

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	6
1.2.1. Integral exterior	6
1.2.2. Fubini exterior	13
1.3. Convergencia débil	16
1.3.1. Redes	18
1.4. Espacios producto	27
1.4.1. Topología y σ -álgebras en productos finitos	27
1.4.2. Tensión y medibilidad en productos finitos	28
1.4.3. Productos numerables	28
1.5. Espacios de funciones acotadas	29
1.6. σ -álgebra generada por las bolas	39
1.7. Convergencias	41
1.8. Uniformidad y metrización	49
2. Proceso empírico	52
2.1. Medidas empíricas y clases de Donsker	52
2.2. Desigualdades maximales y números de recubrimiento	54
2.2.1. Desigualdades subgaussianas	59
2.3. Simetrización y medibilidad	61
2.3.1. Simetrización	61
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	68
2.5.1. Entropía uniforme	68
2.5.2. Acotamientos (<i>bracketing</i>)	72
2.6. Cotas para la entropía y clases VC	73
2.6.1. Clases VC de conjuntos	74
2.6.2. Clases VC de funciones	76
2.7. Números de bracketing	81
2.7.1. Funciones y conjuntos suaves	81
3. Aplicaciones estadísticas	84
3.1. M-estimadores	84
3.1.1. Tasa de convergencia	87

Estas notas están muy fuertemente basadas en [4]. Tienen como requisitos probabilidad y medida. Una referencia clásica para los conceptos de probabilidad es [3]. Para teoría de la medida, véase [2]; 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

Sea $X_1, \dots, X_n$ una muestra i.i.d. con ley común P sobre un espacio medible $(\mathcal{X}, \mathcal{A})$. Para una función medible $f : \mathcal{X} \rightarrow \mathbb{R}$, escribimos

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

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,

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

Estos resultados son puntuales en f . Sin embargo, en muchas aplicaciones estadísticas no se estudia un único promedio $P_n f$, sino toda una familia de promedios $(P_n f)_{f \in \mathcal{F}}$. Por eso aparecen naturalmente las cantidades

$$\|P_n - P\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |P_n f - Pf|, \quad \mathbb{G}_n f := \sqrt{n}(P_n - P)f, \quad f \in \mathcal{F}.$$

El objeto $\mathbb{G}_n$ se llama proceso empírico indexado por $\mathcal{F}$.

El caso clásico: la función de distribución empírica

Si $\mathcal{X} = \mathbb{R}$ y F es la función de distribución de X_1 , la función de distribución empírica es

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

Por lo tanto, estudiar $F_n - F$ equivale a estudiar $P_n - P$ sobre la clase $\mathcal{F}_0 = \{\mathbb{1}_{(-\infty, t]} : t \in \mathbb{R}\}$. Para cada t fijo, la ley fuerte da $F_n(t) \rightarrow F(t)$ c.s. El teorema de Glivenko–Cantelli refuerza esta afirmación puntual y establece

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

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

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

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

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

Para cada t fijo, el teorema central del límite da $\mathbb{G}_n(t) \Rightarrow N(0, F(t)(1 - F(t)))$. 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 $\Rightarrow$ indica convergencia en distribución (que definiremos más adelante), $\mathbb{U}$ es un puente browniano estándar, es decir, un proceso gaussiano centrado con covarianza $\text{Cov}(\mathbb{U}(s), \mathbb{U}(t)) = s \wedge t - st$.

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

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

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.

Del caso clásico al proceso empírico general

La formulación anterior se extiende reemplazando la clase de semirrectas $\mathcal{F}_0$ por una clase general $\mathcal{F}$ de funciones $\mathcal{A}$ -medibles tales que $P|f| < \infty$, $f \in \mathcal{F}$, de modo que Pf esté definido como número real. El proceso empírico general es

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

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

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

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. La primera es una ley uniforme de grandes números: $\|P_n - P\|_{\mathcal{F}} \rightarrow 0$. 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

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

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 $f, g \in \mathcal{F}$.

Estas definiciones no son meramente abstractas. Si

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

la consistencia de $\hat{\theta}_n$ suele requerir

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

Es decir, se necesita que la clase $\{m_\theta : \theta \in \Theta\}$ sea Glivenko–Cantelli. Para estudiar la distribución límite de $\hat{\theta}_n$, en cambio, aparecen procesos empíricos locales. Esta es una de las razones por las que los procesos empíricos son centrales en M -estimación, estimación semiparamétrica, bootstrap y método delta funcional.

Por qué la convergencia puntual no alcanza

Es importante distinguir entre convergencia puntual y convergencia uniforme. Sea P una distribución continua en $\mathbb{R}$, y sea $\mathcal{C}$ la clase de todos los subconjuntos finitos de $\mathbb{R}$. Para cada $C \in \mathcal{C}$ fijo, $PC = 0$, y por lo tanto $P_n C \rightarrow 0$ c.s. Sin embargo,

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

En efecto, si $C_n = \{X_1, \dots, X_n\}$, entonces $C_n \in \mathcal{C}$, $P_n C_n = 1$ 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 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 $\mathcal{D}[0, 1] = \{z : [0, 1] \rightarrow \mathbb{R} : z \text{ es càdlàg}\}$. Aquí, càdlàg significa continua por la derecha y con límites por la izquierda. 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 $\lambda : [0, 1] \rightarrow [0, 1]$ estrictamente crecientes y continuas, con $\lambda(0) = 0$ y $\lambda(1) = 1$. Una forma de la métrica J_1 es

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

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, $\mathcal{D}[0, 1]$ tiene mejores propiedades de separabilidad 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 no suelen estar indexados por $t \in [0, 1]$, sino por una clase arbitraria de funciones $\mathcal{F}$. El espacio natural pasa a ser

$$\ell^\infty(\mathcal{F}) = \{z : \mathcal{F} \rightarrow \mathbb{R} : z \text{ acotada}\}, \quad \|z\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |z(f)|.$$

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 el espacio de probabilidad $([0, 1], \mathcal{B}([0, 1]), \lambda)$, sea $X(x) = x$ y definamos $Z_X(t) = \mathbb{1}_{[X, 1]}(t)$, $t \in [0, 1]$. Para cada t fijo,

$Z_X(t) = \mathbb{1}_{\{X \leq t\}}$ es una variable aleatoria medible. Sin embargo, la aplicación

$$x \mapsto Z_x := \mathbb{1}_{[x,1]} \quad \text{con valores en } (\ell^\infty[0,1], \|\cdot\|_\infty)$$

no es Borel-medible.

En efecto, si $x \neq y$, entonces $\|\mathbb{1}_{[x,1]} - \mathbb{1}_{[y,1]}\|_\infty = 1$. Sea $A \subset [0,1]$ un conjunto no boreliano¹ y pongamos

$$U_A := \bigcup_{x \in A} B_{\ell^\infty}(Z_x, 1/2).$$

El conjunto U_A es abierto en $\ell^\infty[0,1]$, y la separación anterior implica que $\{x \in [0,1] : Z_x \in U_A\} = A$.² Por consiguiente, la preimagen de un abierto puede no ser boreliana, y la aplicación $x \mapsto Z_x$ no es Borel-medible.

Este ejemplo es importante porque muestra que no basta con que todas las coordenadas $Z_X(t)$ sean medibles. La medibilidad coordenada no implica la medibilidad del proceso como variable aleatoria con valores en $\ell^\infty(T)$. En procesos empíricos generales ocurre exactamente este fenómeno: para cada $f \in \mathcal{F}$, $\mathbb{G}_n f$ es una variable aleatoria real, pero la aplicación $\omega \mapsto \mathbb{G}_n(\omega) \in \ell^\infty(\mathcal{F})$ 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 $\|P_n - P\|_{\mathcal{F}} \rightarrow 0$ 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

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

$$\mathbb{E}^*(T) = \inf \left\{ \mathbb{E}U : U : \Omega \rightarrow \overline{\mathbb{R}} \text{ medible}, U(\omega) \geq T(\omega) \quad \forall \omega, \mathbb{E}U \text{ existe} \right\},$$

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

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

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

¹un conjunto de Vitali no es medible Lebesgue y, por tanto, no es boreliano

²La inclusión $\supset$ vale porque cada Z_x , $x \in A$, es el centro de una de las bolas que forman U_A . Para $\subset$: si $Z_x \in B_{\ell^\infty}(Z_{x'}, 1/2)$ con $x' \in A$, entonces $\|Z_x - Z_{x'}\|_\infty < 1/2$, y la separación $\|Z_x - Z_{x'}\|_\infty = 1$ para $x \neq x'$ fuerza $x = x' \in A$.

Advertencia 1.2

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 $\mathcal{A}$).

Observación 1.3

Es fácil ver que^a

$$\mathbb{E}_*(T) = \sup \left\{ \mathbb{E}U : U : \Omega \rightarrow \overline{\mathbb{R}} \text{ medible}, U(\omega) \leq T(\omega) \forall \omega, \mathbb{E}U \text{ existe} \right\},$$

y que $\mathbb{P}_*(B) = \sup \{ \mathbb{P}(A) : A \subset B, A \in \mathcal{A} \}$.

^aPara la primera identidad: U es un minorante medible de T con esperanza existente si y sólo si $-U$ es un mayorante medible de $-T$ con esperanza existente, y $\mathbb{E}(-U) = -\mathbb{E}U$; así, el supremo sobre minorantes es $-\inf$ sobre los mayorantes de $-T$. Para la segunda: $A \in \mathcal{A}$ cumple $A \subset B$ si y sólo si $A^c \supset B^c$, y $\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$, de modo que el supremo se corresponde con $1 - \mathbb{P}^*(B^c)$.

El siguiente lema prueba la existencia de un mayorante medible mínimo de una función arbitraria $T : \Omega \rightarrow \overline{\mathbb{R}}$. 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 . En particular, este mayorante es único $\mathbb{P}$ -c.s. 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 convergencia dominada.

Lema 1.4. Para todo $T : \Omega \rightarrow \overline{\mathbb{R}}$ existe $T^* : \Omega \rightarrow \overline{\mathbb{R}}$ medible tal que:

- (i) $T^* \geq T$ punto a punto;
- (ii) si U es medible y $U \geq T$ $\mathbb{P}$ -c.s.³, 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 $\psi : [-\pi/2, \pi/2] \rightarrow \overline{\mathbb{R}}$ 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 $\tilde{V}_m : \Omega \rightarrow [-\pi/2, \pi/2]$ tal que $\tilde{V}_m \geq \phi(T)$ y $\mathbb{E}\tilde{V}_m \leq a + 1/m$.⁴ Definimos $V_m := \min_{1 \leq k \leq m} \tilde{V}_k$. Entonces (V_m) es decreciente, cada V_m es un mayorante medible de $\phi(T)$, y $a \leq \mathbb{E}V_m \leq \mathbb{E}\tilde{V}_m \leq a + 1/m$. Por lo tanto, V_m converge puntualmente a una función medible $V : \Omega \rightarrow [-\pi/2, \pi/2]$, que satisface $V \geq \phi(T)$.⁵ Como la sucesión

³Cuando escribimos $A \leq B$ c.s. o $A = B$ c.s., queremos decir que la relación vale fuera de un conjunto nulo medible. Es decir, existe $N \in \mathcal{A}$, con $\mathbb{P}(N) = 0$, tal que $A(\omega) \leq B(\omega)$, respectivamente $A(\omega) = B(\omega)$, para todo $\omega \notin N$. En particular, si alguna de las funciones involucradas no es medible, la afirmación se entiende en este sentido: la posible falla de la desigualdad o igualdad queda contenida en un conjunto medible de probabilidad cero.

⁴Nótese 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 $\mathbb{E}\tilde{V}_m \leq a + 1/m$: es la caracterización del ínfimo finito.

⁵Una sucesión decreciente y acotada inferiormente (aquí por $-\pi/2$) converge en todo punto, y el límite puntual de funciones medibles es medible.

es uniformemente acotada, el teorema de convergencia acotada da

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

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),

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

Todas las desigualdades son, por tanto, igualdades.⁶ 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 $\mathbb{E}T^* \leq \mathbb{E}U$. 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). $\square$

Definición 1.5. T^* se llama el **mayorante medible mínimo** de T . Por el punto ii del Lema anterior, es único salvo en conjuntos de probabilidad $\mathbb{P}$ nula).

Advertencia 1.6

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

Definición 1.7. Sean $S, T : \Omega \rightarrow \mathbb{R}$ funciones arbitrarias, no necesariamente medibles, y sean S^*, T^* sus mayorantes medibles mínimos. Escribimos $T_* := -(-T)^*$. Todas las igualdades siguientes se entienden $\mathbb{P}$ -c.s.; además, cuando aparezcan operaciones con valores extendidos, se supone que las expresiones involucradas están bien definidas.

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

- (i) $\mathbb{P}^*(B) = \mathbb{E}^*(\mathbf{1}_B)$ y $\mathbb{P}_*(B) = \mathbb{E}((\mathbf{1}_B)_*)$.
- (ii) Existe $B^* \in \mathcal{A}$ tal que $B \subset B^*$ y $\mathbb{P}(B^*) = \mathbb{P}^*(B)$. Más aún, si $\tilde{B} \in \mathcal{A}$, $B \subset \tilde{B}$ y $\mathbb{P}(\tilde{B}) = \mathbb{P}^*(B)$, entonces $\mathbf{1}_{\tilde{B}} = (\mathbf{1}_B)^*$ $\mathbb{P}$ -c.s. En particular, $(\mathbf{1}_{\tilde{B}})^* = (\mathbf{1}_B)^*$ $\mathbb{P}$ -c.s.
- (iii) $(\mathbf{1}_B)^* + (\mathbf{1}_{\Omega \setminus B})_* = 1$ $\mathbb{P}$ -c.s.

⁶La 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 $M := (\mathbb{1}_B)^*$. Probamos primero simultáneamente la primera identidad de (i) y la existencia afirmada en (ii).

Si $A \in \mathcal{A}$ y $B \subset A$, entonces $\mathbb{1}_A \geq \mathbb{1}_B$, de modo que $\mathbb{E}^*(\mathbb{1}_B) \leq \mathbb{P}(A)$. Tomando ínfimo sobre tales A , obtenemos

$$\mathbb{E}^*(\mathbb{1}_B) \leq \mathbb{P}^*(B). \quad (1)$$

Por otra parte, $M \geq \mathbb{1}_B \geq 0$ punto a punto y, como la constante 1 es un mayorante medible de $\mathbb{1}_B$, la minimalidad de M implica $M \leq 1$, $\mathbb{P}$ -c.s. Por tanto, $\mathbb{E}M$ existe y, por el Lema 1.4, $\mathbb{E}M = \mathbb{E}^*(\mathbb{1}_B)$.

Definamos $B^* := \{M \geq 1\}$. Entonces $B^* \in \mathcal{A}$ y $B \subset B^*$.⁷ Además, $M \geq \mathbb{1}_{B^*}$ punto a punto, por lo que

$$\mathbb{P}^*(B) \leq \mathbb{P}(B^*) = \mathbb{E}\mathbb{1}_{B^*} \leq \mathbb{E}M = \mathbb{E}^*(\mathbb{1}_B). \quad (2)$$

De (1) y (2) se sigue que $\mathbb{P}^*(B) = \mathbb{E}^*(\mathbb{1}_B)$ y $\mathbb{P}(B^*) = \mathbb{P}^*(B)$. Asimismo, $M \geq \mathbb{1}_{B^*}$ y ambas variables tienen la misma esperanza,⁸ de modo que

$$(\mathbb{1}_B)^* = M = \mathbb{1}_{B^*} \quad \mathbb{P}\text{-c.s.} \quad (3)$$

Sea ahora $\tilde{B} \in \mathcal{A}$ tal que $B \subset \tilde{B}$ y $\mathbb{P}(\tilde{B}) = \mathbb{P}^*(B)$. Como $\mathbb{1}_{\tilde{B}}$ es un mayorante medible de $\mathbb{1}_B$, la minimalidad de M da $M \leq \mathbb{1}_{\tilde{B}}$, $\mathbb{P}$ -c.s. Además, $\mathbb{E}M = \mathbb{E}^*(\mathbb{1}_B) = \mathbb{P}^*(B) = \mathbb{P}(\tilde{B}) = \mathbb{E}\mathbb{1}_{\tilde{B}}$. Por consiguiente, $\mathbb{1}_{\tilde{B}} = (\mathbb{1}_B)^*$ $\mathbb{P}$ -c.s. Como $\mathbb{1}_{\tilde{B}}$ es medible, su mayorante medible mínimo coincide con ella, $\mathbb{P}$ -c.s.;⁹ por tanto, $(\mathbb{1}_{\tilde{B}})^* = (\mathbb{1}_B)^*$ $\mathbb{P}$ -c.s. Esto prueba (ii).

Probemos ahora (iii). De $(\mathbb{1}_B)^* \geq \mathbb{1}_B$ se deduce $1 - (\mathbb{1}_B)^* \leq 1 - \mathbb{1}_B = \mathbb{1}_{B^c}$. Así, $1 - (\mathbb{1}_B)^*$ es un minorante medible de $\mathbb{1}_{B^c}$, y por la maximalidad del minorante medible,

$$1 - (\mathbb{1}_B)^* \leq (\mathbb{1}_{B^c})_* \quad \mathbb{P}\text{-c.s.} \quad (4)$$

Recíprocamente, si h es medible y $h \leq \mathbb{1}_{B^c}$, $\mathbb{P}$ -c.s., entonces $1 - h \geq \mathbb{1}_B$, $\mathbb{P}$ -c.s. Por la minimalidad de $(\mathbb{1}_B)^*$, $(\mathbb{1}_B)^* \leq 1 - h$, $\mathbb{P}$ -c.s., es decir, $h \leq 1 - (\mathbb{1}_B)^*$, $\mathbb{P}$ -c.s. Aplicando esto a $h = (\mathbb{1}_{B^c})_*$ y usando (4), obtenemos $(\mathbb{1}_B)^* + (\mathbb{1}_{B^c})_* = 1$ $\mathbb{P}$ -c.s. Esto prueba (iii).

Finalmente, aplicando (iii) a B^c , $(\mathbb{1}_{B^c})^* + (\mathbb{1}_B)_* = 1$ $\mathbb{P}$ -c.s. Por la definición de probabilidad interior, la primera identidad de (i) y el Lema 1.4,

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

Queda probado (i). □

En el lema que sigue cada desigualdad e igualdad es $\mathbb{P}$ -c.s. y bajo la hipótesis de que las expresiones que aparecen en el ítem correspondiente están bien definidas punto a punto (sin formas $\infty - \infty$ ni productos $0 \cdot \infty$ no convenidos),¹⁰

Lema 1.9. Sean $S, T : \Omega \rightarrow \overline{\mathbb{R}}$ 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, $(S + T)^* = S^* + T$.

⁷ $B \subset B^*$ porque $M \geq \mathbb{1}_B$ punto a punto: si $\omega \in B$, entonces $M(\omega) \geq 1$. Además $M \geq \mathbb{1}_B \geq 0$ en todo punto; por tanto, en B^* se tiene $M \geq 1 = \mathbb{1}_{B^*}$ y, fuera de B^* , $M \geq 0 = \mathbb{1}_{B^*}$.

⁸ Se usa que una variable no negativa con esperanza nula es nula c.s., aplicada a $M - \mathbb{1}_{B^*} \geq 0$; ambas esperanzas coinciden por la cadena (1)–(2).

⁹ Si h es medible, h es un mayorante medible de sí misma y la parte (ii) del Lema 1.4 da $h^* \leq h$ c.s.; junto con $h^* \geq h$ punto a punto, resulta $h^* = h$ c.s. Además, dos funciones iguales c.s. tienen mayorantes medibles mínimos iguales c.s.

¹⁰ Cuando S y T son finitas, $S^* \geq S > -\infty$ y $T^* \geq T > -\infty$ punto a punto, de modo que las sumas de envolventes que aparecen en (i) y (ii) están automáticamente definidas. En (iii) todavía puede ocurrir $S^* = T^* = +\infty$; por eso la desigualdad se afirma únicamente en el conjunto donde $S^* - T^*$ está definida. El caso de S y T finitas es el que se usa en el resto del capítulo.

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

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

(iv) Si S es medible y finita, entonces, definiendo $ST = 0$ en $\{S = 0\}$,

$$(ST)^* = ST^* \mathbf{1}_{\{S > 0\}} + ST_* \mathbf{1}_{\{S < 0\}}$$

(v) Sin hipótesis de medibilidad sobre S ni T ,

$$(ST)^* \leq S^* T^* \mathbf{1}_{\{S^* > 0, T^* > 0\}} + S^* T_* \mathbf{1}_{\{S^* < 0, T_* > 0\}} + S_* T^* \mathbf{1}_{\{S_* > 0, T^* < 0\}} + S_* T_* \mathbf{1}_{\{S_* < 0, T_* < 0\}}.$$

donde cada sumando se entiende igual a 0 fuera del conjunto que lo acompaña.

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

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

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

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

Demostración. Todas las relaciones se entienden $\mathbb{P}$ -c.s. y en los conjuntos donde las operaciones involucradas están bien definidas. Recordemos que $S_* \leq S \leq S^*$ y $T_* \leq T \leq T^*$ punto a punto.

Usaremos repetidamente el siguiente *principio de localización*: si $A \in \mathcal{A}$, $v : A \rightarrow \overline{\mathbb{R}}$ 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 de la envolvente 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 una de las envolventes vale $+\infty$ y la otra $-\infty$, y cubren todo el conjunto donde $S^* - T^*$ está definida.

(iv). Supongamos que S es medible y finita, y definamos $R := ST^* \mathbf{1}_{\{S > 0\}} + ST_* \mathbf{1}_{\{S < 0\}}$. 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). Denotemos por R el miembro derecho de (v). Los cuatro conjuntos que aparecen en su definición son medibles. Además, sus soportes son disjuntos salvo, posiblemente, los del primero y el cuarto. Por ejemplo, los soportes del segundo y del tercero no pueden intersectarse, pues ello exigiría $S^* < 0 < S_*$, en contradicción con $S_* \leq S^*$; las demás incompatibilidades se comprueban del mismo modo usando también $T_* \leq T^*$. En la posible intersección de los soportes primero y cuarto, ambos sumandos son no negativos. Por otra parte, los factores tienen signo estricto en sus respectivos soportes, por lo que no aparece ningún producto de la forma $0 \cdot \infty$. En consecuencia, R está bien definida y es medible.

Probaremos que $ST \leq R$ punto a punto. Como $S_* \leq S \leq S^*$ y $T_* \leq T \leq T^*$, distinguimos los posibles signos de S y T .

Si $S, T > 0$, entonces $S^*, T^* > 0$, por lo que el primer sumando está presente. Además, $ST \leq S^* T^* \leq R$. Si $S, T < 0$, entonces $S_*, T_* < 0$, por lo que el cuarto sumando está presente. En este caso, $S_* T_* \geq ST$, y por tanto $ST \leq S_* T_* \leq R$.

Supongamos que $S > 0 > T$. El segundo sumando no puede aparecer porque $S^* \geq S > 0$, y el único sumando negativo que podría aparecer es el tercero. Si no aparece, todos los sumandos presentes son no negativos, de modo que $R \geq 0 > ST$. Si aparece, entonces $S_* > 0$ y $T^* < 0$. Estas condiciones hacen que los otros tres sumandos se anulen. Como $S_* \leq S$ y $T < 0$, se tiene $ST \leq S_* T$; como $T \leq T^*$ y $S_* > 0$, resulta $S_* T \leq S_* T^*$. Por tanto, $ST \leq S_* T^* = R$.

El caso $S < 0 < T$ es simétrico, pero lo detallamos para evitar ambigüedades. El tercer sumando no puede aparecer porque $T^* \geq T > 0$, y el único sumando negativo que podría aparecer es el segundo. Si no aparece, $R \geq 0 > ST$. Si aparece, entonces $S^* < 0$ y $T_* > 0$, y los otros tres sumandos se

anulan. Como $S \leq S^*$ y $T > 0$, se tiene $ST \leq S^*T$; como $T_* \leq T$ y $S^* < 0$, resulta $S^*T \leq S^*T_*$. Por tanto, $ST \leq S^*T_* = R$.

Finalmente, si $ST = 0$, los sumandos segundo y tercero no pueden aparecer. En efecto, si $S = 0$, no pueden cumplirse $S^* < 0$ ni $S_* > 0$, mientras que, si $T = 0$, no pueden cumplirse $T_* > 0$ ni $T^* < 0$. Así, $R \geq 0 = ST$.

Hemos probado que $ST \leq R$ punto a punto en todos los casos. Como R es medible, la minimalidad de $(ST)^*$ da $(ST)^* \leq R$.

(vi). Fijemos $c \in \mathbb{R}$ y sean $C := \{T > c\}$ y $C_c^* := \{T^* > c\}$. Como $T \leq T^*$, se tiene $C \subset C_c^*$. Además, $C_c^* \in \mathcal{A}$, pues T^* es medible.

Sea $A \in \mathcal{A}$ cualquier recubrimiento medible de C . En A^c se tiene $T \leq c$ punto a punto. Aplicando el principio de localización a T , con la función constante $v \equiv c$, obtenemos $T^* \leq c$, $\mathbb{P}$ -c.s. en A^c . Por tanto, $C_c^* \cap A^c$ es nulo, es decir, $C_c^* \subset A$ salvo un conjunto nulo. En consecuencia, $\mathbb{P}(C_c^*) \leq \mathbb{P}(A)$.

Como esta desigualdad vale para todo recubrimiento medible A de C , se tiene $\mathbb{P}(C_c^*) \leq \mathbb{P}^*(C)$. La desigualdad opuesta se sigue de que C_c^* es un recubrimiento medible de C . Por tanto, $\mathbb{P}(C_c^*) = \mathbb{P}^*(C)$, de modo que C_c^* es un recubrimiento medible óptimo de C .

Por el Lema 1.8(ii), aplicado a este recubrimiento óptimo, $(1_C)^* = 1_{C_c^*}$, es decir, $(1_{\{T > c\}})^* = 1_{\{T^* > c\}}$. Finalmente, por el Lema 1.8(i), $\mathbb{P}^*(T > c) = \mathbb{P}(T^* > c)$.

(viii). Probamos (viii) antes de (vii), pues lo usaremos en su demostración. 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, $|T^*| \leq |T|^*$.

(ix). Como $S \wedge T \leq S$ y $S \wedge T \leq T$, la monotonía de la envolvente 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. $\square$

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

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

Demostración. Como $T^* \geq T \geq 0$ punto a punto, $\mathbb{E}T^*$ existe y, por el Lema 1.4(iii), vale $\mathbb{E}^*T$. Para $0 < \varepsilon < c$, el Lema 1.9(vi) y la desigualdad de Markov clásica dan $\mathbb{P}^*(T > c - \varepsilon) = \mathbb{P}(T^* > c - \varepsilon)$; como $\{T \geq c\} \subset \{T > c - \varepsilon\}$, la monotonía de $\mathbb{P}^*$ y $\varepsilon \downarrow 0$ dan $\mathbb{P}^*(T \geq c) \leq \mathbb{P}(T^* \geq c) \leq \mathbb{E}T^*/c = \mathbb{E}^*T/c$.¹¹ $\square$

¹¹El paso al límite usa la continuidad desde arriba de $\mathbb{P} : \{T^* > c - 1/k\} \downarrow \{T^* \geq c\}$ cuando $k \rightarrow \infty$. La última desigualdad es Markov clásica para la variable no negativa T^* .

Advertencia 1.11

- La probabilidad exterior puede no ser aditiva. En $([0, 1], \mathcal{L}, \lambda)$, sea $V \subset [0, 1]$ tal que $\lambda_*(V) = 0$ y $\lambda^*(V) = 1$. Si $A = V$ y $B = [0, 1] \setminus V$, entonces $A \cap B = \emptyset$, $A \cup B = [0, 1]$, pero

$$\mathbb{P}^*(A) + \mathbb{P}^*(B) = 2 > 1 = \mathbb{P}^*(A \cup B).$$

Así, la aditividad puede fallar aun para conjuntos disjuntos.^a

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

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

siempre que los lados derechos estén bien definidos. La igualdad no tiene por qué valer si las variables no son medibles. 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).

^aTal V existe (usando el axioma de elección): por ejemplo, un conjunto de Bernstein, que interseca a todo cerrado no numerable de $[0, 1]$ sin contener ninguno; todo medible contenido en V o en V^c resulta de medida nula. Y $\mathbb{P}^*(B) = 1$ sale de la dualidad: $\lambda^*([0, 1] \setminus V) = 1 - \lambda_*(V) = 1$.

Observación 1.12

El mayorante medible mínimo depende de la σ -álgebra. Por ejemplo, si $\Omega = \{0, 1\}$, $T = \mathbb{1}_{\{0\}}$, entonces respecto de la σ -álgebra discreta 2^Ω (todas las partes de Ω) se tiene $T^* = T$, mientras que respecto de la σ -álgebra trivial $\{\emptyset, \Omega\}$ se tiene $T^* \equiv 1$.

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. Sea $T : \Omega_1 \times \Omega_2 \rightarrow \overline{\mathbb{R}}$ arbitrario, definido sobre $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$. Escribimos $\mathbb{E}$ para la esperanza respecto de $\mathbb{P}_1 \otimes \mathbb{P}_2$,

Definición 1.13. Para $\omega_1 \in \Omega_1$ escribimos $T_{\omega_1}(\omega_2) = T(\omega_1, \omega_2)$. Definimos

$$\mathbb{E}_2^*T(\omega_1) := \inf \left\{ \int_{\Omega_2} U d\mathbb{P}_2 : U : \Omega_2 \rightarrow \overline{\mathbb{R}} \text{ medible, } U \geq T_{\omega_1}(\cdot), \int_{\Omega_2} U d\mathbb{P}_2 \text{ existe} \right\}.$$

Así, $\mathbb{E}_2^*T : \Omega_1 \rightarrow \overline{\mathbb{R}}$ es una función no necesariamente medible. Análogamente, para $\omega_2 \in \Omega_2$, escribimos $T_{\omega_2}(\omega_1) = T(\omega_1, \omega_2)$ y definimos

$$\mathbb{E}_1^*T(\omega_2) := \inf \left\{ \int_{\Omega_1} U d\mathbb{P}_1 : U : \Omega_1 \rightarrow \overline{\mathbb{R}} \text{ medible, } U \geq T_{\omega_2}(\cdot), \int_{\Omega_1} U d\mathbb{P}_1 \text{ existe} \right\}.$$

Observación 1.14

- Para cada $\omega_1 \in \Omega_1$, $\mathbb{E}_2^*T(\omega_1) = \mathbb{E}_{\mathbb{P}_2}^*(T_{\omega_1})$, y, para cada $\omega_2 \in \Omega_2$, $\mathbb{E}_1^*T(\omega_2) = \mathbb{E}_{\mathbb{P}_1}^*(T_{\omega_2})$.
- Con la definición anterior,

$$\mathbb{E}_1^*\mathbb{E}_2^*T := \inf \left\{ \int_{\Omega_1} V d\mathbb{P}_1 : V : \Omega_1 \rightarrow \overline{\mathbb{R}} \text{ medible, } V \geq \mathbb{E}_2^*T, \int_{\Omega_1} V d\mathbb{P}_1 \text{ existe} \right\}.$$

Lema 1.15. Sea $T : \Omega_1 \times \Omega_2 \rightarrow \overline{\mathbb{R}}$ arbitrario, definido sobre $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$. Entonces,

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

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

Demostración. Probemos primero que $\mathbb{E}_1^* \mathbb{E}_2^* T \leq \mathbb{E}^* T$. Si $\mathbb{E}^* T = +\infty$, la afirmación es inmediata. Supongamos entonces que $\mathbb{E}^* T < +\infty$, y sea $U \geq T$ un mayorante medible cuya esperanza exista y satisfaga¹² $\mathbb{E}U < +\infty$. En particular, $\mathbb{E}U^+ < \infty$.

Para $\omega_1 \in \Omega_1$, definamos $a(\omega_1) := \mathbb{E}_2 U^+(\omega_1, \cdot)$ y $b(\omega_1) := \mathbb{E}_2 U^-(\omega_1, \cdot)$. Por Tonelli, a y b son medibles y $\mathbb{E}_1 a = \mathbb{E}U^+ < \infty$. Por tanto, $N := \{a = +\infty\}$ es medible y $\mathbb{P}_1(N) = 0$. Pongamos

$$\phi(\omega_1) := \begin{cases} a(\omega_1) - b(\omega_1), & \omega_1 \notin N, \\ +\infty, & \omega_1 \in N. \end{cases}$$

La función ϕ es medible.¹³ Si $\omega_1 \notin N$, la sección $U(\omega_1, \cdot)$ es un mayorante medible de $T(\omega_1, \cdot)$ cuya esperanza existe, pues su parte positiva es integrable. En consecuencia, $\mathbb{E}_2^* T(\omega_1) \leq \phi(\omega_1)$; en N , la misma desigualdad es trivial. Así, ϕ es un mayorante medible de $\mathbb{E}_2^* T$. Además, $\phi^+ \leq a$, por lo que $\mathbb{E}_1 \phi^+ < \infty$ y, en consecuencia, $\mathbb{E}_1 \phi$ está bien definida. Como $\phi = a - b$, $\mathbb{P}_1$ -c.s., y $\mathbb{E}_1 a < \infty$, Tonelli da¹⁴

$$\mathbb{E}_1 \phi = \mathbb{E}_1 a - \mathbb{E}_1 b = \mathbb{E}U^+ - \mathbb{E}U^- = \mathbb{E}U.$$

Ahora bien, ϕ es un mayorante medible de $\mathbb{E}_2^* T$. Por la definición de esperanza exterior,

$$\mathbb{E}_1^* \mathbb{E}_2^* T \leq \mathbb{E}_1 \phi = \mathbb{E}U.$$

Tomando ínfimo sobre todos los mayorantes medibles $U \geq T$ cuya esperanza existe y es menor que $+\infty$, restricción que no altera el ínfimo cuando $\mathbb{E}^* T < +\infty$, obtenemos $\mathbb{E}_1^* \mathbb{E}_2^* T \leq \mathbb{E}^* T$.

Para la desigualdad central, usamos $\mathbb{E}_{2*} T \leq \mathbb{E}_2^* T$, la monotonía de $\mathbb{E}_1^*$ y la desigualdad general¹⁵ $\mathbb{E}_{1*} X \leq \mathbb{E}_1^* X$. Así, $\mathbb{E}_{1*} \mathbb{E}_{2*} T \leq \mathbb{E}_1^* \mathbb{E}_{2*} T \leq \mathbb{E}_1^* \mathbb{E}_2^* T$.

Finalmente, aplicando la desigualdad ya probada a $-T$, $\mathbb{E}_1^* \mathbb{E}_2^*(-T) \leq \mathbb{E}^*(-T)$. Al multiplicar por -1 y usar las definiciones de esperanza interior,¹⁶ obtenemos $\mathbb{E}_* T \leq -\mathbb{E}_1^*(-\mathbb{E}_{2*} T) = \mathbb{E}_{1*} \mathbb{E}_{2*} T$. En consecuencia,

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

Intercambiando los papeles de 1 y 2 se obtiene la segunda cadena. □

Advertencia 1.16

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.¹⁷ Este fenómeno se formula mejor mediante una noción general, que reaparecerá en el teorema de representación casi segura y, sobre todo, en la simetrización del Capítulo 2.

¹²Tal U existe por la definición del ínfimo, pues $\mathbb{E}^* T < +\infty$. Y de $\mathbb{E}U < +\infty$ se sigue $\mathbb{E}U^+ < \infty$: si fuera $\mathbb{E}U^+ = \infty$, la existencia de $\mathbb{E}U$ exigiría $\mathbb{E}U^- < \infty$, y entonces $\mathbb{E}U = +\infty$, contradicción.

¹³Sobre N , la resta $a - b$ sería la forma indeterminada $\infty - \infty$; el valor $+\infty$ la esquiva sin costo alguno, pues $\mathbb{P}_1(N) = 0$ y agrandar el candidato en un conjunto nulo no cambia su esperanza ni estropea la condición de mayorante.

¹⁴La resta de esperanzas es lícita porque $\mathbb{E}_1 a < \infty$. Si $\mathbb{E}_1 b = +\infty$ (equivalentemente, $\mathbb{E}U^- = \infty$, es decir $\mathbb{E}U = -\infty$), ambos miembros valen $-\infty$ y la identidad subsiste.

¹⁵Monotonía de $\mathbb{E}_1^*$: si $X \leq Y$ punto a punto, todo mayorante medible de Y lo es de X , y el ínfimo sólo puede bajar. Y $\mathbb{E}_{2*} T \leq \mathbb{E}_2^* T$ vale sección a sección: si $V \leq T_{\omega_1} \leq U$ con V, U medibles y esperanzas existentes, la monotonía de la esperanza extendida da $\int V \leq \int U$; tomando supremo e ínfimo, $\mathbb{E}_{2*} \leq \mathbb{E}_2^*$.

¹⁶Sección a sección, $\mathbb{E}_2^*(-T)(\omega_1) = -\mathbb{E}_{2*} T(\omega_1)$, por la definición $\mathbb{E}_{2*} = -\mathbb{E}_2^*(-\cdot)$; y en Ω_1 , $-\mathbb{E}_1^*(-V) = \mathbb{E}_{1*} V$, aplicado a $V = \mathbb{E}_{2*} T$.

¹⁷En 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 $\mathbb{E}_1^* \mathbb{E}_2^* T = \mathbb{E}_1^* g$: las esperanzas iteradas colapsan a la marginal. 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 1.17. Sea $\phi : (\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbb{P}}) \rightarrow (\Omega, \mathcal{A}, \mathbb{P})$ medible y tal que $\mathbb{P} = \tilde{\mathbb{P}} \circ \phi^{-1}$. Diremos que ϕ es **perfecta** si, para todo $B \subset \Omega$,

$$\mathbb{P}^*(B) = \tilde{\mathbb{P}}^*(\phi^{-1}(B)) \quad \text{y} \quad \mathbb{P}_*(B) = \tilde{\mathbb{P}}_*(\phi^{-1}(B)),$$

donde $\mathbb{P}^*, \mathbb{P}_*$ (resp. $\tilde{\mathbb{P}}^*, \tilde{\mathbb{P}}_*$) son la probabilidad exterior e interior asociadas a $\mathbb{P}$ (resp. $\tilde{\mathbb{P}}$).

Lema 1.18. Sea $\phi : (\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbb{P}}) \rightarrow (\Omega, \mathcal{A}, \mathbb{P})$ medible y tal que $\mathbb{P} = \tilde{\mathbb{P}} \circ \phi^{-1}$. Entonces son equivalentes:

- (i) Para todo $B \subset \Omega$, $\mathbb{P}^*(B) = \tilde{\mathbb{P}}^*(\phi^{-1}(B))$ y $\mathbb{P}_*(B) = \tilde{\mathbb{P}}_*(\phi^{-1}(B))$.
- (ii) Para toda función acotada $g : \Omega \rightarrow \mathbb{R}$, $\mathbb{E}^*g = \tilde{\mathbb{E}}^*(g \circ \phi)$.

Demostración. Sea $g : \Omega \rightarrow \mathbb{R}$ con $|g| \leq M$. Como $g^* \geq g \geq -M$ punto a punto y la constante M es un mayorante medible de g , la minimalidad da $-M \leq g^* \leq M$, $\mathbb{P}$ -c.s.; en particular $\mathbb{E}g^*$ existe y, por el Lema 1.4(iii), vale $\mathbb{E}^*g$. Por la fórmula de capas aplicada a $g^* + M \geq 0$ y por el Lema 1.9(vi),

$$\mathbb{E}^*g = \mathbb{E}g^* = -M + \int_{-M}^M \mathbb{P}(g^* > t) dt = -M + \int_{-M}^M \mathbb{P}^*(g > t) dt. \quad (1.2)$$

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

(i) $\Rightarrow$ (ii). Sea $|g| \leq M$; entonces $|g \circ \phi| \leq M$. Como $\{g \circ \phi > t\} = \phi^{-1}(\{g > t\})$, la hipótesis da $\tilde{\mathbb{P}}^*(g \circ \phi > t) = \mathbb{P}^*(g > t)$ para todo t . Aplicando (1.2) en ambos espacios, $\tilde{\mathbb{E}}^*(g \circ \phi) = \mathbb{E}^*g$.

(ii) $\Rightarrow$ (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 $\mathbb{P}^*(B) = \mathbb{E}^*(\mathbb{1}_B) = \tilde{\mathbb{E}}^*(\mathbb{1}_{\phi^{-1}(B)}) = \tilde{\mathbb{P}}^*(\phi^{-1}(B))$. Para las interiores, $\mathbb{P}_*(B) = 1 - \mathbb{P}^*(B^c)$ y $\phi^{-1}(B^c) = \phi^{-1}(B)^c$. $\square$

Lema 1.19 (Las proyecciones coordenadas son perfectas). Sea $\pi_1 : \Omega_1 \times \Omega_2 \rightarrow \Omega_1$ 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 $T : \Omega_1 \rightarrow \bar{\mathbb{R}}$,

$$(T \circ \pi_1)^* = T^* \circ \pi_1 \quad (\mathbb{P}_1 \otimes \mathbb{P}_2)\text{-c.s.},$$

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 $g : \Omega_1 \rightarrow \mathbb{R}$ acotada, $\mathbb{E}^*(g \circ \pi_1) = \mathbb{E}_1^*g$; en consecuencia (Lema 1.18), π_1 es perfecta en el sentido de la Definición 1.17: para todo $B \subset \Omega_1$,

$$(\mathbb{P}_1 \otimes \mathbb{P}_2)^*(\pi_1^{-1}(B)) = \mathbb{P}_1^*(B), \quad (\mathbb{P}_1 \otimes \mathbb{P}_2)_*(\pi_1^{-1}(B)) = \mathbb{P}_{1*}(B).$$

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, $(T \circ \pi_1)^* \leq T^* \circ \pi_1$ c.s.

