1}_{C_i^+} - \mathbb{1}_{C_i^-}\|_{Q,r} = Q(C_i^+ \setminus C_i^-)^{1/r} \leq (2\eta\|q\|_\infty)^{1/r}$. En consecuencia, si $2\eta\|q\|_\infty < \varepsilon^r$, estos brackets tienen tamaño estrictamente menor que ε, como exige la definición, y

$$N_{[]}(\varepsilon, \mathcal{C}_{\alpha,d}, L_r(Q)) \leq p = N\left(\frac{\eta}{2}, C_1^\alpha([0,1]^{d-1}), \|\cdot\|_\infty\right).$$

Tomando $\eta = \frac{\varepsilon^r}{4\|q\|_\infty}$ se obtiene

$$\log N_{[]}(\varepsilon, \mathcal{C}_{\alpha,d}, L_r(Q)) \leq K_0 \left(\frac{4\|q\|_\infty}{\varepsilon^r}\right)^{(d-1)/\alpha}.$$

Es decir,

$$\log N_{[]}(\varepsilon, \mathcal{C}_{\alpha,d}, L_r(Q)) \leq K \, \|q\|_\infty^{(d-1)/\alpha} \left(\frac{1}{\varepsilon}\right)^{(d-1)r/\alpha}, \qquad K = 4^{(d-1)/\alpha} K_0.$$

□

[73] La raíz de la cota del corolario es $K^{1/2}\varepsilon^{-d/(2\alpha)}$. Para $\varepsilon > 2$, un único bracket $[-1,1]$ cubre la clase ($\|f\|_\infty \leq \|f\|_\alpha \leq 1$) y tiene tamaño $\|2\|_{Q,2} = 2 < \varepsilon$; la integral sobre $(2,\infty)$ se anula y sólo importa el comportamiento cerca de 0.

Observación 2.55 (El caso $d = 1$)**.** Si $d = 1$, el dominio horizontal es un punto y la clase se reduce a una familia de intervalos iniciales de $[0, 1]$. Es una clase VC de dimensión 1 y, en particular, es Donsker. La cota de potencia del corolario anterior no debe extrapolarse poniendo formalmente $d - 1 = 0$: en ese caso la entropía puede crecer de manera logarítmica en $1/\varepsilon$, y no existe una cota uniforme independiente de ε.

Consecuencia para subgráficos. Para $d \geq 2$, del Corolario 2.54 y el teorema de Donsker por bracketing, los subgráficos de funciones en $C_1^\alpha([0, 1]^{d-1})$ son P-Donsker para toda P dominada por Lebesgue con densidad acotada si $\alpha > d - 1$. En efecto, aplicando el corolario con $r = 2$, la integral de bracketing queda dominada cerca de 0 por $\int_0^1 \varepsilon^{-(d-1)/\alpha}\, d\varepsilon$, que es finita exactamente cuando $\alpha > d - 1$. En el caso de conjuntos del plano recortados por funciones $f : [0, 1] \to [0, 1]$, esto se reduce a $\alpha > 1$; es decir, basta una derivada con regularidad Hölder positiva.

Capítulo 3

Aplicaciones estadísticas

3.1. M-estimadores

El método más importante para construir estimadores estadísticos consiste en elegir el estimador que maximiza cierta función de criterio. Denominaremos a tales estimadores *M-estimadores* (de "maximum" o "minimum"). En el caso de observaciones i.i.d. $X_1, \ldots, X_n$, un tipo común de función de criterio es de la forma $\theta \mapsto \mathbb{P}_n m_\theta = \frac{1}{n}\sum_{i=1}^n m_\theta(X_i)$, para funciones m_θ conocidas y definidas sobre el espacio muestral. En particular, el método de máxima verosimilitud corresponde a la elección $m_\theta = \log p_\theta$, donde p_θ es la densidad de las observaciones.

La teoría de procesos empíricos aparece de manera natural al estudiar las propiedades asintóticas de estos estimadores. En este capítulo presentamos varios resultados que proporcionan la distribución asintótica de los M-estimadores. Algunos resultados son de carácter general, mientras que otros presuponen el esquema de observaciones i.i.d.

En muchas situaciones, los estimadores que maximizan cierta aplicación también resuelven un sistema de ecuaciones. En particular, en el caso de observaciones i.i.d., muchos estimadores son un *cero* de una aplicación del tipo $\theta \mapsto \mathbb{P}_n \psi_\theta$, para aplicaciones dadas ψ_θ en el espacio muestral. Nos referiremos a las soluciones de ecuaciones de estimación como *Z-estimadores* (de "zero", cero).[1] Obsérvese que en parte de la literatura el nombre *M-estimador* se utiliza (también) para lo que aquí llamamos *Z-estimador*, y no siempre se hace la distinción entre los distintos tipos de estimadores.

El comportamiento asintótico de los Z-estimadores se estudia más adelante. El enfoque allí presentado ofrece una alternativa al enfoque de este capítulo, aunque en general las propiedades asintóticas de los M-estimadores se estudian más fácilmente a partir de su caracterización como máximos. En este capítulo discutimos dos enfoques que utilizan directamente esta caracterización.

El parámetro natural (o funcional) se denota θ y se supone que recorre un subconjunto Θ de un espacio métrico; por ejemplo, un espacio euclídeo. A menudo resulta útil separar la derivación del comportamiento límite de una sucesión de estimadores $\hat{\theta}_n$ en subproblemas, tales como demostrar la consistencia, obtener la tasa de convergencia y, finalmente, establecer la distribución límite. En este esquema ciertos teoremas se aplican a parámetros locales (pseudo-parámetros) (como $h = \sqrt{n}\,(\theta - \theta_0)$ si θ_0 es el parámetro verdadero). Por conveniencia, los teoremas que típicamente se aplican a parámetros locales se enuncian en términos de un parámetro h.

El teorema del argmax

El teorema del argmax formaliza una idea sencilla: bajo condiciones adecuadas, la convergencia de las funciones de criterio permite deducir la convergencia de sus maximizadores. Si el funcional argmax fuera continuo con respecto a alguna métrica en el espacio de funciones de criterio, entonces la convergencia en distribución de las funciones de criterio implicaría la convergencia en distribución de sus puntos de máximo—los M-estimadores—al punto de máximo de la función de criterio límite.

[1] Para algunos propósitos, como demostrar consistencia o tasas de convergencia, particularmente en el caso de modelos convexos en el parámetro, es técnicamente conveniente elegir una función diferente.

Esto es simplemente un caso particular del teorema de la aplicación continua para la convergencia débil.

En consonancia con el desarrollo principal de estas notas —la convergencia de procesos estocásticos en espacios del tipo $\ell^\infty(T)$, como el proceso empírico—desarrollaremos esta idea utilizando la topología uniforme en el espacio de funciones de criterio.

Considérese una sucesión $\{M_n(h) : h \in H\}$ de procesos estocásticos indexados por un espacio métrico H. Para cada n, sea el estimador $\hat{h}_n$ un punto de (casi) máximo de la función de criterio $h \mapsto M_n(h)$. Formalmente, $\hat{h}_n$ puede ser cualquier aplicación con valores en H definida sobre el mismo espacio de probabilidad que el proceso M_n, sin que supongamos necesariamente su medibilidad de Borel. Sea $\{M(h) : h \in H\}$ un proceso límite y sea $\hat{h}$ un elemento aleatorio Borel-medible con valores en H. Por esta razón, en los enunciados y pruebas usaremos probabilidades exteriores $\mathbb{P}^*$ cuando intervengan los estimadores $\hat{h}_n$ o supremos potencialmente no medibles.

