



CIMAT

---

Centro de Investigación  
en Matemáticas, A.C.

# Teoría Asintótica para probabilidad y estadística

## Notas del curso

Autor:

inm

Profesor:

Emilien Antoine Marie Joly

CIMAT

Guanajuato

18 de mayo de 2023

# Índice general

---

<b>I. Teoremas límites uniformes</b>	<b>4</b>
1. Convergencias con dependencias débiles	5
1.1. Ley fuerte de grandes números . . . . .	5
1.2. Teoremas de límite central . . . . .	5
2. Convergencia de procesos empíricos	7
2.1. Ley de grandes números . . . . .	11
3. Introducción a la dimensión de Vapnik–Chervonenkis	20
4. Teorema central del límite	26
 <b>II. Temas relacionados de geometría convexa y probabilidad</b>	<b>29</b>
5. Desigualdad de Brunn–Minkowski	31
5.1. Consecuencias en concentración de la medida . . . . .	33
5.2. Log-concavidad . . . . .	35
5.3. Primeros ejemplos de convexos . . . . .	41
5.3.1. Ejemplos fundamentales . . . . .	41
5.4. La gran conjetura . . . . .	43
6. Concentración de la medida	45
6.1. Concentración clásica . . . . .	45
6.2. Desigualdades de entropía (teoría de la información) . . . . .	48
6.3. Desigualdades Log-Sobolev . . . . .	49
 <b>III. Tareas</b>	<b>51</b>
 <b>IV. Referencias</b>	<b>53</b>
Referencias	54

## Índice de fechas

---

25 de enero de 2023	5	8 de marzo de 2023	23
30 de enero de 2023	7	15 de marzo de 2023	26
1 de febrero de 2023	9	22 de marzo de 2023	31
8 de febrero de 2023	11	27 de marzo de 2023	32
13 de febrero de 2023	13	17 de abril de 2023	34
15 de febrero de 2023	15	19 de abril de 2023	36
20 de febrero de 2023	17	24 de abril de 2023	38
1 de marzo de 2023	20	26 de abril de 2023	41
6 de marzo de 2023	21	17 de mayo de 2023	45

**Parte I.**

**Teoremas límites uniformes**

# Convergencias con dependencias débiles

1

25/01/2023

## 1.1. Ley fuerte de grandes números

Veremos la ley fuerte de grandes números usando teoría de martingalas.

### 1. Proposición

Sean  $X_1, \dots, X_n$  variables aleatorias reales integrables centradas. Supongamos:

- Existe  $r \geq 0$  tal que si  $|i - j| \geq r$ , entonces  $X_i$  y  $X_j$  son independientes.
- Para cualesquiera  $i, j \in \llbracket 1, r \rrbracket$ ,  $\mathbb{E}[\text{Var}(X_i | \mathcal{F}_{i-j})] \leq \sigma_i^2$  con  $\sum_{i=1}^{\infty} \sigma_i^2 < \infty$ .

Entonces

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{\text{c.s.}} 0.$$

Demostración: Consideremos  $Y_i = X_i - \mathbb{E}[X_i | \mathcal{F}_{i-1}]$  y definamos  $M_n = \sum_{i=1}^n Y_i$ . Es fácil ver que  $\{M_n\}$  es una martingala. Entonces

$$\mathbb{E}[M_n^2] = \sum_{i=1}^n \mathbb{E}[(M_i - M_{i-1})^2] = \sum_{i=1}^n \mathbb{E}[Y_i^2] = \sum_{i=1}^n \mathbb{E}[\text{Var}(X_i | \mathcal{F}_{i-1})] < \sum_{i=1}^{\infty} \sigma_i^2 < \infty.$$

Por ende  $\{M_n\}$  es una martingala acotada en  $L^2$ , de modo que es casi seguramente convergente. De igual manera, si definimos  $Z_i = \mathbb{E}[X_{i+1} | \mathcal{F}_i] - \mathbb{E}[X_{i+1} | \mathcal{F}_{i-1}]$  y  $N_n = \sum_{i=1}^n Z_i$ , entonces  $\{N_n\}$  es otra martingala acotada en  $L^2$ .

Continuando con el procedimiento, podemos escribir  $S_n = \sum_{i=1}^n X_i$  como una suma de  $r$  martingalas acotadas en  $L^2$ . Por lo tanto  $\{S_n\}$  converge casi seguramente a un límite  $S \in L^2$  y en consecuencia  $\{S_n/n\}$  converge casi seguramente a 0. ■

## 1.2. Teoremas de límite central

Comenzaremos considerando las variables aleatorias  $\{X_n\}$ , las cuales repartiremos en bloques de diferentes tamaños. Para ello tomaremos sucesiones  $\{p_n\}$  y  $\{q_n\}$  con las propiedades  $p_n, q_n \rightarrow \infty$ ,  $q_n = o(p_n)$  y  $p_n = o(n)$ .

Tomando  $k = \lfloor n/(p+q) \rfloor$  definamos, para  $1 \leq i \leq k$  a

$$\xi_i = \sum_{j=(i-1)p+(1-i)q+1}^{ip+(i-1)q} X_j \quad \text{y} \quad \eta_i = \sum_{j=ip+(i-1)q+1}^{ip+iq} X_j,$$

y

$$\eta_{k+1} = \sum_{j=k(p+q)+1}^n X_j.$$

Con ello escribimos  $S_n = Z_k + Z'_{k+1}$  con

$$Z_k = \sum_{i=1}^k \xi_i \quad \text{y} \quad Z'_{k+1} = \sum_{i=1}^{k+1} \nu_i.$$

A este tipo de particiones se le conoce como *bloques de Bernstein*.

### 1. Teorema

Sean  $\{X_n\}$  variables aleatorias cuadrado integrables,  $S_n = \sum_{i=1}^n X_i$  y  $\sigma_n^2 = \text{Var}(S_n)$ . Supongamos:

- (i)  $\sigma_n^{-2} \mathbb{E}[(Z'_{k+1})^2] \rightarrow 0$  conforme  $n \rightarrow \infty$ .
- (ii) Para cualesquiera  $t \in \mathbb{R}$  y  $g, h \in \{\text{sen}, \text{cos}\}$ ,

$$C_{k,g,h}(t) = \sum_{j=2}^k \text{Cov}\left(g\left(\frac{t}{\sigma_n} \sum_{i=1}^{j-1} \xi_i\right), h\left(\frac{t}{\sigma_n} \xi_j\right)\right) \xrightarrow{n \rightarrow \infty} 0.$$

- (iii) Para toda  $\varepsilon > 0$ ,

$$\frac{1}{\sigma_n^2} \sum_{i=1}^k \mathbb{E}[\xi_i^2 \mathbb{1}_{\{|\xi_i| \geq \varepsilon \sigma_n\}}] \xrightarrow{n \rightarrow \infty} 0.$$

$$\sigma_n^{-2} \sum_{i=1}^k \mathbb{E}[\xi_i^2] \rightarrow 1.$$

Entonces  $\{S_n/\sigma_n\}$  converge en distribución a una variable aleatoria cuya distribución es  $N(0, 1)$ .

# Convergencia de procesos empíricos

# 2

30/01/2023

En los cursos clásicos se demuestra que  $\bar{X}_n \rightarrow \mathbb{E}[X]$  con probabilidad uno, donde  $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$  con  $\{X_i : i \geq 1\}$  i.i.d. En adición se tiene que

$$\sqrt{n}(\bar{X}_n - \mathbb{E}[X]) \Rightarrow Z$$

con  $Z$  distribuida acorde a  $N(0, \Sigma)$ .

Ahora consideraremos que  $X_i \in E$ , donde  $E$  es un subconjunto de un espacio vectorial topológico. Algunos casos importantes son:

- $X \in H$  donde  $H$  es un espacio de Hilbert;
- $X \in L^2(\mathbb{R}_+)$ ;
- $X$  una función (aleatoria) sobre gráficas aleatorias.

**Ejemplo 1** Tomemos  $X_i = (Y_i, Z_i)$  una muestra de personas, donde  $Y_i \in \{0, 1\}$  y  $Z_i \in \mathbb{R}_+$ . Una cuestión de interés puede ser

$$F_0(z) = \mathbb{P}(Y = 1 \mid Z = z),$$

donde  $F_0$  es creciente y  $0 \leq F_0(z) \leq 1$ . Luego, el espacio de parámetros es

$$\Lambda := \{F : \mathbb{R} \rightarrow [0, 1] \mid F \text{ creciente}\},$$

es decir que queremos estimar una función, algo que no es de dimensión finita.

Un posible estimador es el de máxima verosimilitud, dado por

$$\hat{F}_n = \arg \max_{F \in \Lambda} \left\{ \sum_{i=1}^n Y_i \log F(z_i) + (1 - Y_i) \log(1 - F(z_i)) \right\}.$$

Para medir qué tan bien se comporta el estimador, dada  $Q$  la distribución de  $Z$ , definimos

$$\|\hat{F}_n - F_0\|_Q = \left( \int (\hat{F}_n(z) - F_0(z))^2 dQ(z) \right)^{1/2}.$$

Con la teoría que se verá, se podrá verificar que  $\|\hat{F}_n - F_0\|_Q = O_p(n^{-1/3})$ .

En caso de que

$$\Lambda = \left\{ F : \mathbb{R} \rightarrow [0, 1] \mid 0 < \frac{dF}{dz} \leq M, F(z) \text{ cóncava} \right\},$$

se puede mostrar que  $\|\hat{F}_n - F_0\|_Q = O_p(n^{-2/5})$ .

Finalmente, si

$$\Lambda = \{F : \mathbb{R} \rightarrow [0, 1] \mid F(z) = F_0(\theta z), \theta \in \mathbb{R}\} \quad \text{donde } F_0(x) = \frac{e^x}{1 + e^x},$$

entonces  $\|\hat{F}_n - F_0\|_Q = O_p(n^{-1/2})$ .

**Ejemplo 2** Sea  $\{\xi_n : n \geq 1\}$  variables aleatorias i.i.d. centradas con  $\mathbb{E}[\xi_1^2] = 1$ . Definimos la caminata aleatoria  $S_n = \sum_{i=1}^n \xi_i$ . Por el teorema de límite central sabemos que  $\{n^{-1}S_n\}$  converge débilmente a una variable aleatoria con distribución normal estándar. Ahora bien, si consideramos  $n_1 < \dots < n_k$  tales que  $n_i/n \rightarrow t_i$  para  $0 = t_0 \leq t_1 \leq \dots \leq t_k$  y definimos  $S_{n_i} - S_{n_{i-1}}$ , entonces el teorema de límite central vectorial nos dice que

$$\left( \frac{S_{n_i} - S_{n_{i-1}}}{\sqrt{n}}, 1 \leq i \leq k \right) \Rightarrow (N_1, \dots, N_k)$$

con  $N_i \sim N(0, t_i - t_{i-1})$ . Así, por mapeo continuo,

$$\left( \frac{S_{n_i}}{n}, 1 \leq i \leq k \right) \Rightarrow (N_1, \dots, N_1 + \dots + N_k).$$

Nuevamente, con la teoría que veremos, tendremos que

$$\{S_t^{[n]} : t \in [0, 1]\} \Rightarrow \{B_t : t \in [0, 1]\},$$

donde

$$S_t^{[n]} = \frac{1}{\sqrt{n}} S_{[nt]}, \quad t \geq 0.$$

En general hay que trabajar más para demostrar que

$$\sup_{0 \leq t \leq 1} S_t^{[n]} \Rightarrow \sup_{0 \leq t \leq 1} B_t.$$

Esto será una consecuencia directa de nuestra teoría.