Recíprocamente, sea $V := (T \circ \pi_1)^*$. Como $V \geq T \circ \pi_1$ 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)¹⁸; por la minimalidad de T^* en Ω_1 , se tiene $T^* \leq V(\cdot, \omega_2)$, $\mathbb{P}_1$ -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, $T^* \circ \pi_1 \leq V$ c.s., y queda probada la igualdad.

¹⁸Aquí 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^* . Si T dependiera de ambas coordenadas, el mayorante mínimo de la sección cambiaría con ω_2 , no habría una función fija en Ω_1 que comparar con V y el argumento colapsaría: esa es exactamente la obstrucción por la que no hay Fubini exterior en general (Advertencia previa).

(ii) Sea g acotada, digamos $|g| \leq M$. Entonces $g^* \geq g \geq -M$ punto a punto y, por minimalidad contra la constante M , $g^* \leq M$ c.s.; en particular, $\mathbb{E}_1 g^*$ existe y es finita, y el Lema 1.4(iii) da $\mathbb{E}_1 g^* = \mathbb{E}_1^* g$. El mismo argumento en el producto da $\mathbb{E}[(g \circ \pi_1)^*] = \mathbb{E}^*(g \circ \pi_1)$. Usando la parte (i) y que la ley de π_1 bajo $\mathbb{P}_1 \otimes \mathbb{P}_2$ es $\mathbb{P}_1$,

$$\mathbb{E}^*(g \circ \pi_1) = \mathbb{E}[(g \circ \pi_1)^*] = \mathbb{E}[g^* \circ \pi_1] = \mathbb{E}_1 g^* = \mathbb{E}_1^* g.$$

Como la ley de π_1 bajo $\mathbb{P}_1 \otimes \mathbb{P}_2$ es $\mathbb{P}_1$, la implicación (ii) $\Rightarrow$ (i) del Lema 1.18 traduce esta identidad de esperanzas en la igualdad de las probabilidades exteriores e interiores. $\square$

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.¹⁹

1.3. Convergencia débil

Sea (D, d) un espacio métrico, denotamos

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

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

Lema 1.20. *La σ -álgebra de Borel $\mathcal{B}(D)$ es la menor σ -álgebra que hace medibles (hacia $\mathcal{B}(\mathbb{R})$) a todas las funciones de $C_b(D)$. Equivalente: $\mathcal{B}(D)$ es la menor σ -álgebra que contiene a los abiertos.*

Demostración. Cada $f \in C_b(D)$ es continua y, por lo tanto, Borel-medible; luego $\sigma(C_b(D)) \subset \mathcal{B}(D)$. Recíprocamente, si $F \subset D$ es cerrado y no vacío, la función $f(x) = d(x, F) \wedge 1$ pertenece a $C_b(D)$ y $F = f^{-1}(\{0\}) \in \sigma(C_b(D))$; ²⁰ el caso $F = \emptyset$ es trivial, pues $\emptyset$ pertenece a toda σ -álgebra. Como los cerrados generan $\mathcal{B}(D)$, se concluye $\mathcal{B}(D) \subset \sigma(C_b(D))$. $\square$

Definición 1.21. Sea $\mathbb{L}$ una probabilidad de Borel sobre $(D, \mathcal{B}(D))$. Decimos que $\mathbb{L}$ es **tensa** si

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

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

Observación 1.22

Toda probabilidad de Borel en un espacio métrico separable y completo^a es tensa, ya que es interiormente regular por compactos (es de Radón).^b

^aUn espacio metrizable se llama polaco si es separable y admite una métrica completa compatible con su topología; en particular, un espacio provisto de una métrica completa y separable es polaco.

^bEste es el teorema de Ulam. Véase, p.ej., R. M. Dudley, *Real Analysis and Probability*, Teorema 7.1.4, o P. Billingsley, *Convergence of Probability Measures*, Teorema 1.3.

¹⁹Obsérvese que la desigualdad difícil, $\mathbb{E}_1^* g \leq \mathbb{E}^*(g \circ \pi_1)$, es un caso particular del Lema 1.15: con $T = g \circ \pi_1$ se tiene $\mathbb{E}_2^* T = g$, y la cadena da $\mathbb{E}_1^* g = \mathbb{E}_1^* \mathbb{E}_2^* T \leq \mathbb{E}^* T$. La opuesta es elemental: todo mayorante medible $U \geq g$ en Ω_1 induce el mayorante $U \circ \pi_1$ en el producto, con la misma esperanza. El contenido genuinamente nuevo del lema es la identidad c.s. del ítem (i) entre los mayorantes medibles mínimos, más fuerte que la igualdad de esperanzas, y ésa es la forma en que la perfección se usa en la práctica.

²⁰ $d(x, F) = 0$ si y sólo si $x \in \bar{F} = F$, por ser F cerrado; el $\wedge 1$ no afecta el conjunto de ceros.

Observación 1.23

$\mathbf{L}$ es tensa si y sólo si existen compactos $K_1 \subset K_2 \subset \dots$ con $\mathbf{L}(K_n) \uparrow 1$. En particular, la unión $\bigcup_{n \geq 1} K_n$ (que es σ -compacta) tiene probabilidad 1.

Definición 1.24. Diremos que una probabilidad de Borel $\mathbf{L}$ es **separable** si existe un conjunto numerable $M \subset D$ tal que $\mathbf{L}(\overline{M}) = 1$.

Advertencia 1.25

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 $X = \ell^\infty = \{x = (x_n)_{n \geq 1} : \|x\|_\infty := \sup_n |x_n| < \infty\}$ con la distancia inducida por $\|\cdot\|_\infty$. X no es separable. Sea $c_0 = \{x : x_n \rightarrow 0\}$, 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 $Z : \Omega \rightarrow \mathbb{R}$ una variable aleatoria real. Definimos $\xi : \Omega \rightarrow X$, $\xi(\omega) = (Z(\omega)/n)_{n \geq 1}$. 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 $(X, \mathcal{B}(X))$ tal que $\mu(c_0) = 1$. Concluimos que μ es *separable* (está concentrada en c_0) mientras que el espacio métrico total $(X, \|\cdot\|_\infty)$ no es separable.

Definición 1.26. En un espacio métrico (D, d) , una probabilidad $\mathbf{L}$ es **pretensa** si, para todo $\varepsilon > 0$, existe $C \subset D$ totalmente acotado²¹ y medible tal que $\mathbf{L}(C) \geq 1 - \varepsilon$.

Lema 1.27. Sea $\mathbf{L}$ una probabilidad de Borel en un espacio métrico (D, d) . Entonces

$$\mathbf{L} \text{ es pretensa} \iff \mathbf{L} \text{ es separable.}$$

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

$$\mathbf{L} \text{ es pretensa} \iff \mathbf{L} \text{ es tensa} \iff \mathbf{L} \text{ es separable.}$$

Demostración. Supongamos $\mathbf{L}$ pretensa. Para $n \geq 1$ sea $C_n \subset D$ totalmente acotado con $\mathbf{L}(C_n) \geq 1 - 2^{-n}$. Cada C_n es separable;²² por tanto $C := \bigcup_n C_n$ es separable y $\mathbf{L}(C) = 1$. Sea $M \subset C$ numerable y denso en C ; entonces $\mathbf{L}(\overline{M}) = 1$, es decir, $\mathbf{L}$ es separable.

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

$$\mathbf{L}\left(\bigcup_{m \geq 1} B(x_m, 1/j)\right) = 1.$$

Luego existe N_j tal que $G_j := \bigcup_{m=1}^{N_j} B(x_m, 1/j)$ satisface $\mathbf{L}(G_j) \geq 1 - \varepsilon 2^{-j}$.²³ Sea $C := \bigcap_{j \geq 1} G_j$. Entonces $C \in \mathcal{B}(D)$ y

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

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, $\mathbf{L}$ es pretensa.

²¹ $A \subset D$ es totalmente acotado si para todo $\varepsilon > 0$ existen $x_1, \dots, x_N \in D$ tales que $A \subset \bigcup_{i=1}^N B(x_i, \varepsilon)$. Nótese que depende de la métrica elegida. Recordemos que todo compacto es totalmente acotado, y todo totalmente acotado es separable. Si D es completo, la clausura de todo conjunto totalmente acotado es compacta; en particular, en espacios completos, pretensión y tensión son equivalentes.

²²Todo 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 .

²³Por continuidad desde abajo de $\mathbf{L}$ 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.

Si D es completo, el cierre de un conjunto totalmente acotado es compacto; 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. $\square$

Definición 1.28. Un par $(A, \leq)$ es un **orden parcial** si, para todo $x, y, z \in A$,

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

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 1.29. 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 1.30. 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 $\{X_\alpha : \Omega_\alpha \rightarrow D\}_{\alpha \in A}$ (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 $\{\mathbb{L}_\alpha\}_{\alpha \in A}$ 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 1.31. Dadas una red X_α y una probabilidad de Borel $\mathbb{L}$ en D , decimos que X_α **converge débilmente** a $\mathbb{L}$, y denotamos $X_\alpha \Rightarrow \mathbb{L}$, si

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

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

Definición 1.32. Sea (D, d) un espacio métrico, $\{\mathbb{L}_\alpha\}_{\alpha \in A}$ una red de probabilidades de Borel en D , y $\mathbb{L}$ una probabilidad de Borel en D . Diremos que $\mathbb{L}_\alpha$ **converge débilmente a** $\mathbb{L}$, y denotamos $\mathbb{L}_\alpha \Rightarrow \mathbb{L}$ si

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

Observación 1.33

Cuando escribimos $X_\alpha \Rightarrow X$ damos por sobreentendido que X es Borel medible, y lo entendemos como $X_\alpha \Rightarrow \mathbb{L}$, donde $\mathbb{L} = \mathbb{P} \circ X^{-1}$ es la ley de X (una probabilidad de Borel).

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

- (i) $X_\alpha \Rightarrow \mathbb{L}$.
- (ii) $\liminf_\alpha \mathbb{P}_{\alpha,*}(X_\alpha \in G) \geq \mathbb{L}(G)$, para todo abierto $G \subset D$.
- (iii) $\limsup_\alpha \mathbb{P}_\alpha^*(X_\alpha \in F) \leq \mathbb{L}(F)$, para todo cerrado $F \subset D$.

(iv) $\liminf_{\alpha} \mathbb{E}_{\alpha,*} f(X_{\alpha}) \geq \int f d\mathbb{L}$, para toda $f : D \rightarrow \mathbb{R}$ l.s.c. ²⁴ y acotada por abajo.

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

(vi) Para todo boreliano B con $\mathbb{L}(\partial B) = 0$,

$$\lim_{\alpha} \mathbb{P}_{\alpha,*}(X_{\alpha} \in B) = \lim_{\alpha} \mathbb{P}_{\alpha}^*(X_{\alpha} \in B) = \mathbb{L}(B).$$

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

Demostración. (ii) $\Leftrightarrow$ (iii). Si F es cerrado y $G = F^c$, entonces $\mathbb{P}_{\alpha}^*(X_{\alpha} \in F) = 1 - \mathbb{P}_{\alpha,*}(X_{\alpha} \in G)$, y $\mathbb{L}(F) = 1 - \mathbb{L}(G)$, y las desigualdades con $\liminf$ y $\limsup$ son equivalentes. ²⁵

(iv) $\Leftrightarrow$ (v). Se usa f u.s.c. $\Leftrightarrow -f$ l.s.c. y $\mathbb{E}_{\alpha,*}(-Y) = -\mathbb{E}_{\alpha}^*(Y)$.

(iv) $\Rightarrow$ (ii). Si $G \subset D$ es abierto, entonces $\mathbb{1}_G$ es semicontinua inferiormente y acotada por abajo. ²⁶ Aplicando (iv) a $f = \mathbb{1}_G$,

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

(i) $\Rightarrow$ (vii). Si $f \geq 0$ es acotada y Lipschitz, entonces $f, -f \in C_b(D)$. Por (i),

$$\mathbb{E}_{\alpha}^* f(X_{\alpha}) \rightarrow \int f d\mathbb{L}, \quad \mathbb{E}_{\alpha}^*(-f(X_{\alpha})) \rightarrow -\int f d\mathbb{L}.$$

Luego

$$\mathbb{E}_{\alpha,*} f(X_{\alpha}) = -\mathbb{E}_{\alpha}^*(-f(X_{\alpha})) \rightarrow \int f d\mathbb{L},$$

y en particular vale (vii).

(vii) $\Rightarrow$ (ii). Sea G abierto. Si $G = D$, es trivial. Si no, sea $f_m(x) := (md(x, G^c)) \wedge 1$. Entonces $f_m \geq 0$, es acotada y Lipschitz (la distancia a un conjunto es 1-Lipschitz), $f_m \leq \mathbb{1}_G$ y $f_m \uparrow \mathbb{1}_G$ ²⁷. Por monotonía de $\mathbb{E}_*$ ²⁸ y (vii),

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

Por convergencia monótona, $\sup_m \int f_m d\mathbb{L} = \mathbb{L}(G)$.

(ii) $\Rightarrow$ (iv). Restando una constante, suponga $f \geq 0$. ²⁹ Para $m \geq 1$, definimos

$$f_m := \frac{1}{m} \sum_{i=1}^{m^2} \mathbb{1}_{G_{i,m}}, \quad G_{i,m} := \{x : f(x) > i/m\}.$$

²⁴Una función $f : D \rightarrow \mathbb{R}$ 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 $\liminf_{y \rightarrow x} f(y) \geq f(x)$ para todo x (resp. $\limsup_{y \rightarrow x} f(y) \leq f(x)$), y también a la secuencial: $x_n \rightarrow x \Rightarrow \liminf_n f(x_n) \geq f(x)$. 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 $\int f d\mathbb{L}$ 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.

²⁵En detalle: $\{X_{\alpha} \in F\} = \{X_{\alpha} \in G\}^c$ y, por definición de probabilidad interior, $\mathbb{P}_{\alpha,*}(E) = 1 - \mathbb{P}_{\alpha}^*(E^c)$ para todo $E \subset \Omega_{\alpha}$; además, $\limsup(1 - z_{\alpha}) = 1 - \liminf z_{\alpha}$.

²⁶ $\mathbb{1}_G$ es s.c.i. porque $\{\mathbb{1}_G > a\}$ es D, G o $\emptyset$ según $a < 0, 0 \leq a < 1$ o $a \geq 1$: siempre abierto.

²⁷Si $f_m(x) > 0$, entonces $d(x, G^c) > 0$ y, al ser G^c cerrado, $x \in G$; esto da $0 \leq f_m \leq \mathbb{1}_G$. Y para $x \in G$, la apertura de G da $d(x, G^c) > 0$, de modo que $md(x, G^c) \geq 1$ para m grande: $f_m(x) \uparrow 1$.

²⁸Si $Y \leq Z$ punto a punto, todo minorante medible de Y lo es de Z ; luego $\mathbb{E}_* Y \leq \mathbb{E}_* Z$.

²⁹Si $f \geq -c$, se aplica lo que sigue a $f + c$: tanto $\mathbb{E}_{\alpha,*}$ como $\int \cdot d\mathbb{L}$ se trasladan por constantes.

Cada $G_{i,m}$ es abierto porque f es s.c.i. Además,

$$0 \leq f_m \leq f, \quad (f \wedge m) - m^{-1} \leq f_m \leq f \wedge m,$$

y por tanto $f_m(x) \rightarrow f(x)$ para todo x .³⁰ Por superaditividad y homogeneidad positiva de $\mathbb{E}_*$ (véase la Advertencia 1.11),³¹

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

Así, usando (ii),

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

Como $f_m \leq f$,

$$\liminf_{\alpha} \mathbb{E}_{\alpha,*} f(X_\alpha) \geq \int f_m d\mathbb{L} \quad \forall m.$$

Finalmente, $f_m \rightarrow f$ y $0 \leq f_m \leq f$; por Fatou aplicado a f_m ,

$$\int f d\mathbb{L} \leq \liminf_m \int f_m d\mathbb{L} \leq \sup_m \int f_m d\mathbb{L},$$

y queda (iv).

(ii), (iii) $\Rightarrow$ (vi). Si $\mathbb{L}(\partial B) = 0$, entonces $\mathbb{L}(B^\circ) = \mathbb{L}(\overline{B}) = \mathbb{L}(B)$. Como $B^\circ \subset B \subset \overline{B}$,

$$\mathbb{L}(B) \leq \liminf_{\alpha} \mathbb{P}_{\alpha,*}(X_\alpha \in B) \leq \limsup_{\alpha} \mathbb{P}_{\alpha}^*(X_\alpha \in B) \leq \mathbb{L}(B),$$

donde se usó (ii) para B° y (iii) para $\overline{B}$. Por tanto ambas probabilidades, interior y exterior, convergen a $\mathbb{L}(B)$.

(vi) $\Rightarrow$ (iii). Sea F cerrado; si $F = \emptyset$, la desigualdad es trivial, así que supongamos $F \neq \emptyset$ y pongamos $F^\delta := \{x : d(x, F) < \delta\}$. Como $\partial F^\delta \subset \{x : d(x, F) = \delta\}$, salvo a lo sumo numerables $\delta > 0$ se tiene $\mathbb{L}(\partial F^\delta) = 0$.³² Tome $\delta_n \downarrow 0$ con esa propiedad. Por (vi),

$$\limsup_{\alpha} \mathbb{P}_{\alpha}^*(X_\alpha \in F) \leq \lim_{\alpha} \mathbb{P}_{\alpha}^*(X_\alpha \in F^{\delta_n}) = \mathbb{L}(F^{\delta_n}).$$

Como $F^{\delta_n} \downarrow F$,³³ se obtiene $\limsup_{\alpha} \mathbb{P}_{\alpha}^*(X_\alpha \in F) \leq \mathbb{L}(F)$.

(ii) $\Rightarrow$ (i). Por (ii) $\Rightarrow$ (iv) y (iv) $\Leftrightarrow$ (v), para $f \in C_b(D)$,

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

Como $\mathbb{E}_{\alpha,*} f(X_\alpha) \leq \mathbb{E}_{\alpha}^* f(X_\alpha)$, se sigue $\mathbb{E}_{\alpha}^* f(X_\alpha) \rightarrow \int f d\mathbb{L}$, que es $X_\alpha \Rightarrow \mathbb{L}$. Con esto, todas las flechas del diagrama están probadas y el grafo es conexo en ambos sentidos: las siete condiciones son equivalentes. $\square$

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

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

³⁰Escribiendo $r = mf(x)$: $mf_m(x) = \#\{1 \leq i \leq m^2 : i < r\}$. Si $r \leq m^2$, ese número está entre $r - 1$ y r ; si $r \geq m^2$, está entre $m^2 - 1$ y m^2 (cuando $r = m^2$, el último índice no cuenta, por ser estricta la desigualdad). En ambos casos, $(f \wedge m)(x) - m^{-1} \leq f_m(x) \leq (f \wedge m)(x) \leq f(x)$.

³¹La superaditividad de $\mathbb{E}_*$, enunciada allí para dos sumandos, se itera a sumas finitas; la homogeneidad positiva saca el factor $1/m$.

³²Los conjuntos $\{x : d(x, F) = \delta\}$, $\delta > 0$, son disjuntos dos a dos, y una probabilidad puede dar masa positiva a lo sumo a numerables conjuntos disjuntos (para cada k , a lo sumo k de ellos tienen masa $> 1/k$). En la cadena siguiente, la primera desigualdad usa $F \subset F^{\delta_n}$ y la monotonía de $\mathbb{P}_{\alpha}^*$.

³³ $\bigcap_n F^{\delta_n} = \{x : d(x, F) = 0\} = F$ porque F es cerrado; se concluye por continuidad de la probabilidad desde arriba.

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

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

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

Teorema 1.36 (Aplicación continua). Sean D, E espacios métricos, $D_0 \subset D$ y $g : D \rightarrow E$ continua en cada punto de D_0 . Sea $X_\alpha : \Omega_\alpha \rightarrow D$ una red no necesariamente medible y sea $X : \Omega \rightarrow D$ 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. La continuidad de g en cada punto de D_0 implica que la restricción $g|_{D_0} : D_0 \rightarrow E$ es continua para la topología inducida y, por lo tanto, medible respecto de la σ -álgebra traza $\mathcal{B}(D) \cap D_0$. Como X es Borel-medible y toma valores en D_0 , la composición $Y := g(X) = (g|_{D_0})(X)$ es Borel-medible.³⁴

Sea ahora $F \subset E$ cerrado y escribamos $A := g^{-1}(F)$. Para cada α , $\{g(X_\alpha) \in F\} = \{X_\alpha \in A\} \subset \{X_\alpha \in \bar{A}\}$. Como $\bar{A}$ es cerrado en D , por Portmanteau para $X_\alpha \Rightarrow X$,

$$\limsup_\alpha \mathbb{P}_\alpha^*(g(X_\alpha) \in F) \leq \limsup_\alpha \mathbb{P}_\alpha^*(X_\alpha \in \bar{A}) \leq \mathbb{P}(X \in \bar{A}).$$

Veamos que el último término coincide con $\mathbb{P}(Y \in F)$. En efecto, $D_0 \cap \bar{A} = D_0 \cap A$. La inclusión $\supset$ es obvia. Para la otra, si $x \in D_0 \cap \bar{A}$, existe $x_n \in A$ con $x_n \rightarrow x$; como g es continua en x , $g(x_n) \rightarrow g(x)$. Pero $g(x_n) \in F$ y F es cerrado, luego $g(x) \in F$, es decir $x \in A$. Como X toma valores en D_0 , los eventos $\{X \in \bar{A}\}$, $\{X \in D_0 \cap \bar{A}\}$, $\{X \in D_0 \cap A\}$ y $\{X \in A\} = \{Y \in F\}$ coinciden como subconjuntos de Ω . Así, para todo cerrado $F \subset E$, $\limsup_\alpha \mathbb{P}_\alpha^*(g(X_\alpha) \in F) \leq \mathbb{P}(Y \in F)$, y el teorema de Portmanteau da $g(X_\alpha) \Rightarrow g(X)$. $\square$

Corolario 1.37 (Continuidad casi segura). Sean D, E espacios métricos, $g : D \rightarrow E$ arbitraria y D_g su conjunto de discontinuidades. Sea $X : \Omega \rightarrow D$ Borel-medible con $\mathbb{P}(X \in D_g) = 0$. Si $X_\alpha \Rightarrow X$, entonces $g(X_\alpha) \Rightarrow g(X)$, donde $g(X)$ se entiende mediante una versión Borel.

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 $D_0 \neq \emptyset$; fijemos $x_0 \in D_0$ y definamos

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

Como $\{X \in D_0\} \in \mathcal{A}$, $\widetilde{X}$ es Borel-medible, toma valores en D_0 y coincide con X salvo en el evento nulo $\{X \in D_g\}$; en particular tiene la misma ley que X , de modo que $X_\alpha \Rightarrow \widetilde{X}$. Por el Teorema 1.36, $g(X_\alpha) \Rightarrow g(\widetilde{X})$, y $g(\widetilde{X})$ es una versión Borel de $g(X)$, pues coincide con $g(X)$ en $\{X \in D_0\}$. $\square$

³⁴En detalle: si $V \subset E$ es abierto, la continuidad de $g|_{D_0}$ da $(g|_{D_0})^{-1}(V) = G \cap D_0$ para algún abierto $G \subset D$; como X toma valores en D_0 , $\{Y \in V\} = \{X \in G \cap D_0\} = \{X \in G\} \in \mathcal{A}$.

Observación 1.38

Ni en el Teorema 1.36 ni en el Corolario 1.37 se requiere que g sea medible (ni que $g^{-1}(F)$ lo sea). La prueba sólo usa que $\overline{g^{-1}(F)}$ es cerrado (Borel) y la continuidad de g en los puntos visitados por X . Si g es continua en todo D , puede tomarse $D_0 = D$ y las condiciones sobre X se cumplen trivialmente.

Definición 1.39. 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

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

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

Definición 1.40. 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

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

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

Lema 1.41. 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. Sea $f \in C_b(D)$. Por $X_\alpha \Rightarrow X$, $\mathbb{E}_\alpha^* f(X_\alpha) \rightarrow \mathbb{E} f(X)$. Aplicando lo mismo a $-f$,

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

Luego $\mathbb{E}_\alpha^* f(X_\alpha) - \mathbb{E}_{\alpha,*} f(X_\alpha) \rightarrow 0$, y $\{X_\alpha\}$ es *asintóticamente medible*.

Suponga primero que $\{X_\alpha\}$ es *asintóticamente tensa*. Dado $\varepsilon > 0$, tome compacto $K \subset D$ tal que $\liminf_\alpha \mathbb{P}_{\alpha,*}(X_\alpha \in K^\eta) \geq 1 - \varepsilon, \forall \eta > 0$. Para $\delta > 0$, sea $K_\delta := \{x \in D : d(x, K) \leq \delta\}$. Entonces K_δ es cerrado y $K^{\delta/2} \subset K_\delta$.³⁵ Por tanto, $1 - \varepsilon \leq \liminf_\alpha \mathbb{P}_{\alpha,*}(X_\alpha \in K^{\delta/2}) \leq \limsup_\alpha \mathbb{P}_\alpha^*(X_\alpha \in K_\delta)$. Por Portmanteau aplicado al cerrado K_δ , $\limsup_\alpha \mathbb{P}_\alpha^*(X_\alpha \in K_\delta) \leq \mathbb{P}(X \in K_\delta)$. Así, $\mathbb{P}(X \in K_\delta) \geq 1 - \varepsilon$ para todo $\delta > 0$. Como $K_\delta \downarrow K$ al tomar $\delta \downarrow 0$,³⁶ se obtiene $\mathbb{P}(X \in K) \geq 1 - \varepsilon$. Luego X es *tenso*.

Recíprocamente, si X es *tenso*, dado $\varepsilon > 0$ tome compacto $K \subset D$ con $\mathbb{P}(X \in K) \geq 1 - \varepsilon$. Para todo $\delta > 0$, K^δ es abierto y contiene a K ; por Portmanteau,

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

Esto es la tensión asintótica de $\{X_\alpha\}$. □

El teorema de Prokhorov usa el siguiente resultado, que generaliza la representación de Riesz–Markov y que tomaremos como conocido.³⁷

Teorema 1.42 (Daniell–Stone). Sea (D, d) un espacio métrico y sea $I : C_b(D) \rightarrow \mathbb{R}$ un funcional tal que:

- (i) I es *lineal*;

³⁵ K_δ es la preimagen del cerrado $[0, \delta]$ por la función continua $x \mapsto d(x, K)$. En la cadena siguiente se usa además $\mathbb{P}_{\alpha,*} \leq \mathbb{P}_\alpha^*$ y la monotonía en el evento.

³⁶ $\bigcap_{\delta > 0} K_\delta = \{x : d(x, K) = 0\} = \overline{K} = K$, por ser K compacto (cerrado); se concluye por continuidad de la probabilidad desde arriba a lo largo de $\delta = 1/k$.

³⁷ Para el teorema de Daniell–Stone, véase R. M. Dudley, *Real Analysis and Probability*, Cambridge University Press, 2002, §4.5.

- (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 $\mathbb{L}$ sobre D tal que

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

Teorema 1.43 (Prokhorov). Sea (D, d) métrico y $X_\alpha : \Omega_\alpha \rightarrow D$ 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 $\mathbb{L}$ en D tales que $X_{\alpha(\beta)} \Rightarrow \mathbb{L}$.
- (ii) Si $\{X_n\}$ es una sucesión asintóticamente tensa y asintóticamente medible, entonces existe una subsucesión $X_{n_j} \Rightarrow \mathbb{L}$, para alguna probabilidad de Borel tensa $\mathbb{L}$.

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,³⁸ existe una subred $\alpha(\beta)$ tal que, para todo $f \in C_b(D)$, $\mathbb{E}_{\alpha(\beta)}^* f(X_{\alpha(\beta)}) \rightarrow L(f)$. Como la red es asintóticamente medible,³⁹ también $\mathbb{E}_{\alpha(\beta),*} f(X_{\alpha(\beta)}) \rightarrow L(f)$, $f \in C_b(D)$. Probemos que L es lineal y positivo. Para $f, g \in C_b(D)$, usando subaditividad de $\mathbb{E}^*$ y superaditividad de $\mathbb{E}_*$,

$$\mathbb{E}_* f(X) + \mathbb{E}_* g(X) \leq \mathbb{E}_*(f+g)(X) \leq \mathbb{E}^*(f+g)(X) \leq \mathbb{E}^* f(X) + \mathbb{E}^* g(X),$$

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

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

luego $L(f+g) = L(f) + L(g)$. La homogeneidad para $a \geq 0$ viene de $\mathbb{E}^*(af(X)) = a\mathbb{E}^*f(X)$; por aditividad, $L(-f) = -L(f)$, y entonces $L(af) = aL(f)$ para todo $a \in \mathbb{R}$. Si $f \geq 0$, entonces $\mathbb{E}_* f(X_{\alpha(\beta)}) \geq 0$, luego $L(f) \geq 0$. Además $L(1) = 1$.

Resta verificar la continuidad monótona. Sea $f_m \in C_b(D)$ con $f_m \downarrow 0$; en particular $0 \leq f_m \leq f_1$. Fijado $\varepsilon > 0$, por tensión asintótica de la subred⁴⁰ existe compacto $K \subset D$ tal que, para todo $\delta > 0$,

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

Por Dini⁴¹, $f_m \downarrow 0$ uniformemente en K ; tome m tal que $\sup_K f_m \leq \varepsilon$. Por continuidad y compacidad, existe $\delta > 0$ tal que $\sup_{K^\delta} f_m \leq 2\varepsilon$.⁴² Con $B_\beta = \{X_{\alpha(\beta)} \in K^\delta\}$ y $M = \|f_1\|_\infty$, $f_m(X_{\alpha(\beta)}) \leq 2\varepsilon + M\mathbb{1}_{B_\beta^c}$,⁴³ y por tanto

$$\mathbb{E}_{\alpha(\beta)}^* f_m(X_{\alpha(\beta)}) \leq 2\varepsilon + MP_{\alpha(\beta)}^*(B_\beta^c).$$

³⁸El 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 coordinada a coordinada, es decir, para cada f por separado.

³⁹La 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.

⁴⁰Las subredes heredan la tensión asintótica: para toda red numérica, $\liminf_{\beta} z_{\alpha(\beta)} \geq \liminf_{\alpha} z_{\alpha}$, de modo que la cota inferior sólo puede mejorar.

⁴¹Teorema de Dini: si K es compacto y $f_m \in C(K)$ converge monótonamente punto a punto a una función continua f , entonces $f_m \rightarrow f$ uniformemente en K .

⁴²Si no, habría puntos x_n con $d(x_n, K) \rightarrow 0$ y $f_m(x_n) > 2\varepsilon$; tomando $y_n \in K$ con $d(x_n, y_n) \rightarrow 0$ y una subsucesión $y_n \rightarrow y \in K$ (compacidad), resulta $x_n \rightarrow y$ y, por continuidad de f_m , $f_m(y) \geq 2\varepsilon > \varepsilon \geq \sup_K f_m$, contradicción.

⁴³La cota puntual: en B_β , $X_{\alpha(\beta)} \in K^\delta$ y allí $f_m \leq 2\varepsilon$; fuera, $f_m \leq f_1 \leq M$. 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).

Como $\limsup_{\beta} P_{\alpha(\beta)}^*(B_{\beta}^c) \leq \varepsilon$, se obtiene $0 \leq L(f_m) \leq (2 + M)\varepsilon$ 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 $\ell \geq 0$. La cota anterior y la arbitrariedad de ε implican $\ell = 0$. Así, $L(f_m) \downarrow 0$.

Por Daniell–Stone, existe una probabilidad de Borel $\mathbb{L}$ sobre D tal que

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

Luego

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

es decir, $X_{\alpha(\beta)} \Rightarrow \mathbb{L}$. Como la subred sigue siendo asintóticamente tensa, el lema anterior implica que $\mathbb{L}$ es tensa.⁴⁴

Para cada $m \geq 1$, sea $K_m \subset D$ 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.⁴⁵ Por Tietze,⁴⁶ 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 $\mathbb{E}^* h_2(X)$, 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 $\|f\|_\infty \leq 1$.⁴⁷ 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 $\sup_{x \in K_m^\delta} |f(x) - h(x)| \leq 2\eta$.⁴⁸ 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

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

Si j, k son grandes,

$$|\mathbb{E}_{n_j}^* f(X_{n_j}) - \mathbb{E}_{n_k}^* f(X_{n_k})| \leq 4\eta + \frac{2}{m} + |\mathbb{E}_{n_j}^* h(X_{n_j}) - \mathbb{E}_{n_k}^* h(X_{n_k})|.$$

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

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

Como la subsucesión sigue siendo asintóticamente medible, $\mathbb{E}_{n_j,*} f(X_{n_j}) \rightarrow L(f)$, $f \in C_b(D)$. 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 $\mathbb{L}$ tal que

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

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

⁴⁴Lema 1.41(b), aplicado a la subred y a la variable canónica $X = \text{id}$ sobre $(D, \mathcal{B}(D), \mathbb{L})$, cuya ley es $\mathbb{L}$.

⁴⁵Para 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.

⁴⁶Teorema de extensión de Tietze: si A es cerrado en un espacio métrico D y $h : A \rightarrow \mathbb{R}$ es continua y acotada, entonces existe $\tilde{h} \in C_b(D)$ tal que $\tilde{h}|_A = h$ y $\|\tilde{h}\|_{\infty, D} = \|h\|_{\infty, A}$.

⁴⁷Para $f \neq 0$, 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.

⁴⁸Mismo argumento de compacidad de la parte (i): puntos x_n con $d(x_n, K_m) \rightarrow 0$ 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 $\|f - h\|_\infty \leq 2$.

Definición 1.44. Sea $\mathcal{F} \subset C_b(D)$.

- $\mathcal{F}$ 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, $f \vee g, f \wedge g \in \mathcal{F}$ 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 cerrado bajo productos puntuales: si $f, g \in \mathcal{F}$, entonces $fg \in \mathcal{F}$.

Lema 1.45. Sean $\mathbf{L}_1$ y $\mathbf{L}_2$ medidas de Borel finitas en D .

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

(ii) Si $\mathbf{L}_1$ y $\mathbf{L}_2$ son probabilidades de Borel tensas en D y

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

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

Demostración. Sea $F \subset D$ cerrado y no vacío (el caso $F = \emptyset$ es trivial) y, para $m \geq 1$, $f_m(x) := (1 - md(x, F))^+$. Entonces $f_m \in C_b(D)$, $0 \leq f_m \leq 1$ y $f_m \downarrow \mathbf{1}_F$ puntualmente, pues F es cerrado.⁴⁹ Por convergencia dominada,

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

Por el teorema de Dynking (ver [1, Teorema 1.1])) $\mathbf{L}_1 = \mathbf{L}_2$.

Sea $g \in C_b(D)$ y $G := \|g\|_\infty$. Fijado $\varepsilon > 0$, por tensión existen compactos $K_1, K_2 \subset D$ con $\mathbf{L}_i(K_i) \geq 1 - \varepsilon$; el compacto $K := K_1 \cup K_2$ satisface $\mathbf{L}_i(K) \geq 1 - \varepsilon$ para $i = 1, 2$. La clase de restricciones $\{f|_K : f \in \mathcal{F}\}$ es un retículo vectorial de $C(K)$ que contiene las constantes y separa puntos de K ; por el teorema de Stone–Weierstrass⁵⁰, es densa en $C(K)$. Existe entonces $f_0 \in \mathcal{F}$ con $\sup_K |f_0 - g| < \varepsilon$. Definiendo $f := (f_0 \vee (-G - 1)) \wedge (G + 1) \in \mathcal{F}$, se conserva $\sup_K |f - g| < \varepsilon$ (el truncamiento acerca f_0 al rango de g)⁵¹ y además $\|f\|_\infty \leq G + 1$. Por lo tanto, para $i = 1, 2$,

$$\left| \int g d\mathbf{L}_i - \int f d\mathbf{L}_i \right| \leq \int_K |g - f| d\mathbf{L}_i + \int_{K^c} |g - f| d\mathbf{L}_i \leq \varepsilon + (2G + 1)\varepsilon.$$

Como $\int f d\mathbf{L}_1 = \int f d\mathbf{L}_2$, se concluye $\left| \int g d\mathbf{L}_1 - \int g d\mathbf{L}_2 \right| \leq 2(2G + 2)\varepsilon$. Haciendo $\varepsilon \downarrow 0$, $\int g d\mathbf{L}_1 = \int g d\mathbf{L}_2$ para toda $g \in C_b(D)$, y la parte (i) da $\mathbf{L}_1 = \mathbf{L}_2$. $\square$

Observación 1.46

Exigir sólo *medibilidad asintótica* de una red $\{X_\alpha\}$ (en lugar de medibilidad Borel de cada X_α) amplía notablemente la aplicabilidad de la convergencia débil. No obstante, la medibilidad asintótica suele ser difícil de verificar directamente. Para redes tensas el cuadro es favorable: la tensión asintótica más un poco de *medibilidad* proporciona *medibilidad asintótica*. El siguiente lema da una versión abstracta de este principio.

⁴⁹Para $x \in F$, $f_m(x) = 1$ para todo m . Para $x \notin F$, $d(x, F) > 0$ precisamente por ser F cerrado, y $f_m(x) = 0$ en cuanto $md(x, F) \geq 1$. La monotonía en m es inmediata de la fórmula.

⁵⁰acá se usa que es un retículo vectorial que separa puntos y contiene a las constantes

⁵¹Dos verificaciones: la pertenencia $f \in \mathcal{F}$ usa que un retículo vectorial con constantes es cerrado bajo truncamientos, pues $f_0 \vee c = c + (f_0 - c)^+$ y $f_0 \wedge c = -((-f_0) \vee (-c))$. Y $\sup_K |f - g| < \varepsilon$ se conserva porque $\tau(t) := (t \vee (-G - 1)) \wedge (G + 1)$ es 1-Lipschitz y deja fijo a g (ya que $|g| \leq G$): $|f - g| = |\tau(f_0) - \tau(g)| \leq |f_0 - g|$ en K .

Lema 1.47. Sea $X_\alpha : \Omega_\alpha \rightarrow D$ 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

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

entonces X_α es asintóticamente medible, es decir,

$$\mathbb{E}_\alpha^* g(X_\alpha) - \mathbb{E}_{\alpha,*} g(X_\alpha) \rightarrow 0 \quad \text{para toda } g \in C_b(D).$$

Demostración. Para $h \in C_b(D)$ escribimos

$$\Delta_\alpha(h) := \mathbb{E}_\alpha^* h(X_\alpha) - \mathbb{E}_{\alpha,*} h(X_\alpha) \geq 0,$$

de modo que la hipótesis es $\Delta_\alpha(f) \rightarrow 0$ para toda $f \in \mathcal{F}$, y la conclusión, $\Delta_\alpha(g) \rightarrow 0$ para toda $g \in C_b(D)$.

Podemos suponer que $\mathcal{F}$ contiene las constantes: sea $\tilde{\mathcal{F}} := \{f + c : f \in \mathcal{F}, c \in \mathbb{R}\}$. Es un subespacio vectorial de $C_b(D)$, contiene las constantes y es cerrado bajo productos, pues $(f + c)(f' + c') = ff' + c'f + cf' + cc'$ y $\mathcal{F}$ es un álgebra; separa puntos porque ya lo hacía $\mathcal{F} \subset \tilde{\mathcal{F}}$; y satisface la hipótesis, ya que $\Delta_\alpha(f + c) = \Delta_\alpha(f) \rightarrow 0$.⁵² Como la conclusión no involucra a la clase, basta probar el lema para $\tilde{\mathcal{F}}$. Renombrando, suponemos de aquí en adelante que $\mathcal{F}$ contiene las constantes.

Fijemos $g \in C_b(D)$ y $\eta \in (0, 1]$, y pongamos $G := \|g\|_\infty$ y $C := 2G + 2$. Por tensión asintótica existe un compacto $K \subset D$ tal que

$$\limsup_\alpha \mathbb{P}_\alpha^*(X_\alpha \notin K^\delta) \leq \frac{\eta}{C} \quad \text{para todo } \delta > 0,$$

donde $K^\delta := \{x : d(x, K) < \delta\}$. Nótese que $K \neq \emptyset$, pues $\eta/C < 1$.

Las restricciones $\{f|_K : f \in \mathcal{F}\}$ forman una subálgebra de $C(K)$ que contiene las constantes y separa puntos de K —si $x \neq y$ en K , alguna $f \in \mathcal{F}$ los separa en D , y su restricción los separa en K —; por el teorema de Stone–Weierstrass es densa en $C(K)$, de modo que existe $f_0 \in \mathcal{F}$ con $\sup_K |f_0 - g| < \eta$.