El funcional argmax es continuo en aquellas funciones M que tienen un punto de máximo único y "bien separado": la función $h \mapsto M(h)$ debe ser estrictamente menor que $M(\hat{h})$ en el complemento de todo entorno del punto $\hat{h}$. Este requisito, que parece natural, es la primera condición del siguiente lema.

El lema siguiente aísla el argumento básico; en aplicaciones concretas puede ser necesario adaptarlo a la geometría particular del criterio.

Lema 3.1. *Sean M_n, M procesos estocásticos indexados por un espacio métrico H, y sean $A, B \subset H$, con $B \neq \varnothing$. Supóngase que existe un elemento aleatorio Borel-medible $\hat{h}$, con valores en H, tal que, casi seguramente, $M(\hat{h}) > \sup_{h \in A \setminus G} M(h)$ para todo abierto $G \ni \hat{h}$. Supóngase además que $M_n(\hat{h}_n) \geq \sup_{h \in H} M_n(h) - o_{\mathbb{P}^*}(1)$. Si $M_n \Rightarrow M$ en $\ell^\infty(A \cup B)$, entonces, para todo cerrado $F \subset H$, $\limsup_{n \to \infty} \mathbb{P}^*(\hat{h}_n \in F \cap A) \leq \mathbb{P}(\hat{h} \in F \cup B^c)$.*

Demostración. Si $F \cap A = \varnothing$, no hay nada que probar. En otro caso, por el teorema de la aplicación continua, $Y_n := \sup_{h \in F \cap A} M_n(h) - \sup_{h \in B} M_n(h) \Rightarrow Y := \sup_{h \in F \cap A} M(h) - \sup_{h \in B} M(h)$, pues $f \mapsto \sup_{F \cap A} f - \sup_B f$ es continuo para $\|\cdot\|_\infty$.

Sea $\varepsilon_n = o_{\mathbb{P}^*}(1)$ tal que $M_n(\hat{h}_n) \geq \sup_{h \in H} M_n(h) - \varepsilon_n$. Entonces $\{\hat{h}_n \in F \cap A\} \subset \{Y_n \geq -\varepsilon_n\}$. Como $Y_n + \varepsilon_n \Rightarrow Y$, por Slutsky y Portmanteau aplicado al cerrado $[0, \infty)$, $\limsup_n \mathbb{P}^*(\hat{h}_n \in F \cap A) \leq \mathbb{P}(Y \geq 0) = \mathbb{P}(\sup_{F \cap A} M \geq \sup_B M)$.

Resta acotar este último evento. En efecto, si $\hat{h} \notin F$ y $\hat{h} \in B$, entonces $G = F^c$ es un abierto que contiene a $\hat{h}$, y por la separación del máximo, $\sup_{h \in F \cap A} M(h) \leq \sup_{h \in A \setminus G} M(h) < M(\hat{h}) \leq \sup_{h \in B} M(h)$. Por tanto, $\{\sup_{F \cap A} M \geq \sup_B M\} \subset \{\hat{h} \in F\} \cup \{\hat{h} \notin B\}$, y se obtiene $\limsup_{n \to \infty} \mathbb{P}^*(\hat{h}_n \in F \cap A) \leq \mathbb{P}(\hat{h} \in F \cup B^c)$. □

Si $M_n \Rightarrow M$ en $\ell^\infty(H)$, el lema puede aplicarse con $A = B = H$. Entonces, para todo cerrado $F \subset H$, $\limsup_n \mathbb{P}^*(\hat{h}_n \in F) \leq \mathbb{P}(\hat{h} \in F)$, y, por Portmanteau, $\hat{h}_n \Rightarrow \hat{h}$. Esta hipótesis suele ser demasiado fuerte: la convergencia uniforme de las funciones de criterio en todo H es mucho más exigente que la convergencia de las localizaciones de sus máximos. Por eso se prueba primero que los maximizadores no escapan al infinito. Bajo tensión asintótica, basta controlar los criterios sobre compactos.

Teorema 3.2. *Sean M_n, M procesos estocásticos indexados por un espacio métrico H tales que $M_n \Rightarrow M$ en $\ell^\infty(K)$ para todo compacto $K \subset H$. Supóngase que, casi seguramente, $h \mapsto M(h)$ es semicontinua superiormente y posee un máximo único en un punto aleatorio Borel-medible $\hat{h}$, el cual es tenso como aplicación aleatoria en H. Si $(\hat{h}_n)$ es asintóticamente tensa —es decir, para todo $\varepsilon > 0$ existe un compacto $K \subset H$ tal que $\limsup_n \mathbb{P}^*(\hat{h}_n \notin K) < \varepsilon$; en $H = \mathbb{R}^k$ esto es exactamente $\hat{h}_n = O_{\mathbb{P}^*}(1)$— y $M_n(\hat{h}_n) \geq \sup_{h \in H} M_n(h) - o_{\mathbb{P}^*}(1)$, entonces $\hat{h}_n \Rightarrow \hat{h}$ en H.*

Demostración. Fijemos un compacto $K \subset H$. Verifiquemos la condición de separación del Lema 3.1 con $A = B = K$. Trabajamos en el evento de probabilidad uno en el cual $h \mapsto M(h)$ es semicontinua superiormente y tiene máximo único en $\hat{h}$. Sea G un abierto tal que $\hat{h} \in G$. Si $G^c \cap K = \varnothing$, la

separación es trivial. Si $G^c \cap K \neq \varnothing$, entonces, como $G^c \cap K$ es compacto y M es semicontinua superiormente, existe $h_G \in G^c \cap K$ tal que $M(h_G) = \sup_{h \in K \setminus G} M(h)$. Como $h_G \notin G$ y $\hat{h} \in G$, se tiene $h_G \neq \hat{h}$. Por unicidad del máximo global de M, $\sup_{h \in K \setminus G} M(h) = M(h_G) < M(\hat{h})$. Por tanto, casi seguramente, $M(\hat{h}) > \sup_{h \in K \setminus G} M(h)$ para todo abierto $G \ni \hat{h}$.