### 1. Definición

Diremos que una colección de elementos aleatorios  $\{X_n\}$  con valores en  $E$  converge débilmente a  $X$  si para toda función  $h : E \rightarrow \mathbb{R}$  Lipschitz y acotada,

$$\mathbb{E}[h(X_n)] \xrightarrow{n \rightarrow \infty} \mathbb{E}[h(X)].$$

### 2. Definición (Entropía métrica)

Sea  $E$  un espacio métrico con distancia  $d$ . Denotamos por  $\mathcal{N}(\epsilon, E, d)$  el *número de cobertura*, esto es el menor número de bolas  $B_d(x, \epsilon)$  que se requieren para cubrir  $E$ . La  $\epsilon$ -*entropía métrica* será la cantidad

$$H(\epsilon, E, d) = \log \mathcal{N}(\epsilon, E, d).$$

### 3. Definición

Diremos que  $E$  es *totalmente acotado* si para todo  $\epsilon > 0$  se cumple  $H(\epsilon, E, d) < \infty$ .

### 4. Definición (Entropía de funciones)



Si  $E = \mathfrak{F}$  es una clase de funciones y  $d = \|\cdot\|_{p,Q}$  donde

$$\|\cdot\|_{p,Q} = \left( \int |\cdot|^p dQ \right)^{1/p}.$$

La entropía será

$$H_p(\epsilon, \mathfrak{F}, Q) := H(\epsilon, \mathfrak{F}, \|\cdot\|_{p,Q}).$$

### 5. Definición (Entropía horquillada)

Definimos a  $\mathcal{N}_{p,B}(\epsilon, \mathfrak{F}, Q)$  como el menor número  $N$  para el cual existan parejas de funciones  $(f_i^L, f_i^R)_{i=1}^N$  tales que

- $\|f_i^L - f_i^R\|_{p,Q} \leq \epsilon$  para toda  $i$ ,
- para cada  $f \in \mathfrak{F}$  existe  $i \in \{1, \dots, N\}$  para el cual se satisface  $f_i^L \leq f \leq f_i^R$ .

En este caso la entropía horquillada o euclidiana de  $\mathfrak{F}$  será

$$H_{p,B}(\epsilon, \mathfrak{F}, Q) = \log \mathcal{N}_{p,B}(\epsilon, \mathfrak{F}, Q).$$

### 6. Definición (Entropía por la norma sup)

Si  $d = \|\cdot\|_\infty$  denotamos por  $H_\infty(\epsilon, \mathfrak{F})$  a la entropía. De manera equivalente,  $\mathcal{N}_\infty(\epsilon, \mathfrak{F})$  es el menor entero  $N$  para el cual existe una colección de funciones  $(f_i)_{i=1}^N$  para la cual

$$\sup_{f \in \mathfrak{F}} \min_{1 \leq i \leq N} \|f - f_i\|_\infty \leq \epsilon.$$

01/02/2023

Dada  $B = B_{\|\cdot\|_2}(0, R)$ , queremos encontrar una cota superior de  $\mathcal{N}(\epsilon, B, \|\cdot\|_2)$ .

#### 1. Lema

Una bola  $B_{\|\cdot\|_2}(0, R)$  puede cubrirse con  $\lceil (1 + 2R/\epsilon)^d \rceil$  bolas de radio  $\epsilon > 0$ .

Demostración: Sea  $\{c_j : 1 \leq j \leq N\}$  un conjunto de tamaño máximo en  $B_{\|\cdot\|_2}(0, R)$  tal que cada dos puntos satisfacen  $\|c_{j_1} - c_{j_2}\|_2 \geq \epsilon$ . Entonces

$$\bigcup_{j=1}^N B_{\|\cdot\|_2}(c_j, \epsilon)$$

es una cobertura de  $B_{\|\cdot\|_2}(0, R)$ .

Denotemos por  $\{B_j\}_{j=1}^N$  a las bolas centradas en  $\{c_j\}_{j=1}^N$  de radio  $\epsilon/2$ , las cuales son disjuntas. Entonces

$$\bigcup_{j=1}^N B_j \subset B_{\|\cdot\|_2}(0, R + \epsilon/2),$$

de modo que

$$\text{Vol}\left(\bigcup_{j=1}^N B_j\right) \leq \text{Vol}(B_{\|\cdot\|_2}(0, R + \epsilon/2)),$$

y por lo tanto, poniendo  $\mu_d = \text{Vol}(B(0, 1))$ ,

$$\sum_{j=1}^N \left(\frac{\epsilon}{2}\right)^d \mu_d \leq \left(R + \frac{\epsilon}{2}\right)^d.$$

De esto último deducimos que

$$N \leq \left(1 + \frac{2R}{\varepsilon}\right)^d,$$

de donde se sigue el resultado. ■

### 2. Lema

Para cada  $p \in [1, \infty)$  se cumple

$$H_p(\varepsilon, \mathfrak{F}, Q) \leq H_{p,B}(\varepsilon, \mathfrak{F}, Q)$$

para todo  $\varepsilon > 0$ . En adición, si  $Q$  es una medida de probabilidad, entonces para cualquier  $\varepsilon > 0$  tendremos que

$$H_{p,B}(\varepsilon, \mathfrak{F}, Q) \leq H_\infty\left(\frac{\varepsilon}{2}, \mathfrak{F}\right).$$

### 3. Lema

Si  $d$  y  $d'$  son dos distancias tales que

$$d(x, y) \leq d'(x, y)$$

para cualesquiera  $x, y \in E$ , entonces

$$H(\varepsilon, E, d) \leq H(\varepsilon, E, d').$$

**Ejemplo 1** Si  $E = \{x \in \mathbb{R} : |x| \leq k\}$  entonces  $\mathcal{N}(\varepsilon, E, |\cdot|) \leq \lceil k/\varepsilon \rceil$ .

**Ejemplo 2** Sean  $\Phi_1, \dots, \Psi_d \in L^2(Q)$ . Definamos

$$\mathfrak{F} = \left\{ \sum_{k=1}^d \theta_k \Psi_k : \theta \in \mathbb{R}^d, \|f\|_{2,Q} \leq R \right\}.$$

Entonces  $H_2(\varepsilon, \mathfrak{F}, Q) \leq d_Q \log(1 + 2R/\varepsilon)$  para cualquier  $\varepsilon > 0$ , donde  $d_Q$  es el rango de la matriz

$$\Sigma_Q = \int \Psi \Psi^T dQ \quad \Psi = (\Psi_1, \dots, \Psi_d).$$

En efecto, si  $\Psi'_1, \dots, \Psi'_{d_Q}$  es una base ortonormal de  $\mathfrak{F}$ , entonces

$$f = \sum_{i=1}^d \theta'_i \Psi'_i \mapsto (\theta'_1, \dots, \theta'_{d_Q})$$

es la biyección que da la isometría.

**Ejemplo 3** Sea  $\mathfrak{F} = \{f : \mathcal{X} \rightarrow [0, 1] \text{ creciente}\}$  para  $\mathcal{X}$  un conjunto finito con  $|\mathcal{X}| = n$ . Entonces

$$H_\infty(\varepsilon, \mathfrak{F}) \leq \left\lfloor \frac{1}{\varepsilon} \right\rfloor \log\left(n + \left\lfloor \frac{1}{\varepsilon} \right\rfloor\right).$$

Podemos suponer que  $\mathcal{X} = \{x_1, \dots, x_n\}$  con  $x_1 < \dots < x_n$  y para  $f \in \mathfrak{F}$  sea

$$M_i^f = \left\lfloor \frac{f(x_i)}{\varepsilon} \right\rfloor, \quad 1 \leq i \leq n.$$

Así, al definir  $\tilde{f} = \varepsilon M^f$ , deducimos que  $\|f - \tilde{f}\|_\infty \leq \varepsilon$ . Denotamos  $\tilde{\mathfrak{F}} = \{\tilde{f} : f \in \mathfrak{F}\}$ , el cual es un conjunto finito debido a que

$$0 \leq M^f \leq \left\lfloor \frac{1}{\varepsilon} \right\rfloor \quad \text{y} \quad M^f \in \mathbb{N}^n.$$

Más aún, como

$$0 \leq M_1^f \leq M_2^f \leq \dots \leq M_n^f \leq \left\lfloor \frac{1}{\varepsilon} \right\rfloor,$$

obtenemos

$$|\tilde{\mathfrak{F}}| \leq \binom{n + \lfloor 1/\varepsilon \rfloor}{\lfloor 1/\varepsilon \rfloor} \leq \left( n + \left\lfloor \frac{1}{\varepsilon} \right\rfloor \right)^{\lfloor 1/\varepsilon \rfloor}.$$

**Ejemplo 4** Sea  $\mathfrak{F} = \{f : [0, 1] \rightarrow [0, 1] \mid |f'| \leq 1\}$ . Entonces existe A constante tal que

$$H_\infty(\varepsilon, \mathfrak{F}) \leq \frac{A}{\varepsilon}$$

para cada  $\varepsilon > 0$ .

## 2.1. Ley de grandes números

08/02/2023

### 1. Definición (Proceso empírico)

Sean  $X_1, \dots, X_n$  elementos aleatorios en un espacio medible  $\mathcal{X}$  de ley  $\mathbb{P}$ . Sea  $\mathfrak{F}$  una clase de funciones  $f : \mathcal{X} \rightarrow \mathbb{R}$ . El proceso

$$\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n (f(X_i) - \mathbb{E}[f(X_i)]) \right\}_{f \in \mathfrak{F}}$$

es conocido como el *proceso empírico* sobre  $\mathfrak{F}$ .

Abusando de la notación, diremos que un proceso  $X^n = \{X_t^n : t \in \mathcal{T}\}$  como proceso empírico.

### 2. Definición ( $\mathbb{P}$ -Glivenko–Cantelli)

Decimos que un proceso empírico, o la clase  $\mathfrak{F}$ ,

$$\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n (f(X_i) - \mathbb{E}[f(X_i)]) \right\}_{f \in \mathfrak{F}}$$

es  $\mathbb{P}$ -Glivenko–Cantelli si

$$\sup_{f \in \mathfrak{F}} \frac{1}{n} \left| \sum_{i=1}^n (f(X_i) - \mathbb{E}[f(X_i)]) \right| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

### 1. Observación

Es posible usar argumentos similares a los que permiten demostrar la ley fuerte de grandes números que la convergencia en probabilidad o casi segura son equivalentes para un proceso empírico.

En el siguiente resultado generalizamos el teorema de Glivenko–Cantelli bajo el supuesto de una entropía horquillada finita. La prueba es esencialmente la misma.

### 1. Teorema

Sea  $\mathfrak{F}$  un espacio métrico de funciones. Supongamos que  $H_{1,B}(\varepsilon, \mathfrak{F}, \mathbb{P}) < \infty$  para toda  $\varepsilon > 0$ . Entonces  $\mathfrak{F}$  es  $\mathbb{P}$ -Glivenko–Cantelli.

Demostración: Sea  $\varepsilon > 0$ . Por hipótesis tenemos que  $H_{1,B}(\varepsilon, \mathfrak{F}, \mathbb{P}) < \infty$ , es decir que existe una colección finita  $(f_i^L, f_i^R)_{i \leq N}$  tal que para toda  $f \in \mathfrak{F}$  existe  $i$  con  $f_i^L \leq f \leq f_i^R$  y  $\|f_i^L - f_i^R\|_{1,\mathbb{P}} \leq \varepsilon$ . Entonces

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)] &\leq \frac{1}{n} \sum_{i=1}^n f_i^R(X_i) - \mathbb{E}[f_i^L(X)] \\ &\leq \frac{1}{n} \sum_{i=1}^n f_i^R(X_i) - \mathbb{E}[f_i^R(X)] + \varepsilon. \end{aligned}$$

De manera similar,

$$\frac{1}{n} \sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)] \geq \frac{1}{n} \sum_{i=1}^n f_i^L(X_i) - \mathbb{E}[f_i^L(X)] - \varepsilon.$$

Ahora, por la ley de grandes números,

$$M^L = \max_{1 \leq j \leq N} \left| \frac{1}{n} \sum_{i=1}^n f_j^L(X_i) - \mathbb{E}[f_j^L(X)] \right| \xrightarrow[n \rightarrow \infty]{\text{c.s.}} 0$$

y

$$M^R = \max_{1 \leq j \leq N} \left| \frac{1}{n} \sum_{i=1}^n f_j^R(X_i) - \mathbb{E}[f_j^R(X)] \right| \xrightarrow[n \rightarrow \infty]{\text{c.s.}} 0.$$

Concluimos de ello y la desigualdad

$$\sup_{f \in \mathfrak{F}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)] \right| \leq M^L \vee M^R + \varepsilon. \quad \blacksquare$$

### 3. Definición (Función envolvente)

Dada  $\mathfrak{F}$  definimos la *envolvente* de  $\mathfrak{F}$  como la función

$$F = \sup_{f \in \mathfrak{F}} |f|.$$

### 1. Lema

Si  $H_{1,B}(\varepsilon, \mathfrak{F}, \mathbb{P}) < \infty$  para toda  $\varepsilon > 0$ , entonces  $F \in L^1(\mathbb{P})$ .