El problema es que f_0 puede ser grande fuera de K , y necesitamos controlar $\|f - g\|_\infty$ globalmente. Truncamos: sea $\tau(t) := (t \vee (-G - 1)) \wedge (G + 1)$, y sea p un polinomio con $|p - \tau| < \eta$ en $[-\|f_0\|_\infty, \|f_0\|_\infty]$, que existe por el teorema de Weierstrass. Como $\mathcal{F}$ es un álgebra que contiene las constantes, $f := p(f_0) \in \mathcal{F}$, y valen

$$\|f\|_\infty \leq G + 2, \quad \sup_{x \in K} |f(x) - g(x)| < 2\eta.$$

La primera, porque $|p| \leq |\tau| + \eta \leq (G + 1) + \eta \leq G + 2$, usando $\eta \leq 1$. La segunda, porque en K se tiene $|f_0| \leq |g| + \eta \leq G + 1$, luego $\tau(f_0) = f_0$ y

$$|f - g| \leq |p(f_0) - \tau(f_0)| + |f_0 - g| < \eta + \eta = 2\eta.$$

Del compacto a un entorno suyo. Como $f - g$ es continua y $|f - g| < 2\eta$ en el compacto K , existe $\delta > 0$ tal que $|f - g| \leq 3\eta$ en K^δ . Si no, habría puntos x_n con $d(x_n, K) \rightarrow 0$ y $|f - g|(x_n) > 3\eta$; tomando $y_n \in K$ con $d(x_n, y_n) \rightarrow 0$ y una subsucesión $y_n \rightarrow y \in K$, resultaría $x_n \rightarrow y$ y, por continuidad, $|f - g|(y) \geq 3\eta > 2\eta$, contradicción.

Además $\|f - g\|_\infty \leq (G + 2) + G = C$. Por lo tanto, punto a punto,

$$|f - g|(X_\alpha) \leq 3\eta + C \mathbf{1}_{\{X_\alpha \notin K^\delta\}},$$

y por monotonía y subaditividad de la esperanza exterior,

$$\limsup_\alpha \mathbb{E}_\alpha^* |f - g|(X_\alpha) \leq 3\eta + C \cdot \frac{\eta}{C} = 4\eta.$$

⁵² $\mathbb{E}^*(Y + c) = \mathbb{E}^*Y + c$ para toda constante real c , pues U es mayorante medible de Y si y sólo si $U + c$ lo es de $Y + c$, con esperanzas trasladadas en c ; lo mismo vale para $\mathbb{E}_*$, y en la diferencia Δ_α las constantes se cancelan.

Cierre. De $g = f + (g - f)$ y $-g = (-f) + (f - g)$, la subaditividad de $\mathbb{E}_\alpha^*$ da

$$\mathbb{E}_\alpha^* g(X_\alpha) \leq \mathbb{E}_\alpha^* f(X_\alpha) + \mathbb{E}_\alpha^* |f - g|(X_\alpha), \quad \mathbb{E}_\alpha^* (-g)(X_\alpha) \leq \mathbb{E}_\alpha^* (-f)(X_\alpha) + \mathbb{E}_\alpha^* |f - g|(X_\alpha).$$

Sumando, y usando que $\Delta_\alpha(h) = \mathbb{E}_\alpha^* h(X_\alpha) + \mathbb{E}_\alpha^* (-h)(X_\alpha)$ —pues $\mathbb{E}_{\alpha,*} h = -\mathbb{E}_\alpha^* (-h)$ —,

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

Como $f \in \mathcal{F}$, se tiene $\Delta_\alpha(f) \rightarrow 0$, y por lo tanto $\limsup_\alpha \Delta_\alpha(g) \leq 8\eta$. Siendo $\eta \in (0, 1]$ arbitrario y $\Delta_\alpha(g) \geq 0$, concluimos $\Delta_\alpha(g) \rightarrow 0$. $\square$

1.4. Espacios producto

1.4.1. Topología y σ -álgebras en productos finitos

Definición 1.48: Métricas de producto

Sean (D, d) y (E, e) espacios métricos. En $D \times E$, las métricas $c_\infty((x, y), (x', y')) := d(x, x') \vee e(y, y')$, $c_1((x, y), (x', y')) := d(x, x') + e(y, y')$, y $c_2((x, y), (x', y')) := (d(x, x')^2 + e(y, y')^2)^{1/2}$ son equivalentes y generan la topología producto.^a

^aSe tiene $c_\infty \leq c_2 \leq c_1 \leq 2c_\infty$, de modo que las tres métricas generan la misma topología; y las bolas de c_∞ son productos de bolas, base de la topología producto.

Advertencia 1.49

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.

Lema 1.50. Si D y E son espacios métricos separables, entonces $\mathcal{B}(D) \otimes \mathcal{B}(E) = \mathcal{B}(D \times E)$.

Demostración. Sean $\mathcal{G}_D, \mathcal{G}_E$ bases numerables de D y E .⁵³ Entonces

$$\mathcal{G} := \{G_D \times G_E : G_D \in \mathcal{G}_D, G_E \in \mathcal{G}_E\}$$

es una base numerable de $D \times E$. Por tanto, todo abierto de $D \times E$ es unión numerable de rectángulos abiertos, y así $\mathcal{B}(D \times E) \subset \sigma\{G_D \times G_E : G_D \subset D, G_E \subset E \text{ abiertos}\} \subset \mathcal{B}(D) \otimes \mathcal{B}(E)$.

Para la inclusión opuesta, las proyecciones $\pi_D : D \times E \rightarrow D$ y $\pi_E : D \times E \rightarrow E$ son continuas. Luego, si $A \in \mathcal{B}(D)$ y $B \in \mathcal{B}(E)$, $A \times B = \pi_D^{-1}(A) \cap \pi_E^{-1}(B) \in \mathcal{B}(D \times E)$. Como $\mathcal{B}(D) \otimes \mathcal{B}(E)$ está generada por tales rectángulos, $\mathcal{B}(D) \otimes \mathcal{B}(E) \subset \mathcal{B}(D \times E)$. $\square$

Lema 1.51. Sea $\mathbb{L}$ una probabilidad de Borel separable en $\mathcal{X} \times \mathcal{Y}$. Entonces $\mathbb{L}$ queda unívocamente determinada por los valores

$$\int f(x)g(y)d\mathbb{L}(x, y),$$

cuando f, g recorren las funciones Lipschitz acotadas y no negativas en $\mathcal{X}$ y $\mathcal{Y}$, respectivamente.

⁵³Existen por separabilidad: en un espacio métrico separable, las bolas $B(x_j, q)$, con x_j en un denso numerable y $q \in \mathbb{Q}^+$, forman una base numerable.

1.4.2. Tensión y medibilidad en productos finitos

Lema 1.52. 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 1.53. 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 1.54. Sea $\mathbb{L}$ una probabilidad de Borel separable en $\mathcal{X} \times \mathcal{Y}$. Si, para toda f, g Lipschitz acotadas y no negativas,

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

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

Ejemplo 1.55 (Slutsky). Si $X_\alpha \Rightarrow X$ (con X separable) y $Y_\alpha \Rightarrow c$ (constante), entonces $(X_\alpha, Y_\alpha) \Rightarrow (X, c)$.⁵⁴ 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:

- 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.

1.4.3. Productos numerables

Definición 1.56

Sea $\{(\mathcal{X}_i, d_i)\}_{i \geq 1}$ una familia de espacios métricos. En $\prod_{i \geq 1} \mathcal{X}_i$, definimos

$$d_\infty((x_i), (z_i)) := \sup_{i \geq 1} 2^{-i} (d_i(x_i, z_i) \wedge 1),$$

$$d_1((x_i), (z_i)) := \sum_{i \geq 1} 2^{-i} (d_i(x_i, z_i) \wedge 1).$$

Estas métricas son topológicamente equivalentes y generan la topología producto.

Observación 1.57

En productos numerables, la tensión de cada coordenada equivale a la tensión conjunta; por otra parte si la red es tensa, la medibilidad asintótica de cada coordenada equivale a la conjunta.

Teorema 1.58. Sea $X_\alpha : \Omega_\alpha \rightarrow \prod_{i \geq 1} \mathcal{X}_i$ una red arbitraria y sea X una aplicación Borel-medible, con valores en el producto, cuya ley es separable. Entonces

$$X_\alpha \Rightarrow X \iff (X_{\alpha,1}, \dots, X_{\alpha,m}) \Rightarrow (X_1, \dots, X_m) \text{ para todo } m \in \mathbb{N}.$$

⁵⁴Esbozo vía el Corolario 1.54: para f, g Lipschitz acotadas no negativas, $|\mathbb{E}_\alpha^*[f(X_\alpha)g(Y_\alpha)] - g(c)\mathbb{E}_\alpha^*f(X_\alpha)| \leq \|f\|_\infty \mathbb{E}_\alpha^*[g(Y_\alpha) - g(c)] \rightarrow 0$, usando subaditividad, que $Y_\alpha \Rightarrow c$ equivale a $Y_\alpha \xrightarrow{\mathbb{P}^*} c$ (véase el lema de la sección de Convergencias) y la continuidad de g en c . La ley límite, producto de la ley de X y δ_c , es separable por serlo la de X .

1.5. Espacios de funciones acotadas

Definición 1.59. Sea T un conjunto arbitrario. Definimos

$$\ell^\infty(T) := \left\{ z : T \rightarrow \mathbb{R} : \sup_{t \in T} |z(t)| < \infty \right\}, \quad \|z\|_T := \sup_{t \in T} |z(t)|.$$

Con la distancia uniforme $d(z_1, z_2) = \|z_1 - z_2\|_T$, $\ell^\infty(T)$ es un espacio métrico.

Observación 1.60. Se deja como ejercicio verificar que $\ell^\infty(T)$ es separable si y sólo si T es finito.

Definición 1.61. 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 $X(t) : \Omega \rightarrow \mathbb{R}$ 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

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

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 $X : \Omega \rightarrow D$ es medible como aplicación con valores en D si y solo si lo es como aplicación con valores en $\ell^\infty(T)$.

Si, además, D es cerrado, tampoco cambia la noción de convergencia débil: cuando X_n y X toman valores en D , se tiene $X_n \rightsquigarrow X$ en D si y solo si $X_n \rightsquigarrow X$ en $\ell^\infty(T)$.⁵⁵

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 1.62

Sea (T, ρ) un espacio semimétrico compacto. Por ejemplo, T puede ser un intervalo compacto de la recta real o la recta real extendida $[-\infty, \infty]$, provista de la métrica $\rho(s, t) = |\arctan s - \arctan t|$, donde $\arctan(\pm\infty) = \pm\pi/2$. Denotemos por $C(T)$ el conjunto de las funciones ρ -continuas $z : T \rightarrow \mathbb{R}$, equipado con la norma uniforme $\|z\|_T := \sup_{t \in T} |z(t)|$. Como T es compacto, toda función continua sobre T es acotada, de modo que $C(T) \subset \ell^\infty(T)$.

El espacio $C(T)$ es un subespacio cerrado y separable de $\ell^\infty(T)$; en particular, es completo.^a

La σ -álgebra de Borel de $C(T)$ coincide con la σ -álgebra cilíndrica, es decir,

$$\mathcal{B}(C(T)) = \sigma(\pi_t : t \in T), \quad \pi_t(z) := z(t).$$

En efecto, para cada $t \in T$, la proyección coordenada π_t es continua, pues

$$|\pi_t(z) - \pi_t(w)| = |z(t) - w(t)| \leq \|z - w\|_T.$$

Por tanto, $\sigma(\pi_t : t \in T) \subset \mathcal{B}(C(T))$.

Para probar la inclusión opuesta, observemos primero que, por la compacidad semimétrica de T , existe un subconjunto numerable y denso $T_0 \subset T$. Si $z, z_0 \in C(T)$, la función $t \mapsto |z(t) - z_0(t)|$ es continua y, en consecuencia,

$$\|z - z_0\|_T = \sup_{t \in T} |z(t) - z_0(t)| = \sup_{t \in T_0} |z(t) - z_0(t)|.$$

⁵⁵Una de las implicaciones es inmediata: toda función continua y acotada sobre $\ell^\infty(T)$ se restringe a una función continua y acotada sobre D . Para la recíproca, como D es cerrado y $\ell^\infty(T)$ es un espacio métrico, el teorema de extensión de Tietze asegura que toda función continua y acotada sobre D admite una extensión continua y acotada a $\ell^\infty(T)$. Por tanto, ambas convergencias se caracterizan mediante la misma clase de funciones de prueba.

Por tanto, para todo $r \geq 0$, la bola cerrada de centro z_0 y radio r puede escribirse como

$$\overline{B}(z_0, r) = \bigcap_{t \in T_0} \left\{ z \in C(T) : |z(t) - z_0(t)| \leq r \right\}.$$

Cada conjunto de esta intersección depende de una sola coordenada y es, por tanto, un conjunto cilíndrico. Como T_0 es numerable, la bola cerrada $\overline{B}(z_0, r)$ pertenece a $\sigma(\pi_t : t \in T)$.^b

Las bolas abiertas también pertenecen a la σ -álgebra cilíndrica, pues, para $r > 0$,

$$B(z_0, r) = \bigcup_{\substack{q \in \mathbb{Q} \\ 0 \leq q < r}} \overline{B}(z_0, q).$$

Finalmente, como $C(T)$ es separable, posee una base numerable formada por bolas abiertas. Todo conjunto abierto es, por tanto, una unión numerable de bolas abiertas y pertenece a la σ -álgebra cilíndrica. Esto prueba que $\mathcal{B}(C(T)) \subset \sigma(\pi_t : t \in T)$, y concluye la igualdad.

En particular, para toda aplicación $X : \Omega \rightarrow C(T)$,

X es medible respecto de $\mathcal{B}(C(T)) \iff X(t) := \pi_t \circ X : \Omega \rightarrow \mathbb{R}$ es medible para todo $t \in T$.

En efecto, la implicación directa se sigue de la continuidad de cada π_t . Recíprocamente, si todas las coordenadas $X(t) = \pi_t \circ X$ son medibles, entonces X es medible respecto de la σ -álgebra generada por las proyecciones, que acabamos de probar que coincide con $\mathcal{B}(C(T))$.

^aPara ver que es cerrado, sea $(z_n) \subset C(T)$ una sucesión que converge uniformemente a $z \in \ell^\infty(T)$. Como el límite uniforme de funciones continuas es continuo, $z \in C(T)$. Por tanto, $C(T)$ es cerrado en el espacio completo $\ell^\infty(T)$, y en consecuencia también es completo. Para probar la separabilidad, considérese la relación $s \sim t$ si $\rho(s, t) = 0$. El cociente $\tilde{T} := T/\sim$, provisto de la métrica inducida por ρ , es un espacio métrico compacto. Toda función ρ -continua sobre T es constante en las clases de equivalencia y, por tanto, factoriza de manera única a través de $\tilde{T}$. Esto identifica isométricamente $C(T)$ con $C(\tilde{T})$. Como $\tilde{T}$ es un espacio métrico compacto, $C(\tilde{T})$ es separable.

^bEn general, la bola cerrada no es ella misma un conjunto cilíndrico, pues no depende de una cantidad finita de coordenadas. Lo correcto es que pertenece a la σ -álgebra generada por los conjuntos cilíndricos, ya que es una intersección numerable de ellos.

Lema 1.63. Sea $X_\alpha : \Omega_\alpha \rightarrow \ell^\infty(T)$ 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. Para $t \in T$, sea $\pi_t : \ell^\infty(T) \rightarrow \mathbb{R}$ la evaluación $\pi_t(z) = z(t)$. Esta aplicación es continua, de hecho $|\pi_t(z_1) - \pi_t(z_2)| \leq \|z_1 - z_2\|_T$.

Supongamos que X_α es asintóticamente medible como aplicación con valores en $\ell^\infty(T)$. Si $\varphi \in C_b(\mathbb{R})$, entonces $\varphi \circ \pi_t \in C_b(\ell^\infty(T))$. Por lo tanto

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

Así, para todo $t \in T$, la coordenada $X_\alpha(t)$ es asintóticamente medible.

Supongamos ahora que $X_\alpha(t)$ es asintóticamente medible para todo $t \in T$. Como X_α es asintóticamente tensa en $\ell^\infty(T)$ y cada π_t es continua, cada coordenada $X_\alpha(t)$ es asintóticamente tensa.⁵⁶ Fijados $t_1, \dots, t_k \in T$, consideremos

$$Y_\alpha^{(t_1, \dots, t_k)} := (X_\alpha(t_1), \dots, X_\alpha(t_k)).$$

Por el Lema 1.53 (productos finitos),⁵⁷ $Y_\alpha^{(t_1, \dots, t_k)}$ es asintóticamente medible. En consecuencia, para toda $g \in C_b(\mathbb{R}^k)$,

$$\mathbb{E}_\alpha^* g(X_\alpha(t_1), \dots, X_\alpha(t_k)) - \mathbb{E}_{\alpha,*} g(X_\alpha(t_1), \dots, X_\alpha(t_k)) \rightarrow 0.$$

⁵⁶Si $K \subset \ell^\infty(T)$ es compacto, $\pi_t(K) \subset \mathbb{R}$ es compacto y, al ser π_t 1-Lipschitz, $\{X_\alpha \in K^\delta\} \subset \{X_\alpha(t) \in (\pi_t(K)^\delta)\}$; la monotonía de $\mathbb{P}_{\alpha,*}$ transfiere la cota de tensión.

⁵⁷Enunciado para dos factores, el caso de k factores sale por inducción, tratando $(X_\alpha(t_1), \dots, X_\alpha(t_{k-1}))$ como una sola coordenada y usando el Lema 1.52 para la tensión asintótica del vector.

Sea $\mathcal{C}$ la clase de funciones cilindro

$$f(z) = g(z(t_1), \dots, z(t_k)), \quad g \in C_b(\mathbb{R}^k), \quad t_1, \dots, t_k \in T.$$

Entonces $\mathcal{C} \subset C_b(\ell^\infty(T))$. Además, $\mathcal{C}$ es una subálgebra y un retículo vectorial, contiene las constantes y separa puntos. En efecto, las constantes se obtienen tomando $g \equiv c$. Si $f_1, f_2 \in \mathcal{C}$, reuniendo en una sola lista finita las coordenadas que aparecen en ambas funciones, se ve que $af_1 + bf_2$, f_1f_2 , $f_1 \vee f_2$ y $f_1 \wedge f_2$ vuelven a ser funciones cilindro, porque suma, producto, máximo y mínimo son operaciones continuas en dimensión finita. Finalmente, si $z_1 \neq z_2$ en $\ell^\infty(T)$, existe $t \in T$ tal que $z_1(t) \neq z_2(t)$; tomando $h(x) = \arctan x$,⁵⁸ la función $f(z) = h(z(t))$ pertenece a $\mathcal{C}$ y separa z_1 de z_2 .

Por lo probado para vectores finito-dimensionales,

$$\mathbb{E}_\alpha^* f(X_\alpha) - \mathbb{E}_{\alpha,*} f(X_\alpha) \rightarrow 0 \quad \text{para toda } f \in \mathcal{C}.$$

Como X_α es asintóticamente tensa en $\ell^\infty(T)$, el Lema 1.47, aplicado con $D = \ell^\infty(T)$ y $\mathcal{F} = \mathcal{C}$, implica que

$$\mathbb{E}_\alpha^* f(X_\alpha) - \mathbb{E}_{\alpha,*} f(X_\alpha) \rightarrow 0 \quad \text{para toda } f \in C_b(\ell^\infty(T)).$$

Por definición, X_α es asintóticamente medible. □

Observación 1.64

Si T fuese numerable, la Observación en “Productos numerables” implicaría directamente que, bajo tensión conjunta, la medibilidad asintótica por coordenadas equivale a la conjunta en el producto $\mathbb{R}^T$; el argumento anterior muestra cómo obtener el mismo resultado en $\ell^\infty(T)$ para T arbitrario pasando por funciones cilindro.

Lema 1.65. Sean $\mathbb{L}_1$ y $\mathbb{L}_2$ probabilidades de Borel tensas en $\ell^\infty(T)$. Entonces son equivalentes:

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

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

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

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

Demostración. Claramente (1) implica (2), porque las proyecciones

$$\pi_{t_1, \dots, t_k} : \ell^\infty(T) \rightarrow \mathbb{R}^k, \quad \pi_{t_1, \dots, t_k}(z) = (z(t_1), \dots, z(t_k)),$$

son continuas y, en particular, Borel-medibles.

Veamos que (2) implica (1). Consideremos la clase de funciones

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

Como $\pi_{t_1, \dots, t_k}$ es continua, toda $f \in \mathcal{F}$ pertenece a $C_b(\ell^\infty(T))$.

Por la hipótesis de igualdad de todas las marginales finito-dimensionales, para todo $k \in \mathbb{N}$, todo $t_1, \dots, t_k \in T$ y toda $g \in C_b(\mathbb{R}^k)$,

$$\int g d(\mathbb{L}_1 \circ \pi_{t_1, \dots, t_k}^{-1}) = \int g d(\mathbb{L}_2 \circ \pi_{t_1, \dots, t_k}^{-1}).$$

⁵⁸Se compone con $\arctan$ sólo para que la función sea acotada (la proyección $z \mapsto z(t)$ no lo es); la inyectividad de $\arctan$ garantiza que siga separando.

Por lo tanto, para toda $f \in \mathcal{F}$,

$$\int f d\mathbb{L}_1 = \int f d\mathbb{L}_2.$$

La clase $\mathcal{F}$ es una subálgebra y un retículo vectorial de $C_b(\ell^\infty(T))$, y contiene las funciones constantes. Además, separa puntos: si $z_1 \neq z_2$, existe $t \in T$ tal que $z_1(t) \neq z_2(t)$; entonces $f(z) = \arctan(z(t))$ pertenece a $\mathcal{F}$ y satisface $f(z_1) \neq f(z_2)$.

Como $\mathbb{L}_1$ y $\mathbb{L}_2$ son probabilidades de Borel tensas, el punto (ii) del Lema 1.45 aplicado en $D = \ell^\infty(T)$ implica que $\mathbb{L}_1 = \mathbb{L}_2$ en $\mathcal{B}(\ell^\infty(T))$.

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

Teorema 1.66. *Sea $X_\alpha : \Omega_\alpha \rightarrow \ell^\infty(T)$ una red arbitraria. Entonces X_α converge débilmente hacia una probabilidad de Borel tensa $\mathbb{L}$ en $\ell^\infty(T)$ si y sólo si:*

1. X_α es asintóticamente tensa;
2. para toda elección finita $t_1, \dots, t_k \in T$, existe una ley $\mathbb{L}_{t_1, \dots, t_k}$ en $\mathbb{R}^k$ tal que

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

En tal caso, $\mathbb{L}_{t_1, \dots, t_k}$ es la marginal finito-dimensional de $\mathbb{L}$ correspondiente a $t_1, \dots, t_k$; es decir,

$$\mathbb{L}_{t_1, \dots, t_k} = \mathbb{L} \circ (\pi_{t_1}, \dots, \pi_{t_k})^{-1},$$

donde $\pi_t(z) = z(t)$.

Más aún, supongamos que se cumplen (i) y (ii) y que las leyes límite $\mathbb{L}_{t_1, \dots, t_k}$ son las distribuciones finito-dimensionales de un proceso $X = \{X(t) : t \in T\}$. Entonces existe un proceso $\widetilde{X} = \{\widetilde{X}(t) : t \in T\}$, posiblemente definido en otro espacio de probabilidad, con trayectorias en $\ell^\infty(T)$, cuyas distribuciones finito-dimensionales coinciden con las de X , y tal que $X_\alpha \Rightarrow \widetilde{X}$ en $\ell^\infty(T)$.

Demostración. Para $t \in T$, denotemos por $\pi_t : \ell^\infty(T) \rightarrow \mathbb{R}$, $\pi_t(z) = z(t)$, la proyección de evaluación. Más generalmente, para $t_1, \dots, t_k \in T$, escribimos

$$\pi_{t_1, \dots, t_k}(z) = (z(t_1), \dots, z(t_k)).$$

Estas aplicaciones son continuas.

Supongamos que $X_\alpha \Rightarrow X$ en $\ell^\infty(T)$ y que la ley de X es tensa. Entonces X_α es asintóticamente medible y asintóticamente tensa.⁵⁹ Además, como $\pi_{t_1, \dots, t_k}$ es continua, el teorema de la aplicación continua da

$$(X_\alpha(t_1), \dots, X_\alpha(t_k)) = \pi_{t_1, \dots, t_k}(X_\alpha) \Rightarrow \pi_{t_1, \dots, t_k}(X).$$

Por lo tanto, todas las marginales finito-dimensionales convergen.

Supongamos ahora que valen (i) y (ii). Para cada $t \in T$, la convergencia débil $X_\alpha(t) \Rightarrow \mathbb{L}_t$ en $\mathbb{R}$ implica que $X_\alpha(t)$ es asintóticamente medible.⁶⁰ Como X_α es asintóticamente tensa en $\ell^\infty(T)$, el Lema 1.63 implica que X_α es asintóticamente medible como mapa con valores en $\ell^\infty(T)$.

Por el teorema de Prokhorov,⁶¹ toda subred de X_α contiene una sub-subred que converge débilmente hacia alguna probabilidad de Borel tensa en $\ell^\infty(T)$. Sea $X_{\alpha(\beta)} \Rightarrow \mathbb{L}$ una de esas subredes convergentes.

⁵⁹Por el Lema 1.41: la parte (a) da la medibilidad asintótica, y la parte (b), junto con la tensión de la ley de X , la tensión asintótica.

⁶⁰Lema 1.41(a), aplicado en $\mathbb{R}$ con la variable canónica de ley $\mathbb{L}_t$ (el caso $k = 1$ de la hipótesis (ii)) como límite.

⁶¹Toda subred hereda la tensión y la medibilidad asintóticas: los $\limsup$ y $\liminf$ involucrados sólo mejoran al pasar a subredes. Por eso Prokhorov (i) puede aplicarse a cada subred.

Fijemos $t_1, \dots, t_k \in T$. Por el teorema de la aplicación continua,

$$\pi_{t_1, \dots, t_k}(X_{\alpha(\beta)}) \Rightarrow \mathbf{L} \circ \pi_{t_1, \dots, t_k}^{-1}.$$

Pero, por la hipótesis (ii), la misma subred satisface

$$\pi_{t_1, \dots, t_k}(X_{\alpha(\beta)}) = (X_{\alpha(\beta)}(t_1), \dots, X_{\alpha(\beta)}(t_k)) \Rightarrow \mathbf{L}_{t_1, \dots, t_k}.$$

Por unicidad del límite débil en $\mathbb{R}^k$, $\mathbf{L} \circ \pi_{t_1, \dots, t_k}^{-1} = \mathbf{L}_{t_1, \dots, t_k}$. Así, las marginales finito-dimensionales de cualquier punto límite tenso están determinadas por las medidas $\mathbf{L}_{t_1, \dots, t_k}$.

Veamos ahora que hay un único punto límite. Sean $\mathbf{L}'$ y $\mathbf{L}''$ dos puntos límite tensos de la red. Por el argumento anterior, para toda elección finita $t_1, \dots, t_k \in T$,

$$\mathbf{L}' \circ \pi_{t_1, \dots, t_k}^{-1} = \mathbf{L}_{t_1, \dots, t_k} = \mathbf{L}'' \circ \pi_{t_1, \dots, t_k}^{-1}.$$

Por el Lema 1.65 (en su forma para medidas, aplicado a las probabilidades tensas $\mathbf{L}'$ y $\mathbf{L}''$), obtenemos $\mathbf{L}' = \mathbf{L}''$ en $\mathcal{B}(\ell^\infty(T))$. Por lo tanto todos los puntos límite son iguales. Como, por Prokhorov, toda subred tiene una sub-subred convergente, se concluye que la red completa converge débilmente hacia esa única probabilidad de Borel tensa $\mathbf{L}$.⁶² Esto prueba la equivalencia, y muestra que las $\mathbf{L}_{t_1, \dots, t_k}$ son las marginales finito-dimensionales de $\mathbf{L}$.

Para la última afirmación, supongamos que valen (i) y (ii) y que las medidas límite $\mathbf{L}_{t_1, \dots, t_k}$ son las distribuciones finito-dimensionales de un proceso $X = \{X(t) : t \in T\}$. Por la equivalencia ya demostrada, la hipótesis (i)–(ii) garantiza que X_α converge débilmente hacia una probabilidad de Borel tensa $\mathbf{L}$ en $\ell^\infty(T)$, cuyas marginales finito-dimensionales son precisamente

$$\mathbf{L} \circ \pi_{t_1, \dots, t_k}^{-1} = \mathbf{L}_{t_1, \dots, t_k} = \mathcal{L}(X(t_1), \dots, X(t_k)).$$

(En particular, la existencia de la $\mathbf{L}$ tensa proviene de (i)–(ii) y no meramente de disponer de un proceso X con esas marginales.)

Tomemos el espacio canónico $(\ell^\infty(T), \mathcal{B}(\ell^\infty(T)), \mathbf{L})$ y definamos $\widetilde{X}(z) = z$. Entonces $\widetilde{X}$ toma valores en $\ell^\infty(T)$, luego tiene trayectorias uniformemente acotadas. Además, para toda elección finita $t_1, \dots, t_k \in T$,

$$(\widetilde{X}(t_1), \dots, \widetilde{X}(t_k)) \stackrel{d}{=} (X(t_1), \dots, X(t_k)),$$

ya que la ley de $\widetilde{X}$ es $\mathbf{L}$ y $\mathbf{L} \circ \pi_{t_1, \dots, t_k}^{-1} = \mathcal{L}(X(t_1), \dots, X(t_k))$. Es decir, $\widetilde{X}$ es una versión de X en el sentido de tener las mismas distribuciones finito-dimensionales. Como la ley de $\widetilde{X}$ es $\mathbf{L}$ y $X_\alpha \Rightarrow \mathbf{L}$, concluimos $X_\alpha \Rightarrow \widetilde{X}$ en $\ell^\infty(T)$. $\square$

La convergencia de las marginales puede establecerse mediante cualquiera de los métodos bien conocidos para probar convergencia débil en espacios euclídeos. 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

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

⁶²Se usa aquí el siguiente hecho general: una red en un espacio topológico converge a un punto x si y sólo si toda subred posee una subred ulterior que converge a x . Se aplica, para cada $f \in C_b(\ell^\infty(T))$, a la red numérica $\mathbb{E}_\alpha^* f(X_\alpha)$: toda subred de X_α tiene, por Prokhorov, una sub-subred débilmente convergente, cuyo límite es necesariamente $\mathbf{L}$ por la unicidad recién probada; luego toda subred de $\mathbb{E}_\alpha^* f(X_\alpha)$ tiene una subred ulterior que converge a $\int f d\mathbf{L}$, y por lo tanto $\mathbb{E}_\alpha^* f(X_\alpha) \rightarrow \int f d\mathbf{L}$.

tal que

$$\limsup_{\alpha} \mathbb{P}_{\alpha}^* \left(\sup_{1 \leq i \leq k} \sup_{s, t \in T_i} |X_{\alpha}(s) - X_{\alpha}(t)| > \varepsilon \right) < \eta, \quad (1.3)$$

donde $\mathbb{P}_{\alpha}^*$ denota probabilidad exterior.

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 $(X_{\alpha}(t_1), \dots, X_{\alpha}(t_k))$ 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 1.67. *Una red $X_{\alpha} : \Omega_{\alpha} \rightarrow \ell^{\infty}(T)$ es asintóticamente tensa si y sólo si*

1. $X_{\alpha}(t)$ es asintóticamente tensa en $\mathbb{R}$ para todo $t \in T$, y
2. para todo $\varepsilon, \eta > 0$ existe una partición finita $T = \bigcup_{i=1}^k T_i$ tal que se verifica (1.3).

Este resultado es el Teorema 1.5.6 de [vdV–W]. En el libro, la necesidad se deduce del criterio de equicontinuidad (Teorema 1.70, más adelante), tomando la partición que se obtiene al disjuntar bolas de una semimétrica que hace a T totalmente acotado, y sólo la suficiencia se demuestra directamente; aquí incluimos, además, una demostración directa de la necesidad.

Demostración. Supongamos que X_{α} es asintóticamente tensa en $\ell^{\infty}(T)$. Como $\pi_t : \ell^{\infty}(T) \rightarrow \mathbb{R}$, $\pi_t(z) = z(t)$, es continua, la tensión asintótica se preserva por π_t ⁶³ y, por tanto, $X_{\alpha}(t)$ es asintóticamente tensa para todo $t \in T$.

Probemos la condición de particiones. Sean $\varepsilon, \eta > 0$. Por tensión asintótica, aplicada con nivel $\eta/2$, existe un compacto $K \subset \ell^{\infty}(T)$ tal que

$$\limsup_{\alpha} \mathbb{P}_{\alpha}^*(X_{\alpha} \notin K^{\varepsilon/6}) \leq \frac{\eta}{2} < \eta.$$

Como K es totalmente acotado, existen $z_1, \dots, z_N \in K$ tales que $K \subset \bigcup_{j=1}^N B(z_j, \varepsilon/6)$. Para cada j , cubrimos el rango acotado $z_j(T)$ por finitos intervalos $I_{j,1}, \dots, I_{j,L_j}$ de longitud $\leq \varepsilon/3$, y tomamos la partición común de T dada por las intersecciones no vacías

$$\bigcap_{j=1}^N z_j^{-1}(I_{j,\ell_j}), \quad 1 \leq \ell_j \leq L_j,$$

renumeradas como $T_1, \dots, T_k$ (los intervalos pueden tomarse disjuntos dos a dos para cada j , de modo que las intersecciones formen efectivamente una partición). Así obtenemos una partición finita $T = \bigcup_{i=1}^k T_i$ tal que

$$\max_{1 \leq j \leq N} \sup_{s, t \in T_i} |z_j(s) - z_j(t)| \leq \varepsilon/3, \quad i = 1, \dots, k.$$

Si $X_{\alpha} \in K^{\varepsilon/6}$, existen $z \in K$ y j tales que $\|X_{\alpha} - z\|_T < \varepsilon/6$ y $\|z - z_j\|_T < \varepsilon/6$, luego $\|X_{\alpha} - z_j\|_T < \varepsilon/3$. Por tanto, para $s, t \in T_i$,

$$|X_{\alpha}(s) - X_{\alpha}(t)| \leq 2\|X_{\alpha} - z_j\|_T + |z_j(s) - z_j(t)| < \varepsilon.$$

Luego

$$\left\{ \sup_i \sup_{s, t \in T_i} |X_{\alpha}(s) - X_{\alpha}(t)| > \varepsilon \right\} \subset \{X_{\alpha} \notin K^{\varepsilon/6}\},$$

y tomando $\limsup_{\alpha} \mathbb{P}_{\alpha}^*$ se obtiene (1.3).

⁶³Por el mismo argumento usado en la prueba del Lema 1.63 para las proyecciones coordenadas (π_t es 1-Lipschitz).

Supongamos ahora que valen 1 y 2. Primero vemos que $\|X_\alpha\|_T$ es asintóticamente tensa. Sea $\zeta > 0$. Por 2, con $\varepsilon = 1$ y $\eta = \zeta/2$, existe una partición finita $T = \bigcup_{i=1}^k T_i$ tal que

$$\limsup_{\alpha} \mathbb{P}_{\alpha}^* \left(\sup_i \sup_{s,t \in T_i} |X_{\alpha}(s) - X_{\alpha}(t)| > 1 \right) < \zeta/2.$$

Eligiendo $t_i \in T_i$ y escribiendo $Y_{\alpha} = \max_{1 \leq i \leq k} |X_{\alpha}(t_i)|$, la tensión asintótica de las coordenadas y la finitud de k implican que Y_{α} es asintóticamente tensa.⁶⁴ Aplicando la tensión asintótica de Y_{α} con nivel $\zeta/4$, tomamos M tal que

$$\limsup_{\alpha} \mathbb{P}_{\alpha}^*(Y_{\alpha} > M - 1) \leq \frac{\zeta}{4} < \frac{\zeta}{2}.$$

En el evento de oscilación ≤ 1 se cumple $\|X_{\alpha}\|_T \leq Y_{\alpha} + 1$,⁶⁵ y por tanto

$$\limsup_{\alpha} \mathbb{P}_{\alpha}^*(\|X_{\alpha}\|_T > M) < \zeta.$$

Fijemos ahora $\gamma > 0$. Aplicando la tensión asintótica de $\|X_{\alpha}\|_T$, ya establecida, con nivel $\gamma/4$, tomamos M tal que

$$\limsup_{\alpha} \mathbb{P}_{\alpha}^*(\|X_{\alpha}\|_T > M) \leq \frac{\gamma}{4} < \frac{\gamma}{2}.$$

Sea $r_m \downarrow 0$ y $\eta_m = \gamma 2^{-m-1}$. Para cada m , por 2 existe una partición finita $T = \bigcup_{i=1}^{k(m)} T_i^{(m)}$ tal que

$$\limsup_{\alpha} \mathbb{P}_{\alpha}^* \left(\sup_i \sup_{s,t \in T_i^{(m)}} |X_{\alpha}(s) - X_{\alpha}(t)| > r_m/3 \right) < \eta_m.$$

Definimos

$$\mathcal{G}_m = \{jr_m/3 : |j| \leq \lceil 3M/r_m \rceil\}$$

y sea $\mathcal{Z}_m$ el conjunto finito de funciones constantes en cada bloque $T_i^{(m)}$ y con valores en $\mathcal{G}_m$. Ponemos

$$K_m = \bigcup_{z \in \mathcal{Z}_m} \overline{B}(z, r_m), \quad K = \bigcap_{m \geq 1} K_m.$$

Cada K_m es cerrado, luego K es cerrado. Además, K es totalmente acotado: si $\delta > 0$, elegimos m con $r_m < \delta$; como $K \subset K_m$ y K_m es unión finita de bolas de radio r_m , K se cubre por finitas bolas de radio δ . Como $\ell^\infty(T)$ es completo, K es compacto.⁶⁶

Veamos ahora la inclusión clave. Para todo $\delta > 0$ existe $m(\delta)$ tal que $\bigcap_{i=1}^{m(\delta)} K_i \subset K^\delta$ (con la lectura $K^\delta = \emptyset$ si $K = \emptyset$, de modo que la expresión no requiere dar sentido a la distancia al conjunto vacío). Si no, existirían $\delta > 0$ y

$$x_m \in \left(\bigcap_{i=1}^m K_i \right) \setminus K^\delta, \quad m \geq 1.$$

Fijado ℓ , para todo $m \geq \ell$ se tiene $x_m \in K_\ell$, y K_ℓ es unión finita de bolas de radio r_ℓ ; pasando sucesivamente a subsucesiones y diagonalizando, obtenemos una subsucesión (x_{m_q}) tal que, para cada ℓ , sus términos eventualmente pertenecen a una misma bola de radio r_ℓ . Por tanto, para q, q' grandes, $\|x_{m_q} - x_{m_{q'}}\|_T \leq 2r_\ell$; como $r_\ell \downarrow 0$, la subsucesión es de Cauchy. Sea x su límite. Para cada ℓ ,

⁶⁴Para cada i , la tensión asintótica de $X_{\alpha}(t_i)$ da un compacto $K_i \subset \mathbb{R}$ al nivel $\zeta/(4k)$; elíjase después $M_i > 1 + \sup_{x \in K_i} |x|$. Como $K_i^1 \subset [-M_i, M_i]$, se tiene $\{|X_{\alpha}(t_i)| > M_i\} \subset \{X_{\alpha}(t_i) \notin K_i^1\}$ y, por la formulación dual de la tensión, $\limsup_{\alpha} \mathbb{P}_{\alpha}^*(|X_{\alpha}(t_i)| > M_i) \leq \zeta/(4k)$. Con $M' = \max_i M_i$ y la subaditividad de $\mathbb{P}^*$ sobre los k eventos, $\limsup_{\alpha} \mathbb{P}_{\alpha}^*(Y_{\alpha} > M') \leq \zeta/4$.

⁶⁵Para $s \in T_i$, $|X_{\alpha}(s)| \leq |X_{\alpha}(s) - X_{\alpha}(t_i)| + |X_{\alpha}(t_i)| \leq 1 + Y_{\alpha}$. Así, $\{\|X_{\alpha}\|_T > M\} \subset \{Y_{\alpha} > M - 1\} \cup \{\text{oscilación} > 1\}$, y se concluye por subaditividad de $\mathbb{P}^*$.

⁶⁶ $\ell^\infty(T)$ es completo: una sucesión uniformemente de Cauchy converge puntualmente, el límite es acotado y la convergencia es uniforme. Y en un espacio métrico completo, un conjunto cerrado y totalmente acotado es compacto.

eventualmente $x_{m_q} \in K_\ell$ y K_ℓ es cerrado, luego $x \in K_\ell$. Así $x \in K$ (en particular, $K \neq \emptyset$ en este escenario). Pero entonces, como $x_{m_q} \rightarrow x \in K$, se tiene $x_{m_q} \in K^\delta$ para q suficientemente grande, contradiciendo la elección de x_{m_q} .