Aplicando el Lema 3.1 con $A = B = K$, para todo cerrado $F \subset H$, $\limsup_n \mathbb{P}^*(\hat{h}_n \in F \cap K) \leq \mathbb{P}(\hat{h} \in F \cup K^c)$. Luego,

$$\begin{aligned}\limsup_n \mathbb{P}^*(\hat{h}_n \in F) &\leq \limsup_n \mathbb{P}^*(\hat{h}_n \in F \cap K) + \limsup_n \mathbb{P}^*(\hat{h}_n \notin K) \\ &\leq \mathbb{P}(\hat{h} \in F) + \mathbb{P}(\hat{h} \notin K) + \limsup_n \mathbb{P}^*(\hat{h}_n \notin K).\end{aligned}$$

Dado $\varepsilon > 0$, por tensión de $\hat{h}$ y tensión asintótica de $(\hat{h}_n)$, existe un compacto $K \subset H$ tal que $\mathbb{P}(\hat{h} \notin K) \leq \varepsilon$, $\limsup_n \mathbb{P}^*(\hat{h}_n \notin K) \leq \varepsilon$. Así, $\limsup_n \mathbb{P}^*(\hat{h}_n \in F) \leq \mathbb{P}(\hat{h} \in F) + 2\varepsilon$ para todo cerrado F. Como $\varepsilon > 0$ es arbitrario, el teorema de Portmanteau implica $\hat{h}_n \Rightarrow \hat{h}$ en H. □

Corolario 3.3. *Sea $(M_n)_{n \geq 1}$ una sucesión de procesos estocásticos indexados por un espacio métrico Θ, y sea $M : \Theta \to \mathbb{R}$ determinista.*

(I) Supóngase que $\|M_n - M\|_\Theta \xrightarrow{\mathbb{P}^*} 0$ y que existe $\theta_0 \in \Theta$ tal que

$$M(\theta_0) > \sup_{\theta \notin G} M(\theta) \quad \text{para todo abierto } G \ni \theta_0.$$

Entonces toda sucesión $\hat{\theta}_n$ que satisface $M_n(\hat{\theta}_n) \geq \sup_{\theta \in \Theta} M_n(\theta) - o_{\mathbb{P}^*}(1)$ verifica $\hat{\theta}_n \xrightarrow{\mathbb{P}^*} \theta_0$.

(II) Supóngase que $\|M_n - M\|_K \xrightarrow{\mathbb{P}^*} 0$ para todo compacto $K \subset \Theta$, y que $\theta \mapsto M(\theta)$ es semicontinua superiormente y tiene un único máximo en θ_0. Entonces la misma conclusión vale si $(\hat{\theta}_n)$ es asintóticamente tensa; en particular, si $\Theta = \mathbb{R}^d$ (o, más generalmente, si Θ es cerrado en $\mathbb{R}^d$), basta que $(\hat{\theta}_n)$ sea acotada en probabilidad. Si Θ no es cerrado, la acotación en probabilidad no basta: los cerrados acotados de Θ no son necesariamente compactos de Θ, y el maximizador puede fugarse hacia la frontera.

Criterio reescalado y distribución límite. Los resultados anteriores están escritos para un parámetro abstracto h, porque en las aplicaciones a distribución límite se aplican típicamente al parámetro local $h = r_n(\theta - \theta_0)$, donde θ_0 es el valor verdadero y $r_n \to \infty$ es la tasa de convergencia. Si $\hat{\theta}_n$ maximiza $\theta \mapsto M_n(\theta)$, entonces $\hat{h}_n := r_n(\hat{\theta}_n - \theta_0)$ casi maximiza, en el espacio local correspondiente, un criterio de la forma $h \longmapsto a_n\Big\{M_n\Big(\theta_0 + \frac{h}{r_n}\Big) - M_n(\theta_0)\Big\}$, $a_n > 0$. El factor a_n no cambia el punto de máximo, pero permite obtener un límite no degenerado. En cambio, el Corolario 3.3 se aplica normalmente al criterio original M_n, o a un múltiplo positivo suyo, para probar primero consistencia.

Un esquema estándar para obtener la distribución límite de M-estimadores euclidianos consiste en:

- probar la consistencia $\hat{\theta}_n \xrightarrow{\mathbb{P}^*} \theta_0$;
- probar la tasa r_n, equivalentemente $\hat{h}_n = r_n(\hat{\theta}_n - \theta_0) = O_{\mathbb{P}^*}(1)$, es decir la tensión asintótica de $\hat{h}_n$;
- probar que los criterios locales reescalados convergen en distribución a un proceso límite M en $\ell^\infty(\{h : \|h\| \leq K\})$, para todo $K < \infty$.

Si las trayectorias $h \mapsto M(h)$ son semicontinuas superiormente y poseen un máximo único $\hat{h}$, entonces el teorema del argmax da $r_n(\hat{\theta}_n - \theta_0) = \hat{h}_n \Rightarrow \hat{h}$.

Ejemplo 3.4

Sea $X_1, \dots, X_n$ una muestra i.i.d. de P_{θ_0}, con densidades $(p_\theta : \theta \in \Theta)$ respecto de una medida dominante, donde $\Theta \subset \mathbb{R}^k$ es abierto y $\theta_0 \in \Theta$. Sea $L_n(\theta) := \sum_{i=1}^n \log p_\theta(X_i)$ la log-verosimilitud. Para $h \in \mathbb{R}^k$, definimos el criterio local $g_n(h) := L_n\Big(\theta_0 + \frac{h}{\sqrt{n}}\Big) - L_n(\theta_0), \qquad h \in H_n := \sqrt{n}(\Theta - \theta_0)$. Como θ_0 es interior, todo compacto $K \subset \mathbb{R}^k$ está contenido en H_n para n suficientemente grande. Para trabajar sobre un dominio fijo, extendemos el criterio a todo $\mathbb{R}^k$ mediante

$$\widetilde{g}_n(h) = \begin{cases} g_n(h), & h \in H_n, \\ -n, & h \notin H_n. \end{cases}$$

Como $0 \in H_n$ y $g_n(0) = 0 > -n$, se tiene la identidad exacta $\sup_{h \in \mathbb{R}^k} \widetilde{g}_n = \sup_{h \in H_n} g_n \geq 0$. Además, todo casi-maximizador de $\widetilde{g}_n$ con error $o_{\mathbb{P}^*}(1)$ pertenece a H_n con probabilidad que tiende a uno, pues fuera de H_n el criterio vale $-n$, mientras que el supremo es no negativo. Además, sobre cada compacto fijo se tiene $\widetilde{g}_n = g_n$ para todo n suficientemente grande.
Supóngase que el modelo es LAN en θ_0, con información de Fisher I_{θ_0} positiva definida, en la siguiente forma secuencial: existe una sucesión de vectores aleatorios Δ_n tal que $\Delta_n \Rightarrow \Delta, \qquad \Delta \sim N_k(0, I_{\theta_0})$, y, para toda sucesión $h_n \to h$,

$$L_n\left(\theta_0 + \frac{h_n}{\sqrt{n}}\right) - L_n(\theta_0) = h_n^\top \Delta_n - \frac{1}{2} h_n^\top I_{\theta_0} h_n + o_{\mathbb{P}_{\theta_0}}(1).$$