Demostración: Es claro que  $\mathfrak{F} \subset L^1(\mathbb{P})$ . Puesto que  $L^1(\mathbb{P})$  es completo,  $\overline{\mathfrak{F}}$  es completo. En adición  $\mathfrak{F}$  es pre-compacto, de modo que  $\overline{\mathfrak{F}}$  es compacto. Entonces existe  $M > 0$  tal que

$$\sup_{f \in \mathfrak{F}} \|f\|_{1,\mathbb{P}} \leq M.$$

Ahora, para  $f \in \mathfrak{F}$ ,

$$|f| \leq \sum_{i=1}^N |f_i^L| + |f_i^R - f_i^L|,$$

y por consiguiente,

$$\|F\|_{1,p} \leq N(M + \varepsilon).$$

## 2. Observación

La condición  $F \in L^1(\mathbb{P})$  es necesaria en el caso  $\mathfrak{F}$  acotado en  $L^1(\mathbb{P})$  para obtener que  $\mathfrak{F}$  es  $\mathbb{P}$ -Glivenko–Cantelli.

Demostración: En efecto, si  $\mathfrak{F}$  es  $\mathbb{P}$ -Glivenko–Cantelli,

$$\sup_{f \in \mathfrak{F}} \frac{1}{n} \left| \sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)] \right| \xrightarrow[n \rightarrow \infty]{\text{c.s.}} 0,$$

entonces por Borel–Cantelli,

$$\mathbb{E} \left[ \sup_{f \in \mathfrak{F}} |f(X) - \mathbb{E}[f(X)]| \right] < \infty,$$

lo cual nos da  $F \in L^1(\mathbb{P})$ .

## 2. Teorema

Si  $F \in L^1(\mathbb{P})$  y para toda  $\varepsilon > 0$  se cumple

$$\frac{1}{n} H_1(\varepsilon, \mathcal{F}, P_n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0,$$

donde  $P_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$  entonces  $\mathfrak{F}$  es  $\mathbb{P}$ -Glivenko–Cantelli.

### 1. Afirmación

Se puede mostrar que el segundo supuesto en el teorema 2 es necesario por la *minoración de Sudakov*.

13/02/2023

Para la demostración del teorema 2, usaremos las nociones de *simetrización* y *chaining*.

**Simetrización:** Decimos que una variable aleatoria  $X$  es simétrica cuando ella tiene la misma distribución que  $\varepsilon X$  donde  $\varepsilon$  tiene distribución Rademacher de parámetro  $1/2$  y es independiente de  $X$ .

## 4. Definición

Si  $Y$  es una variable aleatoria, entonces definimos su *simetrización* por  $\varepsilon Y$ .

## 2. Lema

Sean  $\{X^n : n \geq 1\}$  procesos aleatorios independientes, i.e.  $X^i = \{X_s^i : s \in T\}$  para cada  $i$ . Supongamos que para cualesquiera  $i$  y  $s$ ,  $\mathbb{E}[X_s^i] = 0$ . Sea  $\{\varepsilon_n : n \geq 1\}$  una sucesión

de variables aleatorias independientes con distribución Rademacher de parámetro 1/2 independiente de  $\{X^n : n \geq 1\}$ . Entonces

$$\frac{1}{2} \mathbb{E} \left[ \sup_{s \in \mathcal{T}} \left| \sum_{i=1}^n \varepsilon_i X_s^i \right| \right] \leq \mathbb{E} \left[ \sup_{s \in \mathcal{T}} \left| \sum_{i=1}^n X_s^i \right| \right] \leq 2 \mathbb{E} \left[ \sup_{s \in \mathcal{T}} \left| \sum_{i=1}^n \varepsilon_i X_s^i \right| \right].$$

Demostración: Iniciamos con la segunda desigualdad. Para ello notemos que

$$\mathbb{E} \left[ \sup_{s \in \mathcal{T}} \left| \sum_{i=1}^n X_s^i \right| \right] = \mathbb{E} \left[ \sup_{s \in \mathcal{T}} \left| \sum_{i=1}^n (X_s^i - \mathbb{E}[X_s^{i'}]) \right| \right]$$

donde  $X_s^{i'}$  es una copia de  $X_s^i$  independiente. Por consiguiente,

$$\begin{aligned} \mathbb{E} \left[ \sup_{s \in \mathcal{T}} \left| \sum_{i=1}^n (X_s^i - \mathbb{E}[X_s^{i'}]) \right| \right] &= \mathbb{E}_X \left[ \sup_{s \in \mathcal{T}} \left| \mathbb{E}_{X'} \left[ \sum_{i=1}^n (X_s^i - X_s^{i'}) \right] \right| \right] \\ &\leq \mathbb{E} \left[ \sup_{s \in \mathcal{T}} \left| \sum_{i=1}^n (X_s^i - X_s^{i'}) \right| \right] \\ &= \mathbb{E} \left[ \sup_{s \in \mathcal{T}} \left| \sum_{i=1}^n \varepsilon_i (X_s^i - X_s^{i'}) \right| \right] \\ &\leq 2 \mathbb{E} \left[ \sup_{s \in \mathcal{T}} \left| \sum_{i=1}^n \varepsilon_i X_s^i \right| \right], \end{aligned}$$

con la segunda igualdad dándose debido a que  $X - X'$  tiene la misma distribución que  $\varepsilon(X - X')$ . La desigualdad que resta se prueba de manera análoga. ■

### 3. Observación

Es posible demostrar que

$$\mathbb{E} \left[ \sup_{s \in \mathcal{T}} \sum_{i=1}^n X_s^i \right] \leq 2 \mathbb{E} \left[ \sup_{s \in \mathcal{T}} \sum_{i=1}^n \varepsilon_i X_s^i \right].$$

También se tienen resultados de la forma:

### 3. Lema

Supongamos que para toda  $f \in \mathfrak{F}$ ,

$$\mathbb{P} \left( |\langle P_n - P, f \rangle| > \frac{\delta}{2} \right) \leq \frac{1}{2}.$$

Entonces

$$\mathbb{P} \left( \sup_{f \in \mathfrak{F}} |\langle P_n - P, f \rangle| > \delta \right) \leq 2 \mathbb{P} \left( \sup_{f \in \mathfrak{F}} |\langle P_n - P'_n, f \rangle| > \frac{\delta}{2} \right),$$

donde  $P'_n$  es una copia independiente de  $P_n$ .

(Ejercicio) Buscar referencias para la demostración.

Caos de orden  $j$ :

$$\sum_{i_1 \neq i_2 \neq \dots \neq i_j} \varepsilon_{i_1} \cdots \varepsilon_{i_j} X_{i_1} \cdots X_{i_j}.$$

**Chaining:** Algunos preliminares:

### 3. Teorema

Sean  $X_1, \dots, X_N$  variables aleatorias gaussianas centradas y con varianza  $v$ . Entonces se cumple,

$$\mathbb{E} \left[ \max_{1 \leq i \leq N} X_i \right] \leq \sqrt{2v \log N}.$$

Demostración: Podemos suponer sin pérdida de generalidad que  $v = 1$ . Luego, por Jensen,

$$\exp \left\{ \lambda \mathbb{E} \left[ \max_{1 \leq i \leq N} X_i \right] \right\} \leq \mathbb{E} \left[ \max_{1 \leq i \leq N} e^{\lambda X_i} \right] \leq \sum_{i=1}^N \mathbb{E} [e^{\lambda X_i}] = N \exp \left\{ \frac{\lambda^2}{2} \right\}.$$

De manera equivalente,

$$\mathbb{E} \left[ \max_{1 \leq i \leq N} X_i \right] \leq \frac{\log N}{\lambda} + \frac{\lambda}{2}.$$

Como esta desigualdad es válida para todo  $\lambda > 0$ , la desigualdad AM–GM nos da el resultado. ■

### 4. Observación

El resultado sigue siendo cierto si suponemos que las variables aleatorias  $X_1, \dots, X_N$  son *subgaussianas de constante  $v$* , es decir que para toda  $\lambda > 0$ ,

$$\mathbb{E} [e^{\lambda X_i}] \leq \exp \left\{ \frac{\lambda^2 v}{2} \right\}.$$

### 4. Teorema (Dudley)

Sea  $\mathcal{T}$  un espacio métrico. Sea  $\{X_t : t \in \mathcal{T}\}$  una colección de variables aleatorias con la propiedad

$$\log \mathbb{E} [\exp \{\lambda (X_t - X_{t'})\}] \leq \frac{\lambda^2 d(t, t')^2}{2}$$

para cualesquiera  $t, t' \in \mathcal{T}$  y  $\lambda > 0$ . Entonces, para cualquier  $t_0 \in \mathcal{T}$ ,

$$\mathbb{E} \left[ \sup_{t \in \mathcal{T}} (X_t - X_{t_0}) \right] \leq 12 \int_0^{\delta/2} \sqrt{H(\epsilon, \mathcal{T}, d)} d\epsilon,$$

donde

$$\delta = \sup_{t \in \mathcal{T}} d(t, t_0).$$

Demostración: Supondremos que  $\mathcal{T}$  es totalmente acotado. Para cada  $j \in \mathbb{N}$  sea  $\delta_j = \delta 2^{-j}$  y sea

$$\mathcal{T}_j = \left\{ x_i : \bigcup_{i=1}^{N(\delta_j, \mathcal{T}, d)} B(x_i, \delta_j) \supset \mathcal{T} \right\}.$$

15/02/2023

Además definimos  $\mathcal{T}_0 = \{t_0\}$ . Para cada  $j$  definimos  $\Pi_j : \mathcal{T} \rightarrow \mathcal{T}_j$  el mapeo que a cada  $t$  le asigna un valor  $\Pi_j(t)$  tal que  $d(t, \Pi_j(j)) \leq \delta_j$ . Es claro que existe tal función pero no es necesariamente única. Entonces, con probabilidad uno,

$$X_t = X_{\Pi_0(t)} + \sum_{j \geq 0} (X_{\Pi_{j+1}(t)} - X_{\Pi_j(t)})$$

en vista del supuesto dado. Por consiguiente,

$$\mathbb{E} \left[ \sup_{t \in \mathcal{T}} (X_t - X_{t_0}) \right] \leq \sum_{j \geq 0} \mathbb{E} \left[ \sup_{t \in \mathcal{T}} (X_{\Pi_{j+1}(t)} - X_{\Pi_j(t)}) \right].$$

Ahora bien, observamos que

$$\# \{(\Pi_j(t), \Pi_{j+1}(t)) : t \in \mathcal{T}\} \leq \mathcal{N}(\delta_j, \mathcal{T}, d) \mathcal{N}(\delta_{j+1}, \mathcal{T}, d) \leq e^{2H(\delta_{j+1}, \mathcal{T}, d)}.$$