Finalmente, para m fijo,

$$\left\{ X_\alpha \notin \bigcap_{i=1}^m K_i \right\} \subset \{ \|X_\alpha\|_T > M \} \cup \bigcup_{i=1}^m \left\{ \sup_j \sup_{s,t \in T_j^{(i)}} |X_\alpha(s) - X_\alpha(t)| > r_i/3 \right\}.$$

En efecto, supongamos que $\|X_\alpha\|_T \leq M$ y que todas esas oscilaciones son $\leq r_i/3$. Fijado i , elijamos $t_j^{(i)} \in T_j^{(i)}$ en cada bloque y sea $z \in \mathcal{Z}_i$ la función que en el bloque $T_j^{(i)}$ vale el punto de la grilla $\mathcal{G}_i$ más próximo a $X_\alpha(t_j^{(i)})$. Como $|X_\alpha(t_j^{(i)})| \leq M$, la grilla tiene paso $r_i/3$ y alcanza hasta $\lceil 3M/r_i \rceil r_i/3 \geq M$, el error de redondeo es $\leq r_i/6$. Entonces, para $s \in T_j^{(i)}$,

$$|X_\alpha(s) - z(s)| \leq |X_\alpha(s) - X_\alpha(t_j^{(i)})| + |X_\alpha(t_j^{(i)}) - z(s)| \leq \frac{r_i}{3} + \frac{r_i}{6} = \frac{r_i}{2} < r_i,$$

es decir, $\|X_\alpha - z\|_T < r_i$ y $X_\alpha \in K_i$. Por subaditividad de la probabilidad exterior,

$$\limsup_\alpha \mathbb{P}_\alpha^*(X_\alpha \notin \bigcap_{i=1}^m K_i) \leq \frac{\gamma}{4} + \sum_{i=1}^m \eta_i < \frac{\gamma}{4} + \frac{\gamma}{2} = \frac{3\gamma}{4} < \gamma.$$

Dado $\delta > 0$, elegimos $m = m(\delta)$. Como $\bigcap_{i=1}^{m(\delta)} K_i \subset K^\delta$, se sigue que $\limsup_\alpha \mathbb{P}_\alpha^*(X_\alpha \notin K^\delta) < \gamma$. Como $\gamma > 0$ era arbitrario, X_α es asintóticamente tensa en $\ell^\infty(T)$. $\square$

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. Recordemos primero la noción determinista correspondiente.

Definición 1.68

Sean (D, d) y (E, e) espacios métricos, y sea $\mathcal{F} \subset \{f : D \rightarrow E\}$. Decimos que $\mathcal{F}$ es **equicontinua en** $x_0 \in D$ si, para todo $\varepsilon > 0$, existe $\delta > 0$ tal que $d(x, x_0) < \delta$ implica $e(f(x), f(x_0)) < \varepsilon$ para toda $f \in \mathcal{F}$. Decimos que $\mathcal{F}$ es equicontinua en $A \subset D$ si lo es en cada punto de A .

Definición 1.69. Sea ρ una semimétrica en T . Una red $X_\alpha : \Omega_\alpha \rightarrow \ell^\infty(T)$ 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,

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

Teorema 1.70. Una red $X_\alpha : \Omega_\alpha \rightarrow \ell^\infty(T)$ 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.

Teorema 1.71. Supongamos que $X_\alpha \Rightarrow X$ en $\ell^\infty(T)$. Si X_α es asintóticamente tensa, entonces existe una semimétrica ρ en T tal que (T, ρ) es totalmente acotado y casi todas las trayectorias $t \mapsto X(t, \omega)$ son uniformemente ρ -continuas.

Más aún, si para alguna semimétrica ρ se tiene que (T, ρ) es totalmente acotado y casi todas las trayectorias de X son uniformemente ρ -continuas, entonces X_α es asintóticamente uniformemente ρ -equicontinua en probabilidad.

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, pero la convergencia débil en $\ell^\infty(T)$ exige además una condición de tensión. Por el Teorema 1.71, esa condición puede formularse buscando una semimétrica ρ que haga totalmente acotado a T y uniformemente continuas las trayectorias de X . De hecho, si existe alguna semimétrica con estas propiedades, la semimétrica intrínseca

$$\rho_0(s, t) = \mathbb{E} \arctan |X(s) - X(t)|$$

también las tiene, en el siguiente sentido: (T, ρ_0) es totalmente acotado y X admite una versión cuyas trayectorias son uniformemente ρ_0 -continuas casi seguramente.

Que ρ_0 es una semimétrica se comprueba así: la simetría y $\rho_0(t, t) = 0$ son inmediatas, y la desigualdad triangular se sigue de que $\arctan$ es creciente y subaditiva en $[0, \infty)$ —por ser cóncava y nula en el origen, $\arctan(a+b) \leq \arctan a + \arctan b$ —: aplicando ambas propiedades a $|X(s) - X(t)| \leq |X(s) - X(u)| + |X(u) - X(t)|$ y tomando esperanzas se obtiene $\rho_0(s, t) \leq \rho_0(s, u) + \rho_0(u, t)$.

La total acotación de (T, ρ_0) puede verificarse directamente. Sea $w(\delta) := \sup_{\rho(s, t) \leq \delta} |X(s) - X(t)|$. La función $z \mapsto \sup_{\rho(s, t) \leq \delta} |z(s) - z(t)|$ es 2-Lipschitz en $\ell^\infty(T)$ —cada incremento $|z(s) - z(t)|$ cambia a lo sumo en $2\|z - z'\|_T$ al pasar de z a z' —, de modo que $w(\delta)$, composición de esa función continua con la versión Borel-medible de X , es una variable aleatoria, y la continuidad uniforme de las trayectorias da $w(\delta) \downarrow 0$ c.s. cuando $\delta \downarrow 0$. Como $\arctan$ es creciente y acotada, la convergencia dominada da $\eta(\delta) := \mathbb{E} \arctan w(\delta) \rightarrow 0$. Si $\rho(s, t) \leq \delta$, entonces $\rho_0(s, t) \leq \eta(\delta)$; por lo tanto, toda δ -red finita para ρ es una $\eta(\delta)$ -red finita para ρ_0 , y la total acotación de (T, ρ) se transfiere a (T, ρ_0) .

La existencia de una versión con trayectorias uniformemente ρ_0 -continuas requiere más trabajo; la sutileza está en los pares con $\rho_0(s, t) = 0$, para los cuales $X(s) = X(t)$ sólo casi seguramente, con un conjunto nulo que depende del par. De todos modos, ρ_0 no siempre es la semimétrica más cómoda para las estimaciones.

Otra familia natural de candidatas es

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

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

Proposición 1.73. Sea $X = \{X(t) : t \in T\}$ un proceso gaussiano y, para $0 < p < \infty$, definamos

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

Sea $X_\alpha : \Omega_\alpha \rightarrow \ell^\infty(T)$ 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 $t_1, \dots, t_k \in T$, $(X_\alpha(t_1), \dots, X_\alpha(t_k)) \Rightarrow (X(t_1), \dots, X(t_k))$;
 - X_α es asintóticamente uniformemente ρ_p -equicontinuo en probabilidad;
 - (T, ρ_p) es totalmente acotado.

Observación 1.74

La proposición anterior no reemplaza al criterio general de convergencia en $\ell^\infty(T)$, sino que lo especializa al caso en que el posible límite es gaussiano. En el criterio general, además de la convergencia de las marginales finito-dimensionales, hay que encontrar alguna semimétrica ρ que haga a T totalmente acotado y respecto de la cual la red sea asintóticamente uniformemente equicontinua.

La hipótesis de gaussianidad permite elegir una semimétrica canónica, determinada por el propio límite:

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

En particular, suele tomarse

$$\rho_2(s, t) = \left(\mathbb{E}|X(s) - X(t)|^2 \right)^{1/2},$$

la semimétrica *cuadrática media* de los incrementos; cuando el proceso está centrado (o, más generalmente, es de media constante), ρ_2 coincide con la desviación estándar de los incrementos. La cláusula “para algún p , y entonces para todo p ” de la proposición se deduce de la equivalencia uniforme de las normas L^p sobre la familia de las distribuciones gaussianas reales: para cada $p > 0$ existen constantes $0 < a_p \leq b_p < \infty$, que dependen sólo de p , tales que, para toda variable gaussiana real Z , *no necesariamente centrada*,

$$a_p \left(\mathbb{E}Z^2 \right)^{1/2} \leq \left(\mathbb{E}|Z|^p \right)^{1/p} \leq b_p \left(\mathbb{E}Z^2 \right)^{1/2}.$$

La verificación es elemental. Con $Z = \mu + \sigma N$, $N \sim N(0, 1)$ y $m_q := \mathbb{E}|N|^q$: para la cota superior, por Lyapunov y la desigualdad triangular, $(\mathbb{E}|Z|^p)^{1/p} \leq (\mathbb{E}|Z|^{p \vee 1})^{1/(p \vee 1)} \leq |\mu| + m_{p \vee 1}^{1/(p \vee 1)} \sigma$; para la inferior, si $|\mu| \geq \sigma$ entonces $\mathbb{P}(|Z| \geq |\mu|/2) \geq \mathbb{P}(|N| \leq 1/2)$, y si $\sigma > |\mu|$ entonces $\mathbb{P}(|Z| \geq \sigma/2) \geq \mathbb{P}(|N| \geq 3/2)$; en ambos casos $\mathbb{E}|Z|^p \geq c 2^{-p} (|\mu| \vee \sigma)^p$ con $c > 0$ universal, y $(|\mu| \vee \sigma)^2 \geq \frac{1}{2} \mathbb{E}Z^2$. Aplicado a $Z = X(s) - X(t)$, esto da

$$a_p \rho_2 \leq \rho_p \leq b_p \rho_2 \quad (p \geq 1), \quad a_p^p \rho_2^p \leq \rho_p \leq b_p^p \rho_2^p \quad (0 < p < 1) :$$

todas las ρ_p quedan atrapadas entre transformaciones monótonas fijas de ρ_2 , de modo que la total acotación de (T, ρ_p) y la equicontinuidad asintótica respecto de ρ_p valen para algún p si y sólo si

valen para todo p . Si el proceso está centrado, las comparaciones se convierten en identidades exactas: con $m_p := \mathbb{E}|N(0, 1)|^p$,

$$\rho_p = m_p^{1/p} \rho_2 \quad (p \geq 1), \quad \rho_p = m_p \rho_2^p \quad (0 < p < 1).$$

Estas identidades *no* valen para procesos gaussianos con media no constante, y por eso el argumento general debe pasar por la equivalencia uniforme: por ejemplo, el proceso determinista $X(s) = 0$, $X(t) = 1$ sobre $T = \{s, t\}$ es gaussiano (degenerado) según la definición, su incremento tiene desviación estándar nula, y sin embargo $\rho_p(s, t) = 1$ para todo $p > 0$. Así, en el caso gaussiano, una vez verificada la convergencia de las marginales finito-dimensionales hacia las de X , basta elegir un mismo $p \in (0, \infty)$ y probar la total acotación de (T, ρ_p) y la equicontinuidad asintótica respecto de ρ_p . La elección usual es $p = 2$.

1.6. σ -álgebra generada por las bolas

Sea (D, d) un espacio métrico. La σ -álgebra de bolas en D es la menor σ -álgebra que contiene todas las bolas abiertas o, equivalentemente, todas las bolas cerradas. En general puede ser estrictamente más pequeña que la σ -álgebra de Borel: aunque todo abierto es unión de bolas abiertas, esa unión puede ser no numerable, y una σ -álgebra sólo es estable por uniones numerables. Si D es separable, en cambio, todo abierto es unión numerable de bolas con centros en un subconjunto denso numerable, y por eso ambas σ -álgebras coinciden.

En espacios no separables puede ocurrir que una aplicación $X : \Omega \rightarrow D$ no sea Borel-medible, pero sí sea medible respecto de la σ -álgebra de bolas. Esto motiva preguntarse si puede formularse una teoría de convergencia débil usando esa σ -álgebra más pequeña. La respuesta es afirmativa, pero no da lugar a una teoría esencialmente distinta: la clase de funciones $f \in C_b(D)$ que son medibles respecto de bolas es suficientemente rica para detectar convergencia, porque contiene funciones continuas acotadas como $x \mapsto d(x, s) \wedge 1$, que separan puntos; y, al mismo tiempo, esa riqueza permite verla como un caso particular de la teoría general de convergencia débil ya desarrollada mediante funciones test continuas y acotadas.

Más concretamente, las aplicaciones $x \mapsto d(x, s)$, $s \in D$, generan la σ -álgebra de bolas. Por tanto, una aplicación $X : \Omega \rightarrow D$ es medible respecto de bolas si y sólo si, para todo $s \in D$, la aplicación real $\omega \mapsto d(X(\omega), s)$ es medible. En la práctica, a veces sólo se puede verificar esta propiedad para s en un subconjunto $D_0 \subset D$. Si ese subconjunto es suficientemente grande para el problema considerado—por ejemplo, si es denso en una parte separable donde se concentra el límite o donde se localizan los compactos relevantes—, esa información puede alcanzar para controlar las funciones test necesarias y, por ende, para estudiar convergencia débil sin exigir plena Borel-medibilidad de X .

Lema 1.75. *Sea $X_\alpha : \Omega_\alpha \rightarrow D$ una red asintóticamente tensa. Entonces X_α es asintóticamente medible si y sólo si $f(X_\alpha)$ es asintóticamente medible para toda $f \in C_b(D)$ medible respecto de la σ -álgebra de bolas.*

Demostración. Si X_α es asintóticamente medible en D , entonces, para toda $f \in C_b(D)$ medible respecto de bolas y toda $\varphi \in C_b(\mathbb{R})$, se tiene $\varphi \circ f \in C_b(D)$. Por lo tanto

$$\mathbb{E}_\alpha^*(\varphi \circ f)(X_\alpha) - \mathbb{E}_{\alpha,*}(\varphi \circ f)(X_\alpha) \rightarrow 0.$$

Esto significa que $f(X_\alpha)$ es asintóticamente medible.

Supongamos que $f(X_\alpha)$ es asintóticamente medible para toda $f \in C_b(D)$ medible respecto de bolas. Consideremos la clase

$$\mathcal{A} = \left\{ x \mapsto g(d(x, s_1), \dots, d(x, s_k)) : k \in \mathbb{N}, s_i \in D, g \in C_b(\mathbb{R}^k) \right\}.$$

Entonces $\mathcal{A} \subset C_b(D)$, y todos sus elementos son medibles respecto de la σ -álgebra de bolas.

Además, $\mathcal{A}$ es un álgebra y un retículo vectorial, contiene las constantes y separa puntos. En efecto, si $x \neq y$, tomando $p > 1/d(x, y)$, la función $z \mapsto (1 - pd(z, x))^+$ pertenece a $\mathcal{A}$ y separa x de y .

Por hipótesis, para toda $f \in \mathcal{A}$, la variable real $f(X_\alpha)$ es asintóticamente medible. Como f es acotada, podemos elegir $\varphi \in C_b(\mathbb{R})$ tal que $\varphi(u) = u$ para $|u| \leq \|f\|_\infty$.⁶⁷ Entonces

$$\mathbb{E}_\alpha^* f(X_\alpha) - \mathbb{E}_{\alpha,*} f(X_\alpha) = \mathbb{E}_\alpha^* \varphi(f(X_\alpha)) - \mathbb{E}_{\alpha,*} \varphi(f(X_\alpha)) \rightarrow 0.$$

Como X_α es asintóticamente tensa, el Lema 1.47 implica que la misma convergencia vale para toda $f \in C_b(D)$. Por lo tanto X_α es asintóticamente medible. $\square$

Teorema 1.76. *Sea $X_\alpha : \Omega_\alpha \rightarrow D$ una red arbitraria y sea $X : \Omega \rightarrow D$ Borel-medible y separable. Entonces:*

(i) $X_\alpha \Rightarrow X$ si y sólo si

$$\mathbb{E}_\alpha^* f(X_\alpha) \longrightarrow \mathbb{E} f(X) \quad \text{para toda } f \in C_b(D) \text{ medible respecto de bolas.} \quad (1.4)$$

(ii) $X_\alpha \Rightarrow X$ si y sólo si existe una sucesión $(s_i)_{i \in \mathbb{N}}$ en D tal que

$$(d(X_\alpha, s_1), d(X_\alpha, s_2), \dots) \Rightarrow (d(X, s_1), d(X, s_2), \dots) \quad \text{en } \mathbb{R}^\mathbb{N}, \quad (1.5)$$

y además se cumple $\mathbb{P}(X \in \overline{\{s_1, s_2, \dots\}}) = 1$.

Demostración. Las implicaciones “sólo si” son inmediatas.

Si $X_\alpha \Rightarrow X$, entonces, por definición, $\mathbb{E}_\alpha^* f(X_\alpha) \rightarrow \mathbb{E} f(X)$ para toda $f \in C_b(D)$; en particular, para toda $f \in C_b(D)$ medible respecto de bolas. Esto prueba la parte necesaria de (i).

Para la parte necesaria de (ii), como X es separable, existe un conjunto separable, cerrado (y por tanto boreliano) $D_0 \subset D$ tal que $\mathbb{P}(X \in D_0) = 1$: basta tomar $D_0 = \overline{M}$ para el numerable M de la definición de separabilidad. Tomamos una sucesión $S = \{s_1, s_2, \dots\}$ densa en D_0 . Entonces $\mathbb{P}(X \in \overline{S}) = 1$. Además, la aplicación $x \mapsto (d(x, s_1), d(x, s_2), \dots)$ de D en $\mathbb{R}^\mathbb{N}$ es continua. Por el teorema de la aplicación continua,

$$(d(X_\alpha, s_1), d(X_\alpha, s_2), \dots) \Rightarrow (d(X, s_1), d(X, s_2), \dots) \quad \text{en } \mathbb{R}^\mathbb{N}.$$

Veamos ahora que la condición de (ii) implica $X_\alpha \Rightarrow X$. Sea $S = \{s_1, s_2, \dots\}$ la sucesión dada por la hipótesis, con $\mathbb{P}(X \in \overline{S}) = 1$.

Sea $G \subset D$ abierto. Consideremos la familia numerable

$$\mathcal{F} = \{f_{p,i}(x) = (1 - pd(x, s_i))^+ : p \in \mathbb{Q}^+, i \in \mathbb{N}\}.$$

Cada $f_{p,i}$ pertenece a $C_b(D)$ y es medible respecto de bolas.

Afirmamos que, para todo $x \in \overline{S}$, $\mathbb{1}_G(x) = \sup\{f(x) : 0 \leq f \leq \mathbb{1}_G, f \in \mathcal{F}\}$. Si $x \notin G$, esto es claro. Si $x \in G$, tomamos $r > 0$ tal que $B(x, r) \subset G$. Dado $\gamma > 0$, elegimos $s_i \in S$ suficientemente cercano a x como para que exista $p \in \mathbb{Q}^+$ con

$$\frac{1}{r - d(x, s_i)} < p < \frac{\gamma}{d(x, s_i)}$$

(interpretando $\gamma/0 = +\infty$ si $x \in S$ y $s_i = x$). Tal elección es posible: la condición $d(x, s_i) < \gamma r / (1 + \gamma)$ equivale a $1/(r - d(x, s_i)) < \gamma/d(x, s_i)$, de modo que el intervalo es abierto y no vacío, y contiene un racional positivo; y existen puntos de S así de cercanos a x porque $x \in \overline{S}$. Entonces $f_{p,i} \leq \mathbb{1}_G$ y $f_{p,i}(x) = 1 - pd(x, s_i) > 1 - \gamma$. Como $\gamma > 0$ es arbitrario, el supremo vale 1. Esto prueba la afirmación.

⁶⁷Por ejemplo, $\varphi(u) = (u \vee (-\|f\|_\infty)) \wedge \|f\|_\infty$. Como $|f(X_\alpha)| \leq \|f\|_\infty$ en todo punto, $\varphi(f(X_\alpha)) = f(X_\alpha)$ idénticamente, y la medibilidad asintótica de la variable real $f(X_\alpha)$ (aplicada con esta $\varphi \in C_b(\mathbb{R})$) da exactamente la convergencia mostrada.

Enumeremos las funciones de $\mathcal{F}$ que satisfacen $0 \leq f \leq \mathbb{1}_G$, incluyendo también la función cero, como $(g_j)_{j \geq 1}$, y definamos $f_m = \max\{g_1, \dots, g_m\}$. Entonces $0 \leq f_m \leq \mathbb{1}_G$, cada f_m es continua y $f_m(x) \uparrow \mathbb{1}_G(x)$ para todo $x \in \overline{S}$. Como $\mathbb{P}(X \in \overline{S}) = 1$, $f_m(X) \uparrow \mathbb{1}_G(X)$ c.s. Por monotonía de la esperanza interior,

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

Fijado m , la función f_m depende sólo de un número finito de distancias: existe una función continua acotada ψ_m tal que

$$f_m(x) = \psi_m(d(x, s_{i_1}), \dots, d(x, s_{i_k})).$$

Por la hipótesis de (ii),

$$(d(X_{\alpha}, s_{i_1}), \dots, d(X_{\alpha}, s_{i_k})) \Rightarrow (d(X, s_{i_1}), \dots, d(X, s_{i_k})).$$

Aplicando la definición de convergencia débil a ψ_m y a $-\psi_m$,

$$\begin{aligned} \mathbb{E}_{\alpha,*} f_m(X_{\alpha}) &= -\mathbb{E}_{\alpha}^* \left[-\psi_m(d(X_{\alpha}, s_{i_1}), \dots, d(X_{\alpha}, s_{i_k})) \right] \\ &\longrightarrow -\mathbb{E}(-f_m(X)) = \mathbb{E} f_m(X). \end{aligned}$$

Luego

$$\liminf_{\alpha} \mathbb{P}_{\alpha,*}(X_{\alpha} \in G) \geq \mathbb{E} f_m(X) \quad \text{para todo } m.$$

Haciendo $m \rightarrow \infty$ y usando convergencia monótona, $\mathbb{E} f_m(X) \uparrow \mathbb{P}(X \in G)$. Así,

$$\liminf_{\alpha} \mathbb{P}_{\alpha,*}(X_{\alpha} \in G) \geq \mathbb{P}(X \in G) \quad \text{para todo abierto } G.$$

Por el teorema de Portmanteau en la versión con probabilidades interiores, concluimos que $X_{\alpha} \Rightarrow X$.

Finalmente veamos que (i) implica la convergencia débil. Por separabilidad de X , elegimos una sucesión $S = \{s_1, s_2, \dots\}$ tal que $\mathbb{P}(X \in \overline{S}) = 1$. Para todo k y toda $\psi \in C_b(\mathbb{R}^k)$, la función

$$x \mapsto \psi(d(x, s_1), \dots, d(x, s_k))$$

pertenece a $C_b(D)$ y es medible respecto de bolas. Por la hipótesis de (i),

$$\mathbb{E}_{\alpha}^* \psi(d(X_{\alpha}, s_1), \dots, d(X_{\alpha}, s_k)) \rightarrow \mathbb{E} \psi(d(X, s_1), \dots, d(X, s_k)).$$

Por lo tanto, los vectores finito-dimensionales de distancias convergen débilmente. La aplicación de distancias del límite tiene ley separable: si $M \subset D$ es numerable y $\mathbb{P}(X \in \overline{M}) = 1$, su imagen continua queda concentrada en la clausura de la imagen numerable de M .⁶⁸ Aplicando entonces el criterio de convergencia en productos numerables (Teorema 1.58), obtenemos

$$(d(X_{\alpha}, s_1), d(X_{\alpha}, s_2), \dots) \Rightarrow (d(X, s_1), d(X, s_2), \dots) \quad \text{en } \mathbb{R}^{\mathbb{N}}.$$

Ya probamos que esta condición implica $X_{\alpha} \Rightarrow X$. Esto cierra la prueba de (i). $\square$

1.7. 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.

⁶⁸Con $\Phi(x) = (d(x, s_1), d(x, s_2), \dots)$, continua: $\{X \in \overline{M}\} \subset \{\Phi(X) \in \overline{\Phi(M)}\} = \{\Phi(X) \in \overline{\Phi(M)}\}$, pues por continuidad $\Phi(\overline{M}) \subset \overline{\Phi(M)}$; y $\Phi(M)$ es numerable.

Definición 1.77. Sean $X_\alpha, X : \Omega \rightarrow D$ 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,⁶⁹ para todo $\varepsilon > 0$,

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

- (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 $d(X_\alpha, X) \rightarrow 0$ *uniformemente* en A . Se denota $X_\alpha \xrightarrow{c.u.} X$.
- (iii) X_α **converge casi seguramente exterior** a X si $d(X_\alpha, X)^* \rightarrow 0$ 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

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

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

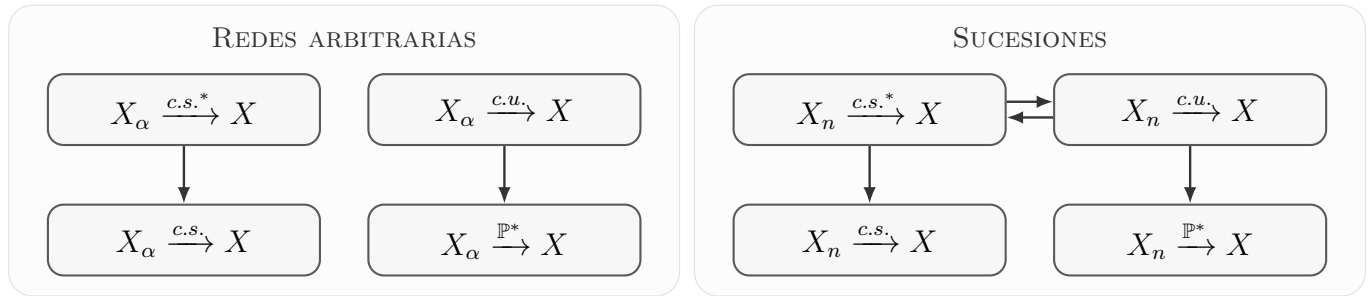


Figura 1.1: Relaciones básicas entre nociones de convergencia para mapas posiblemente no medibles. Sólo se dibujan implicancias válidas. En redes arbitrarias se muestran $c.s.* \Rightarrow c.s.$ y $c.u. \Rightarrow \mathbb{P}^*$. En sucesiones, $c.s.*$ y $c.u.$ son equivalentes, además $c.s.* \Rightarrow c.s.$ 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.

Ejemplo 1.78

Sea $\Omega = \mathbb{N} \times [0, 1]$, provisto de la σ -álgebra trivial $\mathcal{A} = \{\emptyset, \Omega\}$ y de su única probabilidad. Para $n \geq 1$, sea $A_n = \{n\} \times [0, 1]$ y definamos $X_n = \mathbb{1}_{A_n}$. Entonces $X_n(\omega) \rightarrow 0$ para todo ω , pues cada punto pertenece a un único A_n . Sin embargo, $A_n \neq \emptyset$ y la única envolvente medible de A_n es Ω ; ^a por tanto,

$$\mathbb{P}^*(|X_n| > 1/2) = \mathbb{P}^*(A_n) = 1 \quad \text{para todo } n.$$

Así, X_n no converge a 0 en probabilidad exterior; tampoco converge casi uniformemente ni casi seguramente exterior.

⁶⁹La 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.

^aCon $\mathcal{A} = \{\emptyset, \Omega\}$, el único medible que contiene a $A_n \neq \emptyset$ es Ω . Por la misma razón fallan las otras dos convergencias: las funciones medibles son las constantes, así que $(\mathbb{1}_{A_n})^* \equiv 1$ (el mínimo mayorante constante), lo que impide la convergencia c.s.*; y el único conjunto medible con probabilidad $\geq 1 - \varepsilon$, $\varepsilon < 1$, es Ω , donde $\sup_{\omega} |X_n(\omega)| = 1$ para todo n , lo que impide la convergencia c.u.

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.⁷⁰ 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 1.79. *Sea (D, d) un espacio métrico, sea $X : \Omega \rightarrow D$ Borel-medible y sean $X_n : \Omega \rightarrow D$ aplicaciones arbitrarias. Entonces:*

- (i) Si $X_n \xrightarrow{\text{c.s.}^*} X$, entonces $X_n \xrightarrow{P^*} X$.
- (ii) $X_n \xrightarrow{P^*} X$ si y sólo si toda subsucesión admite una subsucesión ulterior que converge a X c.s.*.
- (iii) $X_n \xrightarrow{\text{c.s.}^*} X$ si y sólo si $X_n \xrightarrow{\text{c.u.}} X$.

Lema 1.80. *Sea (D, d) un espacio métrico, sea $X : \Omega \rightarrow D$ Borel-medible y sea (X_α) una red de aplicaciones arbitrarias $\Omega \rightarrow D$. 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 $R_\alpha \xrightarrow{P^*} 0$, es decir, si y sólo si $R_\alpha^* \xrightarrow{P} 0$.
- (ii) Si $X_\alpha \xrightarrow{\text{c.u.}} X$, entonces $X_\alpha \xrightarrow{P^*} X$.

Lema 1.81 (Aplicación continua para c.s.*). *Sean (D, d) y (E, e) espacios métricos, $g : D \rightarrow E$ continua, $X : \Omega \rightarrow D$ Borel-medible y separable, y $X_n : \Omega \rightarrow D$ aplicaciones arbitrarias. Si $X_n \xrightarrow{\text{c.s.}^*} X$, entonces $g(X_n) \xrightarrow{\text{c.s.}^*} g(X)$. Nótese que este resultado es distinto del teorema de aplicación continua para convergencia débil, Teorema 1.36.)*

Se puede ver que la convergencia casi segura no implica convergencia casi segura exterior ni convergencia en probabilidad exterior. El problema es una posible falta de medibilidad. Como la convergencia casi segura es muy cómoda de manejar, interesa saber cuánta medibilidad es suficiente para eliminar la diferencia. Un resultado trivial pero útil es que $X_n \xrightarrow{\text{c.s.}} X$ junto con la medibilidad de $d(X_n, X)$ implica $X_n \xrightarrow{\text{c.s.}^*} X$. El siguiente teorema da una respuesta exacta al problema, aunque, admitidamente, a menudo no es más útil que la condición suficiente trivial recién mencionada.

Lo que se necesita es una cierta forma de medibilidad asintótica de X_n . La medibilidad asintótica introducida antes es ligeramente demasiado débil.

Definición 1.82

Diremos que X_n es **fuertemente asintóticamente medible** si

$$f(X_n)^* - f(X_n)_* \xrightarrow{\text{c.s.}} 0, \quad \text{para toda } f \in C_b(D).$$

⁷⁰Un 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 $X_F \rightarrow 1$ en todo punto ω , al ser cada X_F medible, también c.s.*; sin embargo, $\mathbb{P}(X_F > 1/2) = 0$ para todo F , así que $X_F \rightarrow 0$ en probabilidad.

Esto es más fuerte que la (ordinaria) medibilidad asintótica, pues esta última equivale a la convergencia *en probabilidad* de las mismas diferencias, en lugar de casi segura:

Lema 1.83. *La sucesión X_n es asintóticamente medible si y sólo si*

$$f(X_n)^* - f(X_n)_* \xrightarrow{\mathbb{P}} 0, \quad \text{para toda } f \in C_b(D).$$

Demostración. La medibilidad asintótica fue definida por la condición numérica $\mathbb{E}^* f(X_n) - \mathbb{E}_* f(X_n) \rightarrow 0$ para toda $f \in C_b(D)$. Fijada f , escribese $Y_n := f(X_n)$ con $\|f\|_\infty \leq M$. Como Y_n es acotada, $\mathbb{E}^* Y_n = \mathbb{E} Y_n^*$ y $\mathbb{E}_* Y_n = \mathbb{E} Y_{n*}$, de modo que $\mathbb{E}^* Y_n - \mathbb{E}_* Y_n = \mathbb{E}(Y_n^* - Y_{n*})$. Además, $0 \leq Y_n^* - Y_{n*} \leq 2M$ c.s.: la primera desigualdad porque $Y_{n*} \leq Y_n \leq Y_n^*$ punto a punto, y la segunda por minimalidad contra las constantes $\pm M$. Para una sucesión de variables no negativas y uniformemente acotadas, la convergencia a cero en L^1 equivale a la convergencia a cero en probabilidad, lo que da la equivalencia para cada f y, por tanto, la del enunciado. $\square$

Por supuesto, ambas propiedades están implicadas por la medibilidad para cada n . Se puede mostrar que $X_n \xrightarrow{\text{c.s.}} X$ junto con medibilidad asintótica implica $X_n \xrightarrow{\mathbb{P}^*} X$. Para la conclusión más fuerte $X_n \xrightarrow{\text{c.s.}^*} X$, es necesaria la medibilidad asintótica fuerte. Quizá la mejor parte del siguiente teorema es (iv), que implica que la convergencia casi segura más medibilidad por bolas conlleva convergencia casi uniforme.

Teorema 1.84. *Sea X Borel-medible y separable. Entonces son equivalentes las siguientes afirmaciones:*

- (i) $X_n \xrightarrow{\text{c.s.}^*} X$;
- (ii) $X_n \xrightarrow{\text{c.s.}} X$ y $d(X_n, X)^* - d(X_n, X)_* \xrightarrow{\text{c.s.}} 0$;
- (iii) $X_n \xrightarrow{\text{c.s.}} X$ y X_n es fuertemente asintóticamente medible;
- (iv) $X_n \xrightarrow{\text{c.s.}} X$ y existe un conjunto $S \subset D$ con $\mathbb{P}(X \in \overline{S}) = 1$ tal que, para todo $s \in S$, la función $d(X_n, s)$ es fuertemente asintóticamente medible.

En particular, si $X_n \xrightarrow{\text{c.s.}} X$ y cada X_n y X son medibles por bolas, entonces $X_n \xrightarrow{\text{c.s.}^*} X$.

Demostración. Escribamos $Z_n = d(X_n, X)$. La equivalencia (i) $\Leftrightarrow$ (ii) es inmediata: si $Z_n^* \rightarrow 0$ c.s., entonces $Z_n \leq Z_n^*$ da $X_n \rightarrow X$ c.s. y $0 \leq Z_n^* - Z_{n*} \leq Z_n^* \rightarrow 0$ c.s. Recíprocamente, si $X_n \rightarrow X$ c.s. y $Z_n^* - Z_{n*} \rightarrow 0$ c.s., entonces, en un conjunto medible de probabilidad uno, $0 \leq Z_{n*} \leq Z_n \rightarrow 0$, y por tanto $Z_n^* = Z_{n*} + (Z_n^* - Z_{n*}) \rightarrow 0$.

Probemos (i) $\Rightarrow$ (iii). Por (i) $\Rightarrow$ (ii), $X_n \rightarrow X$ c.s. Además, por el Lema 1.81 (aplicación continua para c.s.*; X es Borel-medible y separable por hipótesis), para toda $f \in C_b(D)$, $f(X_n) \rightarrow f(X)$ c.s.*. Como $f(X)$ es medible y finita, la regla (i) del Lema 1.9 da

$$\begin{aligned} f(X_n)^* - f(X_n)_* &= (f(X_n) - f(X))^* - (f(X_n) - f(X))_* \\ &\leq 2|f(X_n) - f(X)|^* \rightarrow 0 \quad \text{c.s.} \end{aligned}$$

Así, X_n es fuertemente asintóticamente medible.⁷¹

La implicación (iii) $\Rightarrow$ (iv) es directa: si $g \in C_b(\mathbb{R})$ y $s \in D$, entonces $x \mapsto g(d(x, s))$ pertenece a $C_b(D)$. Luego $g(d(X_n, s))^* - g(d(X_n, s))_* \rightarrow 0$ c.s. para todo $s \in D$; tomando $S = D$, que satisface $\mathbb{P}(X \in \overline{D}) = \mathbb{P}(X \in D) = 1$, se obtiene (iv).

⁷¹En el cálculo anterior, la regla (i) del Lema 1.9 se usó con la función medible y finita $S := -f(X)$: da $(f(X_n) - f(X))^* = f(X_n)^* - f(X)$ y, con $-S$, $(f(X_n) - f(X))_* = f(X_n)_* - f(X)$, de modo que $f(X)$ se cancela en la resta. La cota final usa $h^* - h_* \leq 2|h|^*$, pues $h \leq |h|$ y $-h \leq |h|$ dan $h^* \leq |h|^*$ y $-h_* = (-h)^* \leq |h|^*$.

Resta demostrar $(iv) \Rightarrow (i)$. Podemos hacer las siguientes reducciones.⁷² Reemplazar d por $d \wedge 1$ no cambia las nociones de convergencia consideradas. También podemos pasar a la completación de D , pues las distancias entre los puntos originales no cambian y X_n, X siguen tomando valores en D . Finalmente, la hipótesis de (iv) es exactamente la concentración $\mathbb{P}(X \in \bar{S}) = 1$; como X es separable, podemos reemplazar S por un subconjunto numerable $S_0 \subset S$ tal que $\mathbb{P}(X \in \bar{S}_0) = 1$. Esta reducción merece justificación, pues S es arbitrario y la separabilidad de la ley de X no implica la de S . Como X es separable, existe $M \subset D$ numerable con $\mathbb{P}(X \in \bar{M}) = 1$. Sea $A := \bar{S} \cap \bar{M}$; entonces A es boreliano, $\mathbb{P}(X \in A) = 1$ y A es separable, por ser subespacio del espacio separable $\bar{M}$. Sea $Q = \{q_1, q_2, \dots\}$ numerable y denso en A . Como $q_j \in \bar{S}$, para cada $j, k \geq 1$ puede elegirse $s_{j,k} \in S$ con $d(s_{j,k}, q_j) < 1/k$. El conjunto $S_0 := \{s_{j,k} : j, k \geq 1\}$ es numerable, está contenido en S y satisface $Q \subset \bar{S}_0$, de donde $A \subset \bar{Q} \subset \bar{S}_0$ y $\mathbb{P}(X \in \bar{S}_0) \geq \mathbb{P}(X \in A) = 1$. En lo que sigue escribimos nuevamente S para ese subconjunto numerable. Así podemos suponer que S es numerable, que $d \leq 1$ y que D es completo.

Enumeramos

$$\{x \mapsto (1 - md(x, s))^+ : m \in \mathbb{N}, s \in S\} = \{f_1, f_2, \dots\},$$

y definamos $e(x, y) = \sup_{j \geq 1} j^{-1} |f_j(x) - f_j(y)|$. Como supremo de pseudométricas, e es una semimétrica en D (en general no separa todos los puntos de D , pero eso no hará falta).

Afirmamos que, para $x \in \bar{S}$ y una sucesión $(x_n) \subset D$,

$$d(x_n, x) \rightarrow 0 \iff e(x_n, x) \rightarrow 0.$$

$(\Rightarrow)$ Cada f_j es continua y $|f_j| \leq 1$; dado $\gamma > 0$, tomando J con $2/(J+1) < \gamma/2$ se tiene $e(x_n, x) \leq \max_{j \leq J} j^{-1} |f_j(x_n) - f_j(x)| + 2/(J+1)$, y la convergencia de las primeras J coordenadas da $e(x_n, x) < \gamma$ para n grande. $(\Leftarrow)$ Supongamos $e(x_n, x) \rightarrow 0$ pero $d(x_{n'}, x) \geq \varepsilon$ a lo largo de una subsucesión. Fijemos $m \geq 2/\varepsilon$ y elijamos $s \in S$ con $d(x, s) < 1/(2m)$ (posible pues $x \in \bar{S}$). La función $f(y) = (1 - md(y, s))^+$ es algún f_{j_0} de la enumeración, y $f(x) = 1 - md(x, s) > 1/2$, mientras que $d(x_{n'}, s) \geq \varepsilon - d(x, s) \geq \varepsilon - 1/(2m) \geq \varepsilon/2 \geq 1/m$, de modo que $f(x_{n'}) = 0$. Luego $e(x_{n'}, x) \geq j_0^{-1}/2 > 0$ a lo largo de la subsucesión, contradicción.

En consecuencia, para todo compacto $K \subset \bar{S}$ y todo $\varepsilon > 0$ existe $\delta > 0$ tal que

$$x \in K, e(y, x) < \delta \implies d(y, x) < \varepsilon.$$

En efecto, si no, habría $x_k \in K$ e $y_k \in D$ con $e(y_k, x_k) < 1/k$ y $d(y_k, x_k) \geq \varepsilon$; pasando a una subsucesión, $x_k \rightarrow x \in K$ en d , luego $e(x_k, x) \rightarrow 0$ por la equivalencia recién probada, y $e(y_k, x) \leq e(y_k, x_k) + e(x_k, x) \rightarrow 0$; de nuevo por la equivalencia, $d(y_k, x) \rightarrow 0$, y entonces $d(y_k, x_k) \rightarrow 0$, contradicción.

Fijado j , de $X_n \rightarrow X$ c.s. se obtiene $f_j(X_n) \rightarrow f_j(X)$ c.s. Además, por (iv) , $f_j(X_n)$ es fuertemente asintóticamente medible; como $f_j(X)$ es medible y finita, la regla (i) del Lema 1.9 da

$$(f_j(X_n) - f_j(X))^* - (f_j(X_n) - f_j(X))_* = f_j(X_n)^* - f_j(X_n)_* \rightarrow 0 \quad \text{c.s.}$$

Escribamos $W_n := f_j(X_n) - f_j(X)$. En un conjunto medible de probabilidad uno se tiene $W_n \rightarrow 0$ puntualmente (pues f_j es Lipschitz y $d(X_n, X) \rightarrow 0$ allí) y $W_n^* - W_{n*} \rightarrow 0$. De $W_{n*} \leq W_n \leq W_n^* = W_{n*} + (W_n^* - W_{n*})$ se deduce $W_n^* \rightarrow 0$ y $W_{n*} \rightarrow 0$; por la regla (vii) del Lema 1.9,

$$|W_n|^* = W_n^* \vee (-W_n)_* = W_n^* \vee (-W_{n*}) \rightarrow 0 \quad \text{c.s.,}$$

es decir, $|f_j(X_n) - f_j(X)|^* \rightarrow 0$ c.s. Sea $\eta > 0$. Por Egorov,⁷³ para cada j existe $A_j \in \mathcal{A}$ tal que $\mathbb{P}(A_j^c) < \eta 2^{-j}$ y $|f_j(X_n) - f_j(X)|^* \rightarrow 0$ uniformemente en A_j . Con $A = \bigcap_j A_j$, $\mathbb{P}(A^c) < \eta$, y en A se

⁷²Sobre las dos primeras: para $\varepsilon \leq 1$, $\{d > \varepsilon\} = \{d \wedge 1 > \varepsilon\}$, y todas las nociones de convergencia sólo miran valores pequeños de d ; además $(d \wedge 1)(X_n, s)$ es imagen de $d(X_n, s)$ por $u \mapsto u \wedge 1$, continua y acotada en los rangos relevantes, así que la hipótesis (iv) se conserva. Al pasar a la completación, las distancias entre los puntos originales no cambian, X_n y X siguen tomando valores en D , y la separabilidad de X se conserva.