En el caso regular i.i.d., $\Delta_n = n^{-1/2} \sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i)$, con $\dot{\ell}_{\theta_0} := \partial_\theta \log p_\theta|_{\theta=\theta_0}$, con $P_{\theta_0} \dot{\ell}_{\theta_0} = 0, \qquad P_{\theta_0} \dot{\ell}_{\theta_0} \dot{\ell}_{\theta_0}^\top = I_{\theta_0}$. Como $h_n \to h$ y $\Delta_n = O_{\mathbb{P}_{\theta_0}}(1)$, en la expansión anterior el miembro derecho puede reemplazarse por $h^\top \Delta_n - \frac{1}{2} h^\top I_{\theta_0} h + o_{\mathbb{P}_{\theta_0}}(1)$. La formulación con sucesiones $h_n \to h$ es ligeramente más fuerte que la propiedad LAN puntual (la expansión con h fijo); es la que se obtiene, por ejemplo, bajo diferenciabilidad en media cuadrática, y es la que usaremos para obtener la convergencia en $\ell^\infty(K)$. La propiedad LAN implica, en particular, la convergencia finito-dimensional del criterio local: $(g_n(h_1), \dots, g_n(h_m)) \Rightarrow (g(h_1), \dots, g(h_m)), \qquad g(h) := h^\top \Delta - \frac{1}{2} h^\top I_{\theta_0} h$. Esta convergencia finito-dimensional, por sí sola, no basta para aplicar el teorema del argmax. Para ello suponemos además la versión uniforme en compactos: $g_n \Rightarrow g \quad$ en $\ell^\infty(K) \qquad$ para todo compacto $K \subset \mathbb{R}^k$.
Sea $\hat{\theta}_n$ un estimador de máxima verosimilitud, o más generalmente un casi-maximizador, tal que $L_n(\hat{\theta}_n) \geq \sup_{\theta \in \Theta} L_n(\theta) - o_{\mathbb{P}^*_{\theta_0}}(1)$, y supóngase que ya se probó la tasa $\hat{h}_n := \sqrt{n}(\hat{\theta}_n - \theta_0) = O_{\mathbb{P}^*_{\theta_0}}(1)$. Entonces $\hat{h}_n$ casi maximiza $\widetilde{g}_n$ sobre $\mathbb{R}^k$, pues $g_n(\hat{h}_n) = L_n(\hat{\theta}_n) - L_n(\theta_0) \geq \sup_{h \in \mathbb{R}^k} \widetilde{g}_n(h) - o_{\mathbb{P}^*_{\theta_0}}(1)$. El proceso límite g tiene trayectorias continuas y, como I_{θ_0} es positiva definida, $\nabla g(h) = \Delta - I_{\theta_0} h, \qquad \nabla^2 g(h) = -I_{\theta_0} \prec 0$. Por lo tanto g tiene un máximo único en $\hat{h} = I_{\theta_0}^{-1} \Delta$. La tensión asintótica de $(\hat{h}_n)$, la convergencia $\widetilde{g}_n \Rightarrow g$ en $\ell^\infty(K)$ para todo compacto $K \subset \mathbb{R}^k$, y la versión localizada del teorema del argmax implican $\sqrt{n}(\hat{\theta}_n - \theta_0) = \hat{h}_n \Rightarrow \hat{h} = I_{\theta_0}^{-1} \Delta$. Finalmente, como $\Delta \sim N_k(0, I_{\theta_0})$, $\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow N_k(0, I_{\theta_0}^{-1})$.

3.1.1. Tasa de convergencia

En el caso regular euclidiano, si M es dos veces diferenciable, θ_0 es interior y M tiene un máximo local en θ_0, entonces $\nabla M(\theta_0) = 0$ y el Hessiano es semidefinido negativo. Si, además, el Hessiano es definido negativo, la expansión de Taylor motiva la condición cuadrática local $M(\theta) - M(\theta_0) \lesssim -d^2(\theta, \theta_0)$ para θ cerca de θ_0. La tasa de un casi-maximizador de M_n se obtiene comparando esta caída determinista con el módulo de continuidad de $M_n - M$.

Teorema 3.5. *Sea M_n una sucesión de procesos estocásticos indexados por un espacio semimétrico (Θ, d), y sea $M : \Theta \to \mathbb{R}$ determinista. Supóngase que existen $c > 0$ y una vecindad U de θ_0 tales*

que $M(\theta) - M(\theta_0) \le -c\, d^2(\theta, \theta_0), \qquad \theta \in U$. *Supóngase además que, para todo* n *y todo* $\delta > 0$ *suficientemente pequeño,*

$$\mathbb{E}^* \sup_{d(\theta,\theta_0)\le\delta} |(M_n - M)(\theta) - (M_n - M)(\theta_0)| \lesssim \frac{\phi_n(\delta)}{\sqrt{n}},$$

donde, para algún $\alpha < 2$ *independiente de* n, $\delta \mapsto \frac{\phi_n(\delta)}{\delta^\alpha}$ *es no creciente. Sea* $r_n > 0$, *con* $r_n \to \infty$, *tal que* $r_n^2 \phi_n(1/r_n) \le \sqrt{n}$ *para todo* n. *Supóngase que existe una sucesión de variables aleatorias* $\Delta_n \ge 0$, *con* $\Delta_n = O_{\mathbb{P}^*}(r_n^{-2})$, *tal que* $M_n(\hat\theta_n) \ge M_n(\theta_0) - \Delta_n$ *y* $\hat\theta_n \xrightarrow{\mathbb{P}^*} \theta_0$. *Entonces* $r_n\, d(\hat\theta_n, \theta_0) = O_{\mathbb{P}^*}(1)$. *Si las condiciones anteriores valen para todo* θ *y todo* $\delta > 0$, *la hipótesis de consistencia no es necesaria.*

Demostración. Escribamos $W_n = M_n - M$. Por la hipótesis local, existen $\eta > 0$ y $\kappa > 0$ (por ejemplo, $\kappa = c$ y η tal que la bola de radio η centrada en θ_0 esté contenida en U) tales que, si $d(\theta, \theta_0) \le \eta$, $M(\theta) - M(\theta_0) \le -\kappa d^2(\theta, \theta_0)$, y la cota del módulo estocástico vale para todo $0 < \delta \le \eta$. Sea Δ_n la sucesión de la hipótesis de casi-maximización.

Para $j \in \mathbb{Z}$, definamos las coronas diádicas $S_{j,n} := \{\theta : 2^{j-1} < r_n d(\theta, \theta_0) \le 2^j\}$. Fijado un entero m, si $r_n d(\hat\theta_n, \theta_0) > 2^m$ y $2d(\hat\theta_n, \theta_0) < \eta$, entonces $\hat\theta_n \in S_{j,n}$ para algún $j > m$ con $2^j < \eta r_n$. Por tanto,

$$\begin{aligned}\mathbb{P}^*\{r_n d(\hat\theta_n, \theta_0) > 2^m\} &\le \mathbb{P}^*\left(\bigcup_{\substack{j>m\\ 2^j<\eta r_n}} \{\hat\theta_n \in S_{j,n}\}\right) + \mathbb{P}^*\{2d(\hat\theta_n, \theta_0) \ge \eta\} \\ &\le \mathbb{P}^*(A_{n,C}^c) + \sum_{\substack{j>m\\ 2^j<\eta r_n}} \mathbb{P}^*(\hat\theta_n \in S_{j,n}, A_{n,C}) + \mathbb{P}^*\{2d(\hat\theta_n, \theta_0) \ge \eta\}.\end{aligned} \tag{1}$$

Sea $C > 0$ y pongamos $A_{n,C} := \{r_n^2 \Delta_n \le C\}$.