Por otra parte,

$$d(\Pi_j(t), \Pi_{j+1}(t)) \leq \delta_j + \delta_{j+1} = 3\delta_{j+1}.$$

En adición, las variables aleatorias  $\{X_{\Pi_{j+1}(t)} - X_{\Pi_j(t)} : t \in \mathcal{T}\}$  son sub-gaussianas de constante  $9\delta_{j+1}^2$  y hay a lo más  $\exp\{2H(\delta_{j+1}, \mathcal{T}, d)\}$ , de modo que

$$\begin{aligned} \mathbb{E} \left[ \sup_{t \in \mathcal{T}} (X_{\Pi_{j+1}(t)} - X_{\Pi_j(t)}) \right] &\leq \sqrt{2 \times (9\delta_{j+1}^2) \times 2H(\delta_{j+1}, \mathcal{T}, d)} \\ &= 6\delta_{j+1} \sqrt{H(\delta_{j+1}, \mathcal{T}, d)}. \end{aligned}$$

Al juntar los resultados obtenemos

$$\begin{aligned} \mathbb{E} \left[ \sup_{t \in \mathcal{T}} (X_t - X_{t_0}) \right] &\leq 6 \sum_{j \geq 1} \delta_j \sqrt{H(\delta_j, \mathcal{T}, d)} \\ &= 12 \sum_{j \geq 1} (\delta_j - \delta_{j+1}) \sqrt{H(\delta_j, \mathcal{T}, d)} \\ &\leq 12 \int_0^{\delta/2} \sqrt{H(\epsilon, \mathcal{T}, d)} d\epsilon. \end{aligned}$$

**Minoración de Sudakov** La cota inferior, en el caso finito, se encuentra dada por el siguiente resultado:

### 5. Teorema

Sean  $X_1, \dots, X_N$  variables aleatorias con distribución normal estándar. Entonces

$$\mathbb{E} \left[ \max_{1 \leq i \leq N} X_i \right] \geq \frac{1}{2} \min_{i \neq j} \sqrt{\mathbb{E}[(X_i - X_j)^2] \log N}.$$

### 4. Lema

Sean  $X_1, \dots, X_N$  variables aleatorias gaussianas estándares, con  $N \geq 6$ . Existe una constante universal  $C > 0$  con la propiedad de que

$$\mathbb{E} \left[ \max_{1 \leq i \leq N} X_i \right] \geq C \sqrt{\log N}.$$



En adición,

$$\frac{1}{\sqrt{2 \log N}} \mathbb{E} \left[ \max_{1 \leq i \leq N} X_i \right] \xrightarrow{N \rightarrow \infty} 1.$$

El siguiente lema nos permitirá obtener sub-gaussianidad.

### 5. Lema (Hoeffding)

Sea  $Y$  una variable aleatoria centrada y acotada, esto es  $Y \in [a, b]$ . Entonces

$$\mathbb{E}[e^{\lambda Y}] \leq e^{\lambda^2(b-a)^2/8}.$$

### 1. Corolario

Sea  $S_n = n^{-1} \sum_{i=1}^n (X_i - \mathbb{E}[X_i])$  con  $X_i \in [a_i, b_i]$ . Entonces

$$\mathbb{E}[e^{\lambda S_n}] \leq \exp \left\{ \frac{\lambda^2}{8n^2} \sum_{i=1}^n (b_i - a_i)^2 \right\} = \exp \left\{ \frac{\lambda^2}{2} \sum_{i=1}^n \left( \frac{b_i - a_i}{2n} \right)^2 \right\}.$$

En particular, si las variables aleatorias son idénticamente distribuidas y acotadas en  $[a, b]$ ,

$$\mathbb{E}[e^{\lambda S_n}] = \exp \left\{ \frac{\lambda^2}{2} \frac{(b-a)^2}{4n} \right\}.$$

20/02/2023

Condicionalmente a las variables aleatorias  $\{X_i : 1 \leq i \leq n\}$ ,

$$\frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i)$$

es sub-gaussiana de constante  $n^{-2} \sum_{i=1}^n f(X_i)^2$ .

### 6. Lema

Sean  $X_1, \dots, X_n$  variables independientes e idénticamente distribuidas. Sean además  $\{\varepsilon_i\}_i$  independientes de las  $\{X_i\}_i$  y tales que tengan distribución Rademacher de parámetro  $1/2$ . Entonces

$$\mathbb{E}_\varepsilon \left[ \sup_{f \in \mathfrak{F}} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \right| \right] \leq 12 \int_0^\infty \sqrt{\frac{H(\varepsilon, \mathfrak{F}, \|\cdot\|_{2, P_n})}{n}} d\varepsilon,$$

donde

$$\|f\|_{2, P_n}^2 = \frac{1}{n} \sum_{i=1}^n f(X_i)^2.$$

Demostración: Dadas  $f, g \in \mathfrak{F}$ , condicionalmente en  $\{X_i\}_i$ , tenemos que la suma  $n^{-1} \sum_{i=1}^n \varepsilon_i (f(X_i) - g(X_i))$  es sub-gaussiana de constante  $n^{-1} \|f - g\|_{2, P_n}^2$ . Por el teorema de Dudley,

$$\mathbb{E}_\varepsilon \left[ \sup_{f \in \mathfrak{F}} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \right| \right] \leq 12 \int_0^\infty \sqrt{H(\varepsilon, \mathfrak{F}, \|\cdot\|_{2, P_n} / \sqrt{n})} d\varepsilon.$$

Al notar que

$$H(\varepsilon, \mathfrak{F}, \|\cdot\|_{2, P_n} / \sqrt{n}) = H(\varepsilon \sqrt{n}, \mathfrak{F}, \|\cdot\|_{2, P_n})$$

obtenemos el resultado al usar el cambio de variable  $\eta = \varepsilon\sqrt{n}$ . ■

### 5. Observación

En el resultado anterior,

$$\int_0^\infty \sqrt{\frac{H(\varepsilon, \mathfrak{F}, \|\cdot\|_{2, P_n})}{n}} d\varepsilon = \int_0^D \sqrt{\frac{H(\varepsilon, \mathfrak{F}, \|\cdot\|_{2, P_n})}{n}} d\varepsilon,$$

donde  $D = \sup_{f \in \mathfrak{F}} \|f\|_{n, \infty}$ , con

$$\|f\|_{n, \infty} = \max_{1 \leq i \leq n} |f(X_i)|.$$

### 7. Lema

Sea  $R > 0$ . Supongamos que  $\sup_{f \in \mathfrak{F}} \|f\|_\infty < R$  y

$$\frac{1}{n} H(\varepsilon, \mathfrak{F}, \|\cdot\|_{2, P_n}) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0 \quad \text{para toda } \varepsilon > 0.$$

Entonces  $\mathfrak{F}$  es  $\mathbb{P}$ -Glivenko–Cantelli.

Demostración: Es suficiente demostrar que

$$\mathbb{E} \left[ \sup_{f \in \mathfrak{F}} \left| \frac{1}{n} \sum_{i=1}^n (f(X_i) - \mathbb{E}[f(X_i)]) \right| \right] \xrightarrow[n \rightarrow \infty]{} 0.$$

Usaremos la notación  $P_n f = n^{-1} \sum_{i=1}^n f(X_i)$  y  $Pf = \mathbb{E}[f(X)]$ . Observemos que

$$\begin{aligned} \mathbb{E} \left[ \sup_{f \in \mathfrak{F}} |P_n f - Pf| \right] &\leq 2 \mathbb{E} \left[ \sup_{f \in \mathfrak{F}} \frac{1}{n} \left| \sum_{i=1}^n \varepsilon_i f(X_i) \right| \right] \\ &\leq 24 \mathbb{E} \left[ \int_0^D \sqrt{\frac{H(\varepsilon, \mathfrak{F}, \|\cdot\|_{2, P_n})}{n}} d\varepsilon \right]. \end{aligned}$$

Por hipótesis  $D \leq R$ , de modo que Tonelli y Jensen implican que

$$\mathbb{E} \left[ \sup_{f \in \mathfrak{F}} |P_n f - Pf| \right] \leq 24 \int_0^R \sqrt{\mathbb{E} \left[ \frac{H(\varepsilon, \mathfrak{F}, \|\cdot\|_{2, P_n})}{n} \right]} d\varepsilon.$$

Ahora bien,

$$\mathcal{N}(\varepsilon, \mathfrak{F}, \|\cdot\|_{2, P_n}) \leq \mathcal{N}(\varepsilon, \mathfrak{F}, \|\cdot\|_{n, \infty}) \leq \left( \frac{R}{\varepsilon} \right)^n,$$

y en consecuencia

$$H(\varepsilon, \mathfrak{F}, \|\cdot\|_{2, P_n}) \leq n \log \left( \frac{R}{\varepsilon} \right).$$

Por lo tanto, por el teorema de convergencia dominada,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[ \sup_{f \in \mathfrak{F}} |P_n f - Pf| \right] = 0. \quad \text{■}$$

Demostración del teorema 2: Dada  $R > 0$  arbitraria consideremos

$$\mathfrak{F}_R = \{f \mathbb{1}_{\{F \leq R\}} : f \in \mathfrak{F}\}.$$

Observemos que para cualesquiera  $f_1, f_2 \in \mathfrak{F}$

$$\int_{\{F \leq R\}} (f_1 - f_2)^2 dP_n \leq 2R \int |f_1 - f_2| dP_n,$$

lo cual nos dice que para cualquier  $\varepsilon > 0$  y  $R > 0$ ,

$$\frac{1}{n} H_2(\varepsilon, \mathfrak{F}_R, P_n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0,$$

debido a la hipótesis

$$\frac{1}{n} H_1(\varepsilon, \mathfrak{F}, P_n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0 \quad \text{para cualquier } \varepsilon > 0.$$

Así, por el lema previo, para cualquier  $R > 0$ ,  $\mathfrak{F}_R$  es  $\mathbb{P}$ -Glivenko–Cantelli. Por otra parte, por el supuesto  $F \in L^1(\mathbb{P})$ , para toda  $\delta > 0$  existe  $R > 0$  para la cual

$$\int_{\{F > R\}} F dP \leq \delta.$$

Entonces, debido a la ley fuerte de grandes números y lo anterior, para  $n$  suficientemente grande,

$$\int_{\{F > R\}} F dP_n \leq 2\delta.$$

Mientras tanto, usando que  $\mathfrak{F}_R$  es  $\mathbb{P}$ -Glivenko–Cantelli,

$$\sup_{f \in \mathfrak{F}} \left| \int_{\{F \leq R\}} f d(P_n - P) \right| \leq \delta$$

con probabilidad 1. Por consiguiente deducimos

$$\sup_{f \in \mathfrak{F}} \left| \int f d(P_n - P) \right| \leq 4\delta.$$

Esto muestra que  $\mathfrak{F}$  es  $\mathbb{P}$ -Glivenko–Cantelli. ■

# Introducción a la dimensión de Vapnik–Chervonenkis

3  
01/03/2023

Para entender las entropías  $H_1(\epsilon, \mathfrak{F}, P_n)$  es necesario entender las bolas de  $\|\cdot\|_{1, P_n}$ . En este caso  $P_n$  tiene masa únicamente en los puntos  $X_1, \dots, X_n$ . Es decir, necesitamos entender las aproximaciones de funciones sobre conjuntos finitos de valores.

## 1. Definición

Sea  $\mathcal{D}$  una colección de subconjuntos de  $\mathcal{X}$ . Definamos

$$\Delta^{\mathcal{D}}(\xi_1, \dots, \xi_n) = \#\{D \cap \{\xi_1, \dots, \xi_n\} : D \in \mathcal{D}\},$$

$$m^{\mathcal{D}}(n) = \sup\{\Delta^{\mathcal{D}}(\xi_1, \dots, \xi_n) : (\xi_1, \dots, \xi_n) \in \mathcal{X}^n\}$$

y

$$V(\mathcal{D}) = \inf\{n \geq 1 : m^{\mathcal{D}}(n) < 2^n\}.$$

La cantidad  $V(\mathcal{D}) - 1$  es conocida como la *dimensión de Vapnik–Chervonenkis* de  $\mathcal{D}$ .