⁷³Teorema de Egorov: si $Z_n \rightarrow Z$ c.s. para variables medibles en un espacio de probabilidad, entonces, para todo $\eta > 0$, existe A medible con $\mathbb{P}(A^c) < \eta$ y $Z_n \rightarrow Z$ uniformemente en A . Aquí se aplica a las variables medibles $|f_j(X_n) - f_j(X)|^*$, que convergen a 0 c.s. por lo recién probado.

tiene $e(X_n, X) \rightarrow 0$ uniformemente: dado $\gamma > 0$, tome J con $2/(J+1) < \gamma/2$ y use la convergencia uniforme de las primeras J coordenadas.

Como X es separable y $\mathbb{P}(X \in \overline{S}) = 1$, existe un compacto $K \subset \overline{S}$ con $\mathbb{P}(X \in K) > 1 - \eta$.⁷⁴

En $B = A \cap \{X \in K\}$ se tiene $\mathbb{P}(B^c) < 2\eta$ y $e(X_n, X) \rightarrow 0$ uniformemente; por la propiedad del compacto K probada más arriba, $d(X_n, X) \rightarrow 0$ uniformemente en B . Como $\eta > 0$ es arbitrario, $X_n \rightarrow X$ c.u.; para sucesiones, c.u. equivale a c.s.*. $\square$

Observación 1.85

El enunciado anterior sigue al Teorema 1.9.6 de [vdV–W], con la hipótesis de concentración de (iv) formulada directamente sobre la clausura, que es la única forma en que interviene en la demostración; nótese que $\{X \in \overline{S}\}$ es un evento medible por ser $\overline{S}$ cerrado. Condiciones más fuertes, como $\mathbb{P}_*(X \in S) = 1$ (“ $X \in S$ casi seguramente”), la implican.

Consideraremos la relación entre los conceptos de convergencia introducidos en la sección anterior y la convergencia débil. Primero seremos un poco formales: la convergencia en probabilidad exterior hacia una constante también puede definirse para aplicaciones con dominios diferentes $(\Omega_\alpha, \mathcal{A}_\alpha, \mathbb{P}_\alpha)$, de modo que no está cubierta por la definición anterior.

Definición 1.86

Sea $X_\alpha : \Omega_\alpha \rightarrow D$ una red arbitraria de aplicaciones y sea $c \in D$. Decimos que X_α **converge en probabilidad exterior** a c si, para todo $\varepsilon > 0$, $\mathbb{P}_\alpha^*(d(X_\alpha, c) > \varepsilon) \rightarrow 0$. Lo denotamos por $X_\alpha \xrightarrow{\mathbb{P}^*} c$.

En general, la convergencia en probabilidad exterior es más fuerte que la convergencia débil, aunque ambas son equivalentes cuando el límite es constante.

Lema 1.87. Sean $X_\alpha, Y_\alpha : \Omega_\alpha \rightarrow D$ aplicaciones arbitrarias y sea $X : \Omega \rightarrow D$ Borel-medible.

- (i) Si $X_\alpha \Rightarrow X$ y $d(X_\alpha, Y_\alpha) \xrightarrow{\mathbb{P}^*} 0$, entonces $Y_\alpha \Rightarrow X$.
- (ii) Si, además, X_α y X están definidos en un mismo espacio de probabilidad y $X_\alpha \xrightarrow{\mathbb{P}^*} X$, entonces $X_\alpha \Rightarrow X$.
- (iii) Para toda constante $c \in D$, $X_\alpha \xrightarrow{\mathbb{P}^*} c$ si y sólo si $X_\alpha \Rightarrow c$.

Demostración. Sea $F \subset D$ cerrado. Si $F = \emptyset$, la desigualdad de Portmanteau que buscamos es trivial; supongamos entonces $F \neq \emptyset$ y, para $\varepsilon > 0$, sea $F^\varepsilon := \{x \in D : d(x, F) \leq \varepsilon\}$. Entonces F^ε es cerrado. Definimos $B_{\alpha, \varepsilon} := \{d(X_\alpha, Y_\alpha)^* \leq \varepsilon\}$. Como $d(X_\alpha, Y_\alpha)^* \rightarrow 0$ en probabilidad, se tiene

$$\mathbb{P}_\alpha(B_{\alpha, \varepsilon}^c) = \mathbb{P}_\alpha(d(X_\alpha, Y_\alpha)^* > \varepsilon) \rightarrow 0. \text{ Además,}$$

$$\{Y_\alpha \in F\} \subset (\{Y_\alpha \in F\} \cap B_{\alpha, \varepsilon}) \cup B_{\alpha, \varepsilon}^c.$$

Por subaditividad de la probabilidad exterior y porque $B_{\alpha, \varepsilon}$ es medible,

$$\mathbb{P}_\alpha^*(Y_\alpha \in F) \leq \mathbb{P}_\alpha^*(\{Y_\alpha \in F\} \cap B_{\alpha, \varepsilon}) + \mathbb{P}_\alpha(B_{\alpha, \varepsilon}^c).$$

Si $\omega \in \{Y_\alpha \in F\} \cap B_{\alpha, \varepsilon}$, entonces $d(X_\alpha(\omega), Y_\alpha(\omega)) \leq d(X_\alpha, Y_\alpha)^*(\omega) \leq \varepsilon$, y como $Y_\alpha(\omega) \in F$, resulta $X_\alpha(\omega) \in F^\varepsilon$. Por tanto $\{Y_\alpha \in F\} \cap B_{\alpha, \varepsilon} \subset \{X_\alpha \in F^\varepsilon\}$. Luego

$$\mathbb{P}_\alpha^*(Y_\alpha \in F) \leq \mathbb{P}_\alpha^*(X_\alpha \in F^\varepsilon) + \mathbb{P}_\alpha(B_{\alpha, \varepsilon}^c).$$

⁷⁴Aquí se usa que D es completo. En efecto, $\overline{S}$ es cerrado en D , luego completo con la métrica inducida, y es separable por ser S numerable; la restricción de la ley de X a $\overline{S}$ es una probabilidad de Borel sobre ese espacio, de modo que la Observación 1.22 la da tensa. La compacidad es intrínseca a la topología, así que un compacto de $\overline{S}$ lo es también de D .

Tomando $\limsup_{\alpha}$ y usando $X_{\alpha} \Rightarrow X$, por Portmanteau,

$$\limsup_{\alpha} \mathbb{P}_{\alpha}^{*}(Y_{\alpha} \in F) \leq \limsup_{\alpha} \mathbb{P}_{\alpha}^{*}(X_{\alpha} \in F^{\varepsilon}) \leq \mathbb{P}(X \in F^{\varepsilon}).$$

Como F es cerrado, $F^{\varepsilon} \downarrow F$ cuando $\varepsilon \downarrow 0$, y por continuidad de la probabilidad desde arriba, $\mathbb{P}(X \in F^{\varepsilon}) \downarrow \mathbb{P}(X \in F)$. Concluimos que

$$\limsup_{\alpha} \mathbb{P}_{\alpha}^{*}(Y_{\alpha} \in F) \leq \mathbb{P}(X \in F)$$

para todo cerrado $F \subset D$. Por Portmanteau, $Y_{\alpha} \Rightarrow X$.

Sea $Z_{\alpha} := X$. Entonces $Z_{\alpha} \Rightarrow X$, pues Z_{α} tiene la misma ley que X para todo α . Además,

$$d(Z_{\alpha}, X_{\alpha}) = d(X, X_{\alpha}) \xrightarrow{\mathbb{P}^{*}} 0.$$

Aplicando (i) con Z_{α} en el papel de X_{α} y X_{α} en el papel de Y_{α} , obtenemos $X_{\alpha} \Rightarrow X$.

Supongamos primero que $X_{\alpha} \xrightarrow{\mathbb{P}^{*}} c$. Sea $Z_{\alpha} \equiv c$. Entonces $Z_{\alpha} \Rightarrow c$ y $d(Z_{\alpha}, X_{\alpha}) = d(c, X_{\alpha}) \xrightarrow{\mathbb{P}^{*}} 0$. Por (i), $X_{\alpha} \Rightarrow c$. Recíprocamente, supongamos que $X_{\alpha} \Rightarrow c$. Fijemos $\varepsilon > 0$ y tomemos el cerrado $F_{\varepsilon} := \{x \in D : d(x, c) \geq \varepsilon\}$. Como $c \notin F_{\varepsilon}$, se tiene $\mathbb{P}(c \in F_{\varepsilon}) = 0$. Por Portmanteau,

$$\limsup_{\alpha} \mathbb{P}_{\alpha}^{*}(d(X_{\alpha}, c) \geq \varepsilon) = \limsup_{\alpha} \mathbb{P}_{\alpha}^{*}(X_{\alpha} \in F_{\varepsilon}) \leq 0.$$

Por tanto $\mathbb{P}_{\alpha}^{*}(d(X_{\alpha}, c) \geq \varepsilon) \rightarrow 0$. Como $\{d(X_{\alpha}, c) > \varepsilon\} \subset \{d(X_{\alpha}, c) \geq \varepsilon\}$, también $\mathbb{P}_{\alpha}^{*}(d(X_{\alpha}, c) > \varepsilon) \rightarrow 0$. Luego $X_{\alpha} \xrightarrow{\mathbb{P}^{*}} c$. $\square$

La implicación “convergencia en probabilidad $\Rightarrow$ convergencia débil” no se invierte en un espacio de probabilidad dado: una red débilmente convergente puede incluso estar definida en espacios distintos. Los teoremas de representación muestran, sin embargo, que es posible construir copias en un espacio común que converjan casi uniformemente. Comenzamos con la versión medible. Para cada α , sea $\mathcal{D}_{\alpha} \subset \mathcal{B}(D)$ una σ -álgebra. Escribimos el límite como X_{∞} , y los enunciados válidos para todo α incluyen $\alpha = \infty$.

Teorema 1.88. Sean $X_{\alpha} : \Omega_{\alpha} \rightarrow D$ aplicaciones $\mathcal{D}_{\alpha}$ -medibles. Si $X_{\alpha} \Rightarrow X_{\infty}$ y X_{∞} es separable, entonces existen aplicaciones $\mathcal{D}_{\alpha}$ -medibles $\tilde{X}_{\alpha} : \tilde{\Omega} \rightarrow D$, definidas en algún espacio de probabilidad $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$, tales que:

- (i) $\tilde{X}_{\alpha} \xrightarrow{c.u.} \tilde{X}_{\infty}$;
- (ii) $\tilde{X}_{\alpha}$ y X_{α} tienen la misma ley sobre $\mathcal{D}_{\alpha}$, para todo α .

Advertencia 1.89

Usualmente, este teorema se aplicará con cada $\mathcal{D}_{\alpha}$ igual a la σ -álgebra de Borel o a la σ -álgebra de bolas. La afirmación “ $\tilde{X}_{\alpha}$ y X_{α} tienen la misma ley sobre $\mathcal{D}_{\alpha}$ ” significa que sus probabilidades coinciden sólo en los conjuntos de $\mathcal{D}_{\alpha}$. Por eso, cuanto más pequeña sea $\mathcal{D}_{\alpha}$, menos información conserva el teorema sobre la ley original. En el caso extremo $\mathcal{D}_{\alpha} = \{\emptyset, D\}$, cualquier par de aplicaciones con valores en D tiene la misma ley sobre $\mathcal{D}_{\alpha}$, pues sólo se comparan las probabilidades de $\emptyset$ y de D . En ese sentido, el teorema sigue siendo verdadero pero la representación puede no reflejar la distribución original de X_{α} . Por ello vale la pena buscar una formulación más fuerte para aplicaciones no medibles.

En la versión no medible se conserva más información: se exige

$$\tilde{\mathbb{E}}^{*}f(\tilde{X}_{\alpha}) = \mathbb{E}_{\alpha}^{*}f(X_{\alpha}) \quad \text{para toda } f : D \rightarrow \mathbb{R} \text{ acotada.}$$

En particular, coinciden las probabilidades exteriores e interiores de $\{X_{\alpha} \in B\}$ y $\{\tilde{X}_{\alpha} \in B\}$ 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 $\delta_\alpha \rightarrow 0$. En particular, $\mathbb{N}$, con el orden usual, es no trivial.⁷⁵

Teorema 1.90. *Sea $X_\alpha : \Omega_\alpha \rightarrow D$ 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 $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbb{P}})$ y aplicaciones $\tilde{X}_\alpha : \tilde{\Omega} \rightarrow D$ tales que:*

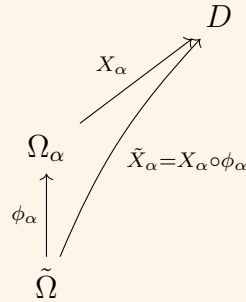
- (i) $\tilde{X}_\alpha \xrightarrow{\text{c.u.}} \tilde{X}_\infty$;
- (ii) $\tilde{\mathbb{E}}^* f(\tilde{X}_\alpha) = \mathbb{E}_\alpha^* f(X_\alpha)$, para toda $f : D \rightarrow \mathbb{R}$ acotada y todo α .

Observación 1.91

Recordemos (Definición 1.17) que una aplicación medible $\phi : (\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbb{P}}) \rightarrow (\Omega, \mathcal{A}, \mathbb{P})$ con $\mathbb{P} = \tilde{\mathbb{P}} \circ \phi^{-1}$ es *perfecta* si conserva las probabilidades exteriores e interiores de todos los subconjuntos de Ω ; equivalentemente, por el Lema 1.18, si $\mathbb{E}^* g = \tilde{\mathbb{E}}^*(g \circ \phi)$ para toda $g : \Omega \rightarrow \mathbb{R}$ acotada.

Observación 1.92

Además de (i) y (ii) del Teorema 1.90, puede elegirse $\tilde{X}_\alpha$ de acuerdo con el siguiente diagrama, con aplicaciones ϕ_α medibles y perfectas, y $\mathbb{P}_\alpha = \tilde{\mathbb{P}} \circ \phi_\alpha^{-1}$:



La perfección de ϕ_α afirmada por la observación anterior mejora la parte (ii) del teorema. Sin embargo, el interés principal de la observación es la implicación de que cada $\tilde{X}_\alpha$ puede forzarse a no tomar otros valores que los originales X_α . Así, $\tilde{X}_\alpha$ y X_α también comparten otras propiedades que quizá no se puedan expresar directamente en términos de sus “leyes”. El ejemplo siguiente explica por qué esta propiedad adicional es útil.

Sea $(\mathcal{X}, \mathcal{B})$ un espacio medible, P una probabilidad en él y $X_1, X_2, \dots$ i.i.d. con ley P . Recordemos que la medida empírica es $P_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$. Sea $\mathcal{H}$ una clase de funciones $\mathcal{B}$ -medibles $h : \mathcal{X} \rightarrow \mathbb{R}$ tales que $P|h| < \infty$; el proceso empírico es

$$\mathbb{G}_n(h) := \sqrt{n}(P_n h - P h), \quad h \in \mathcal{H}.$$

Cuando las trayectorias $h \mapsto \mathbb{G}_n(h)$ son uniformemente acotadas sobre $\mathcal{H}$, consideramos $\mathbb{G}_n$ como una aplicación con valores en $\ell^\infty(\mathcal{H})$; nótese que esto no se deduce de las hipótesis anteriores y debe verificarse (o forzarse mediante una modificación admisible) en cada aplicación concreta. Con estas hipótesis, la representación mediante la medida finita con signo $\mu_{n,\omega}$ que aparece más abajo está bien definida.

⁷⁵La 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 [vdV–W, §1.10].

Qué da el Teorema 1.90. Si $\mathbb{G}_n \Rightarrow \mathbb{G}$ en $\ell^\infty(\mathcal{H})$, el Teorema 1.90 permite construir, en un espacio de probabilidad común $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$, procesos $\tilde{\mathbb{G}}_n, \tilde{\mathbb{G}} : \tilde{\Omega} \rightarrow \ell^\infty(\mathcal{H})$ tales que $\tilde{\mathbb{G}}_n \xrightarrow{\text{c.u.}} \tilde{\mathbb{G}}$ y, para toda función acotada $f : \ell^\infty(\mathcal{H}) \rightarrow \mathbb{R}$,

$$\tilde{\mathbb{E}}^* f(\tilde{\mathbb{G}}_n) = \mathbb{E}^* f(\mathbb{G}_n).$$

Es decir, la representación conserva la ley exterior de $\mathbb{G}_n$ como elemento de $\ell^\infty(\mathcal{H})$ y además realiza la convergencia en un único espacio de probabilidad. Pero esta información es todavía puramente distribucional: no dice, por sí sola, cómo se obtiene cada trayectoria $\tilde{\mathbb{G}}_n(\tilde{\omega})$ a partir de una muestra original. En particular, no fuerza una identidad punto a punto del tipo $\tilde{\mathbb{G}}_n(\tilde{\omega}) = \mathbb{G}_n(\omega)$ para algún ω elegido de manera compatible con el espacio original.

Qué añade la observación. La observación posterior al teorema permite elegir la representación de forma más rígida:

$$\tilde{\mathbb{G}}_n = \mathbb{G}_n \circ \phi_n,$$

donde $\phi_n : \tilde{\Omega} \rightarrow \Omega$ es medible y perfecta, y $\tilde{P} \circ \phi_n^{-1} = P$. Por lo tanto, para todo $\tilde{\omega}$,

$$\tilde{\mathbb{G}}_n(\tilde{\omega}) = \mathbb{G}_n(\phi_n(\tilde{\omega})) \in \text{Im}(\mathbb{G}_n).$$

Esta es la ganancia esencial: cada trayectoria representada es literalmente una trayectoria original, evaluada en otro punto del espacio de probabilidad. Así, si

$$\mathbb{G}_n(\omega)(h) = \int h d\mu_{n,\omega}, \quad h \in \mathcal{H},$$

entonces

$$\tilde{\mathbb{G}}_n(\tilde{\omega})(h) = \mathbb{G}_n(\phi_n(\tilde{\omega}))(h) = \int h d\mu_{n,\phi_n(\tilde{\omega})}.$$

En consecuencia, la representación no sólo conserva la distribución exterior en $\ell^\infty(\mathcal{H})$, sino también la estructura trayectoria por trayectoria: las trayectorias siguen viniendo de medidas finitas con signo del tipo

$$\mu_{n,\omega} = \sqrt{n}(P_n(\omega) - P).$$

La medida subyacente siempre tiene masa total cero; si $1 \in \mathcal{H}$, esto además se ve en la trayectoria mediante $\mathbb{G}_n(1) = 0$.

1.8. Uniformidad y metrización

En principio, la convergencia débil es la convergencia puntual de los “operadores” X_α o L_α sobre $C_b(D)$, esto es, de las aplicaciones $f \mapsto \mathbb{E}_\alpha^* f(X_\alpha)$ o $f \mapsto L_\alpha f$. Sin embargo, hay convergencia uniforme automática sobre ciertos subconjuntos. Estos subconjuntos pueden ser bastante grandes: basta que sean acotados y equicontinuos. Por otro lado, existen también subconjuntos pequeños, incluso numerables, tales que la convergencia puntual en uno de ellos es automáticamente uniforme y equivale a la convergencia débil. Para D separable, incluso es posible elegir un único subconjunto numerable que funcione para todos los X_α a la vez.

Teorema 1.93. Sea (D, d) un espacio métrico y sea $D_0 \subset D$. Sea $\mathcal{F} \subset C_b(D)$ una clase uniformemente acotada y equicontinua en cada punto de D_0 . Si $X_\alpha \Rightarrow X$, donde X es separable y toma valores en D_0 , entonces

$$\sup_{f \in \mathcal{F}} |\mathbb{E}_\alpha^* f(X_\alpha) - \mathbb{E} f(X)| \rightarrow 0.$$

Demostración. Probaremos el teorema bajo la hipótesis más débil $\mathbb{P}^*(X \notin D_0) = 0$, que se cumple trivialmente cuando X toma valores en D_0 ; véase la Observación posterior a la demostración.

Agrandamos $\mathcal{F}$ agregando todas las funciones 1-Lipschitz para d y acotadas por 1.⁷⁶ Esto no afecta la conclusión, pues basta probarla para la clase agrandada. La nueva clase sigue siendo acotada y equicontinua en cada $x \in D_0$, y separa puntos. Definimos

$$e(x, y) := \sup_{f \in \mathcal{F}} |f(x) - f(y)|, \quad x, y \in D.$$

Entonces e es una métrica (separa puntos porque, para $x \neq y$, la función $d(\cdot, y) \wedge 1$ está en la clase agrandada) y toda $f \in \mathcal{F}$ es 1-Lipschitz respecto de e . Además, si $x_n \rightarrow x$ en d y $x \in D_0$, la equicontinuidad de $\mathcal{F}$ en x da $e(x_n, x) \rightarrow 0$. Por tanto, toda función acotada y continua para e es continua para d en cada punto de D_0 . Para cada tal g , su conjunto de d -discontinuidad D_g es boreliano (Lema 1.35) y está contenido en $D \setminus D_0$; de $\mathbb{P}^*(X \notin D_0) = 0$ se sigue $\mathbb{P}(X \in D_g) = 0$, y el Corolario 1.37 (aplicación continua) implica $\mathbb{E}_\alpha^* g(X_\alpha) \rightarrow \mathbb{E} g(X)$ para toda $g \in C_b(D, e)$. En consecuencia, $X_\alpha \Rightarrow X$ también cuando D se equipa con la métrica e (aquí, y en lo que sigue, X denota la versión Borel dada por el teorema de aplicación continua).

Sea $(\bar{D}, e)$ la completación. Identificamos $D \subset \bar{D}$. Toda $f \in \mathcal{F}$, al ser 1-Lipschitz para e , se extiende de modo único a una función 1-Lipschitz y acotada en $\bar{D}$. Vistos como mapas en $\bar{D}$, todavía $X_\alpha \Rightarrow X$.⁷⁷

Veamos que la ley de X es tensa en $(\bar{D}, e)$. Como X es separable para d , existe $M \subset D$ numerable con $\mathbb{P}(X \in \bar{M}^d) = 1$, donde $\bar{M}^d$ denota la d -clausura (que es boreliana). Si $x \in D_0 \cap \bar{M}^d$, existe una sucesión $x_m \in M$ con $d(x_m, x) \rightarrow 0$, y la equicontinuidad de $\mathcal{F}$ en x da $e(x_m, x) \rightarrow 0$; por lo tanto $D_0 \cap \bar{M}^d \subset \bar{M}^e$, la e -clausura de M en $\bar{D}$. Fuera de un evento medible de probabilidad nula se tiene $X \in D_0$ (pues $\mathbb{P}^*(X \notin D_0) = 0$); intersecando con el evento medible $\{X \in \bar{M}^d\}$, de probabilidad uno, se obtiene $X \in \bar{M}^e$ c.s. El evento $\{X \in \bar{M}^e\}$ es medible, porque, desde el comienzo de la prueba, X denota la versión Borel dada por el Corolario 1.37, con valores en $(\bar{D}, e)$, y $\bar{M}^e$ es cerrado; luego $\mathbb{P}(X \in \bar{M}^e) = 1$: la ley de X es separable para e . Por el Lema 1.27, al ser $(\bar{D}, e)$ completo, la ley de X en $\bar{D}$ es tensa. Por el Lema 1.41(b), la red X_α es asintóticamente tensa en $\bar{D}$. Por tanto, dado $\varepsilon > 0$, existe compacto $K \subset \bar{D}$ tal que, para todo $\delta > 0$,

$$\limsup_\alpha \mathbb{P}_\alpha^*(X_\alpha \notin K^\delta) \leq \varepsilon, \quad K^\delta := \{x : e(x, K) < \delta\}.$$

En particular, para todo α suficientemente grande, $\mathbb{P}_\alpha^*(X_\alpha \notin K^\delta) \leq 2\varepsilon$. Además, la ley de X en $(\bar{D}, e)$ es tensa, así que existe un compacto $K' \subset \bar{D}$ con $\mathbb{P}(X \in K') \geq 1 - \varepsilon$. Reemplazamos K por $K \cup K'$: la unión sigue siendo compacta; como $K^\delta \subset (K \cup K')^\delta$, la cota del $\limsup$ anterior no empeora; y como $K' \subset (K \cup K')^\delta$, se tiene $\mathbb{P}(X \notin (K \cup K')^\delta) \leq \varepsilon$. Renombrando $K \cup K'$ como K , podemos suponer entonces $\mathbb{P}(X \notin K^\delta) \leq \varepsilon$ para todo $\delta > 0$.

Consideremos $\mathcal{F}_K := \{f|_K : f \in \mathcal{F}\}$. Como $\mathcal{F}$ es uniformemente acotada y 1-Lipschitz para e , la clase $\mathcal{F}_K$ es acotada y equicontinua en el compacto K . Por Arzelà–Ascoli, es totalmente acotada en la norma uniforme. Por tanto, dado $\eta > 0$, existen $f_1, \dots, f_m \in \mathcal{F}$ tales que, para todo $f \in \mathcal{F}$, existe i con

$$\sup_{x \in K} |f(x) - f_i(x)| \leq \eta/3.$$

Tomamos $\delta = \eta/3$. Si $x \in K^\delta$, existe $y \in K$ con $e(x, y) < \delta$; por la elección de f_i ,

$$|f(x) - f_i(x)| \leq |f(x) - f(y)| + |f(y) - f_i(y)| + |f_i(y) - f_i(x)| < \delta + \eta/3 + \delta = \eta.$$

Luego

$$\sup_{x \in K^\delta} |f(x) - f_i(x)| \leq \eta.$$

⁷⁶La clase agrandada sigue siendo uniformemente acotada (por $\max\{1, \sup_{f \in \mathcal{F}} \|f\|_\infty\}$) y equicontinua en cada punto de D_0 , pues las funciones 1-Lipschitz son equicontinuas en todo punto; y ahora separa puntos: para $x \neq y$, $d(\cdot, y) \wedge 1$ las distingue.

⁷⁷Una función 1-Lipschitz sobre un subespacio denso de un métrico se extiende de manera única y 1-Lipschitz a la completación, con la misma cota uniforme. Y la convergencia débil se conserva porque toda $g \in C_b(\bar{D})$ se restringe a una función de $C_b(D, e)$ sin cambiar $\mathbb{E}_\alpha^* g(X_\alpha)$ ni $\mathbb{E} g(X)$.

Sea $M := \sup_{f \in \mathcal{F}} \|f\|_\infty < \infty$. Para $f \in \mathcal{F}$ y su f_i asociado,

$$\begin{aligned} \left| \mathbb{E}_\alpha^* f(X_\alpha) - \mathbb{E} f(X) \right| &\leq \left| \mathbb{E}_\alpha^* f(X_\alpha) - \mathbb{E}_\alpha^* f_i(X_\alpha) \right| \\ &\quad + \left| \mathbb{E}_\alpha^* f_i(X_\alpha) - \mathbb{E} f_i(X) \right| + \left| \mathbb{E} f_i(X) - \mathbb{E} f(X) \right|. \end{aligned}$$

Por subaditividad de la esperanza exterior, para todo α suficientemente grande,

$$\begin{aligned} \left| \mathbb{E}_\alpha^* f(X_\alpha) - \mathbb{E}_\alpha^* f_i(X_\alpha) \right| &\leq \mathbb{E}_\alpha^* |f(X_\alpha) - f_i(X_\alpha)| \\ &\leq \eta + 2M\mathbb{P}_\alpha^*(X_\alpha \notin K^\delta) \leq \eta + 4M\varepsilon. \end{aligned}$$

Análogamente, $|\mathbb{E} f_i(X) - \mathbb{E} f(X)| \leq \eta + 2M\varepsilon$.⁷⁸ Como $f_1, \dots, f_m \in C_b(\bar{D}, e)$ y son finitas en número,

$$\max_{1 \leq i \leq m} |\mathbb{E}_\alpha^* f_i(X_\alpha) - \mathbb{E} f_i(X)| \rightarrow 0.$$

Por lo tanto, para α suficientemente grande,

$$\sup_{f \in \mathcal{F}} \left| \mathbb{E}_\alpha^* f(X_\alpha) - \mathbb{E} f(X) \right| \leq 2\eta + (6M + 1)\varepsilon.$$

Como $\eta, \varepsilon > 0$ son arbitrarios, se concluye la convergencia uniforme. $\square$

Observación 1.94

La demostración usó únicamente $\mathbb{P}^*(X \notin D_0) = 0$ (equivalentemente, $\mathbb{P}_*(X \in D_0) = 1$; si D_0 es boreliano, esto es $\mathbb{P}(X \in D_0) = 1$), condición implicada por la hipótesis del enunciado. La formulación mediante probabilidad exterior es útil cuando D_0 no es necesariamente boreliano.

Teorema 1.95. Sea $(\mathbb{D}, d)$ un espacio métrico. Para todo subconjunto separable $\mathbb{D}_0 \subset \mathbb{D}$, existe una colección $\mathcal{F} \subset C_b(\mathbb{D})$ numerable, uniformemente equicontinua y acotada tal que para toda red arbitraria $X_\alpha : \Omega_\alpha \rightarrow \mathbb{D}$ y toda variable aleatoria Borel-medible $X : \Omega \rightarrow \mathbb{D}$ con $\mathbb{P}^*(X \notin \mathbb{D}_0) = 0$ (si $\mathbb{D}_0$ es boreliano, esto es $\mathbb{P}(X \in \mathbb{D}_0) = 1$), son equivalentes:

- (i) $X_\alpha \Rightarrow X$;
- (ii) $\mathbb{E}_\alpha^* f(X_\alpha) \rightarrow \mathbb{E} f(X)$ para todo $f \in \mathcal{F}$;
- (iii) $\mathbb{E}_\alpha^* f(X_\alpha) \rightarrow \mathbb{E} f(X)$ uniformemente en $f \in \mathcal{F}$.

Teorema 1.96 (Adenda). La colección $\mathcal{F}$ del teorema anterior puede tomarse igual al conjunto de funciones

$$f(x) = q \left((1 - p_1 d(x, s_1))^+ \vee \dots \vee (1 - p_k d(x, s_k))^+ \right),$$

donde $k \in \mathbb{N}$, $s_1, \dots, s_k$ recorren un subconjunto numerable y denso de $\mathbb{D}_0$, $q \in \mathbb{Q}$, $p_1, \dots, p_k \in \mathbb{Q}^+$, $|q| \leq 1$ y $|q|(p_1 \vee \dots \vee p_k) \leq 1$. Aquí $a \vee b = \max\{a, b\}$ y $t^+ = \max\{t, 0\}$.

⁷⁸Con la cota puntual $|f - f_i| \leq \eta \mathbf{1}_{K^\delta} + 2M \mathbf{1}_{(K^\delta)^c}$ y $\mathbb{P}(X \notin K^\delta) \leq \varepsilon$, garantizada por el reemplazo $K \mapsto K \cup K'$ hecho antes.

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, $X_1, X_2, \dots$ 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 $X_1, \dots, X_n$ es la medida aleatoria discreta dada por $P_n(C) = n^{-1} \# \{1 \leq i \leq n : X_i \in C\}$, $C \in \mathcal{A}$. 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 $f : \mathcal{X} \rightarrow \mathbb{R}$, la medida empírica induce una aplicación $\mathcal{F} \rightarrow \mathbb{R}$ 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 $f \mapsto \mathbb{G}_n f = \sqrt{n}(P_n - P)f = \frac{1}{\sqrt{n}} \sum_{i=1}^n (f(X_i) - Pf)$. Frecuentemente identificaremos la medida signada $\mathbb{G}_n = n^{-1/2} \sum_{i=1}^n (\delta_{X_i} - P)$ con el proceso empírico.

Para una función fija f , por la ley de los grandes números y el teorema central del límite se tiene $P_n f \xrightarrow[n \rightarrow \infty]{c.s.} Pf$, $\mathbb{G}_n f \Rightarrow \mathcal{N}(0, P(f - Pf)^2)$, 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 $\mathbb{R}^d \rightarrow \mathbb{R}$ 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 $\|P_n - P\|_{\mathcal{F}} \rightarrow 0$, donde la convergencia es *casi segura exterior*: $\|P_n - P\|_{\mathcal{F}}^* \rightarrow 0$ 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 $\|P_n - P\|_{\mathcal{F}} \rightarrow 0$ 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

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

donde el límite $\mathbb{G}$ es un elemento medible de Borel 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*).

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

La naturaleza del proceso límite $\mathbb{G}$ se deduce de sus marginales. Si $f \in L_2(P)$, el teorema central del límite da $\mathbb{G}_n f \Rightarrow \mathcal{N}(0, P(f - Pf)^2)$; 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 $\{\mathbb{G}f : f \in \mathcal{F}\}$ ha de ser gaussiano de media cero, con función de covarianza

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

Por el Lema de Slutsky, toda clase P -Donsker es Glivenko–Cantelli débil: $\|P_n - P\|_{\mathcal{F}} \rightarrow 0$ 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 2.27, 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, $x^2 \mathbb{P}^*(\sup_{f \in \mathcal{F}} |f - Pf| > x) \rightarrow 0$, y en particular es integrable (véase [vdV–W, 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 $\{\mathbf{1}_{(-\infty, t]} : t \in \mathbb{R}^d\}$. La medida empírica indexada por $\mathcal{F}$ se identifica con la función de distribución empírica $t \mapsto P_n \mathbf{1}_{(-\infty, t]} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{X_i \leq t\}$, donde $X_i \leq t$ es coordenada a coordenada. En este caso es natural identificar $f = \mathbf{1}_{(-\infty, t]}$ 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 [4, 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.

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 $C \mapsto P_n(C) = \frac{1}{n} \#\{X_i \in C\}$. En este caso conviene hacer la identificación $C \leftrightarrow \mathbf{1}_C$, tanto en notación como en terminología. Así, $\mathcal{C}$ se llama *clase de P -Glivenko–Cantelli* si $\|P_n - P\|_{\mathcal{C}}^* \rightarrow 0$ casi seguramente (y satisface la propiedad *débil* si sólo hay convergencia en probabilidad exterior), y $\mathcal{C}$ es *clase de Donsker* si $\sqrt{n}(P_n - P) \Rightarrow$ un límite tenso en $\ell^\infty(\mathcal{F})$.

²Si $Y_1, Y_2, \dots$ 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 [4, Sección 2.1].

³En 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 $n^{-1/2} \rightarrow 0$ (Slutsky), $\|P_n - P\|_{\mathcal{F}} \Rightarrow 0$, y la convergencia débil a una constante equivale a la convergencia en probabilidad exterior.

2.2. Desigualdades maximales y números de recubrimiento

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 $\psi : [0, \infty) \rightarrow [0, \infty)$ una función convexa, no decreciente, no idénticamente nula, con $\psi(0) = 0$. Para una variable aleatoria X , la **norma de Orlicz** $\|X\|_\psi$ se define como $\|X\|_\psi := \inf \left\{ C > 0 : \mathbb{E} \psi \left(\frac{|X|}{C} \right) \leq 1 \right\}$. Aquí el ínfimo sobre el conjunto vacío se interpreta como ∞ .

Usando la desigualdad de Jensen, se verifica que esto define una norma⁴ sobre el conjunto de variables aleatorias para las cuales $\|X\|_\psi < \infty$.

Ejemplo 2.4. ■ $\psi(x) = x^p$, 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 2.5. 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 $\alpha \geq 1$, $\|X\|_r \leq C_\alpha r^{1/\alpha} \|X\|_{\psi_\alpha}$.

Cualquier norma de Orlicz proporciona una cota para la cola de una 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 $\mathbb{P}(|X| > x) \leq \frac{1}{\psi(x/\|X\|_\psi)}$, $x > 0$. 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. 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.

Advertencia 2.6

Los Lemas 2.7 y 2.9 son ejercicios guiados del práctico 3, sólo demostraremos el Lema 2.8.

Lema 2.7. Sea X una variable aleatoria tal que, para constantes $K, C > 0$ y algún $p \geq 1$, $\mathbb{P}(|X| > x) \leq Ke^{-Cx^p}$, $x > 0$. 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

$$\left\| \max_{1 \leq i \leq m} |X_i| \right\|_p = \left(\mathbb{E} \max_{1 \leq i \leq m} |X_i|^p \right)^{1/p} \leq m^{1/p} \max_{1 \leq i \leq m} \|X_i\|_p.$$

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 .

⁴Esbozo: la homogeneidad absoluta es el cambio $C \mapsto |\lambda|C$ en el ínfimo. Para la triangular, si $C_i > \|X_i\|_\psi$, la convexidad da $\psi \left(\frac{|X_1+X_2|}{C_1+C_2} \right) \leq \frac{C_1}{C_1+C_2} \psi \left(\frac{|X_1|}{C_1} \right) + \frac{C_2}{C_1+C_2} \psi \left(\frac{|X_2|}{C_2} \right)$, y la esperanza queda ≤ 1 . Finalmente, $\|X\|_\psi = 0 \Rightarrow X = 0$ c.s.: si $\mathbb{P}(|X| > \delta) > 0$, entonces $1 \geq \mathbb{E} \psi(|X|/C) \geq \psi(\delta/C) \mathbb{P}(|X| > \delta) \rightarrow \infty$ cuando $C \downarrow 0$, pues una ψ convexa, no nula y con $\psi(0) = 0$ tiende a ∞ .

Lema 2.8. Sea $\psi : [0, \infty) \rightarrow [0, \infty)$ convexa, no decreciente, no idénticamente nula, con $\psi(0) = 0$. Supongamos que existe $c > 0$ tal que $\limsup_{x,y \rightarrow \infty} \frac{\psi(x)\psi(y)}{\psi(cxy)} < \infty$. Entonces existe una constante K , que depende sólo de ψ y de c , tal que, para cualesquiera variables aleatorias $X_1, \dots, X_m$, $\|\max_{1 \leq i \leq m} |X_i|\|_\psi \leq K \psi^{-1}(m) \max_{1 \leq i \leq m} \|X_i\|_\psi$.

Demostración. Escribimos $Y = \max_{1 \leq i \leq m} |X_i|$ y $M = \max_i \|X_i\|_\psi$. Si $M = \infty$, no hay nada que probar. Suponemos entonces $M < \infty$.

Primero probamos el resultado bajo la hipótesis adicional de que $\psi(x)\psi(y) \leq \psi(cxy)$ para todo $x, y \geq 1$ y $\psi(1) \leq 1/2$. En ese caso, si $x \geq y \geq 1$, tomando $u = x/y$ y $v = y$, se obtiene $\psi(x/y)\psi(y) \leq \psi(cx)$, es decir, $\psi(x/y) \leq \frac{\psi(cx)}{\psi(y)}$. Sea $M_0 > M$ y tomemos $C = cM_0$. Entonces $\mathbb{E}\psi(c|X_i|/C) = \mathbb{E}\psi(|X_i|/M_0) \leq 1$ para todo i .⁵ Además, para $y \geq 1$, punto a punto,

$$\psi\left(\frac{Y}{Cy}\right) = \max_i \psi\left(\frac{|X_i|}{Cy}\right) \leq \sum_{i=1}^m \frac{\psi(c|X_i|/C)}{\psi(y)} + \psi(1).$$

En efecto, si $|X_i|/(Cy) \geq 1$, se aplica la desigualdad anterior con $x = |X_i|/C$; si $|X_i|/(Cy) < 1$, basta usar que $\psi(|X_i|/(Cy)) \leq \psi(1)$. Tomando esperanza, $\mathbb{E}\psi\left(\frac{Y}{Cy}\right) \leq \frac{m}{\psi(y)} + \psi(1)$. Ahora elegimos $y = \psi^{-1}(2m)$.⁶ Como $\psi(1) \leq 1/2$, se tiene $y \geq 1$, y además $m/\psi(y) \leq 1/2$. Por tanto $\mathbb{E}\psi(Y/(C\psi^{-1}(2m))) \leq 1$, y de la definición de norma de Orlicz, $\|Y\|_\psi \leq cM_0\psi^{-1}(2m)$. Como ψ es convexa y $\psi(0) = 0$, se cumple $\psi(2x) \geq 2\psi(x)$; por lo tanto $\psi^{-1}(2m) \leq 2\psi^{-1}(m)$.⁷ Luego $\|Y\|_\psi \leq 2cM_0\psi^{-1}(m)$. Haciendo $M_0 \downarrow M$, queda probado el caso especial.

Pasamos ahora al caso general. Por la hipótesis del $\limsup$, existen $A < \infty$ y $x_0 \geq 1$ tales que $\psi(x)\psi(y) \leq A\psi(cxy)$ si $x, y \geq x_0$. Elegimos $\tau \geq x_0$ y $\sigma \in (0, 1]$ suficientemente pequeño para que $\sigma A \leq 1$ y $\sigma\psi(\tau) \leq 1/2$. Definimos $\phi(x) = \sigma\psi(\tau x)$. Entonces, para $x, y \geq 1$, $\phi(x)\phi(y) = \sigma^2\psi(\tau x)\psi(\tau y) \leq \sigma^2 A\psi(c\tau^2 xy) \leq \sigma\psi(c\tau^2 xy) = \phi(c\tau xy)$, y además $\phi(1) \leq 1/2$.⁸ Así, la parte ya probada aplicada a ϕ , con c reemplazado por $c\tau$, da $\|Y\|_\phi \leq 2c\tau\phi^{-1}(m) \max_i \|X_i\|_\phi$.