En el evento $A_{n,C} \cap \{\hat\theta_n \in S_{j,n}\}$, $\sup_{\theta\in S_{j,n}}\{M_n(\theta) - M_n(\theta_0)\} \ge -C r_n^{-2}$. Si además $2^j < \eta r_n$, entonces $S_{j,n} \subset \{\theta : d(\theta, \theta_0) \le \eta\}$, y para todo $\theta \in S_{j,n}$, $M(\theta) - M(\theta_0) \le -\kappa\, 2^{2j-2} r_n^{-2}$. Luego, para m suficientemente grande, dependiendo sólo de C y κ, existe $c > 0$ tal que, para todo $j > m$,

$$\mathbb{P}^*(\hat\theta_n \in S_{j,n}, A_{n,C}) \le \mathbb{P}^*\left\{\sup_{\theta\in S_{j,n}} |W_n(\theta) - W_n(\theta_0)| \ge c\, 2^{2j} r_n^{-2}\right\}.$$

Como $S_{j,n} \subset \{\theta : d(\theta, \theta_0) \le 2^j/r_n\}$, se tiene $\sup_{\theta\in S_{j,n}} |W_n(\theta) - W_n(\theta_0)| \le \sup_{d(\theta,\theta_0)\le 2^j/r_n} |W_n(\theta) - W_n(\theta_0)|$.

Luego, por Markov y por la hipótesis del módulo, con $\delta = 2^j/r_n$,

$$\begin{aligned}\mathbb{P}^*(\hat\theta_n \in S_{j,n}, A_{n,C}) &\lesssim \frac{r_n^2}{2^{2j}}\, \mathbb{E}^* \sup_{d(\theta,\theta_0)\le 2^j/r_n} |W_n(\theta) - W_n(\theta_0)| \\ &\lesssim \frac{r_n^2 \phi_n(2^j/r_n)}{2^{2j}\sqrt{n}}.\end{aligned}$$

Para $j \ge m \ge 0$, la monotonía de $\delta \mapsto \phi_n(\delta)/\delta^\alpha$ implica $\phi_n(2^j/r_n) \le 2^{j\alpha}\phi_n(1/r_n)$, siempre que $2^j/r_n \le \eta$. Usando $r_n^2\phi_n(1/r_n) \le \sqrt{n}$, $\frac{r_n^2\phi_n(2^j/r_n)}{2^{2j}\sqrt{n}} \lesssim 2^{j(\alpha-2)}$. Por consiguiente,

$$\sum_{\substack{j>m\\ 2^j<\eta r_n}} \mathbb{P}^*(\hat\theta_n \in S_{j,n}, A_{n,C}) \lesssim \sum_{j>m} 2^{j(\alpha-2)} \lesssim 2^{m(\alpha-2)}. \tag{2}$$

De (1) y (2),

$$\begin{aligned}\limsup_n \mathbb{P}^*\{r_n d(\hat\theta_n, \theta_0) > 2^m\} \lesssim 2^{m(\alpha-2)} &+ \limsup_n \mathbb{P}^*(A_{n,C}^c) \\ &+ \limsup_n \mathbb{P}^*\{2d(\hat\theta_n, \theta_0) \ge \eta\}.\end{aligned}$$

Por consistencia, $\lim\sup_n \mathbb{P}^*\{2d(\hat{\theta}_n, \theta_0) \geq \eta\} = 0$. Además, como $r_n^2 \Delta_n = O_{\mathbb{P}^*}(1)$, el término $\lim\sup_n \mathbb{P}^*(A_{n,C}^c)$ se hace arbitrariamente pequeño tomando C suficientemente grande.

Finalmente, como $\alpha < 2$, $2^{m(\alpha-2)} \to 0 \qquad (m \to \infty)$. Así, $\lim_{m\to\infty} \lim\sup_n \mathbb{P}^*\{r_n d(\hat{\theta}_n, \theta_0) > 2^m\} = 0$, lo cual equivale a $r_n d(\hat{\theta}_n, \theta_0) = O_{\mathbb{P}^*}(1)$.

Si las condiciones son globales, se omite el recorte por η, la suma se toma sobre todos los $j > m$, y desaparece el término $\mathbb{P}^*\{2d(\hat{\theta}_n, \theta_0) \geq \eta\}$. Por eso, en el caso global no se necesita suponer $\hat{\theta}_n \xrightarrow{\mathbb{P}^*} \theta_0$. □

Apéndice

3.2. Resultados auxiliares

Teorema 3.6 (Teorema de extensión de Tietze)**.** *Sea (D, d) un espacio métrico y sea $A \subset D$ un subconjunto cerrado no vacío. Toda función continua y acotada $h : A \to \mathbb{R}$ admite una extensión $\widetilde{h} \in C_b(D)$ tal que*

$$\widetilde{h}|_A = h, \qquad \|\widetilde{h}\|_{\infty,D} = \|h\|_{\infty,A}.$$

Más precisamente, si $a \leq h(x) \leq b$ para todo $x \in A$, la extensión puede elegirse de modo que $a \leq \widetilde{h}(x) \leq b$ para todo $x \in D$.

Véase [3, Sección 4.2].

Teorema 3.7 (Teorema de Dini)**.** *Sea K un espacio métrico compacto no vacío y sean $f_n, f \in C(K)$ funciones reales. Supongamos que $f_n(x) \to f(x)$ para todo $x \in K$ y que la sucesión es monótona: o bien $f_{n+1}(x) \leq f_n(x)$ para todo n y todo $x \in K$, o bien se cumple la desigualdad opuesta para todo n y todo $x \in K$. Entonces la convergencia es uniforme:*

$$\|f_n - f\|_{\infty,K} \longrightarrow 0.$$

En particular, si $f_n \downarrow 0$ puntualmente en K, entonces $\sup_{x\in K} f_n(x) \downarrow 0$.

Véase [1, Apéndice M].

Definición 3.8 (π-sistemas y clases de Dynkin)**.** Sea S un conjunto. Una familia no vacía $\mathcal{P}$ de subconjuntos de S es un **π-sistema** si $A \cap B \in \mathcal{P}$ para cualesquiera $A, B \in \mathcal{P}$.

Una familia $\mathcal{L}$ de subconjuntos de S es una **clase de Dynkin**, o **λ-sistema**, si:

1. $S \in \mathcal{L}$;
2. si $A \in \mathcal{L}$, entonces $S \setminus A \in \mathcal{L}$;
3. si $(A_n)_{n\geq 1} \subset \mathcal{L}$ es una sucesión de conjuntos disjuntos dos a dos, entonces $\bigcup_{n\geq 1} A_n \in \mathcal{L}$.

Teorema 3.9 (Teorema de Dynkin o teorema π–λ)**.** *Sea $\mathcal{P}$ un π-sistema de subconjuntos de un conjunto S. Si $\mathcal{L}$ es una clase de Dynkin que contiene a $\mathcal{P}$, entonces $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

En particular, sean μ y ν dos medidas finitas sobre un espacio medible $(S, \mathcal{A})$ y sea $\mathcal{P} \subset \mathcal{A}$ un π-sistema tal que $S \in \mathcal{P}$ y $\sigma(\mathcal{P}) = \mathcal{A}$. Si $\mu(A) = \nu(A)$ para todo $A \in \mathcal{P}$, entonces $\mu = \nu$ en $\mathcal{A}$.

Véase [2, Teorema 2.1.6].