## 1. Observación

- Hay un resultado, conocido como *lema de Sauer*, que nos dice que si  $m^{\mathcal{D}}(n') < 2^{n'}$  para alguna  $n' \in \mathbb{N}$ , entonces existe un polinomio  $P$  tal que  $m^{\mathcal{D}}(n) \leq P(n)$  siempre que  $n \geq n'$ .
- En vista del punto anterior, es posible definir la dimensión de Vapnik–Chervonenkis como

$$V(\mathcal{D}) = \sup\{n \geq 1 : m^{\mathcal{D}}(n) = 2^n\}.$$

**Ejemplo 1:** Sean  $\mathcal{X} = \mathbb{R}$  y  $\mathcal{D} = \{(-\infty, t] : t \in \mathbb{R}\}$ . Puede verse que  $m^{\mathcal{D}}(n) = n + 1$ , de modo que  $V(\mathcal{D}) = 2$  y en consecuencia la dimensión de Vapnik–Chervonenkis es 1.

**Ejemplo 2:** Tomemos  $\mathcal{X} = \mathbb{R}$  y  $\mathcal{D} = \{[a, b] : a, b \in \mathbb{R}\}$ . En este caso

$$m^{\mathcal{D}}(n) = \frac{n^2 + n + 2}{2},$$

y por ende  $V(\mathcal{D}) = 3$ ; i.e. la dimensión de Vapnik–Chervonenkis es 2.

**Ejemplo 3:** Consideremos  $\mathcal{X} = \mathbb{R}^d$  y  $\mathcal{D} = \{\prod_{i=1}^d (-\infty, t_i] : t_1, \dots, t_d \in \mathbb{R}\}$ . En este caso la dimensión de Vapnik–Chervonenkis es  $d$ .

**Ejemplo 4:** Ahora sean  $\mathcal{X} = \mathbb{R}^d$  y

$$\mathcal{D} = \{x : \langle \theta, x \rangle > y : \theta \in \mathbb{R}^d, y \in \mathbb{R}\}.$$

Entonces, puede mostrarse que  $m^{\mathcal{D}}(n) \leq 2^d \binom{n}{d}$ , lo cual implica que la dimensión de Vapnik–Chervonenkis es  $d + 1$ .

### 1. Proposición

Sean  $\mathcal{A}$  y  $\mathcal{B}$  dos clases de subconjuntos de  $\mathcal{X}$ . Tenemos:

- (i) Si consideramos  $\mathcal{A}^c = \{A^c : A \in \mathcal{A}\}$ , entonces  $m^{\mathcal{A}^c}(n) = m^{\mathcal{A}}(n)$ .
- (ii) Si  $\mathcal{A} \cup \mathcal{B} = \{A \cup B : A \in \mathcal{A}, B \in \mathcal{B}\}$ ,  $m^{\mathcal{A} \cup \mathcal{B}}(n) \leq m^{\mathcal{A}}(n)m^{\mathcal{B}}(n)$ .
- (iii) Lo mismo para  $\mathcal{A} \cap \mathcal{B} = \{A \cap B : A \in \mathcal{A}, B \in \mathcal{B}\}$ .
- (iv) Lo mismo con  $\mathcal{A} \times \mathcal{B} = \{A \times B : A \in \mathcal{A}, B \in \mathcal{B}\}$ .
- (v) Si  $\mathcal{C} = \{C : C \in \mathcal{A} \text{ o } C \in \mathcal{B}\}$  entonces  $m^{\mathcal{C}}(n) \leq m^{\mathcal{A}}(n) + m^{\mathcal{B}}(n)$ .

El resultado anterior nos dice bajo qué operaciones es estable la finitud de la dimensión de Vapnik–Chervonenkis.

Demostración: Los puntos (i) y (v) son elementales. Para demostrar (iii) consideremos la función  $\Phi : \mathcal{P}(X) \times \mathcal{P}(X) \rightarrow \{0, 1\}^n \times \{0, 1\}^n$  donde

$$(A, B) \mapsto ((\mathbb{1}_{\{\xi_1 \in A\}}, \dots, \mathbb{1}_{\{\xi_n \in A\}}), (\mathbb{1}_{\{\xi_1 \in B\}}, \dots, \mathbb{1}_{\{\xi_n \in B\}})).$$

Ahora tomamos  $\Pi : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}^n$  dada por  $(x, y) \mapsto x \odot y$ . Se afirma que las configuraciones posibles se encuentran dadas por

$$\Pi \circ \Phi(\mathcal{A} \times \mathcal{B}).$$

Luego,

$$m^{\mathcal{A} \cap \mathcal{B}}(n) = \sup |\Pi \circ \Phi(\mathcal{A} \times \mathcal{B})| \leq \sup |\Phi(\mathcal{A} \times \mathcal{B})| = m^{\mathcal{A}}(n)m^{\mathcal{B}}(n). \quad \blacksquare$$

06/03/2023

### 1. Lema (de Sauer)

Sea  $\mathcal{D}$  una clase de Vapnik–Chervonenkis finita  $V$ . Entonces para toda  $n \geq 1$ ,

$$m^{\mathcal{D}}(n) \leq \sum_{i=0}^V \binom{n}{i}.$$

Para la demostración del lema 1 necesitaremos algunos conceptos y resultados auxiliares.

### 2. Definición

Decimos que un conjunto  $B \subset \{0, 1\}^n$  rompe  $S = \{s_1, \dots, s_m\} \subset [n]$  si la restricción de  $B$  a los componentes  $s_1, \dots, s_m$  es todo el hipercubo de dimensión  $m \leq n$ ; esto es, si  $B_s = \{b_{s_1}, \dots, b_{s_m} : (b_1, \dots, b_n) \in B\} = \{0, 1\}^m$ .

Definamos  $\Psi_1 : \mathcal{P}(\{0, 1\}^n) \rightarrow \mathcal{P}(\{0, 1\}^n)$  dada por

$$B \mapsto \Psi_1(B) = \{\bar{b} : b \in B\}$$

donde

$$\bar{b} = \begin{cases} (0, b_2, \dots, b_n) & \text{si } b_1 = 1 \text{ y } \bar{b} \notin B, \\ b & \text{en otro caso.} \end{cases}$$

### 1. Afirmación

Para  $B \subset \{0, 1\}^n$  y  $S = \{s_1, \dots, s_n\} \subset [n]$  tenemos  $|\Psi_1(B)| = |B|$ . En adición,  $B$  rompe  $S$  si, y solo si,  $\Psi_1(B)$  rompe  $S$ .

Demostración: La primera parte se sigue de la inyectividad de  $\bar{b}$ . Si  $1 \notin S$ ,  $(\Psi_1(B))_S = (B)_S$ . El otro caso se sigue de observar  $|\Psi_1(B)_S| = |B_S|$ . ■

### 2. Afirmación

Si consideramos  $\Psi_2, \dots, \Psi_n$  de manera análoga y tomamos  $B_0 = \{0, 1\}^n$ . Entonces  $B_0$  rompe  $S$  si y solo si  $B_n = \Psi_n \circ \dots \circ \Psi_1(B_0)$  rompe  $S$ .

A continuación **supondremos** que cada  $S$  que tiene  $m$  índices con  $m > V$  NO están rotos por  $B_0$ .

Para cada  $v \in B_n$  definimos

$$T_v = \{b \in \{0, 1\}^n : b \leq v\}.$$

### 3. Afirmación

$T_v \subset B_n$  y  $v$  no tiene más que  $V$  unos.

Demostración: La primera afirmación es clara por un argumento de ramificación (hay que formalizarlo). La segunda se deduce por reducción al absurdo, considerando que existe  $v \in B_n$  con  $V + 1$  unos. Entonces  $S = \{s_1, \dots, s_{|v|}\}$  está roto por  $B_n$ , de modo que  $B_0$  rompe  $S$ . ■

### 4. Afirmación

Sea  $T = \bigcup_{v \in \mathcal{V}} T_v$ , donde  $\mathcal{V}$  es el conjunto de todos los vectores con máximo  $V$  unos. Entonces

$$B_n \subset T \quad \text{y} \quad |T| = \sum_{i=0}^V \binom{n}{i}.$$

### 5. Afirmación

Consideramos  $n > V$  y  $x_1, \dots, x_n \in \mathbb{R}^k$  para definir

$$B_0 = \left\{ (\mathbb{1}_{\{x_1 \in D\}}, \dots, \mathbb{1}_{\{x_n \in D\}}) : D \in \mathcal{D} \right\}.$$

Entonces  $m^{\mathcal{D}}(n) = |B_0|$ . Si existe un conjunto  $S$  de  $m > V$  índices que rompe  $B_0$ , entonces  $m^{\mathcal{D}}(n) = 2^n$  lo cual no puede ser pues  $m > V$ . Luego,

$$m^{\mathcal{D}}(n) = |B_0| = |B_n| \leq T \sum_{i=0}^V \binom{n}{i}.$$

### 1. Corolario

Sea  $\mathcal{D}$  de dimensión  $V$ , entonces para toda  $n \geq 1$ ,  $m^{\mathcal{D}}(n) \leq (n+1)^V$  y si  $n \geq N$ ,

$$m^{\mathcal{D}}(n) \leq \left(\frac{ne}{V}\right)^V.$$

Demostración: Observamos que

$$\sum_{i=0}^V \binom{n}{i} \leq \sum_{i=0}^V \frac{n^i}{i!} \leq \sum_{i=0}^V \binom{V}{i} n^i = (n+1)^V.$$

Por otra parte, si  $n \geq V$ , entonces  $V/n \leq 1$ , de modo que

$$\left(\frac{V}{n}\right)^V \sum_{i=0}^V \binom{n}{i} \leq \sum_{i=0}^n \binom{n}{i} \left(\frac{V}{n}\right)^i \leq \left(1 + \frac{V}{n}\right)^n \leq e^V.$$

El objetivo de esta parte es demostrar:

### 1. Teorema

Sea  $\mathfrak{F}$  una clase de funciones tal que

$$\mathcal{G} = \{\text{Subgraph}(f) : f \in \mathfrak{F}\},$$

donde

$$\text{Subgraph}(f) = \{(x, y) \in \mathcal{X} \times \mathbb{R} : y < f(x)\},$$

es de dimensión de Vapnik–Chervonenkis finita  $V$  en  $\mathcal{X} \times \mathbb{R}$ . Supongamos que la envolvente  $F \in L^1(Q)$  para una medida de probabilidad  $Q$ . Entonces

$$H_1(\epsilon \|F\|_{L^1(Q)}, \mathfrak{F}, Q) \leq \frac{V}{1 - 1/e} \log\left(\frac{8e}{\epsilon}\right).$$

08/03/2023

Antes de demostrar este resultado, veremos una versión en el hipercubo.