Sólo resta volver de ϕ a ψ . Como $\phi(x) = \sigma\psi(\tau x)$ y ψ es convexa con $\psi(0) = 0$, se tiene, para toda variable Z ,⁹ $\|Z\|_\phi \leq \tau\|Z\|_\psi$, $\|Z\|_\psi \leq \frac{1}{\sigma\tau}\|Z\|_\phi$. Además, $\phi^{-1}(m) = \tau^{-1}\psi^{-1}(m/\sigma) \leq (\sigma\tau)^{-1}\psi^{-1}(m)$, usando nuevamente la convexidad de ψ .¹⁰ Por consiguiente,

$$\|Y\|_\psi \leq \frac{1}{\sigma\tau}\|Y\|_\phi \leq \frac{1}{\sigma\tau} 2c\tau\phi^{-1}(m) \max_i \|X_i\|_\phi \leq \frac{2c}{\sigma^2} \psi^{-1}(m) \max_i \|X_i\|_\psi.$$

Tomando $K = 2c/\sigma^2$, queda probado el resultado. $\square$

Para evitar ambigüedades, diremos que un proceso $\{X_t : t \in T\}$ es *separable respecto de d* si existen un conjunto numerable $T_0 \subset T$ y un conjunto medible nulo N tales que, para todo $\omega \notin N$ y todo $\delta > 0$,

$$\sup_{d(s,t) \leq \delta} |X_s(\omega) - X_t(\omega)| = \sup_{\substack{s,t \in T_0 \\ d(s,t) \leq \delta}} |X_s(\omega) - X_t(\omega)|.$$

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.

⁵La igualdad es $C = cM_0$. Y $\mathbb{E}\psi(|X_i|/M_0) \leq 1$ vale para todo $M_0 > \|X_i\|_\psi$: por definición del ínfimo existe $C' \in (\|X_i\|_\psi, M_0]$ con $\mathbb{E}\psi(|X_i|/C') \leq 1$, y la monotonía de ψ concluye. Esta propiedad del ínfimo se usará repetidamente.

⁶Como ψ es convexa y finita, es continua, y su inversa generalizada $\psi^{-1}(t) = \inf\{x : \psi(x) \geq t\}$ satisface $\psi(\psi^{-1}(t)) \geq t$ para t por debajo de $\sup \psi$. De $\psi(1) \leq 1/2 < 2m$ y la monotonía sale $y \geq 1$; y $\psi(y) \geq 2m$ da $m/\psi(y) \leq 1/2$.

⁷La primera desigualdad sale de la convexidad con $\psi(0) = 0$: $\psi(x) \leq \frac{1}{2}\psi(2x) + \frac{1}{2}\psi(0) = \frac{1}{2}\psi(2x)$. Para la segunda, con $s = \psi^{-1}(m)$: $\psi(2s) \geq 2\psi(s) \geq 2m$, luego $\psi^{-1}(2m) \leq 2s$.

⁸Nótese que para $x, y \geq 1$ se tiene $\tau x, \tau y \geq \tau \geq x_0$, lo que habilita la cota del $\limsup$; y ϕ hereda de ψ la convexidad, la monotonía, $\phi(0) = 0$ y la no nulidad, de modo que el caso especial se le aplica.

⁹Primera desigualdad: si $C > \|Z\|_\psi$ y $D = \tau C$, entonces $\mathbb{E}\phi(|Z|/D) = \sigma \mathbb{E}\psi(|Z|/C) \leq \sigma \leq 1$, luego $\|Z\|_\phi \leq \tau C$. Segunda: la convexidad con $\psi(0) = 0$ da $\psi(\sigma u) \leq \sigma\psi(u)$; si $D > \|Z\|_\phi$, entonces $\mathbb{E}\psi(\sigma\tau|Z|/D) \leq \sigma \mathbb{E}\psi(\tau|Z|/D) = \mathbb{E}\psi(|Z|/D) \leq 1$, luego $\|Z\|_\psi \leq D/(\sigma\tau)$.

¹⁰La igualdad: $\phi(x) \geq m \iff \psi(\tau x) \geq m/\sigma \iff \tau x \geq \psi^{-1}(m/\sigma)$. La desigualdad: con $k = 1/\sigma \geq 1$ y $s = \psi^{-1}(m)$, la desigualdad $\psi(ks) \geq k\psi(s) \geq km$, consecuencia elemental de la convexidad y de $\psi(0) = 0$, da $\psi^{-1}(m/\sigma) \leq s/\sigma$.

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 2.9. Sean $X_1, \dots, X_m$ variables aleatorias tales que, para ciertos $a, b > 0$, $\mathbb{P}(|X_i| > x) \leq \frac{2}{\exp\left\{-\frac{x^2}{a+bx}\right\}}$, $x > 0$, $i = 1, \dots, m$. Entonces $\|\max_{1 \leq i \leq m} |X_i|\|_{\psi_1} \leq K \left(\sqrt{a} \sqrt{\log(1+m)} + b \log(1+m) \right)$ 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 $\psi_p(x) = e^{x^p} - 1$, este crecimiento es de orden logarítmico, porque $\psi_p^{-1}(m) = (\log(1+m))^{1/p}$.

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*. Cada variable aleatoria en el supremo se escribe como una suma de “pequeños eslabones”, y la cota depende del número y tamaño de dichos 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 2.10. Sea (T, d) un espacio semimétrico arbitrario. El **número de recubrimiento** $N(\varepsilon, d)$ es el número mínimo de bolas de radio ε necesarias para recubrir 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 recubrimiento y de empaquetamiento, respectivamente.

Para los propósitos actuales, pueden usarse tanto números de recubrimiento como de empaquetamiento. En los argumentos, uno puede ser reemplazado por el otro mediante las desigualdades

$$N(\varepsilon, d) \leq D(\varepsilon, d) \leq N(\varepsilon/2, d). \quad (2.1)$$

Claramente, los números de recubrimiento y de empaquetamiento aumentan cuando $\varepsilon \downarrow 0$. Por definición, el conjunto semimétrico T está totalmente acotado si y sólo si los números de recubrimiento 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.

Teorema 2.11. Sea ψ una función convexa, no decreciente y no nula, con $\psi(0) = 0$, tal que $\limsup_{x, y \rightarrow \infty} \frac{\psi(x)\psi(y)}{\psi(cxy)} < \infty$ para alguna constante $c > 0$. Sea $\{X_t : t \in T\}$ un proceso estocástico separable, indexado por un espacio semimétrico (T, d) , tal que $\|X_s - X_t\|_\psi \leq Cd(s, t)$, $s, t \in T$, para alguna constante $C < \infty$. Entonces, para todo $\eta, \delta > 0$,

$$\left\| \sup_{d(s, t) \leq \delta} |X_s - X_t| \right\|_\psi \leq K \left\{ \int_0^\eta \psi^{-1}(D(\varepsilon, d)) d\varepsilon + \delta \psi^{-1}(D^2(\eta, d)) \right\},$$

donde K depende solo de ψ y de C .

Corolario 2.12. La constante K puede elegirse de forma que

$$\left\| \sup_{s, t \in T} |X_s - X_t| \right\|_\psi \leq K \int_0^{\text{diam } T} \psi^{-1}(D(\varepsilon, d)) d\varepsilon,$$

donde $\text{diam } T = \sup_{s, t \in T} d(s, t)$, y el lado derecho se interpreta como $+\infty$ si la integral diverge.

Demostración del Teorema 2.11. Escribimos $D(\varepsilon) = D(\varepsilon, d)$. Basta probar la desigualdad con $d(s, t) < \delta$; la versión con $d(s, t) \leq \delta$ se obtiene aplicando la desigualdad a $\delta + \rho$ y dejando $\rho \downarrow 0$.

Sea A_ψ la constante del Lema 2.8: si $Y_1, \dots, Y_m$ son variables aleatorias con $\max_i \|Y_i\|_\psi \leq r$, entonces $\left\| \max_{1 \leq i \leq m} |Y_i| \right\|_\psi \leq A_\psi r \psi^{-1}(m)$. Pongamos $K_1 = 2A_\psi C$. Al final solo aparecerá una constante universal K que depende de ψ y de C .

Si algún $D(\eta 2^{-j})$ es infinito, o si $\int_0^\eta \psi^{-1}(D(\varepsilon)) d\varepsilon = \infty$, no hay nada que probar. Suponemos entonces que todos esos números son finitos y que la integral es finita.

Sea $T'_0 \subset T$ el separante numerable del proceso y N el nulo asociado: fuera de N , para todo $\rho > 0$, el supremo de $|X_s - X_t|$ sobre $\{d(s, t) \leq \rho\}$ coincide con el supremo sobre los pares del separante con $d(s, t) \leq \rho$. Tomando supremo en $\rho < \delta$, lo mismo vale con la restricción estricta $d(s, t) < \delta$. Enumeramos $T'_0 = \{t_1, t_2, \dots\}$ y escribimos $F_k := \{t_1, \dots, t_k\}$. La estrategia es probar la cota, con constante independiente de k , para el máximo sobre pares de F_k ; el caso general se obtendrá al final por convergencia monótona.

Fijamos k y construimos conjuntos anidados $T_0 \subset T_1 \subset T_2 \subset \dots \subset F_k$ ¹¹ de la siguiente manera. Tomamos $T_0 \subset F_k$ maximal η -separado y, una vez elegido T_j , extendemos T_j a un conjunto $T_{j+1} \subset F_k$ maximal $\eta 2^{-j-1}$ -separado en F_k . Como todo conjunto ε -separado de F_k lo es de T , se tiene $|T_j| \leq D(\eta 2^{-j})$, $j \geq 0$. Además, por maximalidad, todo punto de T_{j+1} está a distancia a lo sumo $\eta 2^{-j}$ de algún punto de T_j . Elegimos uno de esos puntos y lo denotamos por $\pi_j u$. Así, si $u \in T_{j+1}$, entonces $d(u, \pi_j u) \leq \eta 2^{-j}$.

La jerarquía agota F_k módulo distancias nulas: sea $j_0 = j_0(k)$ tal que $\eta 2^{-j_0-1}$ es menor que la mínima distancia *positiva* entre puntos de F_k (si todas las distancias en F_k son nulas, sirve cualquier j_0). Si $u \in F_k \setminus T_{j_0+1}$, la maximalidad de T_{j_0+1} da un $v \in T_{j_0+1}$ con $d(u, v) \leq \eta 2^{-j_0-1}$, y por la elección de j_0 necesariamente $d(u, v) = 0$; como $\|X_u - X_v\|_\psi \leq C d(u, v) = 0$, se tiene $X_u = X_v$ c.s. Por lo tanto, fuera de un conjunto nulo N_k ,

$$\max_{\substack{s, t \in F_k \\ d(s, t) < \delta}} |X_s - X_t| = \max_{\substack{s, t \in T_{j_0+1} \\ d(s, t) < \delta}} |X_s - X_t| \quad (0)$$

(los máximos sobre conjuntos vacíos se interpretan como 0). Denotaremos este máximo por S_{j_0} .

Fijado un nivel $l \geq 0$ de la jerarquía, cada $t \in T_{l+1}$ determina una cadena $t = t_{l+1}, t_l, \dots, t_0$, donde $t_j = \pi_j t_{j+1}$. Para $s, t \in T_{l+1}$, escribimos sus cadenas como $s = s_{l+1}, s_l, \dots, s_0$ y $t = t_{l+1}, t_l, \dots, t_0$. Entonces

$$\left| (X_s - X_{s_0}) - (X_t - X_{t_0}) \right| = \left| \sum_{j=0}^l (X_{s_{j+1}} - X_{s_j}) - \sum_{j=0}^l (X_{t_{j+1}} - X_{t_j}) \right| \leq 2 \sum_{j=0}^l M_j,$$

donde $M_j = \max |X_u - X_v|$, y el máximo se toma sobre todos los eslabones¹² (u, v) con $u \in T_{j+1}$, $v = \pi_j u \in T_j$. Para cada eslabón de nivel j , $\|X_u - X_v\|_\psi \leq C d(u, v) \leq C \eta 2^{-j}$.

Además hay a lo sumo $|T_{j+1}| \leq D(\eta 2^{-j-1})$ eslabones. Por el lema del máximo, $\|M_j\|_\psi \leq A_\psi C \eta 2^{-j} \psi^{-1}\{D(\eta 2^{-j-1})\}$. Por tanto, usando la desigualdad triangular de la norma de Orlicz,¹³

$$B_l := \left\| \max_{s, t \in T_{l+1}} \left| (X_s - X_{s_0}) - (X_t - X_{t_0}) \right| \right\|_\psi \leq K_1 \sum_{j=0}^l \eta 2^{-j} \psi^{-1}\{D(\eta 2^{-j-1})\}.$$

¹¹La construcción se hace por extensión sucesiva. Primero se toma $T_0 \subset F_k$ maximal η -separado. Luego, dado T_j , como T_j es automáticamente $\eta 2^{-j-1}$ -separado, se lo extiende agregando puntos de F_k mientras sea posible mantener la propiedad de ser $\eta 2^{-j-1}$ -separado. Como F_k es finito, el procedimiento se detiene, y el conjunto resultante, maximal en F_k , se llama T_{j+1} .

¹²La igualdad previa es una suma telescópica: $X_s - X_{s_0} = \sum_{j=0}^l (X_{s_{j+1}} - X_{s_j})$, pues $s_{l+1} = s$. Cada término satisface $|X_{s_{j+1}} - X_{s_j}| \leq M_j$, porque $(s_{j+1}, \pi_j s_{j+1})$ es un eslabón de nivel j ; lo mismo para la cadena de t , y de ahí el factor 2.

¹³La cota puntual $\max_{s, t} |(X_s - X_{s_0}) - (X_t - X_{t_0})| \leq 2 \sum_{j=0}^l M_j$ vale porque el miembro derecho no depende del par (s, t) ; luego se aplican la monotonía de $\|\cdot\|_\psi$ y la desigualdad triangular a la suma finita, con $K_1 = 2A_\psi C$.

Ahora comparamos la suma diádica con la integral.¹⁴ La función $f(\varepsilon) = \psi^{-1}\{D(\varepsilon)\}$ es no creciente. Para cada j , $\eta 2^{-j} f(\eta 2^{-j-1}) \leq 4 \int_{\eta 2^{-j-2}}^{\eta 2^{-j-1}} f(\varepsilon) d\varepsilon$. Luego

$$B_l \leq 4K_1 \int_0^\eta \psi^{-1}\{D(\varepsilon)\} d\varepsilon. \quad (1)$$

Pasamos ahora al módulo local. Sea

$$S_l = \max_{\substack{s, t \in T_{l+1} \\ d(s, t) < \delta}} |X_s - X_t|.$$

Para $s, t \in T_{l+1}$ con $d(s, t) < \delta$, $|X_s - X_t| \leq |(X_s - X_{s_0}) - (X_t - X_{t_0})| + |X_{s_0} - X_{t_0}|$. El primer término está controlado por el máximo que define B_l . Queda controlar los extremos $s_0, t_0 \in T_0$.

Para cada par de extremos $(a, b) \in T_0^2$ que provenga de al menos un par local $s, t \in T_{l+1}$ con $d(s, t) < \delta$, elegimos exactamente un par representante $(\bar{s}, \bar{t}) \in T_{l+1}^2$ tal que $d(\bar{s}, \bar{t}) < \delta$ y cuyas cadenas terminan en a, b . Hay a lo sumo $|T_0|^2 \leq D^2(\eta)$ representantes.

Para ese representante, $|X_a - X_b| \leq |(X_{\bar{s}} - X_a) - (X_{\bar{t}} - X_b)| + |X_{\bar{s}} - X_{\bar{t}}|$. El máximo del primer término está de nuevo acotado por el mismo máximo de cadenas que aparece en B_l . Si R_l denota el máximo de $|X_{\bar{s}} - X_{\bar{t}}|$ sobre todos los representantes elegidos, obtenemos la cota determinista $S_l \leq 2 \max_{u, v \in T_{l+1}} |(X_u - X_{u_0}) - (X_v - X_{v_0})| + R_l$. Tomando normas y usando (1),

$$\|S_l\|_\psi \leq 8K_1 \int_0^\eta \psi^{-1}\{D(\varepsilon)\} d\varepsilon + \|R_l\|_\psi. \quad (2)$$

Solo falta acotar R_l . Es el máximo de a lo sumo $D^2(\eta)$ variables. Cada una satisface $\|X_{\bar{s}} - X_{\bar{t}}\|_\psi \leq Cd(\bar{s}, \bar{t}) \leq C\delta$. Por el Lema 2.8, $\|R_l\|_\psi \leq A_\psi C \delta \psi^{-1}\{D^2(\eta)\} \leq K_1 \delta \psi^{-1}\{D^2(\eta)\}$. Sustituyendo esto en (2),

$$\|S_l\|_\psi \leq 8K_1 \int_0^\eta \psi^{-1}\{D(\varepsilon)\} d\varepsilon + K_1 \delta \psi^{-1}\{D^2(\eta)\}.$$

En particular, con $K = 9K_1$,

$$\|S_l\|_\psi \leq K \left[\int_0^\eta \psi^{-1}\{D(\varepsilon)\} d\varepsilon + \delta \psi^{-1}\{D^2(\eta)\} \right].$$

La cota recién obtenida para $\|S_l\|_\psi$ vale para todo nivel l de la jerarquía, en particular para $l = j_0(k)$; combinada con (0), acota el máximo sobre pares de F_k . Definiendo

$$Z_k := \max_{\substack{s, t \in F_k \\ d(s, t) < \delta}} |X_s - X_t|$$

(con $Z_k := 0$ si no hay tales pares), se obtiene, para todo k ,

$$\|Z_k\|_\psi \leq B := K \left[\int_0^\eta \psi^{-1}\{D(\varepsilon)\} d\varepsilon + \delta \psi^{-1}\{D^2(\eta)\} \right],$$

con B independiente de k , porque los números de empaquetamiento usados en la construcción son los de T . Como $F_k \uparrow T'_0$, la sucesión Z_k es no decreciente y

$$Z_k \uparrow Z := \sup_{\substack{s, t \in T'_0 \\ d(s, t) < \delta}} |X_s - X_t|,$$

¹⁴El intervalo $[\eta 2^{-j-2}, \eta 2^{-j-1}]$ tiene longitud $\eta 2^{-j-2}$ y, por monotonía, $f \geq f(\eta 2^{-j-1})$ en él; luego $\int_{\eta 2^{-j-2}}^{\eta 2^{-j-1}} f \geq \frac{1}{4} \eta 2^{-j} f(\eta 2^{-j-1})$. Al sumar sobre $j \geq 0$, los intervalos son disjuntos y están contenidos en $(0, \eta/2] \subset (0, \eta]$, de donde la cota (1).

que, fuera del nulo N introducido al comienzo de la prueba, coincide con $\sup_{d(s,t) < \delta} |X_s - X_t|$. Sea $C_0 > B$.¹⁵ Para cada k , $\mathbb{E}\psi(Z_k/C_0) \leq 1$; como ψ es continua y no decreciente, $\psi(Z_k/C_0) \uparrow \psi(Z/C_0)$, y por convergencia monótona $\mathbb{E}\psi(Z/C_0) \leq 1$, es decir $\|Z\|_\psi \leq C_0$. Haciendo $C_0 \downarrow B$, $\left\| \sup_{d(s,t) < \delta} |X_s - X_t| \right\|_\psi \leq B$. Esto da la desigualdad con $d(s,t) < \delta$. Como se dijo al principio, reemplazando δ por $\delta + \rho$ y dejando $\rho \downarrow 0$, se obtiene la formulación con $d(s,t) \leq \delta$. $\square$

Demostración del Corolario 2.12. Tomamos $\eta = \text{diam } T$, que podemos suponer finito y positivo: si es 0, el supremo es nulo c.s. por separabilidad, y si es ∞ , entonces $D(\varepsilon, d) \geq 2$ para todo ε y la integral diverge. En la construcción anterior (dentro de cada F_k), T_0 puede elegirse con un solo punto, porque no hay dos puntos de T separados por más que $\text{diam } T$. Por tanto, para cualesquiera dos cadenas se tiene $s_0 = t_0$, y desaparece el término de los extremos. La prueba anterior da directamente

$$\left\| \sup_{s,t \in T} |X_s - X_t| \right\|_\psi \leq K \int_0^{\text{diam } T} \psi^{-1}\{D(\varepsilon, d)\} d\varepsilon.$$

Si la integral diverge, la desigualdad se interpreta trivialmente con lado derecho $+\infty$. $\square$

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 $t_0 \in T$,¹⁶

$$\left\| \sup_{t \in T} |X_t| \right\|_\psi \leq \|X_{t_0}\|_\psi + K \int_0^{\text{diam } T} \psi^{-1}(D(\varepsilon, d)) d\varepsilon.$$

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 $\int_0^\eta \psi^{-1}(D(\varepsilon, d)) d\varepsilon < \infty$ para algún $\eta > 0$. En efecto, en el Teorema 2.11 el lado derecho puede hacerse arbitrariamente pequeño eligiendo primero η pequeño y luego δ pequeño.¹⁷ De esta desigualdad puede deducirse la continuidad de casi todas las trayectorias.

2.2.1. Desigualdades subgaussianas

Una variable normal estándar tiene colas del orden $x^{-1}e^{-x^2/2}$ y satisface $\mathbb{P}(|X| > x) \leq 2e^{-x^2/2}$, $x > 0$. Por cálculo directo, su norma ψ_2 , con $\psi_2(x) = e^{x^2} - 1$, es $\sqrt{8/3}$.¹⁸ En esta subsección estudiaremos variables y procesos con cotas de cola similares.

La desigualdad de Hoeffding da una cota subgaussiana para sumas de variables independientes, centradas y acotadas. Usaremos el siguiente caso particular.

Lema 2.13. 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 $x > 0$, $\mathbb{P}(|\sum_{i=1}^n \varepsilon_i a_i| > x) \leq 2 \exp\left\{-\frac{x^2}{2\|a\|^2}\right\}$, donde $\|a\|$ denota la norma euclidiana de $a = (a_1, \dots, a_n)$. En consecuencia, $\|\sum_{i=1}^n \varepsilon_i a_i\|_{\psi_2} \leq \sqrt{6} \|a\|$.

Definición 2.14. 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 $x > 0$, $\mathbb{P}(|X_s - X_t| > x) \leq 2 \exp\left\{-\frac{x^2}{2d^2(s,t)}\right\}$, 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.

¹⁵Renombramos la constante como C_0 para no confundirla con la constante de Lipschitz C del enunciado. La desigualdad que sigue vale para todo $C_0 > \|Z_k\|_\psi$: por definición del ínfimo existe $C' \in (\|Z_k\|_\psi, C_0]$ con $\mathbb{E}\psi(Z_k/C') \leq 1$, y la monotonía de ψ permite pasar de C' a C_0 .

¹⁶La desigualdad se obtiene de $|X_t| \leq |X_{t_0}| + \sup_{s,u} |X_s - X_u|$, aplicando la monotonía y la desigualdad triangular de la norma de Orlicz, y luego el Corolario 2.12 al término del supremo.

¹⁷Primero 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.

¹⁸Para $X \sim N(0, 1)$ 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 $C^2 = 8/3$.

Ejemplo 2.15. ■ 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 2.13, este proceso es subgaussiano para la distancia euclidiana $d(a, b) = \|a - b\|$.

Observación 2.16. Los procesos subgaussianos satisfacen la cota de incrementos $\|X_s - X_t\|_{\psi_2} \leq \sqrt{6} d(s, t)$: aplicando el Lema 2.7 con $X = X_s - X_t$, $p = 2$, $K = 2$ y $C = (2d^2(s, t))^{-1}$, se obtiene $\|X_s - X_t\|_{\psi_2} \leq \left(\frac{1+2}{(2d^2(s, t))^{-1}}\right)^{1/2} = \sqrt{6} d(s, t)$.

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 2.17 (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 $\delta > 0$, $\mathbb{E} \sup_{d(s, t) \leq \delta} |X_s - X_t| \leq K \int_0^\delta \sqrt{\log D(\varepsilon, d)} d\varepsilon$, donde K es una constante universal. En particular, para todo $t_0 \in T$,*

$$\mathbb{E} \sup_{t \in T} |X_t| \leq \mathbb{E} |X_{t_0}| + K \int_0^\infty \sqrt{\log D(\varepsilon, d)} d\varepsilon. \quad (2.2)$$

Demostración. Si la integral del lado derecho es infinita, no hay nada que probar. Supongamos entonces que es finita. Si $\text{diam}(T) = 0$, entonces $d(s, t) = 0$ para todo par y , por la hipótesis subgaussiana, $X_s = X_t$ c.s. para cada par fijo; por la separabilidad del proceso, el supremo del lado izquierdo coincide c.s. con un supremo sobre un conjunto numerable de pares, luego es nulo c.s., y la conclusión es inmediata. Suponemos entonces $\text{diam}(T) > 0$. Podemos suponer además $\delta \leq \text{diam}(T)$: como $d(s, t) \leq \text{diam}(T)$ para todo par, si $\delta > \text{diam}(T)$ el supremo del lado izquierdo coincide con el correspondiente a $\bar{\delta} = \text{diam}(T)$, mientras que la integral hasta δ domina a la integral hasta $\bar{\delta}$. Como el proceso es subgaussiano, sus incrementos satisfacen $\|X_s - X_t\|_{\psi_2} \leq C d(s, t)$, $\psi_2(x) = e^{x^2} - 1$, para una constante universal C . Aplicamos la desigualdad maximal del Teorema 2.11 con $\psi = \psi_2$ y $\eta = \delta$. Como $\psi_2^{-1}(m) = \sqrt{\log(1+m)}$, obtenemos

$$\left\| \sup_{d(s, t) \leq \delta} |X_s - X_t| \right\|_{\psi_2} \leq K \left\{ \int_0^\delta \sqrt{\log(1 + D(\varepsilon, d))} d\varepsilon + \delta \sqrt{\log(1 + D^2(\delta, d))} \right\}.$$

El segundo término queda absorbido por el primero.¹⁹ En efecto, para $0 < \varepsilon \leq \delta$ se tiene $D(\varepsilon, d) \geq D(\delta, d)$, y además $\sqrt{\log(1 + D^2(\delta, d))} \leq C \sqrt{\log(1 + D(\delta, d))}$. Por tanto,

$$\left\| \sup_{d(s, t) \leq \delta} |X_s - X_t| \right\|_{\psi_2} \leq K \int_0^\delta \sqrt{\log(1 + D(\varepsilon, d))} d\varepsilon.$$

Si $0 < \varepsilon < \text{diam}(T)$, entonces $D(\varepsilon, d) \geq 2$, y de ahí $\log(1 + D(\varepsilon, d)) \leq 2 \log D(\varepsilon, d)$.²⁰ Como $\delta \leq \text{diam}(T)$, esta comparación vale para todo $\varepsilon \in (0, \delta)$. Aumentando la constante K ,

$$\left\| \sup_{d(s, t) \leq \delta} |X_s - X_t| \right\|_{\psi_2} \leq K \int_0^\delta \sqrt{\log D(\varepsilon, d)} d\varepsilon.$$

Finalmente, $\|Z\|_1 \leq C \|Z\|_{\psi_2}$ para una constante universal C , y por²¹ eso la misma cota vale para la esperanza.

Para probar (2.2), fijamos $t_0 \in T$ y usamos $\sup_{t \in T} |X_t| \leq |X_{t_0}| + \sup_{t \in T} |X_t - X_{t_0}| \leq |X_{t_0}| + \sup_{s, t \in T} |X_s - X_t|$. Aplicando la cota anterior con $\delta = \text{diam}(T)$, o extendiendo la integral hasta ∞ porque $D(\varepsilon, d) = 1$ para $\varepsilon \geq \text{diam}(T)$, obtenemos (2.2). $\square$

¹⁹Dos observaciones: para $D \geq 0$, $1 + D^2 \leq (1 + D)^2$, y por tanto $\log(1 + D^2) \leq \log((1 + D)^2) = 2 \log(1 + D)$. Además, $\delta \sqrt{\log(1 + D(\delta, d))} \leq \int_0^\delta \sqrt{\log(1 + D(\varepsilon, d))} d\varepsilon$, por la monotonía de $\varepsilon \mapsto D(\varepsilon, d)$.

²⁰Si $\varepsilon < \text{diam}(T)$, hay dos puntos a distancia $> \varepsilon$, luego $D(\varepsilon, d) \geq 2$; y para $D \geq 2$ vale $1 + D \leq D^2$, de donde $\log(1 + D) \leq 2 \log D$.

²¹Por ejemplo con $C = 2$: $u \leq \psi_2(u) + 1$ para todo $u \geq 0$ (si $u \geq 1$, $u \leq u^2 \leq e^{u^2} - 1$; si $u < 1$, $u < 1$); así $\mathbb{E}|Z| \leq C'(\mathbb{E}\psi_2(|Z|/C') + 1) \leq 2C'$ para todo $C' > \|Z\|_{\psi_2}$.

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 $(P_n - P)f = \frac{1}{n} \sum_{i=1}^n \{f(X_i) - Pf\}$, $P_n^\circ f = \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i)$. 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 2.17.

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. Aislamos primero un hecho sobre mayorantes medibles y transformaciones convexas monótonas que se usará varias veces.

Lema 2.18. *Sea $\Phi : [0, \infty) \rightarrow [0, \infty)$ convexa, no decreciente y continua, extendida con la convención $\Phi(+\infty) := \lim_{t \rightarrow \infty} \Phi(t)$, y sea $H : \Omega \rightarrow [0, \infty]$ una función arbitraria en un espacio de probabilidad. Entonces $\Phi(H)^* = \Phi(H^*)$ c.s.*

Demostración. Una desigualdad sigue de la minimalidad: $\Phi(H^*)$ es medible y domina a $\Phi(H)$ por monotonía, luego $\Phi(H)^* \leq \Phi(H^*)$ c.s. Para la recíproca, si Φ es constante la igualdad es inmediata; supongamos que no lo es, con lo cual, por convexidad y monotonía, $\Phi(t) \rightarrow \infty$. Sea $W := \Phi(H)^*$. En $\{W < \infty\}$, la inversa generalizada $g := \sup\{u \geq 0 : \Phi(u) \leq W\}$ es finita y medible (sus conjuntos de nivel se describen con W vía la monotonía de Φ), satisface $g \geq H$ punto a punto (pues $\Phi(H) \leq W$, por el Lema 1.4(i), y en particular $H < \infty$ allí) y $\Phi(g) \leq W$, por continuidad de Φ . Extendiendo $g := +\infty$ en $\{W = \infty\}$, la minimalidad da $H^* \leq g$ c.s., y entonces $\Phi(H^*) \leq \Phi(g) \leq W$ c.s. en $\{W < \infty\}$; en $\{W = \infty\}$ la desigualdad es trivial. $\square$

Lema 2.19. *Sea $\Phi : \mathbb{R} \rightarrow \mathbb{R}$ convexa, no decreciente y finita, y sea $\mathcal{F}$ una clase de funciones medibles con $P|f| < \infty$ para todo $f \in \mathcal{F}$. Entonces*

$$\mathbb{E}^* \Phi(\|P_n - P\|_{\mathcal{F}}) \leq \mathbb{E}^* \Phi(2\|P_n^\circ\|_{\mathcal{F}}).$$

Demostración. Como los argumentos de Φ son no negativos, podemos reemplazar Φ por $\Phi - \Phi(0)$. La desigualdad resultante es equivalente a la original.²² Así, sin pérdida de generalidad, suponemos que $\Phi \geq 0$ en $[0, \infty)$.

Sea $Y_1, \dots, Y_n$ una copia independiente de $X_1, \dots, X_n$, definida en un factor producto adicional. Añadir este factor no cambia las esperanzas exteriores de las variables que dependen solo de $X_1, \dots, X_n$, porque las proyecciones coordenadas del espacio producto son perfectas.²³

Para $x = (x_1, \dots, x_n)$ fijo, definimos

$$H_x(y) := \left\| \frac{1}{n} \sum_{i=1}^n \{f(x_i) - f(y_i)\} \right\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n \{f(x_i) - f(y_i)\} \right|.$$

²²Porque $\mathbb{E}^*(\Phi(Z) - \Phi(0)) = \mathbb{E}^* \Phi(Z) - \Phi(0)$: la esperanza exterior se traslada por constantes. Y $\Phi - \Phi(0) \geq 0$ en $[0, \infty)$ por la monotonía de Φ .

²³Este es exactamente el uso anunciado en el Capítulo 1: por el Lema 1.19(i), $(g \circ \pi)^* = g^* \circ \pi$ c.s. para la proyección π sobre el bloque de las X , cualquiera sea g ; tomando esperanzas, $\mathbb{E}^*(g \circ \pi) = \mathbb{E}_X^* g$ siempre que alguna de las dos exista (aquí los integrandos son no negativos).

Entonces, usando que $Pf = \mathbb{E}_Y f(Y_i)$, para cada $f \in \mathcal{F}$,²⁴

$$\left| \frac{1}{n} \sum_{i=1}^n \{f(x_i) - Pf\} \right| = \left| \mathbb{E}_Y \frac{1}{n} \sum_{i=1}^n \{f(x_i) - f(Y_i)\} \right| \leq \mathbb{E}_Y^* H_X(Y).$$

Al tomar supremo en f , obtenemos $\|P_n - P\|_{\mathcal{F}} \leq \mathbb{E}_Y^* H_X(Y)$, donde $X = (X_1, \dots, X_n)$.

Ahora aplicamos Jensen: sea H_X^{*Y} el menor mayorante medible de $H_X(Y)$ respecto de la variable Y , manteniendo X fijo. Entonces²⁵ $\mathbb{E}_Y^* H_X(Y) = \mathbb{E}_Y H_X^{*Y}$. Como Φ es convexa, finita y no decreciente, es continua, y por tanto

$$\Phi(\|P_n - P\|_{\mathcal{F}}) \leq \Phi(\mathbb{E}_Y H_X^{*Y}) \leq \mathbb{E}_Y \Phi(H_X^{*Y}) = \mathbb{E}_Y^* \Phi\left(\left\|\frac{1}{n} \sum_{i=1}^n \{f(X_i) - f(Y_i)\}\right\|_{\mathcal{F}}\right).$$

En la última igualdad usamos que el mayorante medible mínimo de $\Phi(H_X)$ respecto de Y es $\Phi(H_X^{*Y})$: es el Lema 2.18, aplicado en el espacio de Y con X fijo (tras la reducción inicial, Φ es no negativa, convexa, no decreciente y continua en $[0, \infty)$).

Tomando esperanza exterior respecto de X , y usando que la esperanza exterior conjunta domina a la iterada (Lema 1.15: $\mathbb{E}_X^* \mathbb{E}_Y^* \leq \mathbb{E}_{X,Y}^*$), se obtiene

$$\mathbb{E}^* \Phi(\|P_n - P\|_{\mathcal{F}}) \leq \mathbb{E}_{X,Y}^* \Phi\left(\left\|\frac{1}{n} \sum_{i=1}^n \{f(X_i) - f(Y_i)\}\right\|_{\mathcal{F}}\right). \quad (1)$$

Introducimos ahora los signos. Para cada $e = (e_1, \dots, e_n) \in \{-1, 1\}^n$, pongamos $H_e(X, Y) := \left\|\frac{1}{n} \sum_{i=1}^n e_i \{f(X_i) - f(Y_i)\}\right\|_{\mathcal{F}}$. Si $e_i = -1$, cambiar el signo del i -ésimo sumando equivale a intercambiar X_i con Y_i . Como $(X_1, Y_1), \dots, (X_n, Y_n)$ viven en un espacio producto y X_i e Y_i tienen la misma ley, la esperanza exterior conjunta es invariante bajo esos intercambios.²⁶ Por tanto, para todo e , $\mathbb{E}_{X,Y}^* \Phi(H_1(X, Y)) = \mathbb{E}_{X,Y}^* \Phi(H_e(X, Y))$, donde H_1 corresponde a $e_i = 1$ para todo i .

Promediando sobre $e \in \{-1, 1\}^n$, o equivalentemente sobre variables de Rademacher $\varepsilon_1, \dots, \varepsilon_n$, la desigualdad (1) da

$$\mathbb{E}^* \Phi(\|P_n - P\|_{\mathcal{F}}) \leq \mathbb{E}_{\varepsilon} \mathbb{E}_{X,Y}^* \Phi\left(\left\|\frac{1}{n} \sum_{i=1}^n \varepsilon_i \{f(X_i) - f(Y_i)\}\right\|_{\mathcal{F}}\right). \quad (2)$$

Para e fijo, por la desigualdad triangular,

$$H_e(X, Y) \leq \left\|\frac{1}{n} \sum_{i=1}^n e_i f(X_i)\right\|_{\mathcal{F}} + \left\|\frac{1}{n} \sum_{i=1}^n e_i f(Y_i)\right\|_{\mathcal{F}} =: A_e(X) + B_e(Y).$$

Como Φ es no decreciente y convexa,

$$\Phi(H_e(X, Y)) \leq \Phi(A_e(X) + B_e(Y)) = \Phi\left(\frac{2A_e(X) + 2B_e(Y)}{2}\right) \leq \frac{1}{2} \Phi(2A_e(X)) + \frac{1}{2} \Phi(2B_e(Y)).$$

²⁴La hipótesis $P|f| < \infty$ garantiza que $\mathbb{E}_Y f(Y_i)$ existe y es finita, y permite usar la linealidad; luego $|\mathbb{E}_Y g| \leq \mathbb{E}_Y |g|$ y, como $|g| \leq H_x$ punto a punto con g medible en y , la monotonía de la esperanza exterior da $\mathbb{E}_Y |g| \leq \mathbb{E}_Y^* H_x(Y)$.

²⁵Como $H_X \geq 0$, también $H_X^{*Y} \geq 0$ c.s., su esperanza (extendida) existe, y la igualdad es el Lema 1.4(iii), aplicado en el espacio de Y con X fijo. Si $\mathbb{E}_Y H_X^{*Y} = +\infty$, la desigualdad de Jensen del paso siguiente se interpreta con la convención $\Phi(+\infty) := \lim_{t \rightarrow \infty} \Phi(t)$ y es trivial; si es finita, H_X^{*Y} es integrable y Jensen se aplica en su forma usual.

²⁶Con precisión: el intercambio τ_e , que permuta $x_i \leftrightarrow y_i$ exactamente en las coordenadas con $e_i = -1$, es una biyección bimedible del producto que preserva la medida (los pares (X_i, Y_i) son independientes y cada par tiene coordenadas i.i.d., luego intercambiables). Para una τ así, $\mathbb{E}^*(g \circ \tau) = \mathbb{E}^* g$ para toda g : la correspondencia $U \mapsto U \circ \tau$ es una biyección entre los mayorantes medibles de g y los de $g \circ \tau$ que preserva esperanzas. Finalmente, $H_e = H_1 \circ \tau_e$ punto a punto, pues cambiar el signo del sumando i equivale a intercambiar $f(x_i)$ con $f(y_i)$.

²⁷La esperanza en ε es ordinaria, pues sólo promedia sobre el conjunto finito $\{-1, 1\}^n$ los números $\mathbb{E}_{X,Y}^* \Phi(H_e(X, Y))$. La esperanza exterior queda únicamente respecto de (X, Y) , donde el supremo sobre $\mathcal{F}$ puede no ser medible.

Tomando esperanza exterior conjunta y usando nuevamente que las proyecciones coordenadas son perfectas,²⁸

$$\mathbb{E}_{X,Y}^* \Phi(H_e(X, Y)) \leq \frac{1}{2} \mathbb{E}_X^* \Phi(2A_e(X)) + \frac{1}{2} \mathbb{E}_Y^* \Phi(2B_e(Y)) = \mathbb{E}_X^* \Phi(2A_e(X)),$$

porque X e Y tienen la misma distribución. Sustituyendo esto en (2),

$$\mathbb{E}^* \Phi(\|P_n - P\|_{\mathcal{F}}) \leq \mathbb{E}_\varepsilon \mathbb{E}_X^* \Phi\left(2 \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}}\right). \quad (3)$$

Finalmente, la esperanza exterior conjunta domina a la iterada.²⁹ Aplicada al espacio producto de X y ε , la desigualdad (3) implica

$$\mathbb{E}^* \Phi(\|P_n - P\|_{\mathcal{F}}) \leq \mathbb{E}^* \Phi\left(2 \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}}\right) = \mathbb{E}^* \Phi(2\|P_n^\circ\|_{\mathcal{F}}).$$

□

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 $X_1, \dots, X_n$. 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.

Como las variables de Rademacher toman sólo finitos valores, la medibilidad conjunta de $(x_1, \dots, x_n, \varepsilon_1, \dots, \varepsilon_n) \mapsto \left\| \sum_{i=1}^n \varepsilon_i f(x_i) \right\|_{\mathcal{F}}$ equivale a la medibilidad, para todo $e = (e_1, \dots, e_n) \in \{-1, 1\}^n$, de

$$(x_1, \dots, x_n) \mapsto \left\| \sum_{i=1}^n e_i f(x_i) \right\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} \left| \sum_{i=1}^n e_i f(x_i) \right|. \quad (2.3.2)$$

Para esta aplicación concreta de Fubini basta considerar vectores de signos. La Definición 2.3.3 de [vdV–W], que adoptamos a continuación, exige la misma medibilidad para todo vector real y es, por tanto, ligeramente más fuerte.

Definición 2.20. Sea $\mathcal{F}$ una clase de funciones medibles $f : \mathcal{X} \rightarrow \mathbb{R}$ 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 $(x_1, \dots, x_n) \mapsto \sup_{f \in \mathcal{F}} |\sum_{i=1}^n e_i f(x_i)|$ es medible respecto de la completación de $(\mathcal{X}^n, \mathcal{A}^n, P^n)$.