Lema 3.10. *Sean $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$, $i = 1, 2$, espacios de probabilidad y $\mathbb{P} := \mathbb{P}_1 \otimes \mathbb{P}_2$ la probabilidad producto sobre $\mathcal{A}_1 \otimes \mathcal{A}_2$, es decir, la única probabilidad sobre $\mathcal{A}_1 \otimes \mathcal{A}_2$ tal que $\mathbb{P}(A \times B) = \mathbb{P}_1(A)\mathbb{P}_2(B)$ para todo $A \in \mathcal{A}_1$, $B \in \mathcal{A}_2$. Sea $g : \Omega_1 \times \Omega_2 \to [0, \infty]$ medible respecto de $\mathcal{A}_1 \otimes \mathcal{A}_2$. Entonces:*

(a) para todo $\omega_1 \in \Omega_1$, la sección $g(\omega_1, \cdot) : \Omega_2 \to [0, \infty]$ es $\mathcal{A}_2$-medible;

(b) la función $\Phi(\omega_1) := \int_{\Omega_2} g(\omega_1, \omega_2)\, d\mathbb{P}_2(\omega_2)$, que por (a) está definida para todo ω_1 con valores en $[0, \infty]$, es $\mathcal{A}_1$-medible;

(c) (Tonelli)

$$\mathbb{E}g = \int_{\Omega_1\times\Omega_2} g\,d\mathbb{P} = \int_{\Omega_1}\Big(\int_{\Omega_2} g(\omega_1,\omega_2)\,d\mathbb{P}_2(\omega_2)\Big)d\mathbb{P}_1(\omega_1),$$

con ambos miembros en $[0,\infty]$.

Demostración. Para $C\subset\Omega_1\times\Omega_2$ y $\omega_1\in\Omega_1$, escribimos $C_{\omega_1} := \{\omega_2 : (\omega_1,\omega_2)\in C\}$.

Secciones de conjuntos. Sea

$$\mathcal{D} := \{C\in\mathcal{A}_1\otimes\mathcal{A}_2 : C_{\omega_1}\in\mathcal{A}_2\ \forall\omega_1\in\Omega_1\}.$$

Los rectángulos pertenecen a $\mathcal{D}$, pues $(A\times B)_{\omega_1} = B$ si $\omega_1\in A$, y $\varnothing$ en caso contrario. Además,

$$(C^c)_{\omega_1} = (C_{\omega_1})^c, \qquad \Big(\bigcup_n C_n\Big)_{\omega_1} = \bigcup_n (C_n)_{\omega_1},$$

de modo que $\mathcal{D}$ es una σ-álgebra. Como contiene los rectángulos y $\mathcal{D}\subset\mathcal{A}_1\otimes\mathcal{A}_2$, $\mathcal{D} = \mathcal{A}_1\otimes\mathcal{A}_2$.

Fijado ω_1, para todo $c\in\mathbb{R}$, $\{\omega_2 : g(\omega_1,\omega_2) > c\} = (\{g>c\})_{\omega_1}\in\mathcal{A}_2$, pues $\{g>c\}\in\mathcal{A}_1\otimes\mathcal{A}_2$. Como los conjuntos $(c,\infty]$, $c\in\mathbb{R}$, generan $\mathcal{B}([0,\infty])$, esto prueba (a).

Por el primer paso, $C_{\omega_1}\in\mathcal{A}_2$ para todo $C\in\mathcal{A}_1\otimes\mathcal{A}_2$ y todo ω_1, por lo que $\mathbb{P}_2(C_{\omega_1})$ está bien definido. Definamos

$$\mathcal{D}' := \Big\{C\in\mathcal{A}_1\otimes\mathcal{A}_2 : \omega_1\mapsto\mathbb{P}_2(C_{\omega_1}) \text{ es } \mathcal{A}_1\text{-medible}\Big\}.$$

Los rectángulos pertenecen a $\mathcal{D}'$, ya que $\mathbb{P}_2((A\times B)_{\omega_1}) = \mathbb{P}_2(B)\mathbb{1}_A(\omega_1)$. Además, $\mathcal{D}'$ es una clase de Dynkin: contiene $\Omega_1\times\Omega_2$; si $C\subset C'$ están en $\mathcal{D}'$,

$$(C'\setminus C)_{\omega_1} = C'_{\omega_1}\setminus C_{\omega_1}, \qquad \mathbb{P}_2((C'\setminus C)_{\omega_1}) = \mathbb{P}_2(C'_{\omega_1}) - \mathbb{P}_2(C_{\omega_1}),$$

que es medible (aquí se usa que $\mathbb{P}_2$ es finita), y si $C_n\uparrow C$ con $C_n\in\mathcal{D}'$, entonces

$$(C_n)_{\omega_1}\uparrow C_{\omega_1}, \qquad \mathbb{P}_2(C_{\omega_1}) = \lim_n \mathbb{P}_2((C_n)_{\omega_1}),$$

también medible. Como los rectángulos forman un π-sistema que genera $\mathcal{A}_1\otimes\mathcal{A}_2$, el teorema π–λ da $\mathcal{D}' = \mathcal{A}_1\otimes\mathcal{A}_2$.

Si $s = \sum_{k=1}^N a_k\mathbb{1}_{C_k}$ es simple, $a_k\ge 0$, entonces

$$s(\omega_1,\cdot) = \sum_{k=1}^N a_k\mathbb{1}_{(C_k)_{\omega_1}}, \qquad \int_{\Omega_2} s(\omega_1,\cdot)\,d\mathbb{P}_2 = \sum_{k=1}^N a_k\,\mathbb{P}_2((C_k)_{\omega_1}),$$

que es $\mathcal{A}_1$-medible. Para $g\ge 0$ medible, tomamos simples no negativas $s_n\uparrow g$ puntualmente. Fijado ω_1, $s_n(\omega_1,\cdot)\uparrow g(\omega_1,\cdot)$, y por convergencia monótona,

$$\int_{\Omega_2} s_n(\omega_1,\cdot)\,d\mathbb{P}_2 \uparrow \int_{\Omega_2} g(\omega_1,\cdot)\,d\mathbb{P}_2 =: \Phi(\omega_1).$$

Así, Φ es límite puntual de funciones $\mathcal{A}_1$-medibles, y por tanto es $\mathcal{A}_1$-medible. Esto prueba (b).