### 3. Definición (Distancia de Hamming)

Para  $b, c \in \{0, 1\}^n$  se define

$$\rho(b, c) = \sqrt{\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{b_i \neq c_i\}}}.$$

### 2. Teorema

Sea  $\mathcal{D}$  una colección de subconjuntos de  $\mathbb{R}^k$  de dimensión de VC finita. Entonces para cualesquiera  $x_1, \dots, x_n \in \mathbb{R}^k$  y  $\epsilon \in [0, 1]$ ,

$$H(\epsilon, \mathcal{D}(x_1^n), \rho) \leq \frac{V}{1 - 1/e} \log\left(\frac{4e}{\epsilon^2}\right),$$

donde

$$\mathcal{D}(x_1^n) = \{(\mathbb{1}_{\{x_1 \in D\}}, \dots, \mathbb{1}_{\{x_n \in D\}}) : D \in \mathcal{D}\}.$$

Demostración: Usaremos el método probabilista. Sea  $C_\epsilon$  una  $\epsilon$ -red de  $\mathcal{D}(x_1^n)$ . Denotamos por  $c^{(1)}, \dots, c^{(M)}$  a sus elementos. Introducimos

$$A_{i,j} = \{t \in [k] : c_t^{(i)} \neq c_t^{(j)}\}.$$

Sean  $Y_1, \dots, Y_K$  variables aleatorias iide uniformes sobre  $[k]$ . Tenemos que

$$\mathbb{P}(Y_m \in A_{i,j}) = \frac{\#A_{i,j}}{k} \geq \frac{\epsilon^2 k}{k} = \epsilon^2.$$

Entonces

$$\begin{aligned} \mathbb{P}\left(\bigcap_{i=1}^{M-1} \bigcap_{j=2}^M \bigcup_{m=1}^K \{Y_m \in A_{ij}\}\right) &= 1 - \mathbb{P}\left(\bigcup_{i=1}^M \bigcup_{j=1}^M \bigcap_{m=1}^K \{Y_m \notin A_{ij}\}\right) \\ &\geq 1 - M^2 \mathbb{P}\left(\bigcap_{m=1}^K \{Y_m \in A_{1,2}\}\right) \\ &= 1 - M^2 \prod_{m=1}^K \mathbb{P}(Y_m \notin A_{1,2}) \\ &\geq 1 - M^2 (1 - \epsilon^2)^K \geq 1 - M^2 e^{-K\epsilon^2}. \end{aligned}$$

Si consideramos  $K = \lfloor 2 \log(M/\epsilon^2) \rfloor + 1$ , la cota anterior es positiva. Lo anterior implica que existe una realización  $y_1 = Y_1(\omega), \dots, y_K = Y_K(\omega)$  para la cual se cumple que dados  $i, j \in [M]$  distintos entonces existe  $y_m$  con la propiedad  $c_{y_m}^{(i)} \neq c_{y_m}^{(j)}$ . Denotaremos por  $(C_\epsilon)_{y_1, \dots, y_K}$  la restricción de los  $c^{(i)}$  a las coordenadas  $y_1, \dots, y_K$ . Entonces  $\#C_\epsilon = \#(C_\epsilon)_{y_1, \dots, y_K} = M$ . Consecuentemente, usando el lema de Sauer,

$$M = \#(C_\epsilon) \leq \mathcal{D}(y_1^K) \leq m^{\mathcal{D}}(K) \leq \left(\frac{eK}{V}\right)^V.$$

Por ende, si  $\log M \geq V$  entonces  $K \geq V$  y así

$$\begin{aligned} \log M &\leq V \log\left(\frac{eK}{V}\right) \leq V \log\left(\frac{4e \log M}{V\epsilon^2}\right) \\ &= V \left( \log\left(\frac{4e}{\epsilon^2}\right) + \log\left(\frac{\log M}{V}\right) \right) \\ &\leq V \log\left(\frac{4e}{\epsilon^2}\right) + \frac{\log M}{e}. \end{aligned}$$

De lo anterior deducimos

$$\log M \leq \frac{V}{1 - 1/e} \log\left(\frac{4e}{\epsilon^2}\right).$$

En el caso  $\log M < V$  la desigualdad es obvia. El resultado se sigue de la desigualdad

$$H_1(\epsilon, \mathcal{D}(x_1^n), \rho) \leq \log M. \quad \blacksquare$$

Demostración del teorema 1: Tenemos

$$Q(|f - g|) = Q \otimes \lambda(G_f \triangle G_g),$$

donde  $G_\cdot = \text{Subgraph}(\cdot)$ . Tomamos  $P$  la medida de probabilidad  $Q \otimes \lambda$  condicionada a  $\{(x, y) : |y| \leq F(x)\}$ ; esto es

$$P = \frac{Q \otimes \lambda}{2\|F\|_{1,Q}}.$$



Aplicaremos el teorema anterior, primero observando que

$$H_1(2\varepsilon\|F\|_{1,Q}, \mathfrak{F}, Q) = H_1(2\varepsilon\|F\|_{1,Q}, \mathcal{G}, Q \otimes \lambda) = H(\sqrt{\varepsilon}, \mathcal{G}, d_P),$$

donde  $d_P(A, B) = \|\mathbb{1}_A - \mathbb{1}_B\|_{2,P}$ . Así,

$$H(\sqrt{\varepsilon}, \mathcal{G}, d_P) = H(\sqrt{x}, \mathcal{G}(x_1^n), \rho) \leq \frac{V}{1 - 1/e} \log\left(\frac{4e}{\varepsilon}\right). \quad \blacksquare$$

Una generalización del teorema 1 considerando  $L^r$  en lugar de  $L^1$  es la siguiente:

### 3. Teorema

Si  $F \in L^r(Q)$  y  $\varepsilon \in (0, 1)$ ,

$$H_r(\varepsilon\|F\|_{r,Q}, \mathfrak{F}, Q) \leq \frac{V}{1 - 1/e} \log\left(\frac{8e^r}{\varepsilon/2}\right).$$

Demostración: Sea  $R$  un cambio de medida de la forma

$$\frac{dR}{dQ} = \frac{F^{r-1}}{Q(F^{r-1})}.$$

Con ello tenemos

$$Q(|f - g|^r) \leq Q(|f - g|(2F)^{r-1}) = 2^{r-1}Q(F^{r-1})R(|f - g|).$$

Entonces

$$\begin{aligned} H(2\varepsilon\|F\|_{r,Q}, \mathfrak{F}, \|\cdot\|_{r,Q}) &\leq H\left(\frac{\varepsilon^r\|F\|_{1,Q}}{Q(F^{r-1})}, \mathfrak{F}, \|\cdot\|_{1,R}\right) \\ &\leq H_1(\varepsilon^r\|F\|_{1,R}, \mathfrak{F}, R) \\ &\leq \frac{V}{1 - 1/e} \log\left(\frac{8e}{\varepsilon^r}\right). \end{aligned} \quad \blacksquare$$

### 2. Corolario

Sea  $\mathfrak{F}$  de dimensión VC finita y  $F \in L^1(P)$ . Entonces  $\mathfrak{F}$  es  $\mathbb{P}$ -Glivenko–Cantelli.

### 3. Corolario

Sea  $\mathfrak{F} = \text{conv}(\mathfrak{G})$  tal que  $\mathfrak{G}$  es  $\mathbb{P}$ -Glivenko–Cantelli, entonces  $\mathfrak{F}$  es  $\mathbb{P}$ -Glivenko–Cantelli.

Demostración: Observemos que para  $f \in \mathfrak{F}$  existen  $g_1, \dots, g_k \in \mathfrak{G}$  y  $\lambda_1, \dots, \lambda_k$  tales que  $f = g_1, \dots, g_k$  y entonces

$$|(P_n - P)f| \leq \sum_{i=1}^k \lambda_i |(P_n - P)g_i| \leq \sup_{g \in \mathfrak{G}} |(P_n - P)g|.$$

Por lo tanto

$$\sup_{f \in \mathfrak{F}} |(P_n - P)f| = \sup_{g \in \mathfrak{G}} |(P_n - P)g|. \quad \blacksquare$$

# Teorema central del límite

4  
15/03/2023

## 1. Definición

Para cada  $n$  sea  $Z_n = \{Z_n(f) : f \in \mathfrak{F}\}$  dado por

$$Z_n(f) := \sqrt{n}(P_n f - P f).$$

Este proceso es conocido como *proceso empírico renormalizado*.

Una pregunta natural es bajo qué condiciones se tiene que  $Z_n \xrightarrow{d} G$  donde  $G$  es un proceso gaussiano centrado con

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

## 2. Definición

Cuando la convergencia anterior diremos que  $\mathfrak{F}$  es  $\mathbb{P}$ -Donsker.

### 1. Lema

Sea  $\mathfrak{F}$  una clase de funciones en  $L^2(\mathbb{P})$  totalmente acotada y supongamos que para cualquier  $\eta > 0$  existe  $\delta > 0$  tal que

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left( \sup_{f_1, f_2 \in \mathfrak{F}, \|f_1 - f_2\|_2 \leq \delta} |Z_n(f_1) - Z_n(f_2)| > \eta \right) \leq \eta.$$

Entonces  $\mathfrak{F}$  es  $\mathbb{P}$ -Donsker.

Demostración: Sea  $\delta > 0$ . Sea  $\mathfrak{F}_\delta$  el conjunto de los centros de bolas de radio  $\delta$  que cubren  $\mathfrak{F}$ , el cual podemos tomar finito debido a la hipótesis de acotamiento total. De lo anterior, para cada  $f \in \mathfrak{F}$  existe  $f_\delta \in \mathfrak{F}_\delta$  con  $\|f - f_\delta\|_2 \leq \delta$ . Sean  $k = \#\mathfrak{F}_\delta$ ,  $h$  una función Lipschitz y  $\pi_\delta : \mathfrak{F} \rightarrow \mathfrak{F}_\delta$  una función que proyecta una función  $f$  a  $f_\delta$ . Nos encontramos interesados en el proceso  $Z_n \circ \pi_\delta$ . Este resulta ser una clase finita de variables aleatorias que satisfacen

$$(Z_n(f_1), \dots, Z_n(f_k)) \xrightarrow{d} (N(f_1), \dots, N(f_k)),$$

de modo que para  $\tilde{h} : \mathbb{R}^k \rightarrow \mathbb{R}$ ,

$$\mathbb{E}[\tilde{h}(Z_n \circ \pi_\delta)] \rightarrow \mathbb{E}[\tilde{h}(N^\delta)],$$

donde  $N^\delta$  tiene distribución  $N(0, \Sigma)$ , con

$$\Sigma_{f_i, f_j} = P(f_i f_j) - P f_i P f_j.$$

Sea  $G$  es único proceso gaussiano en ley con  $\text{Cov}(G(f), G(g)) = P(fg) - P f P g$ . Entonces  $G \circ \pi_\delta \sim N(0, \Sigma)$  y en consecuencia  $Z_n \circ \pi_\delta \xrightarrow{d} G \circ \pi_\delta$ . Antes de continuar

observemos que para  $\mathcal{G} \subset \mathfrak{F}$  finito, el lema de Portmanteau nos da la primera igualdad en

$$\begin{aligned} & \mathbb{P}\left(\sup_{f_1, f_2 \in \mathcal{G}, \|f_1 - f_2\|_2 \leq \delta} |G(f_1) - G(f_2)| > \eta\right) \\ & \leq \liminf_{n \rightarrow \infty} \mathbb{P}\left(\sup_{f_1, f_2 \in \mathcal{G}, \|f_1 - f_2\|_2 \leq \delta} |Z_n(f_1) - Z_n(f_2)| > \eta\right) \\ & \leq \liminf_{n \rightarrow \infty} \mathbb{P}\left(\sup_{f_1, f_2 \in \mathfrak{F}, \|f_1 - f_2\|_2 \leq \delta} |Z_n(f_1) - Z_n(f_2)| > \eta\right) \\ & \leq \limsup_{n \rightarrow \infty} \mathbb{P}\left(\sup_{f_1, f_2 \in \mathfrak{F}, \|f_1 - f_2\|_2 \leq \delta} |Z_n(f_1) - Z_n(f_2)| > \eta\right) \leq \delta. \end{aligned}$$

Es claro que esto se puede extender a  $\mathcal{G}$  numerable y de ahí a  $\mathfrak{F}$ .