En particular, una clase P -medible satisface (2.3.2) y, con ella, la condición necesaria para aplicar Fubini al proceso de Rademacher y escribir³⁰

$$\mathbb{E}_{X,\varepsilon}^* \left\| \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}} = \mathbb{E}_X \mathbb{E}_\varepsilon^* \left\| \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}}.$$

²⁸Dos pasos: primero la subaditividad de $\mathbb{E}_{X,Y}^*$ sobre la suma $\frac{1}{2} \Phi(2A_e(X)) + \frac{1}{2} \Phi(2B_e(Y))$; luego, como cada término depende de un solo bloque de coordenadas, el Lema 1.19(i) (válido para funciones arbitrarias) da, tomando esperanzas, $\mathbb{E}_{X,Y}^* \Phi(2A_e(X)) = \mathbb{E}_X^* \Phi(2A_e(X))$, y análogamente para $B_e(Y)$. La igualdad final usa que X e Y tienen la misma ley: $\Phi(2B_e)$ es la misma función de una muestra con la misma distribución, así que $\mathbb{E}_Y^* \Phi(2B_e(Y)) = \mathbb{E}_X^* \Phi(2A_e(X))$.

²⁹De nuevo el Lema 1.15, ahora en el producto $\mathcal{X}^n \times \{-1, 1\}^n$: como el factor de los signos es finito, $\mathbb{E}_\varepsilon = \mathbb{E}_\varepsilon^*$, y $\mathbb{E}_\varepsilon \mathbb{E}_X^* \leq \mathbb{E}_{X,\varepsilon}^*$.

³⁰Con la medibilidad respecto de la completación, el supremo simetrizado coincide c.s. con una versión medible para la σ -álgebra producto, y la esperanza exterior es la ordinaria. El teorema de Fubini–Tonelli (integrando no negativo) se aplica a esa versión: las secciones y las integrales iteradas quedan definidas fuera de los conjuntos nulos usuales y producen la misma esperanza.

2.4. Teoremas de Glivenko–Cantelli

2.4.1. Glivenko–Cantelli por bracketing

Definición 2.21. Sean $l, u : \mathcal{X} \rightarrow \mathbb{R}$ medibles, con $l \leq u$. El **bracket** $[l, u]$ es la clase de funciones f tales que $l \leq f \leq u$. 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 2.22. 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, $\|P_n - P\|_{\mathcal{F}}^* \rightarrow 0$ 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 $(P_n - P)f \leq (P_n - P)u_i + P(u_i - f) \leq (P_n - P)u_i + \varepsilon$, pues $0 \leq u_i - f \leq u_i - l_i$.³¹ Análogamente, $(P_n - P)f \geq (P_n - P)l_i - P(f - l_i) \geq (P_n - P)l_i - \varepsilon$. Por tanto, para todo $f \in \mathcal{F}$, $|(P_n - P)f| \leq \max_{1 \leq i \leq m} \max\{|(P_n - P)u_i|, |(P_n - P)l_i|\} + \varepsilon$. El máximo del lado derecho es medible y³², por la ley fuerte aplicada al conjunto finito de extremos $\{l_i, u_i : i = 1, \dots, m\}$, converge c.s. a 0. Así, $\limsup_{n \rightarrow \infty} \|P_n - P\|_{\mathcal{F}}^* \leq \varepsilon$ c.s. Como $\varepsilon > 0$ es arbitrario, tomando una sucesión $\varepsilon_k \downarrow 0$ se obtiene $\|P_n - P\|_{\mathcal{F}}^* \rightarrow 0$ 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. $\square$

Ejemplo 2.23. 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 $-\infty = t_0 < t_1 < \dots < t_m = \infty$ tal que $P((t_i, t_{i+1})) < \varepsilon$ para todo i .³³ Entonces los brackets $[\mathbb{1}_{(-\infty, t_i]}, \mathbb{1}_{(-\infty, t_{i+1})}]$, $i = 0, \dots, m-1$, cubren $\mathcal{F}$ y tienen tamaño $\|\mathbb{1}_{(-\infty, t_{i+1})} - \mathbb{1}_{(-\infty, t_i)}\|_{P,1} = P((t_i, t_{i+1})) < \varepsilon$. Luego $N_{[]}(\varepsilon, \mathcal{F}, L_1(P)) < \infty$, y por el Teorema 2.22, $\mathcal{F}$ 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 2.24. Una **envolvente** de una clase $\mathcal{F}$ es una función $F : \mathcal{X} \rightarrow [0, \infty]$ tal que $|f(x)| \leq F(x)$ 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 2.25. Si Q es una medida de probabilidad, escribimos $\|f - g\|_{Q,1} = Q|f - g|$. El **número de cubrimiento** $N(\varepsilon, \mathcal{F}, L_1(Q))$ es el menor número de bolas de radio ε para la semimétrica $\|\cdot\|_{Q,1}$ que cubren $\mathcal{F}$. En particular, $N(\varepsilon, \mathcal{F}, L_1(P_n)) = N(\varepsilon, \mathcal{F}, d_n)$, $d_n(f, g) = P_n|f - g|$.

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: 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.

³¹En detalle: $P_n f \leq P_n u_i$ por monotonía de P_n (los extremos son medibles e integrables, parte de la definición de bracket), de donde $(P_n - P)f \leq (P_n - P)u_i + P(u_i - f)$, y $P(u_i - f) \leq P(u_i - l_i) < \varepsilon$.

³²Por eso la minimalidad del mayorante da $\|P_n - P\|_{\mathcal{F}}^* \leq \max_i \{\dots\} + \varepsilon$ c.s. para cada n ; la unión en n de los conjuntos excepcionales sigue siendo nula.

³³Esto no exige continuidad de P . Con $K = \lceil 1/\varepsilon \rceil$ y los cortes cuantílicos $t_i := \inf\{t : F(t) \geq i/(K+1)\}$, $i = 1, \dots, K$, donde F es la función de distribución de P , la masa *abierto* entre cortes consecutivos es $P((t_i, t_{i+1})) \leq 1/(K+1) < \varepsilon$: por la continuidad a derecha de F , $F(t_i) \geq i/(K+1)$, mientras que $F(t_{i+1}^-) \leq (i+1)/(K+1)$. Los átomos, en particular los de masa $\geq \varepsilon$, quedan absorbidos en los puntos de corte; si algunos t_i coinciden, se descartan las repeticiones.

Definición 2.26. 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 $\mathbb{E}(Y_n | \mathcal{G}_{n+1}) = Y_{n+1}$ c.s. Es una **submartingala inversa** si $\mathbb{E}(Y_n | \mathcal{G}_{n+1}) \geq Y_{n+1}$ c.s., y una **supermartingala inversa** si la desigualdad es la opuesta.

Lema 2.27. Sea $\mathcal{F}$ una clase de funciones medibles con envolvente F tal que $P^*F < \infty$. Trabajamos en el producto $(\mathcal{X}^\infty, \mathcal{A}^\infty, P^\infty)$ y denotamos por Σ_n la σ -álgebra generada por todas las funciones medibles $h : \mathcal{X}^\infty \rightarrow \mathbb{R}$ invariantes por permutaciones de sus primeros n argumentos. Entonces $\mathbb{E}(\|P_n - P\|_{\mathcal{F}}^* | \Sigma_{n+1}) \geq \|P_{n+1} - P\|_{\mathcal{F}}^*$ c.s. Además, pueden elegirse versiones medibles y simétricas de los mayorantes medibles mínimos, y por tanto Σ_n -medibles; con esa elección forman una submartingala inversa y convergen c.s. a una constante.

Demostración. Podemos suponer sin pérdida que $Pf = 0$ para todo $f \in \mathcal{F}$: basta reemplazar $\mathcal{F}$ por $\{f - Pf : f \in \mathcal{F}\}$. La envolvente sigue siendo integrable exteriormente, por ejemplo $F + P^*F$.³⁴

Sea Σ_n la σ -álgebra generada por las funciones medibles simétricas en las primeras n coordenadas. Para cada n , elegimos una versión Σ_n -medible y simétrica del mayorante medible mínimo de $\|P_n\|_{\mathcal{F}}$, y la denotamos por $Y_n := \|P_n\|_{\mathcal{F}}^*$. Esto es posible porque la función $(x_1, \dots, x_n) \mapsto \left\| n^{-1} \sum_{j=1}^n f(x_j) \right\|_{\mathcal{F}}$ es simétrica, y su mayorante medible puede elegirse simétrico.³⁵ Además,

$$\mathbb{E}Y_n = \mathbb{E}^*\|P_n\|_{\mathcal{F}} \leq \mathbb{E}^*P_n(F + P^*F) \leq 2P^*F < \infty.³⁶$$

Para $i = 1, \dots, n+1$, sea $P_n^{(i)}$ la medida empírica que deja fuera X_i : $P_n^{(i)} := n^{-1} \sum_{j \neq i}^{n+1} \delta_{X_j}$. Como, para todo f , $P_{n+1}f = (n+1)^{-1} \sum_{i=1}^{n+1} P_n^{(i)}f$, tenemos

$$\|P_{n+1}\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} \left| (n+1)^{-1} \sum_{i=1}^{n+1} P_n^{(i)}f \right| \leq (n+1)^{-1} \sum_{i=1}^{n+1} \sup_{f \in \mathcal{F}} |P_n^{(i)}f| = (n+1)^{-1} \sum_{i=1}^{n+1} \|P_n^{(i)}\|_{\mathcal{F}}.$$

Ahora reemplazamos cada término del lado derecho por su mayorante medible: $\|P_n^{(i)}\|_{\mathcal{F}} \leq \|P_n^{(i)}\|_{\mathcal{F}}^*$ c.s. Por tanto, $\|P_{n+1}\|_{\mathcal{F}} \leq R_{n+1} := (n+1)^{-1} \sum_{i=1}^{n+1} \|P_n^{(i)}\|_{\mathcal{F}}^*$ c.s. El lado derecho R_{n+1} es medible, luego es un mayorante medible de $\|P_{n+1}\|_{\mathcal{F}}$. Como $Y_{n+1} = \|P_{n+1}\|_{\mathcal{F}}^*$ es el menor mayorante medible, $Y_{n+1} \leq R_{n+1} = (n+1)^{-1} \sum_{i=1}^{n+1} \|P_n^{(i)}\|_{\mathcal{F}}^*$ c.s.

Como Y_{n+1} es Σ_{n+1} -medible, al condicionar respecto de Σ_{n+1} ,

$$Y_{n+1} \leq (n+1)^{-1} \sum_{i=1}^{n+1} \mathbb{E}(\|P_n^{(i)}\|_{\mathcal{F}}^* | \Sigma_{n+1}) \quad \text{c.s.}$$

Por simetría, los $n+1$ términos condicionales son iguales: dado Σ_{n+1} , ninguna de las $n+1$ coordenadas tiene un rol distinguido.³⁷ Luego el promedio es cualquiera de ellos. Tomando el término que deja fuera X_{n+1} , queda $Y_{n+1} \leq \mathbb{E}(Y_n | \Sigma_{n+1})$ c.s. Por tanto (Y_n, Σ_n) es una submartingala inversa.

Como $Y_n \geq 0$ y $\sup_n \mathbb{E}Y_n \leq 2P^*F < \infty$, el teorema de convergencia para submartingalas inversas da $Y_n \rightarrow Y_\infty$ c.s.³⁸ Además, Y_∞ es medible respecto de $\cap_n \Sigma_n$, que es trivial por la ley 0-1 de

³⁴Para $f \in \mathcal{F}$: $|Pf| \leq P|f| \leq P^*F$ (monotonía de la integral exterior y $|f| \leq F$), de modo que $|f - Pf| \leq F + P^*F$, cuya integral exterior es $\leq 2P^*F < \infty$.

³⁵Sea $\tilde{h}$ un menor mayorante medible de la función simétrica h , y defínase $\bar{h}(x_1, \dots, x_n) := \min_{\pi \in \mathfrak{S}_n} \tilde{h}(x_{\pi(1)}, \dots, x_{\pi(n)})$. Entonces $\bar{h}$ es medible, simétrica y domina a h , pues cada $\tilde{h} \circ \pi$ domina a $h \circ \pi = h$. Por incluir $\pi = \text{id}$, $\bar{h} \leq \tilde{h}$; y por la minimalidad de $\tilde{h}$ (Lema 1.4(ii), siendo $\bar{h}$ un mayorante medible), $\tilde{h} \leq \bar{h}$ c.s. Así, $\bar{h}$ es una versión simétrica del menor mayorante medible.

³⁶La primera desigualdad es $\|P_n\|_{\mathcal{F}} \leq P_n(F + P^*F)$ punto a punto, pues tras el centrado la envolvente de la clase es $F + P^*F$ (no la F original). Para la segunda: $P_n(F + P^*F) = P_nF + P^*F$, con $P_nF = n^{-1} \sum_i F(X_i)$, y la subaditividad de $\mathbb{E}^*$ junto con la perfección de las proyecciones (cada $F(X_i)$ depende de una sola coordenada; Lema 1.19) dan $\mathbb{E}^*P_nF \leq n^{-1} \sum_i \mathbb{E}^*F(X_i) = P^*F$.

³⁷Con precisión: la transposición τ_i de las coordenadas i y $n+1$ preserva P^∞ y fija las funciones Σ_{n+1} -medibles; además $\|P_n^{(i)}\|_{\mathcal{F}}^* = \|P_n^{(n+1)}\|_{\mathcal{F}}^* \circ \tau_i$ c.s. (el mayorante medible mínimo conmuta con biyecciones bimedibles que preservan la medida). De la caracterización de la esperanza condicional se sigue $\mathbb{E}(\|P_n^{(i)}\|_{\mathcal{F}}^* | \Sigma_{n+1}) = \mathbb{E}(\|P_n^{(n+1)}\|_{\mathcal{F}}^* | \Sigma_{n+1})$ c.s.

³⁸El teorema de convergencia para submartingalas inversas afirma que una submartingala inversa no negativa con $\sup_n \mathbb{E}Y_n < \infty$ converge casi seguramente. En el presente caso también converge en L_1 : iterando la desigualdad inversa se obtiene $Y_n \leq \mathbb{E}(Y_1 | \Sigma_n)$, y la familia de esperanzas condicionales de la variable integrable Y_1 es uniformemente integrable, de modo que las Y_n también lo son. Véase, p.ej., Dudley, *Uniform Central Limit Theorems*, o [vdV-W, §2.4].

Hewitt–Savage.³⁹ Así, Y_∞ es constante c.s. Deshaciendo el centrado se obtiene el resultado para⁴⁰
 $\|P_n - P\|_{\mathcal{F}}^*$. $\square$

Teorema 2.28. *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) $\|P_n - P\|_{\mathcal{F}}^* \rightarrow 0$ 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. Antes de comenzar, reducimos la truncación al caso de una envolvente medible. Sea H el mayorante medible mínimo de la envolvente F . Entonces $H \geq F$ punto a punto y $PH = P^*F < \infty$. Para $\mathcal{H}_M := \{f\mathbb{1}_{\{H \leq M\}} : f \in \mathcal{F}\}$ se tiene $\mathcal{H}_M = \{g\mathbb{1}_{\{H \leq M\}} : g \in \mathcal{F}_M\}$, porque $H \leq M$ implica $F \leq M$. Multiplicar todos los elementos por el mismo indicador es una contracción en $L_1(P_n)$; por tanto,⁴¹ $N(\varepsilon, \mathcal{H}_M, L_1(P_n)) \leq N(\varepsilon, \mathcal{F}_M, L_1(P_n))$. La condición de entropía se conserva al reemplazar F por H . Además, $\mathcal{H}_M$ es P -medible: se particiona $\mathcal{X}^n$ según el subconjunto de índices para los que $H(x_i) \leq M$, y en cada pieza la aplicación de la Definición 2.20 aparece con un vector fijo de coeficientes, algunos de ellos iguales a cero.⁴² En adelante relabelamos H como F y $\mathcal{H}_M$ como $\mathcal{F}_M$.

Si procedemos como en la prueba del Lema 2.19 y usamos la P -medibilidad de $\mathcal{F}$, el término simetrizado es medible en (X, ε) . Por tanto, Fubini permite reemplazar la esperanza exterior conjunta por una esperanza iterada ordinaria.⁴³

$$\mathbb{E}^* \|P_n - P\|_{\mathcal{F}} \leq 2\mathbb{E}_X \mathbb{E}_\varepsilon \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}}.$$

Fijado $M > 0$, escribimos $f = f\mathbb{1}_{\{F \leq M\}} + f\mathbb{1}_{\{F > M\}}$. Entonces⁴⁴

$$\left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}} \leq \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}_M} + \frac{1}{n} \sum_{i=1}^n F(X_i) \mathbb{1}_{\{F(X_i) > M\}}.$$

Tomando esperanza, obtenemos

$$\mathbb{E}^* \|P_n - P\|_{\mathcal{F}} \leq 2\mathbb{E}_X \mathbb{E}_\varepsilon \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}_M} + 2P^*(F\mathbb{1}_{\{F > M\}}).⁴⁵$$

Como $P^*F < \infty$, se tiene $P^*(F\mathbb{1}_{\{F > M\}}) \rightarrow 0$ cuando $M \rightarrow \infty$.⁴⁶ Basta probar, para M fijo, que $\mathbb{E}_X \mathbb{E}_\varepsilon \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}_M} \rightarrow 0$.

³⁹Ley 0–1 de Hewitt–Savage: si $X_1, X_2, \dots$ son i.i.d., todo evento invariante bajo permutaciones finitas de las coordenadas tiene probabilidad 0 o 1. Equivalentemente, $\cap_n \Sigma_n$ es P^∞ -trivial.

⁴⁰ $(P_n - P)f = P_n(f - Pf)$, de modo que $\|P_n - P\|_{\mathcal{F}}$ coincide, como función de la muestra, con $\|P_n\|_{\mathcal{F}_c}$ para la clase centrada $\mathcal{F}_c = \{f - Pf\}$; las versiones Y_n elegidas sirven simultáneamente para ambas.

⁴¹ $|g\mathbb{1}_A - h\mathbb{1}_A| \leq |g - h|$ punto a punto; toda ε -red de $\mathcal{F}_M$ induce, multiplicando los centros por $\mathbb{1}_A$, una ε -red de $\mathcal{H}_M$ con la misma cantidad de bolas.

⁴²Sobre la pieza medible $\{x : H(x_i) \leq M \iff i \in I\}$, el supremo de la Definición 2.20 para $\mathcal{H}_M$ coincide con $\sup_f |\sum_{i \in I} \varepsilon_i f(x_i)|$, que es la aplicación de la definición para $\mathcal{F}$ con el vector e' que anula las coordenadas fuera de I ; pegando sobre las 2^n piezas se obtiene la medibilidad.

⁴³El Lema 2.19 con $\Phi(x) = x$ da el factor 2; la P -medibilidad hace medible (respecto de la completación) el supremo simetrizado, la esperanza exterior coincide con la ordinaria y Fubini–Tonelli (integrando no negativo) la factoriza.

⁴⁴El bloque truncado da el supremo sobre $\mathcal{F}_M$; el otro se acota uniformemente en f : $|n^{-1} \sum_i \varepsilon_i (f\mathbb{1}_{\{F > M\}})(X_i)| \leq n^{-1} \sum_i F(X_i) \mathbb{1}_{\{F(X_i) > M\}}$.

⁴⁵El factor 2 es el de la simetrización. Para el término de truncación, otra vez subaditividad y perfección: $\mathbb{E}^* [n^{-1} \sum_i F(X_i) \mathbb{1}_{\{F(X_i) > M\}}] \leq P^*(F\mathbb{1}_{\{F > M\}})$.

⁴⁶Tras la reducción inicial, F es medible; $F\mathbb{1}_{\{F > M\}} \leq F$ y $F\mathbb{1}_{\{F > M\}} \rightarrow 0$ puntualmente cuando $M \rightarrow \infty$ (F es finita c.s. porque $PF < \infty$), y la convergencia dominada concluye.

Fijemos $X_1, \dots, X_n$. Partimos de una $\delta/2$ -red, cuyos centros pueden ser externos. En cada bola que intersecciona $\mathcal{F}_M$ elegimos un punto de $\mathcal{F}_M$. Obtenemos así una red interna $\mathcal{G} \subset \mathcal{F}_M$ de radio δ , con $|\mathcal{G}| \leq N(\delta/2, \mathcal{F}_M, L_1(P_n))$. Para cada $f \in \mathcal{F}_M$, elegimos $g_f \in \mathcal{G}$ con $P_n|f - g_f| \leq \delta$. Entonces, para todo vector de signos, $\left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i \{f - g_f\}(X_i) \right| \leq P_n|f - g_f| \leq \delta$, y por tanto

$$\mathbb{E}_\varepsilon \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}_M} \leq \mathbb{E}_\varepsilon \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i g(X_i) \right\|_{\mathcal{G}} + \delta.$$

Aplicamos ahora la desigualdad maximal finita con $\psi_2(x) = e^{x^2} - 1$.⁴⁷ Condicionalmente en $X_1, \dots, X_n$,

$$\mathbb{E}_\varepsilon \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i g(X_i) \right\|_{\mathcal{G}} \leq C \sqrt{1 + \log |\mathcal{G}|} \sup_{g \in \mathcal{G}} \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i g(X_i) \right\|_{\psi_2|X}.$$

Por el Lema 2.13 obtenemos la primera desigualdad en:

$$\left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i g(X_i) \right\|_{\psi_2|X} \leq C \left(\frac{1}{n^2} \sum_{i=1}^n g^2(X_i) \right)^{1/2} = C n^{-1/2} (P_n g^2)^{1/2} \leq C M n^{-1/2},$$

donde en la última usamos que $|g| \leq M$ para $g \in \mathcal{F}_M$. Así,

$$\mathbb{E}_\varepsilon \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}_M} \leq \delta + CM \sqrt{\frac{1 + \log N(\delta/2, \mathcal{F}_M, L_1(P_n))}{n}}.$$

Por hipótesis, $\frac{1}{n} \log N(\delta/2, \mathcal{F}_M, L_1(P_n)) \rightarrow 0$ en probabilidad exterior, de modo que el segundo término tiende a 0 en probabilidad exterior. Además, el lado izquierdo está acotado por M . Por convergencia dominada en forma exterior,⁴⁸ $\mathbb{E}_X \mathbb{E}_\varepsilon \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}_M} \leq \delta + o(1)$. Como $\delta > 0$ es arbitrario, $\mathbb{E}_X \mathbb{E}_\varepsilon \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}_M} \rightarrow 0$. Luego, haciendo primero $n \rightarrow \infty$ y después $M \rightarrow \infty$, $\mathbb{E}^* \|P_n - P\|_{\mathcal{F}} \rightarrow 0$. Esto prueba la convergencia en media exterior y, por Markov exterior (Corolario 1.10), $\|P_n - P\|_{\mathcal{F}} \xrightarrow{\mathbb{P}^*} 0$.

Resta obtener la convergencia casi segura exterior. Por el Lema 2.27, pueden elegirse versiones medibles Y_n de $\|P_n - P\|_{\mathcal{F}}^*$ que forman una submartingala inversa. El teorema de convergencia para submartingalas inversas da $Y_n \rightarrow Y$ c.s. Como, por lo probado, $\mathbb{E} Y_n = \mathbb{E}^* \|P_n - P\|_{\mathcal{F}} \rightarrow 0$, se sigue que $Y = 0$ c.s. Por tanto,⁴⁹ $\|P_n - P\|_{\mathcal{F}}^* \rightarrow 0$ c.s. $\square$

La necesidad de la condición de entropía, que completa la equivalencia del Teorema 2.28, no se demuestra aquí. La prueba de [vdV–W] usa la desimetrización, desigualdades para multiplicadores gaussianos y la desigualdad de Sudakov; además obtiene la convergencia de la entropía en media exterior.

Proposición 2.29. Sean $\mathcal{F}, \mathcal{F}_1, \mathcal{F}_2$ clases P -Glivenko–Cantelli. Entonces:

1. $\{a_1 f_1 + a_2 f_2 : f_i \in \mathcal{F}_i, |a_i| \leq 1\}$ 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.

⁴⁷Lema 2.8 con $\psi = \psi_2$, aplicado condicionalmente en la muestra (las únicas variables son los signos): $\psi_2^{-1}(|\mathcal{G}|) = \sqrt{\log(1 + |\mathcal{G}|)} \leq \sqrt{1 + \log |\mathcal{G}|}$, y $\mathbb{E}_\varepsilon |Z| \leq C \|Z\|_{\psi_2|X}$ como en la prueba del Corolario 2.17. La notación $\|\cdot\|_{\psi_2|X}$ es la norma de Orlicz respecto de $\mathbb{E}_\varepsilon$, con la muestra fija.

⁴⁸El integrando es $Z_n := \mathbb{E}_\varepsilon \|n^{-1} \sum_{i=1}^n \varepsilon_i f(X_i)\|_{\mathcal{F}_M} \leq \delta + (W_n \wedge M)$, con $W_n = CM \sqrt{(1 + \log N(\delta/2, \mathcal{F}_M, L_1(P_n)))/n}$; además $0 \leq Z_n \leq M$, pues $|f| \mathbb{1}_{\{f \leq M\}} \leq M$. Como $W_n^* \wedge M$ es medible, acotada y tiende a 0 en probabilidad, su esperanza tiende a 0 (convergencia dominada a lo largo de subsucesiones c.s.).

⁴⁹La igualdad $\mathbb{E} Y_n = \mathbb{E}^* \|P_n - P\|_{\mathcal{F}}$ es el Lema 1.4(iii); y de $Y \geq 0$ con $\mathbb{E} Y \leq \liminf_n \mathbb{E} Y_n = 0$ (Fatou) se concluye $Y = 0$ c.s.

Demostración. Para (1), si $\mathcal{H} = \{a_1 f_1 + a_2 f_2 : f_i \in \mathcal{F}_i, |a_i| \leq 1\}$, entonces $\|P_n - P\|_{\mathcal{H}}^* \leq \|P_n - P\|_{\mathcal{F}_1}^* + \|P_n - P\|_{\mathcal{F}_2}^* \rightarrow 0$ c.s.⁵⁰ Para (2),

$$\|P_n - P\|_{\mathcal{F}_1 \cup \mathcal{F}_2}^* \leq \max\{\|P_n - P\|_{\mathcal{F}_1}^*, \|P_n - P\|_{\mathcal{F}_2}^*\} \rightarrow 0 \quad \text{c.s.}$$

Para (3), sea $\overline{\mathcal{F}}$ la clase descripta. Si $g \in \overline{\mathcal{F}}$, existe $f_m \in \mathcal{F}$ tal que $f_m \rightarrow g$ puntualmente y $P|f_m - g| \rightarrow 0$. Para cada muestra fija, $P_n|f_m - g| \rightarrow 0$, 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 $|(P_n - P)g| \leq \|P_n - P\|_{\mathcal{F}} + P_n|f_m - g| + P|f_m - g| \leq \|P_n - P\|_{\mathcal{F}} + \eta$. Tomando supremo en g y luego $\eta \downarrow 0$, $\|P_n - P\|_{\overline{\mathcal{F}}} \leq \|P_n - P\|_{\mathcal{F}}$. Por mayorantes medibles mínimos, $\|P_n - P\|_{\overline{\mathcal{F}}}^* \leq \|P_n - P\|_{\mathcal{F}}^*$ c.s., y el resultado sigue. $\square$

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.

Definición 2.30. Extendemos la definición de número de cubrimiento dada para $L_1(Q)$ a una seminorma cualquiera $\|\cdot\|$: el **número de cubrimiento** $N(\varepsilon, \mathcal{F}, \|\cdot\|)$ es el mínimo número de bolas $\{g : \|g - f\| < \varepsilon\}$ necesarias para cubrir $\mathcal{F}$. Los centros no necesitan pertenecer a $\mathcal{F}$, pero deben tener norma finita; vale también aquí la internalización descrita más arriba. La entropía es $\log N(\varepsilon, \mathcal{F}, \|\cdot\|)$.

Si la norma es de tipo $L_r(Q)$, entonces $N(\varepsilon, \mathcal{F}, L_r(Q)) \leq N_{[]} (2\varepsilon, \mathcal{F}, L_r(Q))$, pues cada bracket de tamaño 2ε queda contenido en una bola de radio ε , centrada en el punto medio de sus extremos.

Sea F una envolvente de $\mathcal{F}$. La *condición de entropía uniforme* se formula mediante la integral truncada que definimos a continuación, con el supremo tomado sobre todas las medidas de probabilidad finitamente discretas Q tales que $0 < \|F\|_{Q,2} < \infty$.

Definición 2.31. Sea $\mathcal{F}$ una clase de funciones medibles con envolvente F . Para $\delta > 0$, definimos la integral de entropía uniforme truncada por

$$J(\delta, \mathcal{F}, L_2) := \int_0^\delta \sup_Q \sqrt{\log N(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q))} d\varepsilon,$$

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

$$J(\infty, \mathcal{F}, L_2) < \infty. \quad (\text{UE})$$

Teorema 2.32. 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 $\mathcal{F}_\delta = \{f - g : f, g \in \mathcal{F}, \|f - g\|_{P,2} < \delta\}$, $\mathcal{F}_\infty = \{(f - g)^2 : f, g \in \mathcal{F}\}$ son P -medibles. Entonces $\mathcal{F}$ es P -Donsker.

Demostración. Escribamos $\mathbb{G}_n = \sqrt{n}(P_n - P)$, $\rho(f, g) = \|f - g\|_{P,2}$, $\mathcal{F}_\delta = \{f - g : f, g \in \mathcal{F}, \rho(f, g) < \delta\}$, $\mathcal{F}_\infty = \{f - g : f, g \in \mathcal{F}\}$. 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

⁵⁰Punto a punto, $|(P_n - P)(a_1 f_1 + a_2 f_2)| \leq |(P_n - P)f_1| + |(P_n - P)f_2|$ para $|a_i| \leq 1$; tomando supremos, $\|P_n - P\|_{\mathcal{H}} \leq \|P_n - P\|_{\mathcal{F}_1} + \|P_n - P\|_{\mathcal{F}_2}$, y la regla $(S + T)^* \leq S^* + T^*$ 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).

uniforme.⁵¹ La convergencia finito-dimensional se obtiene del TCL multivariado, pues $|f| \leq F$ y $P^*F^2 < \infty$.⁵²

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,⁵³

$$\mathbb{P}^*\{\|\mathbb{G}_n\|_{\mathcal{F}_{\delta_n}} > x\} \leq \frac{1}{x} \mathbb{E}^*\|\mathbb{G}_n\|_{\mathcal{F}_{\delta_n}} \leq \frac{2}{x} \mathbb{E}^* \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i h(X_i) \right\|_{\mathcal{F}_{\delta_n}}. \quad (1)$$

Como $\mathcal{F}_{\delta_n}$ es P -medible, para cada $e \in \{-1, 1\}^n$ el mapa $(x_1, \dots, x_n) \mapsto \sup_{h \in \mathcal{F}_{\delta_n}} |\sum_{i=1}^n e_i h(x_i)|$ es medible para la completación de P^n . Como los signos toman finitos valores, el supremo simetrizado es medible en (X, ε) ; por Fubini,

$$\mathbb{E}^* \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i h(X_i) \right\|_{\mathcal{F}_{\delta_n}} = \mathbb{E}_X \mathbb{E}_\varepsilon \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i h(X_i) \right\|_{\mathcal{F}_{\delta_n}}. \quad (2)$$

Condicionando en $X_1, \dots, X_n$, ponemos $\|h\|_n = (P_n h^2)^{1/2}$, $\theta_n^2 = \sup_{h \in \mathcal{F}_{\delta_n}} P_n h^2$. Para $h, k \in \mathcal{F}_{\delta_n}$, $\left\| \frac{1}{\sqrt{n}} \sum_i \varepsilon_i (h - k)(X_i) \right\|_{\psi_2|X} \lesssim \|h - k\|_n$, luego el proceso de Rademacher es subgaussiano para $d_n(h, k) = \|h - k\|_n$.⁵⁴

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 $|Z_h - Z_k| \leq n^{-1/2} \sum_i |h - k|(X_i) \leq \sqrt{n} d_n(h, k)$, 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 2.17 se aplica con $\delta = \text{diam}$, 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: $\sup_h |Z_h| = \sup_h |Z_h - Z_0| \leq \sup_{h,k} |Z_h - Z_k|$. Además, el diámetro de $(\mathcal{F}_{\delta_n}, d_n)$ es a lo sumo $2\theta_n$, y por (2.1), $\int_0^{2\theta_n} \sqrt{\log D(u, d_n)} du \leq 2 \int_0^{\theta_n} \sqrt{\log N(v, d_n)} dv$.⁵⁵ Por la desigualdad maximal subgaussiana (Corolario 2.17 con $\delta = \text{diam}$), absorbiendo el factor 2 en K ,

$$\mathbb{E}_\varepsilon \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i h(X_i) \right\|_{\mathcal{F}_{\delta_n}} \leq K \int_0^{\theta_n} \sqrt{\log N(u, \mathcal{F}_{\delta_n}, L_2(P_n))} du. \quad (3)$$

Además, para toda probabilidad Q ,

$$N(u, \mathcal{F}_\infty, L_2(Q)) \leq N^2(u/2, \mathcal{F}, L_2(Q)), \quad (4)$$

pues si $g_1, \dots, g_N$ cubre $\mathcal{F}$ con radio $u/2$, entonces $\{g_j - g_k : 1 \leq j, k \leq N\}$ cubre $\mathcal{F}_\infty$ con radio u :

⁵¹El criterio es el Teorema 1.66 combinado con el de equicontinuidad (Teorema 1.70): 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 .

⁵²Cada f es medible con $Pf^2 = P^*f^2 \leq P^*F^2 < \infty$, así que el TCL multivariado se aplica a los vectores $(f_1, \dots, f_k)$ centrados.

⁵³Corolario 1.10 y Lema 2.19 con $\Phi(x) = x$, aplicado a la clase $\mathcal{F}_{\delta_n}$. En efecto, para $h = f - g \in \mathcal{F}_{\delta_n}$, $P|h| \leq 2P^*F \leq 2(P^*F^2)^{1/2} < \infty$; la segunda desigualdad es Cauchy-Schwarz aplicada al mayorante F^* , usando $P^*F = PF^*$ y $(F^2)^* = (F^*)^2$ c.s. (Lema 2.18 con $\Phi(t) = t^2$). Además, $\sqrt{n}(P_n - P)h = \mathbb{G}_n h$ y el factor $\sqrt{n}$ convierte P_n° en $n^{-1/2} \sum_i \varepsilon_i h(X_i)$.

⁵⁴Lema 2.13 condicional, con $a_i = n^{-1/2}(h - k)(X_i)$: $\|a\|^2 = n^{-1} \sum_i (h - k)^2(X_i) = \|h - k\|_n^2$, y la constante es $\sqrt{6}$.

⁵⁵El diámetro: $\|h - k\|_n \leq \|h\|_n + \|k\|_n \leq 2\theta_n$, por Minkowski. La integral: con $u = 2v$, $\int_0^{2\theta_n} \sqrt{\log D(u, d_n)} du = 2 \int_0^{\theta_n} \sqrt{\log D(2v, d_n)} dv \leq 2 \int_0^{\theta_n} \sqrt{\log N(v, d_n)} dv$, pues $D(2v, d_n) \leq N(v, d_n)$ por (2.1). Obsérvese que el Corolario 2.17 está enunciado con números de empaquetamiento D ; reemplazar D por N al mismo radio iría en la dirección incorrecta ($N \leq D$), y es el cambio de escala $u = 2v$ el que permite pasar a números de cubrimiento.

$\|(f - f') - (g_j - g_k)\|_{Q,2} \leq \|f - g_j\|_{Q,2} + \|f' - g_k\|_{Q,2} < u$. Como $\mathcal{F}_{\delta_n} \subset \mathcal{F}_\infty$, de (3)–(4),

$$\begin{aligned} \mathbb{E}_\varepsilon \left\| \frac{1}{\sqrt{n}} \sum_i \varepsilon_i h(X_i) \right\|_{\mathcal{F}_{\delta_n}} &\leq K \int_0^{\theta_n} \sqrt{\log N(u/2, \mathcal{F}, L_2(P_n))} du \\ &= K \|F\|_{n,2} \int_0^{\theta_n/(2\|F\|_{n,2})} \sqrt{\log N(\varepsilon\|F\|_{n,2}, \mathcal{F}, L_2(P_n))} d\varepsilon \\ &\leq K \|F\|_{n,2} J\left(\frac{\theta_n}{\|F\|_{n,2}}, \mathcal{F}, L_2\right), \end{aligned} \quad (5)$$

donde usamos que P_n es finitamente discreta y absorbimos constantes en K .⁵⁶ 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 $\|F\|_{n,2} \geq 1$.⁵⁷ Además,

$$\mathbb{E}^* \|F\|_{n,2}^2 = \mathbb{E}^* P_n F^2 \leq P^* F^2 < \infty. \quad (6)$$

Mostramos ahora que $\theta_n \rightarrow 0$ en probabilidad exterior.⁵⁸ Si $h \in \mathcal{F}_{\delta_n}$, entonces $Ph^2 < \delta_n^2$, y por tanto

$$\theta_n^2 = \sup_{h \in \mathcal{F}_{\delta_n}} P_n h^2 \leq \delta_n^2 + \|P_n - P\|_{\mathcal{F}_\infty}^*, \quad \mathcal{F}_\infty^2 = \{h^2 : h \in \mathcal{F}_\infty\}. \quad (7)$$

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 $h, k \in \mathcal{F}_\infty$, $|h^2 - k^2| = |h - k||h + k| \leq |h - k| 4F$, pues $|h| \leq 2F$, $|k| \leq 2F$. Luego, para toda probabilidad discreta Q ,

$$Q|h^2 - k^2| \leq Q(|h - k| 4F) \leq \|h - k\|_{Q,2} \|4F\|_{Q,2}. \quad (8)$$

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 $|k| \leq 2F$; por eso internalizamos primero: dado un cubrimiento de radio $\frac{\varepsilon}{2} \|F\|_{Q,2}$ en $L_2(Q)$, eligiendo en cada bola que corta a $\mathcal{F}_\infty$ un elemento de la clase se obtienen 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

$$N(\varepsilon \|2F\|_{Q,2}^2, \mathcal{F}_\infty^2, L_1(Q)) \leq N\left(\frac{\varepsilon}{2} \|F\|_{Q,2}, \mathcal{F}_\infty, L_2(Q)\right). \quad (9)$$

Usando otra vez (4),

$$N\left(\frac{\varepsilon}{2} \|F\|_{Q,2}, \mathcal{F}_\infty, L_2(Q)\right) \leq N^2\left(\frac{\varepsilon}{4} \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right). \quad (10)$$

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 .⁵⁹ Aplicado con $Q = P_n$, obtenemos

$$\log N(\varepsilon \|2F\|_{n,2}^2, \mathcal{F}_\infty^2, L_1(P_n)) = O_{\mathbb{P}^*}(1) = o_{\mathbb{P}^*}(n). \quad (11)$$

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, $N(\delta, \mathcal{F}_\infty^2, L_1(P_n)) \leq$

⁵⁶El paso intermedio es el cambio de variable $u = 2\varepsilon \|F\|_{n,2}$ (el factor 2 se absorbe en K); 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 $0 < \|F\|_{n,2} < \infty$.

⁵⁷El reemplazo conserva las hipótesis: $F \vee 1$ sigue siendo envolvente, $P^*(F \vee 1)^2 \leq P^*F^2 + 1 < \infty$ y, como $\|F \vee 1\|_{Q,2} \geq \|F\|_{Q,2}$, 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.

⁵⁸En (7): para cada h , $P_n h^2 = Ph^2 + (P_n - P)h^2 < \delta_n^2 + \|P_n - P\|_{\mathcal{F}_\infty}^*$; el paso al mayorante $\|\cdot\|^*$ vale c.s. y da una cota medible, uniforme en h .

⁵⁹La 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 $\varepsilon > 0$.