Para $C\in\mathcal{A}_1\otimes\mathcal{A}_2$, definimos

$$\nu(C) := \int_{\Omega_1}\mathbb{P}_2(C_{\omega_1})\,d\mathbb{P}_1(\omega_1).$$

Entonces $\nu(\Omega_1\times\Omega_2) = 1$. Si C_n son disjuntos dos a dos y $C = \bigcup_n C_n$, entonces las secciones $(C_n)_{\omega_1}$ son disjuntas dos a dos y $C_{\omega_1} = \bigcup_n (C_n)_{\omega_1}$; por tanto,

$$\nu(C) = \int_{\Omega_1}\sum_{n=1}^\infty \mathbb{P}_2((C_n)_{\omega_1})\,d\mathbb{P}_1 = \sum_{n=1}^\infty\int_{\Omega_1}\mathbb{P}_2((C_n)_{\omega_1})\,d\mathbb{P}_1 = \sum_{n=1}^\infty \nu(C_n),$$

por convergencia monótona aplicada a las sumas parciales. Así, ν es una probabilidad. Sobre los rectángulos,

$$\nu(A \times B) = \int_{\Omega_1} \mathbb{P}_2(B)\mathbb{1}_A \, d\mathbb{P}_1 = \mathbb{P}_1(A)\mathbb{P}_2(B) = \mathbb{P}(A \times B).$$

Como los rectángulos forman un π-sistema que genera $\mathcal{A}_1 \otimes \mathcal{A}_2$ y $\nu, \mathbb{P}$ son probabilidades que coinciden sobre él, el teorema de unicidad da $\nu = \mathbb{P}$. Por tanto, para todo $C \in \mathcal{A}_1 \otimes \mathcal{A}_2$,

$$\mathbb{P}(C) = \int_{\Omega_1} \mathbb{P}_2(C_{\omega_1}) \, d\mathbb{P}_1(\omega_1). \tag{3.1}$$

Tonelli. Si $s = \sum_{k=1}^N a_k \mathbb{1}_{C_k}$ es simple y no negativa, por (3.1),

$$\int s \, d\mathbb{P} = \sum_{k=1}^N a_k \mathbb{P}(C_k) = \sum_{k=1}^N a_k \int_{\Omega_1} \mathbb{P}_2((C_k)_{\omega_1}) \, d\mathbb{P}_1 = \int_{\Omega_1} \left(\int_{\Omega_2} s(\omega_1, \cdot) \, d\mathbb{P}_2 \right) d\mathbb{P}_1.$$

Para $g \geq 0$, tomamos nuevamente simples $s_n \uparrow g$ y definimos

$$\Phi_n(\omega_1) := \int_{\Omega_2} s_n(\omega_1, \cdot) \, d\mathbb{P}_2.$$

Por convergencia monótona en Ω_2, $\Phi_n \uparrow \Phi$, y por convergencia monótona en $\Omega_1 \times \Omega_2$ y luego en Ω_1,

$$\int g \, d\mathbb{P} = \lim_n \int s_n \, d\mathbb{P} = \lim_n \int_{\Omega_1} \Phi_n \, d\mathbb{P}_1 = \int_{\Omega_1} \Phi \, d\mathbb{P}_1.$$

Esto prueba (c), con ambos miembros en $[0, \infty]$. □

Completación y productos con un factor finito

Los resultados siguientes se usan en la Sección 2.3 para justificar el teorema de Fubini para el proceso de Rademacher. En todos ellos, $(\Omega, \mathcal{C}, \mu)$ es un espacio de probabilidad.

Definición 3.11 (Completación)**.** La **completación** de $\mathcal{C}$ respecto de μ es

$$\overline{\mathcal{C}} := \Big\{ B \subset \Omega : \ B \triangle C \subset N \text{ para algunos } C, N \in \mathcal{C} \text{ con } \mu(N) = 0 \Big\},$$

donde $\triangle$ denota la diferencia simétrica. Para $B \in \overline{\mathcal{C}}$ se define $\overline{\mu}(B) := \mu(C)$, con C como antes. Cuando haga falta explicitar la medida escribimos $\overline{\mathcal{C}}^{\mu}$. En un producto se completa la σ-álgebra producto respecto de la probabilidad producto.

Es inmediato que $\overline{\mathcal{C}}$ es una σ-álgebra que contiene a $\mathcal{C}$ y que $\overline{\mu}$ es una probabilidad bien definida que extiende a μ.[2]

Lema 3.12. *Sea $h : \Omega \to [0, \infty]$.*

*(i) h es $\overline{\mathcal{C}}$-medible si y sólo si existen una función $\mathcal{C}$-medible $\widetilde{h} : \Omega \to [0, \infty]$ y un conjunto $N \in \mathcal{C}$ con $\mu(N) = 0$ tales que $h = \widetilde{h}$ en $\Omega \setminus N$. Una tal $\widetilde{h}$ se llama una **versión medible** de h.*

(ii) En ese caso, $\int h \, d\overline{\mu} = \int \widetilde{h} \, d\mu$.

(iii) En ese caso, además, $h^* = \widetilde{h}$ c.s. y $\mathbb{E}^* h = \int h \, d\overline{\mu}$, donde el mayorante medible mínimo h^* y la esperanza exterior se calculan en $(\Omega, \mathcal{C}, \mu)$.

[2] Si $B \triangle C \subset N$, también $B^c \triangle C^c = B \triangle C \subset N$; y si $B_k \triangle C_k \subset N_k$ para todo k, entonces $\big(\bigcup_k B_k\big) \triangle \big(\bigcup_k C_k\big) \subset \bigcup_k N_k$, que es nulo. Si $B \triangle C_1 \subset N_1$ y $B \triangle C_2 \subset N_2$, entonces $C_1 \triangle C_2 \subset N_1 \cup N_2$, y por tanto $\mu(C_1) = \mu(C_2)$. Por último, si los B_k son disjuntos dos a dos y $N := \bigcup_k N_k$, los conjuntos $C_k \setminus N = B_k \setminus N$ son disjuntos dos a dos, y $\overline{\mu}\big(\bigcup_k B_k\big) = \mu\big(\bigcup_k C_k \setminus N\big) = \sum_k \mu(C_k \setminus N) = \sum_k \overline{\mu}(B_k)$.

Demostración. (i) Supongamos que h es $\overline{\mathcal{C}}$-medible y sean $s_k \geq 0$ funciones simples $\overline{\mathcal{C}}$-medibles con $s_k \to h$ puntualmente. Cada s_k es combinación lineal finita de indicadores de conjuntos $B \in \overline{\mathcal{C}}$; para cada uno de ellos elegimos $C_B, N_B \in \mathcal{C}$ con $B \triangle C_B \subset N_B$ y $\mu(N_B) = 0$. Sea $\widetilde{s}_k$ la función que se obtiene al reemplazar en s_k cada $\mathbb{1}_B$ por $\mathbb{1}_{C_B}$, y sea N la unión de todos los N_B, que es nula por ser unión numerable de conjuntos nulos. Fuera de N se tiene $\mathbb{1}_B = \mathbb{1}_{C_B}$ para todos los B, y por tanto $\widetilde{s}_k = s_k$ para todo k. Así, $\widetilde{h} := \operatorname{lím\,sup}_k \widetilde{s}_k$ es $\mathcal{C}$-medible, toma valores en $[0, \infty]$ y coincide con h fuera de N.

Recíprocamente, si $h = \widetilde{h}$ fuera de N, para cada $t \in \mathbb{R}$ se tiene $\{h > t\} \triangle \{\widetilde{h} > t\} \subset N$ con $\{\widetilde{h} > t\} \in \mathcal{C}$, luego $\{h > t\} \in \overline{\mathcal{C}}$. Como los conjuntos $(t, \infty]$, $t \in \mathbb{R}$, generan $\mathcal{B}([0, \infty])$, h es $\overline{\mathcal{C}}$-medible.

(ii) Como $h = \widetilde{h}$ fuera del conjunto $\overline{\mu}$-nulo N, $\int h \, d\overline{\mu} = \int \widetilde{h} \, d\overline{\mu}$, y esta última integral es $\int \widetilde{h} \, d\mu$, porque $\widetilde{h}$ es $\mathcal{C}$-medible y $\overline{\mu} = \mu$ en $\mathcal{C}$.