Ahora sí probaremos la convergencia funcional de  $Z_n$  a  $G$ . Para ello vemos que

$$\begin{aligned} & |\mathbb{E}[h(Z_n)] - \mathbb{E}[h(G)]| \\ & \leq |\mathbb{E}[h(Z_n)] - \mathbb{E}[h(Z_n \circ \pi_\delta)]| + |\mathbb{E}[h(Z_n \circ \pi_\delta)] - \mathbb{E}[h(G \circ \pi_\delta)]| \\ & \quad + |\mathbb{E}[h(G \circ \pi_\delta)] - \mathbb{E}[h(G)]|. \end{aligned}$$

Únicamente debemos controlar el primer y tercer sumando en la cota superior dada. Solamente tratamos el primer término, notando que

$$|\mathbb{E}[h(Z_n)] - \mathbb{E}[h(Z_n \circ \pi_\delta)]| \leq \lambda \eta + 2 \|h\|_\infty \mathbb{P}(\|Z_n - Z_n \circ \pi_\delta\|_D > \eta),$$

donde  $\lambda$  es la constante Lipschitz de  $h$  y  $D$  es el espacio métrico de las trayectorias. Esto es lo mismo que

$$|\mathbb{E}[h(Z_n)] - \mathbb{E}[h(Z_n \circ \pi_\delta)]| \leq \lambda \eta + 2 \|h\|_\infty \mathbb{P}\left(\sup_{f \in \mathfrak{F}} |Z_n(f) - Z_n \circ \pi_\delta(f)| > \eta\right),$$

con lo cual obtenemos el resultado. ■

### 1. Teorema

Sea  $\mathfrak{F}$  un subconjunto de funciones de  $L^2(\mathbb{P})$  con envolvente  $F \in L^2(\mathbb{P})$ . Supongamos que existe  $H$  no decreciente tal que

$$\int_0^\infty \sqrt{H(\delta)} d\delta < \infty \quad \gamma \quad \lim_{A \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}\left(\sup_{\delta > 0} \frac{H_2(\delta, \mathfrak{F}, P_n)}{H(\delta)} > A\right) = 0.$$

Entonces para toda  $\eta > 0$  existe  $\delta > 0$  tal que

$$\limsup_{n \rightarrow \infty} \mathbb{P}\left(\sup_{\|f_1 - f_2\|_2 \leq \delta} |Z_n(f_1) - Z_n(f_2)| > \eta\right) \leq \eta.$$

Demostración: Sean

$$A_\delta \left\{ \sup_{\|f_1 - f_2\|_2 \leq \delta} |Z_n(f_1) - Z_n(f_2)| > \eta \right\} \quad \text{y} \quad E_{n,A} = \left\{ \sup_{\delta > 0} \frac{H_2(\delta, \mathfrak{F}, P_n)}{H(\delta)} > A \right\}.$$

Entonces

$$\begin{aligned} \mathbb{P}(A_\delta) &= \mathbb{E}[\mathbb{1}_{A_\delta} \mathbb{1}_{E_{n,A}}] + \mathbb{E}[\mathbb{1}_{A_\delta} \mathbb{1}_{E_{n,A}^c}] \\ &\leq \mathbb{P}(E_{n,A}) + \frac{1}{\eta} \mathbb{E}\left[\sup_{\|f_1 - f_2\|_2 \leq \delta} |Z_n(f_1) - Z_n(f_2)| \mathbb{1}_{E_{n,A}^c}\right] \\ &\leq \mathbb{P}(E_{n,A}) + \frac{24}{\eta} \mathbb{E}\left[\int_0^{\delta_n} \sqrt{H_2(\epsilon, \mathcal{G}, P_n)} d\epsilon \mathbb{1}_{E_{n,A}^c}\right], \end{aligned}$$

donde  $\mathcal{G} = \{g = f_1 - f_2 : f_1, f_2 \in \mathfrak{F}, \|g\|_2 \leq \delta\}$ . Ahora, tenemos que

$$H_2(2\varepsilon, \mathcal{G}, P_n) \leq 2H_2(\varepsilon, \mathfrak{F}, P_n) \quad \text{y} \quad (f_1 - f_2)^2 \leq 2F|f_1 - f_2|.$$

Del segundo hecho,

$$\|f_1 - f_2\|_{n,2}^2 \leq 2\|F\|_{n,2}\|f_1 - f_2\|_{1,n}.$$

Además, por la ley de grandes números,  $\|F\|_{n,2} \rightarrow \|F\|_2$  de manera casi segura. Luego, si  $\mathcal{S} = \{(f_1 - f_2)^2 : f_1, f_2 \in \mathfrak{F}\}$ ,

$$H_2(\varepsilon, \mathcal{S}, P_n) \leq H_2\left(\frac{\varepsilon}{4\|F\|_2}, \mathcal{G}, P_n\right) \leq 2H_2\left(\frac{\varepsilon}{8\|F\|_2}, \mathfrak{F}, P_n\right).$$

La envolvente de  $\mathcal{S}$  es tal que

$$\sup_{g \in \mathcal{S}} g \leq 4F^2 \in L^1(\mathbb{P}).$$

Entonces, como  $\mathcal{S}$  es  $\mathbb{P}$ -Glivenko–Cantelli,

$$\sup_{f_1, f_2} \left| \|f_1 - f_2\|_{2,n}^2 - \|f_1 - f_2\|_2^2 \right| \xrightarrow{\text{c.s.}} 0.$$

Así, para  $n$  suficientemente grande

$$\frac{1}{2}\|f_1 - f_2\|_{2,n} \leq \|f_1 - f_2\|_2 \leq 2\|f_1 - f_2\|_{2,n},$$

lo cual implica

$$H_2(2\varepsilon, \mathfrak{F}, \mathbb{P}) \leq H_2(\varepsilon, \mathfrak{F}, P_n).$$

En particular  $\mathfrak{F}$  es totalmente acotado.

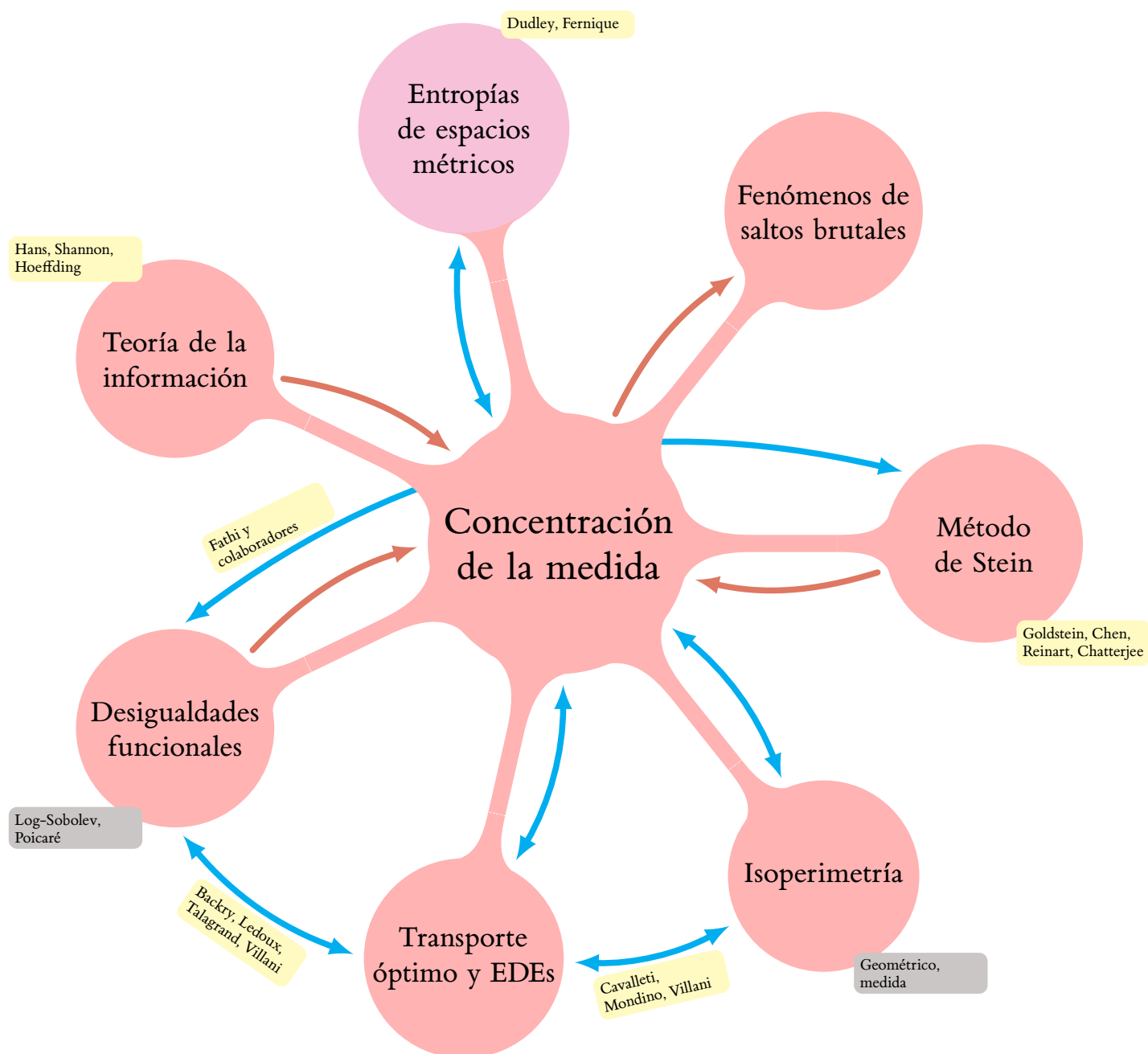
Finalmente, vemos que

$$\begin{aligned} \mathbb{P}(A_\delta) &\leq \mathbb{P}(E_{n,A}) + \frac{L}{\eta} \mathbb{E} \left[ \int_0^\delta \sqrt{H_2(\varepsilon, \mathfrak{F}, P_n)} \, d\varepsilon \mathbf{1}_{E_{n,A}^c} \right] \\ &\leq \mathbb{P}(E_{n,A}) + \frac{L}{\eta} \mathbb{E} \left[ \int_0^\delta \sqrt{AH(\varepsilon)} \, d\varepsilon \right], \end{aligned}$$

cota que converge a cero cuando  $n \rightarrow \infty$ ,  $A \rightarrow \infty$  y  $\delta \rightarrow 0$ . ■

## **Parte II.**

### **Temas relacionados de geometría convexa y probabilidad**



# Desigualdad de Brunn–Minkowski

22/03/2023

**Isoperimetría** La idea central de esta teoría es maximizar el volumen de un compacto  $K$  sujeto a que su perímetro sea dado. Resulta que en  $\mathbb{R}^n$  con la medida de Lebesgue el resultado es la bola.

Denotaremos por  $\text{Vol}_n(A)$  a la medida de Lebesgue de  $A \subset \mathbb{R}^n$ , y

$$A + B = \{x + y : x \in A, y \in B\} \quad \text{y} \quad \lambda A = \{\lambda x : x \in A\}.$$

## 1. Teorema (Brunn–Minkowski)

Sean  $A, B \subset \mathbb{R}^n$  distintos del vacío y compactos. Entonces para cualquier  $\lambda \in [0, 1]$ ,

$$\text{Vol}_n((1 - \lambda)A + \lambda B)^{1/n} \geq (1 - \lambda) \text{Vol}_n(A)^{1/n} + \lambda \text{Vol}_n(B)^{1/n}.$$

Demostración del caso unidimensional: La prueba es esencialmente una aplicación “inteligente” de la desigualdad AM–GM, es decir convexidad.