$N(\varepsilon \|2F\|_{n,2}^2, \mathcal{F}_\infty^2, L_1(P_n))$. Por tanto, (11) implica la condición de entropía aleatoria para cada radio fijo. La misma cota vale para las clases truncadas que aparecen en el Teorema 2.28: multiplicar todos los elementos por un mismo indicador es una contracción en $L_1(P_n)$, y por ello no aumenta sus números de cubrimiento. Como $(2F)^2$ es una envolvente integrable de $\mathcal{F}_\infty^2$, el teorema de Glivenko–Cantelli por entropía aleatoria da

$$\|P_n - P\|_{\mathcal{F}_\infty^2}^* \xrightarrow{\mathbb{P}^*} 0. \quad (12)$$

De (7) se sigue

$$\theta_n \xrightarrow{\mathbb{P}^*} 0. \quad (13)$$

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 $\|F\|_{n,2} \geq 1$,

$$J\left(\frac{\theta_n}{\|F\|_{n,2}}, \mathcal{F}, L_2\right) \xrightarrow{\mathbb{P}^*} 0. \quad (14)$$

Además, está acotada por $J(\infty, \mathcal{F}, L_2)$. Por Cauchy–Schwarz y (6),

$$\mathbb{E}_X^* \left[\|F\|_{n,2} J\left(\frac{\theta_n}{\|F\|_{n,2}}, \mathcal{F}, L_2\right) \right] \leq (\mathbb{E}^* \|F\|_{n,2}^2)^{1/2} \left(\mathbb{E}^* J^2\left(\frac{\theta_n}{\|F\|_{n,2}}, \mathcal{F}, L_2\right) \right)^{1/2} \rightarrow 0, \quad (15)$$

por convergencia dominada exterior. Juntando (1), (2), (5) y (15),⁶⁰

$$\|\mathbb{G}_n\|_{\mathcal{F}_{\delta_n}} \xrightarrow{\mathbb{P}^*} 0 \quad (\delta_n \downarrow 0). \quad (16)$$

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

$$\sup_{h \in \mathcal{F}_\infty} |(Q_m - P)h|^2 \leq m^{-2}. \quad (17)$$

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 $P^* F^2 < \infty$.⁶¹ Fijado $\varepsilon > 0$, elegimos m con $m^{-1} \leq \varepsilon$. Por la condición de entropía uniforme,⁶²

$$N(\varepsilon/2, \mathcal{F}, L_2(Q_m)) < \infty. \quad (18)$$

Partimos de una red de radio $\varepsilon/2$, posiblemente con centros externos, y elegimos un elemento de $\mathcal{F}$ en cada bola que intersecta la clase. Los centros internos así obtenidos forman una red de radio ε . Por tanto, para todo $f \in \mathcal{F}$ existe j tal que $\|f - f_j\|_{Q_m,2} \leq \varepsilon$. Como $f - f_j \in \mathcal{F}_\infty$, por (17),

$$\|f - f_j\|_{P,2}^2 = P(f - f_j)^2 \leq Q_m(f - f_j)^2 + |(P - Q_m)(f - f_j)| \leq \varepsilon^2 + m^{-2} \leq 2\varepsilon^2. \quad (19)$$

Entonces $f_1, \dots, f_N$ 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. $\square$

⁶⁰El Cauchy–Schwarz exterior para factores no negativos se justifica con $(AB)^* \leq A^* B^*$ c.s. (Lema 1.9(v), caso de factores positivos) seguido del Cauchy–Schwarz clásico para A^*, B^* . El segundo factor tiende a 0: $J^2(\theta_n/\|F\|_{n,2}) \leq J^2(\infty)$ está acotada y tiende a 0 en probabilidad exterior por (14), y se razona como en la convergencia dominada del Teorema 2.28.

⁶¹Sea $(P_n F^2)^*$ un mayorante medible mínimo de $P_n F^2$. Como $\mathbb{E}(P_n F^2)^* = \mathbb{E}^* P_n F^2 \leq P^* F^2 < \infty$, 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 $\{\|P_n - P\|_{\mathcal{F}_\infty^2}^* \leq m^{-2}\}$ tiene probabilidad positiva; su intersección con el evento de probabilidad 1 recién descrito es no vacía, y cualquier ω en ella da la medida finitamente discreta $Q_m = P_n(\omega)$ requerida (nótese que $\|P_n - P\|_{\mathcal{F}_\infty^2}^* \leq \|P_n - P\|_{\mathcal{F}_\infty^2}^*$ en todo punto).

⁶²La 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 2.33. 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

$$N(\varepsilon, \mathcal{F}, L_2(Q)) \leq N_{[]}(\varepsilon^2, \mathcal{F}, L_1(Q)) \leq \frac{2}{\varepsilon^2}.$$

⁶³ 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 recubrimiento 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 $L_2(P)$: si

$$\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_2(P))} d\varepsilon < \infty, \quad (2.3)$$

entonces $\mathcal{F}$ es P -Donsker.

Teorema 2.34. Sea $\mathcal{F}$ una clase de funciones medibles. Si se cumple (2.3), 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 [4, 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 (2.3). 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 $P^*(F > x) = o(x^{-2})$, $x \rightarrow \infty$. 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 $\|f\|_{P,2,\infty} := \sup_{x>0} (x^2 P(|f| > x))^{1/2}$. 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,

$$\|f\|_{P,2,\infty} \leq \|f\|_{P,2}.^{64}$$

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, no se usa Hoeffding ni la desigualdad maximal subgaussiana. En su lugar se usa la desigualdad de

⁶³Para 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 2.23, 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 $0 < \varepsilon \leq 1$.

⁶⁴ $x^2 P(|f| > x) \leq \mathbb{E}(f^2 \mathbb{1}_{\{|f|>x\}}) \leq P f^2$, y se toma supremo en x .

Bernstein (véase [vdV–W, Sección 2.2]): para toda función f cuadrado-integrable y uniformemente acotada,

$$\mathbb{P}\{|\mathbb{G}_n f| > x\} \leq 2 \exp \left\{ -\frac{1}{2} \frac{x^2}{P(f - Pf)^2 + \|f - Pf\|_\infty x / (3\sqrt{n})} \right\}, \quad x > 0.$$

En particular, como $P(f - Pf)^2 \leq Pf^2$ y $\|f - Pf\|_\infty \leq 2\|f\|_\infty$, 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,⁶⁵ si $\mathcal{H}$ es finita y $|\mathcal{H}| \geq 2$, entonces

$$\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{H}} \lesssim \max_{h \in \mathcal{H}} \frac{\|h\|_\infty}{\sqrt{n}} \log |\mathcal{H}| + \max_{h \in \mathcal{H}} \|h\|_{P,2} \sqrt{\log |\mathcal{H}|}. \quad (2.4)$$

El encadenamiento se organiza para que los dos términos de (2.4) 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 2.35. *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, $P^(F > x) = o(x^{-2})$, $x \rightarrow \infty$, entonces $\mathcal{F}$ es P -Donsker.*

Demostración del Teorema 2.34 a partir del Teorema 2.35. Verificamos las dos hipótesis. Primero, como $\|f\|_{P,2,\infty} \leq \|f\|_{P,2}$, todo ε -bracket en $L_2(P)$ es un ε -bracket en $L_{2,\infty}(P)$, luego $N_{[]}(\varepsilon, \mathcal{F}, L_{2,\infty}(P)) \leq N_{[]}(\varepsilon, \mathcal{F}, L_2(P))$; y los brackets también controlan los cubrimientos ordinarios, $N(\varepsilon, \mathcal{F}, L_2(P)) \leq N_{[]}(\varepsilon, \mathcal{F}, L_2(P))$, tomando como centro el punto medio de cada bracket. Así, la condición (2.3) 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 $F := \max_{1 \leq i \leq m} (|l_i| \vee |u_i|)$ 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: $x^2 \mathbb{P}(F > x) \leq \mathbb{E}(F^2 \mathbf{1}_{\{F > x\}}) \rightarrow 0$ cuando $x \rightarrow \infty$, por convergencia dominada. Aplicando el Teorema 2.35, $\mathcal{F}$ es P -Donsker. $\square$

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 $0 < \varepsilon < 1$,

$$\sup_Q \log N\left(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) \leq K \left(\frac{1}{\varepsilon}\right)^{2-\delta}, \quad (2.5)$$

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 $|f| \leq F$.⁶⁶ 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.

⁶⁵Es el Lema 2.9 con $a = 2 \max_h P h^2$ y $b = \frac{4}{3\sqrt{n}} \max_h \|h\|_\infty$, seguido de $\mathbb{E}|Z| \leq C \|Z\|_{\psi_1}$ y $\log(1+m) \leq 2 \log m$ para $m \geq 2$.

⁶⁶Para $\varepsilon > 1$, la bola de radio $\varepsilon \|F\|_{Q,2}$ centrada en 0 contiene a toda la clase ($\|f\|_{Q,2} \leq \|F\|_{Q,2}$), 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

$$\sup_Q N\left(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) \leq K \left(\frac{1}{\varepsilon}\right)^V, \quad 0 < \varepsilon < 1, \quad (2.6)$$

para algún $V < \infty$. En efecto, (2.6) implica $\sup_Q \log N\left(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) \leq \log K + V \log(1/\varepsilon)$, y por tanto (2.5) 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 2.36 (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

$$V(\mathcal{C}) := \sup \left\{ n : \max_{\{x_1, \dots, x_n\} \subset \mathcal{X}} \Delta_n(\mathcal{C}, x_1, \dots, x_n) = 2^n \right\}.$$

Es decir, $V(\mathcal{C})$ es el mayor tamaño de un conjunto finito que puede ser fragmentado por $\mathcal{C}$.⁶⁷ 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 recubrimiento de una clase VC crecen a lo sumo polinómicamente en $1/\varepsilon$. Para una clase de conjuntos, vista como clase de indicadores, la envolvente es $F \equiv 1$, luego es cuadrado-integrable para toda probabilidad P . Una vez probada la cota polinómica uniforme, el teorema de Donsker por entropía uniforme implicará que toda clase VC de conjuntos medibles es P -Donsker para toda probabilidad P , bajo las hipótesis de medibilidad requeridas. Las cotas de cubrimiento que siguen están enunciadas en términos de la dimensión $V(\mathcal{C})$; en [4] aparecen formuladas con el índice $V(\mathcal{C}) + 1$, y los exponentes difieren en consecuencia.

Ejemplo 2.37. 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.⁶⁸

El siguiente resultado combinatorio, conocido como *lema de Sauer*, es la base de las cotas polinomiales de entropía para clases VC. El lema de Sauer dice que, para un conjunto finito $S =$

⁶⁷Aquí $V(\mathcal{C})$ denota la dimensión VC. En parte de la literatura se llama *índice VC* al entero $V(\mathcal{C}) + 1$.

⁶⁸Para 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 [vdV–W, §2.6].

$\{x_1, \dots, x_n\}$, el número $\{C \cap S : C \in \mathcal{C}\}$ está acotado por el número de subconjuntos $A \subset S$ que son fragmentados por $\mathcal{C}$. En particular, si $V(\mathcal{C}) < \infty$, sólo pueden ser fragmentados subconjuntos de S de tamaño a lo sumo $V(\mathcal{C})$; por tanto, $\Delta_n(\mathcal{C}, x_1, \dots, x_n) \leq \sum_{j=0}^{V(\mathcal{C})} \binom{n}{j}$.

Ejemplo 2.38. Sea $S = \{1, 2, 3\}$ y $\mathcal{C} = \{\emptyset, S\}$. Sobre todo S , la clase sólo selecciona $\emptyset \cap S = \emptyset$, $S \cap S = S$, 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 $\emptyset, \{1\}, \{2\}, \{3\}$.

Lema 2.39 (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

$$\Delta_n(\mathcal{C}, x_1, \dots, x_n) \leq \sum_{j=0}^{V(\mathcal{C})} \binom{n}{j}. \quad (2.7)$$

Demostración. Escribamos $S = \{x_1, \dots, x_n\}$. Reemplazando cada $C \in \mathcal{C}$ por su trazo $C \cap S$, 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 $C \cap A = B \cup \{x_i\}$, $(C \setminus \{x_i\}) \cap A = B$. 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 $\mathcal{D}$.⁶⁹

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, $\Delta_n(\mathcal{C}, x_1, \dots, x_n) \leq \sum_{j=0}^{V(\mathcal{C})} \binom{n}{j}$. $\square$

⁶⁹Con 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 $B \in \mathcal{D}$.

Así, aunque por definición una clase VC sólo excluye la posibilidad de seleccionar los 2^n subconjuntos cuando $n > V(\mathcal{C})$, el Lema 2.39 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 2.40. *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 $1 \leq V(\mathcal{C}) \leq n$,*

$$\max_{x_1, \dots, x_n} \Delta_n(\mathcal{C}, x_1, \dots, x_n) \leq \left(\frac{en}{V(\mathcal{C})} \right)^{V(\mathcal{C})},$$

de modo que el crecimiento es polinómico de orden a lo sumo $O(n^{V(\mathcal{C})})$.

Demostración. Por el Lema 2.39, $\Delta_n(\mathcal{C}, x_1, \dots, x_n) \leq \sum_{j=0}^{V(\mathcal{C})} \binom{n}{j}$, porque ningún subconjunto de tamaño mayor que $V(\mathcal{C})$ puede ser fragmentado por $\mathcal{C}$. La última desigualdad es la cota combinatoria usual para sumas binomiales parciales.⁷⁰ $\square$

Teorema 2.41. *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 $\{1_C : C \in \mathcal{C}\}$, y $\|1_C - 1_D\|_{Q,r} = Q(C \triangle D)^{1/r}$. 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 2.39 y no se incluye aquí; véase [vdV-W, §2.6].

2.6.2. Clases VC de funciones

Definición 2.42. El **subgráfico** de una función $f : \mathcal{X} \rightarrow \mathbb{R}$ 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 2.43. *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 $0 < \varepsilon < 1$,*

$$N(\varepsilon \|F\|_{Q,r}, \mathcal{F}, L_r(Q)) \leq K V(\mathcal{F}) (16e)^{V(\mathcal{F})} \left(\frac{1}{\varepsilon} \right)^{rV(\mathcal{F})},$$

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 2.44 (Cota elemental con constante relajada). *Bajo las hipótesis del teorema anterior y suponiendo $V(\mathcal{F}) \geq 1$,*

$$N(\varepsilon \|F\|_{Q,r}, \mathcal{F}, L_r(Q)) \leq K V(\mathcal{F}) (16e)^{V(\mathcal{F})} \left(\frac{2}{\varepsilon} \right)^{rV(\mathcal{F})}.$$

⁷⁰Con $V = V(\mathcal{C})$ 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 $1 + x \leq e^x$.

Demostración. Sea $\mathcal{C} = \{C_f : f \in \mathcal{F}\}$, donde $C_f = \{(x, t) : t < f(x)\}$. Para $f, g \in \mathcal{F}$, por Fubini,

$$(Q \times \lambda)(C_f \triangle C_g) = \int \lambda\{t : \mathbf{1}_{\{t < f(x)\}} \neq \mathbf{1}_{\{t < g(x)\}}\} dQ(x) = Q|f - g|. \quad (1)$$

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)

$$P(C_f \triangle C_g) = \frac{Q|f - g|}{2QF}. \quad (2)$$

Los centros de un cubrimiento de $\mathcal{C}$ pueden ser conjuntos externos a la clase; internalizando —en cada bola de radio $\varepsilon/2$ en $L_1(P)$ que corta a $\mathcal{C}$ se elige un subgráfico $C_{f_j} \in \mathcal{C}$ — se obtiene, para todo $f \in \mathcal{F}$, un índice j con $P(C_f \triangle C_{f_j}) < \varepsilon$, y (2) da $Q|f - f_j| < 2\varepsilon QF$. Por tanto,

$$N(2\varepsilon QF, \mathcal{F}, L_1(Q)) \leq N\left(\frac{\varepsilon}{2}, \mathcal{C}, L_1(P)\right). \quad (3)$$

Aplicando el Teorema 2.41 a $\mathcal{C}$, en el radio $\varepsilon/2$,

$$N(2\varepsilon QF, \mathcal{F}, L_1(Q)) \leq K V(\mathcal{F})(4e)^{V(\mathcal{F})} \left(\frac{2}{\varepsilon}\right)^{V(\mathcal{F})}. \quad (4)$$

Reescalando el radio, y aumentando la constante de la base, obtenemos el caso $r = 1$:

$$N(\varepsilon QF, \mathcal{F}, L_1(Q)) \leq K V(\mathcal{F})(16e)^{V(\mathcal{F})} \left(\frac{1}{\varepsilon}\right)^{V(\mathcal{F})}. \quad (5)$$

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 $|f - g| \leq 2F$,

$$Q|f - g|^r \leq Q\{|f - g|(2F)^{r-1}\} = 2^{r-1}(QF^{r-1})R|f - g|. \quad (6)$$

Luego

$$\|f - g\|_{Q,r} \leq 2(QF^{r-1})^{1/r} \|f - g\|_{R,1}^{1/r}. \quad (7)$$

Además,

$$(QF^{r-1})^{1/r} (RF)^{1/r} = (QF^r)^{1/r} = \|F\|_{Q,r}. \quad (8)$$

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

$$N(2\varepsilon \|F\|_{Q,r}, \mathcal{F}, L_r(Q)) \leq N_{\text{int}}(\varepsilon^r RF, \mathcal{F}, L_1(R)), \quad (9)$$

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 ,

$$N_{\text{int}}(\varepsilon^r RF, \mathcal{F}, L_1(R)) \leq K V(\mathcal{F})(16e)^{V(\mathcal{F})} \left(\frac{1}{\varepsilon}\right)^{rV(\mathcal{F})}. \quad (10)$$

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 2.43 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 2.32 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 2.45 (Espacios vectoriales finito-dimensionales). *Sea $\mathcal{F}$ un espacio vectorial finito-dimensional de funciones medibles $f : \mathcal{X} \rightarrow \mathbb{R}$. 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 $(x_1, t_1), \dots, (x_n, t_n) \in \mathcal{X} \times \mathbb{R}$. Consideremos los vectores $v_f := (f(x_1) - t_1, \dots, f(x_n) - t_n) \in \mathbb{R}^n$, $f \in \mathcal{F}$. Como $\{(f(x_1), \dots, f(x_n)) : f \in \mathcal{F}\}$ está contenido en un subespacio de dimensión $\leq d$, el conjunto de vectores v_f está contenido en un subespacio lineal de dimensión $\leq d + 1 = n - 1$.⁷¹ Por tanto, existe $a = (a_1, \dots, a_n) \neq 0$ ortogonal a todos los v_f , es decir, $\sum_{i=1}^n a_i \{f(x_i) - t_i\} = 0$, $f \in \mathcal{F}$. 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 $C_f := \{(x, t) : t < f(x)\}$. 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.⁷² Así, ningún conjunto de $d + 2$ puntos puede ser fragmentado por los subgráficos de $\mathcal{F}$. Luego $V(\mathcal{F}) \leq d + 1$. $\square$

Lema 2.46 (Traslaciones de una función monótona). *Sea $\psi : \mathbb{R} \rightarrow \mathbb{R}$ 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 $h \in \mathbb{R}$, escribimos $G_h := \{(x, t) \in \mathbb{R}^2 : t < \psi(x - h)\}$, el subgráfico de ψ_h . 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, $h_1 \geq h_2 \implies G_{h_1} \subseteq G_{h_2}$. 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.⁷³ 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, $V(\mathcal{F}) = 1$. $\square$

Lema 2.47 (Permanencia de clases VC de conjuntos). *Sean $\mathcal{C}$ y $\mathcal{D}$ clases VC de subconjuntos de $\mathcal{X}$. Sean $\phi : \mathcal{X} \rightarrow \mathcal{Y}$ y $\psi : \mathcal{Z} \rightarrow \mathcal{X}$ 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;

⁷¹Los vectores $(f(x_1), \dots, f(x_n))$ 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 $t = (t_1, \dots, t_n)$, de dimensión $\leq d + 1 < n$, y el complemento ortogonal de ese subespacio contiene algún $a \neq 0$.

⁷²Si $a_i > 0$, 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.

⁷³Para 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.

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 $\mathcal{X} \times \mathcal{Y}$.

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 2.39, 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.⁷⁴ 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 $\mathcal{C}$.⁷⁵

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 $\overline{\mathcal{C}}$ un límite puntual de elementos de $\mathcal{C}$. Fijados $x_1, \dots, x_n$, para cada i se tiene $\mathbb{1}_{C_\alpha}(x_i) \rightarrow \mathbb{1}_{\overline{\mathcal{C}}}(x_i)$; como los valores son 0 o 1, para α suficientemente grande coinciden simultáneamente en los n puntos.⁷⁶ Así, $\overline{\mathcal{C}}$ 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). $\square$

Lema 2.48 (Permanencia de clases VC-subgráfico). *Sean $\mathcal{F}$ y $\mathcal{G}$ clases VC-subgráfico de funciones sobre $\mathcal{X}$. Sean $a : \mathcal{X} \rightarrow \mathbb{R}$ y $\psi : \mathcal{Z} \rightarrow \mathcal{X}$ funciones medibles fijas, y $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ 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.

⁷⁴Porque $(C \cap D) \cap S = (C \cap S) \cap (D \cap S)$: cada traza de $\mathcal{C} \cap \mathcal{D}$ sobre S está determinada por un par de trazas, de modo que $\Delta_n(\mathcal{C} \cap \mathcal{D}) \leq \Delta_n(\mathcal{C}) \Delta_n(\mathcal{D})$; por Sauer, esto es polinómico en n y, para n grande, $< 2^n$.

⁷⁵Si 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.

⁷⁶Una 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.

Demostración. Escribimos $C_f := \{(x, t) : t < f(x)\}$ para el subgráfico abierto de f y $\overline{C}_f := \{(x, t) : t \leq f(x)\}$ para el subgráfico cerrado. Primero observamos que las clases de subgráficos abiertos y cerrados tienen la misma dimensión VC. En efecto, si los cerrados fragmentan $(x_i, t_i)_{i \leq m}$, para cada $S \subset \{1, \dots, m\}$ elegimos f_S tal que $t_i \leq f_S(x_i)$ 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 $(x_i, t_i - \eta)_{i \leq m}$. 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 2.47.

Para (3), si $\iota : \mathcal{X} \rightarrow \mathcal{X} \times \mathbb{R}$ 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.

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.⁷⁷

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.⁷⁸

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, ∞) , $[a_s, \infty)$, $\emptyset$ o $\mathbb{R}$. Supongamos que los subgráficos de $\varphi \circ \mathcal{F}$ fragmentan $(x_i, s_i)_{i \leq m}$. Ninguno de los conjuntos A_{s_i} puede ser $\emptyset$ ni $\mathbb{R}$, pues en esos casos la pertenencia de la coordenada i sería constante.⁷⁹ Por tanto, para cada i , existe a_i tal que la condición de pertenencia es $a_i < f(x_i)$ o $a_i \leq f(x_i)$.

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 $a_i \leq f(x_i)$. Para cada $i \in I_c$, las cantidades $a_i - f_S(x_i)$, $S \subset \{1, \dots, m\}$, $i \notin S$, 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 $(x_i, b_i)_{i \leq m}$. En consecuencia, la dimensión VC-subgráfico de $\varphi \circ \mathcal{F}$ no excede la de $\mathcal{F}$.

Si φ es no creciente, la función $\tilde{\varphi}(u) := \varphi(-u)$ es no decreciente y $\varphi \circ f = \tilde{\varphi} \circ (-f)$. El resultado se sigue del caso anterior y de (4), porque $-\mathcal{F}$ es VC-subgráfico. Esto prueba (8). $\square$

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

⁷⁷En 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 2.47 con la observación inicial.

⁷⁸Formalmente, 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), $E_{\pm} = \{a \gtrless 0\} \times \mathbb{R}$ y K fijos; una clase de un solo conjunto es VC, y se aplican las partes (2) y (3) del Lema 2.47. 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 2.42; la φ monótona de (8) es automáticamente boreliana.

⁷⁹La 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 $A_{s_i} = \emptyset$, 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.

$\mathcal{G} := \{x \mapsto \theta^\top x + c : \theta \in \mathbb{R}^d, c \in \mathbb{R}\}$, un espacio vectorial de funciones sobre $\mathbb{R}^d$ de dimensión $d + 1$. Por el Lema 2.45, $\mathcal{G}$ es VC-subgráfico, y por el Lema 2.48(3) la colección de semiespacios abiertos

$$\mathcal{H} = \{\mathcal{G} > 0\} = \left\{ \{x \in \mathbb{R}^d : \theta^\top x + c > 0\} : \theta \in \mathbb{R}^d, c \in \mathbb{R} \right\}$$

es una clase VC de conjuntos. Su envolvente es $F \equiv 1$, de modo que $P^*F^2 = 1 < \infty$ para toda P ; el Teorema 2.41 da la cota (2.6), que implica (UE), y el Teorema 2.32 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 $\mathcal{G}_0 = \text{span}\{1, x_1, \dots, x_d, \|x\|^2\}$, 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.

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 $k = (k_1, \dots, k_d)$ escribimos $|k| = k_1 + \dots + k_d$ y $D^k = \frac{\partial^{|k|}}{\partial x_1^{k_1} \dots \partial x_d^{k_d}}$. Para funciones $f : \mathcal{X} \rightarrow \mathbb{R}$ suficientemente regulares, definimos

$$\|f\|_\alpha = \max_{|k| \leq m} \sup_{x \in \text{int}(\mathcal{X})} |D^k f(x)| + \max_{|k|=m} \sup_{\substack{x, y \in \text{int}(\mathcal{X}) \\ x \neq y}} \frac{|D^k f(x) - D^k f(y)|}{\|x - y\|^{\alpha - m}}.$$

Sea $C_M^\alpha(\mathcal{X})$ la clase de funciones continuas $f : \overline{\mathcal{X}} \rightarrow \mathbb{R}$ 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 2.49 (Entropía supremo de clases Hölder (Teorema 2.7.1 en [vdV–W])). *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*

$$\log N(\varepsilon, C_1^\alpha(\mathcal{X}), \|\cdot\|_\infty) \leq K \lambda(\mathcal{X}^1) \left(\frac{1}{\varepsilon}\right)^{d/\alpha}, \quad \varepsilon > 0,$$

donde $\mathcal{X}^1 := \{x \in \mathbb{R}^d : \text{dist}(x, \mathcal{X}) < 1\}$.

Corolario 2.50 (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 $\text{diam}(\mathcal{X})$, 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 —en cada bola de radio $\eta/2$ que corta a $C_1^\alpha(\mathcal{X})$ se elige un elemento de la clase—, obtenemos $f_1, \dots, f_p \in C_1^\alpha(\mathcal{X})$ 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,

$$N_{[]} (2\eta, C_1^\alpha(\mathcal{X}), L_r(Q)) \leq N\left(\frac{\eta}{2}, C_1^\alpha(\mathcal{X}), \|\cdot\|_\infty\right).$$

Aplicando el Teorema 2.49 en el radio $\eta/2 = \varepsilon/4$, reemplazando $\eta = \varepsilon/2$ y absorbiendo el factor $4^{d/\alpha}$ en la constante,

$$\log N_{[]}(\varepsilon, C_1^\alpha(\mathcal{X}), L_r(Q)) \leq K \lambda(\mathcal{X}^1) \left(\frac{1}{\varepsilon}\right)^{d/\alpha}.$$

Si $D = \text{diam}(\mathcal{X})$ y $x_0 \in \mathcal{X}$, entonces $\mathcal{X}^1 \subset B(x_0, D+1)$; por tanto $\lambda(\mathcal{X}^1) \leq v_d(D+1)^d$.⁸⁰ Absorbiendo esta cantidad en la constante, obtenemos la cota anunciada. $\square$

Consecuencia. El Corolario 2.50, 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 $d/(2\alpha) < 1$.⁸¹ Por ejemplo, en $[0, 1]$ basta una condición Hölder uniforme de orden $> 1/2$; en $[0, 1]^2$ basta cualquier suavidad $\alpha > 1$.

Subgráficas 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 subgráficas suaves bajo una hipótesis de densidad acotada.

Corolario 2.51 (Subgráficas Hölder). *Sea $d \geq 2$ y sea $\mathcal{C}_{\alpha,d}$ la colección de subgráficas en $[0, 1]^d$ de funciones $f \in C_1^\alpha([0, 1]^{d-1})$, es decir, $\mathcal{C}_{\alpha,d} := \{(x, t) \in [0, 1]^{d-1} \times [0, 1] : t \leq f(x)\} : f \in C_1^\alpha([0, 1]^{d-1})\}$. 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 —en cada bola de radio $\eta/2$ que corta a la clase se elige un elemento de ella—, obtenemos $f_1, \dots, f_p \in C_1^\alpha([0, 1]^{d-1})$ que forman una η -red en norma supremo; en particular, los f_i son continuos y los conjuntos que siguen son medibles. Por el Teorema 2.49, aplicado en dimensión $d-1$ y en el radio $\eta/2$, absorbiendo el factor $2^{(d-1)/\alpha}$ en la constante, $\log p \leq K_0 \eta^{-(d-1)/\alpha}$. Para cada i , definimos los conjuntos inferior y superior dentro del mismo cilindro que aparece en la definición de $\mathcal{C}_{\alpha,d}$:

$$\begin{aligned} C_i^- &:= \{(x, t) \in [0, 1]^{d-1} \times [0, 1] : t \leq f_i(x) - \eta\}, \\ C_i^+ &:= \{(x, t) \in [0, 1]^{d-1} \times [0, 1] : t \leq f_i(x) + \eta\}. \end{aligned}$$

Si $C_f := \{(x, t) \in [0, 1]^{d-1} \times [0, 1] : t \leq f(x)\}$ y $\|f - f_i\|_\infty \leq \eta$, entonces $C_i^- \subset C_f \subset C_i^+$. 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 $\|\mathbb{1}_{C_i^+} - \mathbb{1}_{C_i^-}\|_{Q,r} = Q(C_i^+ \setminus C_i^-)^{1/r} \leq (2\eta \|q\|_\infty)^{1/r}$. Así,

$$N_{[]}((2\eta \|q\|_\infty)^{1/r}, \mathcal{C}_{\alpha,d}, L_r(Q)) \leq p = N\left(\frac{\eta}{2}, C_1^\alpha([0, 1]^{d-1}), \|\cdot\|_\infty\right).$$

⁸⁰ 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 $\|y - x_0\| \leq \|y - x\| + \|x - x_0\| < 1 + D$.

⁸¹La 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.

Tomando $\eta = \frac{\varepsilon^r}{2\|q\|_\infty}$ se obtiene

$$\log N_{[]}(\varepsilon, \mathcal{C}_{\alpha,d}, L_r(Q)) \leq K_0 \left(\frac{2\|q\|_\infty}{\varepsilon^r} \right)^{(d-1)/\alpha}.$$

Absorbiendo $2^{(d-1)/\alpha}$ en la constante, queda

$$\log N_{[]}(\varepsilon, \mathcal{C}_{\alpha,d}, L_r(Q)) \leq K \|q\|_\infty^{(d-1)/\alpha} \left(\frac{1}{\varepsilon} \right)^{(d-1)r/\alpha}.$$

□

Observación 2.52 (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 subgráficas. Para $d \geq 2$, del Corolario 2.51 y el teorema de Donsker por bracketing, las subgráficas 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 $f : [0, 1] \rightarrow [0, 1]$, esto se reduce a $\alpha > 1$; es decir, basta una derivada con regularidad Hölder positiva.

Observación 2.53: Reindexación por la dimensión del dominio

Si en lugar de seguir la convención anterior se consideran subgráficos de funciones $f \in C_1^\alpha([0, 1]^d)$, entonces los subgráficos viven en $\mathbb{R}^{d+1}$. Denotando esa clase por

$$\mathcal{C}'_{\alpha,d} := \{\text{Subg}(f) : f \in C_1^\alpha([0, 1]^d)\}, \quad \text{Subg}(f) := \{(x, t) \in [0, 1]^d \times \mathbb{R} : t \leq f(x)\},$$

el mismo argumento da, para toda probabilidad Q sobre $\mathbb{R}^{d+1}$ con densidad acotada q ,

$$\log N_{[]}(\varepsilon, \mathcal{C}'_{\alpha,d}, L_r(Q)) \leq K \|q\|_\infty^{d/\alpha} \left(\frac{1}{\varepsilon} \right)^{dr/\alpha}.$$

En esta parametrización, la condición de Donsker por bracketing es $\alpha > d$. Esta observación es sólo una reindexación de la versión de [4].

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. $X_1, \dots, X_n$, 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.

¹Para 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 $B \neq \emptyset$. 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 $F \subset H$, $\limsup_{n \rightarrow \infty} \mathbb{P}^*(\hat{h}_n \in F \cap A) \leq \mathbb{P}(\hat{h} \in F \cup B^c)$.*

Demostración. Si $F \cap A = \emptyset$, 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 $\limsup_{n \rightarrow \infty} \mathbb{P}^*(\hat{h}_n \in F \cap A) \leq \mathbb{P}(\hat{h} \in F \cup B^c)$. $\square$

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 tenso —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 $G^c \cap K = \emptyset$, la

separación es trivial. Si $G^c \cap K \neq \emptyset$, 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 . $\square$

Corolario 3.3. Sea $(M_n)_{n \geq 1}$ una sucesión de procesos estocásticos indexados por un espacio métrico Θ , y sea $M : \Theta \rightarrow \mathbb{R}$ determinista.

- (I) Supóngase que $\|M_n - M\|_{\Theta} \xrightarrow{\mathbb{P}^*} 0$ y que existe $\theta_0 \in \Theta$ tal que $M(\theta_0) > \sup_{\theta \notin G} M(\theta)$ 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 $r_n \rightarrow \infty$ 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 $h \mapsto a_n \left\{ M_n\left(\theta_0 + \frac{h}{r_n}\right) - M_n(\theta_0) \right\}$, $a_n > 0$. 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 $g_n(h) := L_n\left(\theta_0 + \frac{h}{\sqrt{n}}\right) - L_n(\theta_0)$, $h \in H_n := \sqrt{n}(\Theta - \theta_0)$. 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

$$\tilde{g}_n(h) = \begin{cases} g_n(h), & h \in H_n, \\ -n, & h \notin H_n. \end{cases}$$

Como $0 \in H_n$ y $g_n(0) = 0 > -n$, se tiene la identidad exacta $\sup_{h \in \mathbb{R}^k} \tilde{g}_n = \sup_{h \in H_n} g_n \geq 0$. Además, todo casi-maximizador de $\tilde{g}_n$ 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 $\tilde{g}_n = g_n$ 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 $\Delta_n \Rightarrow \Delta$, $\Delta \sim N_k(0, I_{\theta_0})$, y, para toda sucesión $h_n \rightarrow h$,

$$L_n\left(\theta_0 + \frac{h_n}{\sqrt{n}}\right) - L_n(\theta_0) = h_n^\top \Delta_n - \frac{1}{2} h_n^\top I_{\theta_0} h_n + o_{\mathbb{P}_{\theta_0}}(1).$$

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 $P_{\theta_0} \dot{\ell}_{\theta_0} = 0$, $P_{\theta_0} \dot{\ell}_{\theta_0} \dot{\ell}_{\theta_0}^\top = I_{\theta_0}$. Como $h_n \rightarrow h$ 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 $h_n \rightarrow h$ 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: $(g_n(h_1), \dots, g_n(h_m)) \Rightarrow (g(h_1), \dots, g(h_m))$, $g(h) := h^\top \Delta - \frac{1}{2} h^\top I_{\theta_0} h$. Esta convergencia finito-dimensional, por sí sola, no basta para aplicar el teorema del argmax. Para ello suponemos además la versión uniforme en compactos: $g_n \Rightarrow g$ en $\ell^\infty(K)$ 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 $\tilde{g}_n$ sobre $\mathbb{R}^k$, pues $g_n(\hat{h}_n) = L_n(\hat{\theta}_n) - L_n(\theta_0) \geq \sup_{h \in \mathbb{R}^k} \tilde{g}_n(h) - o_{\mathbb{P}_{\theta_0}}^*(1)$. El proceso límite g tiene trayectorias continuas y, como I_{θ_0} es positiva definida, $\nabla g(h) = \Delta - I_{\theta_0} h$, $\nabla^2 g(h) = -I_{\theta_0} \prec 0$. 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 $\tilde{g}_n \Rightarrow g$ 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 $M : \Theta \rightarrow \mathbb{R}$ determinista. Supóngase que existen $c > 0$ y una vecindad U de θ_0 tales que $M(\theta) - M(\theta_0) \leq -c d^2(\theta, \theta_0)$, $\theta \in U$. Supóngase además que, para todo n y todo $\delta > 0$*

suficientemente pequeño,

$$\mathbb{E}^* \sup_{d(\theta, \theta_0) \leq \delta} |(M_n - M)(\theta) - (M_n - M)(\theta_0)| \lesssim \frac{\phi_n(\delta)}{\sqrt{n}},$$

donde, para algún $\alpha < 2$ independiente de n , $\delta \mapsto \frac{\phi_n(\delta)}{\delta^\alpha}$ es no creciente. Sea $r_n > 0$, con $r_n \rightarrow \infty$, tal que $r_n^2 \phi_n(1/r_n) \leq \sqrt{n}$ para todo n . Supóngase que existe una sucesión de variables aleatorias $\Delta_n \geq 0$, con $\Delta_n = O_{\mathbb{P}^*}(r_n^{-2})$, tal que $M_n(\hat{\theta}_n) \geq M_n(\theta_0) - \Delta_n$ 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 $d(\theta, \theta_0) \leq \eta$, $M(\theta) - M(\theta_0) \leq -\kappa d^2(\theta, \theta_0)$, y la cota del módulo estocástico vale para todo $0 < \delta \leq \eta$. Sea Δ_n la sucesión de la hipótesis de casi-maximización.

Para $j \in \mathbb{Z}$, definamos las coronas diádicas $S_{j,n} := \{\theta : 2^{j-1} < r_n d(\theta, \theta_0) \leq 2^j\}$. 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,

$$\begin{aligned} \mathbb{P}^*\{r_n d(\hat{\theta}_n, \theta_0) > 2^m\} &\leq \mathbb{P}^*\left(\bigcup_{\substack{j > m \\ 2^j < \eta r_n}} \{\hat{\theta}_n \in S_{j,n}\}\right) + \mathbb{P}^*\{2d(\hat{\theta}_n, \theta_0) \geq \eta\} \\ &\leq \mathbb{P}^*(A_{n,C}^c) + \sum_{\substack{j > m \\ 2^j < \eta r_n}} \mathbb{P}^*(\hat{\theta}_n \in S_{j,n}, A_{n,C}) + \mathbb{P}^*\{2d(\hat{\theta}_n, \theta_0) \geq \eta\}. \end{aligned} \quad (1)$$

Sea $C > 0$ y pongamos $A_{n,C} := \{r_n^2 \Delta_n \leq C\}$.

En el evento $A_{n,C} \cap \{\hat{\theta}_n \in S_{j,n}\}$, $\sup_{\theta \in S_{j,n}} \{M_n(\theta) - M_n(\theta_0)\} \geq -Cr_n^{-2}$. Si además $2^j < \eta r_n$, entonces $S_{j,n} \subset \{\theta : d(\theta, \theta_0) \leq \eta\}$, y para todo $\theta \in S_{j,n}$, $M(\theta) - M(\theta_0) \leq -\kappa 2^{2j-2} r_n^{-2}$. Luego, para m suficientemente grande, dependiendo sólo de C y κ , existe $c > 0$ tal que, para todo $j > m$,

$$\mathbb{P}^*(\hat{\theta}_n \in S_{j,n}, A_{n,C}) \leq \mathbb{P}^*\left\{\sup_{\theta \in S_{j,n}} |W_n(\theta) - W_n(\theta_0)| \geq c 2^{2j} r_n^{-2}\right\}.$$

Como $S_{j,n} \subset \{\theta : d(\theta, \theta_0) \leq 2^j/r_n\}$, se tiene $\sup_{\theta \in S_{j,n}} |W_n(\theta) - W_n(\theta_0)| \leq \sup_{d(\theta, \theta_0) \leq 2^j/r_n} |W_n(\theta) - W_n(\theta_0)|$.

Luego, por Markov y por la hipótesis del módulo, con $\delta = 2^j/r_n$,

$$\begin{aligned} \mathbb{P}^*(\hat{\theta}_n \in S_{j,n}, A_{n,C}) &\lesssim \frac{r_n^2}{2^{2j}} \mathbb{E}^* \sup_{d(\theta, \theta_0) \leq 2^j/r_n} |W_n(\theta) - W_n(\theta_0)| \\ &\lesssim \frac{r_n^2 \phi_n(2^j/r_n)}{2^{2j} \sqrt{n}}. \end{aligned}$$

Para $j \geq m \geq 0$, la monotonía de $\delta \mapsto \phi_n(\delta)/\delta^\alpha$ implica $\phi_n(2^j/r_n) \leq 2^{j\alpha} \phi_n(1/r_n)$, siempre que $2^j/r_n \leq \eta$. Usando $r_n^2 \phi_n(1/r_n) \leq \sqrt{n}$, $\frac{r_n^2 \phi_n(2^j/r_n)}{2^{2j} \sqrt{n}} \lesssim 2^{j(\alpha-2)}$. Por consiguiente,

$$\sum_{\substack{j > m \\ 2^j < \eta r_n}} \mathbb{P}^*(\hat{\theta}_n \in S_{j,n}, A_{n,C}) \lesssim \sum_{j > m} 2^{j(\alpha-2)} \lesssim 2^{m(\alpha-2)}. \quad (2)$$

De (1) y (2),

$$\begin{aligned} \limsup_n \mathbb{P}^*\{r_n d(\hat{\theta}_n, \theta_0) > 2^m\} &\lesssim 2^{m(\alpha-2)} + \limsup_n \mathbb{P}^*(A_{n,C}^c) \\ &\quad + \limsup_n \mathbb{P}^*\{2d(\hat{\theta}_n, \theta_0) \geq \eta\}. \end{aligned}$$

Por consistencia, $\limsup_n \mathbb{P}^*\{2d(\hat{\theta}_n, \theta_0) \geq \eta\} = 0$. Además, como $r_n^2 \Delta_n = O_{\mathbb{P}^*}(1)$, el término $\limsup_n \mathbb{P}^*(A_{n,C}^c)$ se hace arbitrariamente pequeño tomando C suficientemente grande.

Finalmente, como $\alpha < 2$, $2^{m(\alpha-2)} \rightarrow 0$ ($m \rightarrow \infty$). Así, $\lim_{m \rightarrow \infty} \limsup_n \mathbb{P}^*\{r_n d(\hat{\theta}_n, \theta_0) > 2^m\} = 0$, 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 $\hat{\theta}_n \xrightarrow{\mathbb{P}^*} \theta_0$. $\square$

Bibliografía

- [1] P. Billingsley. *Convergence of Probability Measures. 2nd ed.* John Wiley & Sons., 1999.
- [2] G. Folland. *Real Analysis: Modern techniques and their applications.* Wiley, 1999.
- [3] A.N. Shiryaev. *Probability.* Springer-Verlag, 1984.
- [4] 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.