(iii) La función $\widetilde{h}$ es medible y $\widetilde{h} \geq h$ fuera de N, así que el Lema 1.5(ii) da $h^* \leq \widetilde{h}$ c.s. Por otro lado, $h^* \geq h = \widetilde{h}$ fuera de N. Luego $h^* = \widetilde{h}$ c.s. Como $h^* \geq h \geq 0$, la esperanza de h^* existe, y el Lema 1.5(iii) y la parte (ii) dan $\mathbb{E}^* h = \mathbb{E} h^* = \int \widetilde{h} \, d\mu = \int h \, d\overline{\mu}$. □

Lema 3.13. *Sea E un conjunto finito, con la σ-álgebra 2^E de todas sus partes, y sea Q una probabilidad sobre 2^E tal que $Q(\{e\}) > 0$ para todo $e \in E$. Sean $\nu := \mu \otimes Q$ la probabilidad producto sobre $\mathcal{C} \otimes 2^E$, y $\overline{\mathcal{C} \otimes 2^E}$, $\overline{\nu}$ su completación. Para $B \subset \Omega \times E$ y $e \in E$ escribimos $B_e := \{\omega \in \Omega : (\omega, e) \in B\}$. Entonces:*

(a) si $D \in \mathcal{C} \otimes 2^E$ y $\nu(D) = 0$, entonces $\mu(D_e) = 0$ para todo $e \in E$; en particular, $D \subset N \times E$, donde $N := \bigcup_{e \in E} D_e \in \mathcal{C}$ y $\mu(N) = 0$;

(b) un conjunto $B \subset \Omega \times E$ pertenece a $\overline{\mathcal{C} \otimes 2^E}$ si y sólo si $B_e \in \overline{\mathcal{C}}$ para todo $e \in E$;

(c) una función $H : \Omega \times E \to [0, \infty]$ es $\overline{\mathcal{C} \otimes 2^E}$-medible si y sólo si, para cada $e \in E$, la sección $H(\cdot, e)$ es $\overline{\mathcal{C}}$-medible. En ese caso,

$$\int H \, d\overline{\nu} = \int_\Omega \Big(\sum_{e \in E} Q(\{e\}) \, H(\omega, e) \Big) \, d\overline{\mu}(\omega).$$

Demostración. Todo $B \subset \Omega \times E$ es la unión disjunta, finita por serlo E,

$$B = \bigcup_{e \in E} \big(B_e \times \{e\} \big).$$

Por tanto, $B \in \mathcal{C} \otimes 2^E$ si y sólo si $B_e \in \mathcal{C}$ para todo e: si las secciones están en $\mathcal{C}$, B es una unión finita de rectángulos medibles; recíprocamente, los conjuntos con todas sus secciones en $\mathcal{C}$ forman una σ-álgebra que contiene a los rectángulos. En ese caso, por la aditividad de ν,

$$\nu(B) = \sum_{e \in E} \nu\big(B_e \times \{e\} \big) = \sum_{e \in E} Q(\{e\}) \, \mu(B_e). \tag{3.2}$$

(a) Por (3.2), $0 = \nu(D) = \sum_e Q(\{e\}) \mu(D_e)$; como todos los pesos $Q(\{e\})$ son positivos, $\mu(D_e) = 0$ para todo e. (Éste es el único lugar de la prueba donde se usa esa hipótesis.) La inclusión $D \subset N \times E$ es inmediata, y N es una unión finita de conjuntos nulos de $\mathcal{C}$.

(b) Si $B \in \overline{\mathcal{C} \otimes 2^E}$, existen $C, D \in \mathcal{C} \otimes 2^E$ con $\nu(D) = 0$ y $B \triangle C \subset D$. Tomando secciones, $B_e \triangle C_e \subset D_e$, donde $C_e, D_e \in \mathcal{C}$ y, por (a), $\mu(D_e) = 0$; luego $B_e \in \overline{\mathcal{C}}$. Recíprocamente, supongamos que, para cada e, existen $C_e, N_e \in \mathcal{C}$ con $\mu(N_e) = 0$ y $B_e \triangle C_e \subset N_e$. Los conjuntos $C := \bigcup_e (C_e \times \{e\})$ y $D := \bigcup_e (N_e \times \{e\})$ pertenecen a $\mathcal{C} \otimes 2^E$, se tiene $\nu(D) = 0$ por (3.2) y $B \triangle C \subset D$, porque la inclusión vale sección a sección. Luego $B \in \overline{\mathcal{C} \otimes 2^E}$.

(c) Para $t \in \mathbb{R}$, la sección en e de $\{H > t\}$ es $\{H(\cdot, e) > t\}$. Por (b), $\{H > t\} \in \overline{\mathcal{C} \otimes 2^E}$ para todo t si y sólo si $\{H(\cdot, e) > t\} \in \overline{\mathcal{C}}$ para todo t y todo e. Como los conjuntos $(t, \infty]$, $t \in \mathbb{R}$, generan $\mathcal{B}([0, \infty])$, esto prueba la equivalencia.

Para la fórmula, el Lema 3.12(i) da, para cada e, una versión $\mathcal{C}$-medible $\widetilde{h}_e$ de $H(\cdot, e)$ que coincide con ella fuera de un conjunto nulo $N_e \in \mathcal{C}$. Sean $N := \bigcup_e N_e$ y $\widetilde{H}(\omega, e) := \widetilde{h}_e(\omega)$. Las secciones de

$\{\widetilde{H} > t\}$ son los conjuntos $\{\widetilde{h}_e > t\} \in \mathcal{C}$, así que $\widetilde{H}$ es $\mathcal{C} \otimes 2^E$-medible; además, $H = \widetilde{H}$ fuera de $N \times E$, y $\nu(N \times E) = \mu(N) = 0$. Por lo tanto, usando el Lema 3.12(ii) en $\Omega \times E$, el Lema 3.10(c) con $\Omega_1 = \Omega$ y $\Omega_2 = E$, y otra vez el Lema 3.12(ii) en Ω,

$$\int H \, d\overline{\nu} = \int \widetilde{H} \, d\nu = \int_\Omega \Big(\sum_{e \in E} Q(\{e\}) \, \widetilde{h}_e \Big) \, d\mu = \int_\Omega \Big(\sum_{e \in E} Q(\{e\}) \, H(\cdot, e) \Big) \, d\overline{\mu}.$$

En la segunda igualdad se usó que la integral de $\widetilde{H}(\omega, \cdot)$ respecto de Q es $\sum_e Q(\{e\}) \widetilde{h}_e(\omega)$; en la tercera, que $\sum_e Q(\{e\}) \widetilde{h}_e$ es una versión medible de $\sum_e Q(\{e\}) H(\cdot, e)$. □

Bibliografía

[1] P. Billingsley. *Convergence of Probability Measures. 2nd ed.* John Wiley & Sons., 1999.

[2] Rick Durrett. *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, United Kingdom, 5 edition, 2019.

[3] G. Folland. *Real Analysis: Modern techniques and their applications.* Wiley, 1999.

[4] A.N. Shiryaev. *Probability.* Springer-Verlag, 1984.

[5] A. W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer, Cham, Switzerland, 2 edition, 2023.