Notemos que  $\text{Vol}_1(cA) = |c| \text{Vol}_1(A)$ , de modo que es suficiente demostrar que

$$\text{Vol}_1(A + B) \geq \text{Vol}_1(A) + \text{Vol}_1(B).$$

Para verificar esto notemos que el volumen es invariante bajo traslaciones y en consecuencia podemos considerar  $A \subset (-\infty, 0]$ ,  $B \subset [0, \infty)$  con  $A \cap B = \{0\}$ . Pero entonces  $A \cup B \subset A + B$  y el resultado se sigue debido a que

$$\text{Vol}_1(A) + \text{Vol}_1(B) = \text{Vol}_1(A \cup B) \leq \text{Vol}_1(A + B). \quad \blacksquare$$

## 2. Teorema (Prékopa–Leindler)

Dado  $\lambda \in (0, 1)$  sean  $f, g, h : \mathbb{R}^n \rightarrow \mathbb{R}_+$  funciones medibles escogidas de tal manera que para cualesquiera  $x, y \in \mathbb{R}^n$ ,

$$h((1 - \lambda)x + \lambda y) \geq f(x)^{1-\lambda} g(y)^\lambda.$$

Luego

$$\int_{\mathbb{R}^n} h(x) dx \geq \left( \int_{\mathbb{R}^n} f(x) dx \right)^{1-\lambda} \left( \int_{\mathbb{R}^n} g(x) dx \right)^\lambda.$$

Demostración: La prueba se realiza por inducción sobre  $n$ . La idea principal consiste en descomponer

$$\int_{\mathbb{R}^n} f(x) dx = \int_0^\infty \text{Vol}_n(\{x : f(x) > t\}) dt.$$

Por la hipótesis tenemos

$$(1 - \lambda)\{x : f(x) \geq t\} + \lambda\{x : g(x) \geq t\} \subset \{x : h(x) \geq t\}.$$

Por consiguiente, cuando  $n = 1$ , usamos la desigualdad de Brunn–Minkowski en dimensión 1 para obtener la primera desigualdad en

$$\begin{aligned} \int_{\mathbb{R}} h(x) dx &= \int_0^\infty \text{Vol}_1(\{x : h(x) \geq t\}) dt \\ &\geq \int_0^\infty \left( (1-\lambda) \text{Vol}_1(\{x : f(x) \geq t\}) + \lambda \text{Vol}_1(\{x : g(x) \geq t\}) \right) dt \\ &= (1-\lambda) \int_{\mathbb{R}} f(x) dx + \lambda \int_{\mathbb{R}} g(x) dx \\ &\geq \left( \int_{\mathbb{R}} f(x) dx \right)^{1-\lambda} \left( \int_{\mathbb{R}} g(x) dx \right)^\lambda. \end{aligned}$$

Ahora realizaremos el paso inductivo, para lo cual consideraremos  $x, y \in \mathbb{R}^{n-1}$  y  $a, b \in \mathbb{R}$ . Así, por hipótesis del enunciado

$$h((1-\lambda)x + \lambda y, (1-\lambda)a + \lambda b) \geq f(x, a)^{1-\lambda} g(y, b)^\lambda.$$

Entonces obtenemos, para cualesquiera  $a, b \in \mathbb{R}$ , que

$$H((1-\lambda)a + \lambda b) \geq F(a)^{1-\lambda} G(b)^\lambda$$

donde

$$F(a) = \int_{\mathbb{R}^{n-1}} f(x, a) dx, \quad G(a) = \int_{\mathbb{R}^{n-1}} g(x, a) dx \quad \text{y} \quad H(a) = \int_{\mathbb{R}^{n-1}} h(x, a) dx.$$

El resultado se sigue por el caso unidimensional y Fubini. ■

### 1. Lema (Desigualdad multiplicativa de Brunn–Minkowski)

Sean  $A, B \subset \mathbb{R}^n$  compactos y  $\lambda \in [0, 1]$  dados. Entonces

$$\text{Vol}_n((1-\lambda)A + \lambda B) \geq \text{Vol}_n(A)^{1-\lambda} \text{Vol}_n(B)^\lambda.$$

Demostración: Consideremos  $f = \mathbb{1}_A$ ,  $g = \mathbb{1}_B$  y  $h = \mathbb{1}_{(1-\lambda)A + \lambda B}$ . Basta probar que  $f(x) = g(x) = 1$  implica  $h((1-\lambda)x + \lambda y) = 1$ , lo cual es evidente. ■

27/03/2023

Demostración del teorema 1: Basta demostrar que

$$\text{Vol}_n(A + B)^{1/n} \geq \text{Vol}_n(A)^{1/n} + \text{Vol}_n(B)^{1/n}.$$

Para esto consideremos

$$A' = \frac{1}{\text{Vol}_n(A)^{1/n}} A \quad \text{y} \quad B' = \frac{1}{\text{Vol}_n(B)^{1/n}} B,$$

de tal forma que  $\text{Vol}_n(A') = \text{Vol}_n(B') = 1$ . Por la desigualdad multiplicativa tenemos que para  $\lambda \in [0, 1]$  se cumple  $\text{Vol}_n((1-\lambda)A' + \lambda B') \geq 1$ . Considerando

$$\lambda = \frac{\text{Vol}_n(B)^{1/n}}{\text{Vol}_n(A)^{1/n} + \text{Vol}_n(B)^{1/n}}$$

deducimos

$$\text{Vol}_n\left(\frac{1}{\text{Vol}_n(A)^{1/n} + \text{Vol}_n(B)^{1/n}}(A + B)\right) \geq 1. \quad \text{■}$$



### 1. Proposición

Sea  $K \subset \mathbb{R}^n$  un compacto con volumen dado  $\text{Vol}_n(K)$ . Entonces

$$\sigma(\partial K)^n \geq n^n \omega_n \text{Vol}_n(K)^{n-1},$$

donde  $\sigma$  es la medida de superficie dada por

$$\sigma(\partial K) = \liminf_{t \rightarrow 0} \frac{\text{Vol}_n(K + tB) - \text{Vol}_n(K)}{t}$$

y  $\omega_n$  es el volumen de la bola unitaria de  $\mathbb{R}^n$ . En adición, la igualdad se da con las bolas euclidianas.

Demostración: Notemos que

$$\text{Vol}_n(K + tB) \geq \left( \text{Vol}_n(K)^{1/n} + t \omega_n^{1/n} \right)^n \geq \text{Vol}_n(K) \left( 1 + n \text{Vol}_n(K)^{(n-1)/n} t \omega_n^{1/n} \right)^n,$$

con la primera igualdad dándose por el teorema 1. Por consiguiente,

$$\sigma(\partial K) \geq n \text{Vol}_n(K)^{(n-1)/n} \omega_n.$$

Debido a que la igualdad se da cuando  $K = tB$ , obtenemos el resultado. ■

En el caso gaussiano,  $(\mathbb{R}^n, \gamma_n)$ , para un volumen fijo  $\gamma_n(K)$ , los conjuntos de medida  $\sigma_{n-1}(\partial K)$  mínimas son los semi-espacios de la forma

$$K = \{x \in \mathbb{R}^n : \langle x, u \rangle \geq \lambda\}.$$

## 5.1. Consecuencias en concentración de la medida

### 1. Lema (Borell)

Sea  $C$  un compacto convexo con interior no vacío. Sea  $P$  la medida uniforme sobre  $C$  normalizada de tal forma que  $P(C) = 1$ . Sea además  $A$  un conjunto convexo simétrico tal que  $P(A) > 1/2$ . Entonces para cualquier  $t > 1$  tendremos que

$$P(C \setminus tA) \leq P(A) \left( \frac{1 - P(A)}{P(A)} \right)^{(t+1)/2}.$$

Demostración: Comenzamos demostrando que

$$\frac{2}{t+1}(tA)^c + \frac{t-1}{t+1}A \subset A^c.$$

Sea  $u \in \mathbb{S}^{n-1}$  dado y definamos  $g(u) = \lambda u$ , donde  $\lambda = \sup\{\langle a, u \rangle : a \in A \cap C\}$ . Denotemos  $A^* = g(\mathbb{S}^{n-1})$ . Si  $x \in (tA)^c$  existen  $\lambda > t$  y  $a^* \in A^*$  tales que  $x = \lambda a^*$  y sea  $u$  la dirección asociada. Supongamos que  $a \in A$ , entonces

$$\begin{aligned} \left\langle \frac{2}{t+1}x + \frac{t-1}{t+1}a, u \right\rangle &= \frac{2\lambda}{t+1} \langle a^*, u \rangle + \frac{t-1}{t+1} \langle a, u \rangle \\ &> \frac{2t}{t+1} \langle a^*, u \rangle + \frac{t-1}{t+1} (-\langle -a^*, u \rangle) \\ &= \langle a^*, u \rangle, \end{aligned}$$

lo cual implica que

$$\frac{2}{t+1}x + \frac{t-1}{t+1}a \notin A.$$

De lo anterior

$$\text{Vol}_n(A^c) > \text{Vol}_n\left(\frac{2}{t+1}(tA)^c + \frac{t-1}{t+1}A\right) \geq \text{Vol}_n(C \setminus tA)^{2/(t+1)} \text{Vol}_n(A)^{(t-1)/(t+1)}.$$

Consecuentemente, dividiendo entre  $\text{Vol}_n(C)$ ,

$$1 - P(A) \geq P(C \setminus tA)^{2/(t+1)} P(A)^{(t-1)/(t+1)}.$$

17/04/2023

**Ejemplo del lema de Borell en acción** Sea  $K$  un convexo compacto y  $\varphi$  una aplicación lineal. Tomemos  $M$  de tal forma que el conjunto  $T = \{x \in \mathbb{R}^n : |\varphi(x)| \leq M\}$  satisfaga  $P(T) = 2/3$ . Entonces el lema de Borell nos da la desigualdad

$$P(\{x \in K : |\varphi(x)| > tM\}) \leq \frac{1}{3} \frac{1}{2^{(t-1)/2}},$$

lo cual nos dice la probabilidad de que una variable aleatoria uniforme en  $K$  exceda  $tM$  al ser evaluada en  $\varphi$  decrece de forma exponencial.

Ahora, observemos que

$$\|\varphi\|_{L^1(K)} = \frac{1}{\text{Vol}_n(K)} \int_K |\varphi(x)| dx \geq \frac{1}{\text{Vol}_n(K)} \int_{K \setminus T} |\varphi(x)| dx \geq MP(K \setminus T) = \frac{M}{3}.$$

Esta desigualdad será útil para demostrar el siguiente resultado.

### 1. Corolario

Existe una constante universal  $C > 0$  tal que para cualesquiera  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  lineal y cada  $K \subset \mathbb{R}^n$  compacto y convexo,

$$\|\varphi\|_{\Psi_1, K} \leq C \|\varphi\|_{L^1(K)},$$

donde

$$\|\varphi\|_{\Psi_1, K} = \inf \left\{ \lambda > 0 : \int_K \Psi_1 \left( \frac{|\varphi(x)|}{\lambda} \right) dP < 1 \right\}$$

$$\Psi_1(u) = e^u - 1.$$

Dicho de manera coloquial, este resultado nos dice que la aplicación de cualquier funcional lineal sobre una variable aleatoria uniforme en un convexo se concentrará de manera subexponencial.

Idea de la prueba: Basta notar que la cota

$$\text{Vol}_n(\{|\varphi(x)| > 3t \|\varphi\|_{L^1(K)}\}) \leq \text{Vol}_n(\{|\varphi(x)| > tM\}) \leq \frac{1}{3} \frac{1}{2^{(t-1)/2}}$$

no depende de  $n$  ni de  $K$ .

La norma dada en el corolario previo es una norma tipo *Orlicz*, de las cuales se verán algunas propiedades, como que ella es finita si y sólo si  $P$  es subexponencial, más adelante.

En cuanto a la norma  $\|\cdot\|_{\Psi_1}$ , es posible cambiar la función  $\Psi_1$  por cualquier función  $h$  convexa con  $h(0) = 0$ . En particular podemos usar  $\Psi_2 = e^{u^2} - 1$ , para la cual se tendrá finitud precisamente cuando la variable aleatoria sea subgaussiana.

## 5.2. Log-concavidad

### 1. Definición

Dícese que una función  $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$  es log-cóncava si para cualesquiera  $x, y \in \mathbb{R}^n$  y cualquier  $\lambda \in [0, 1]$ ,

$$f(\lambda x + (1 - \lambda)y) \geq f(x)^\lambda f(y)^{1-\lambda}.$$

Algunas propiedades básicas de las funciones log-cóncavas las resumimos en la siguiente proposición.

### 1. Proposición

- (i) Si  $f$  es no negativa y cóncava, entonces  $f$  es log-cóncava.
- (ii) La log-concavidad implica la casi-concavidad, esto es  $\{f(x) \geq t\}$  es cóncavo para cada  $t$ .
- (iii) Si  $f$  y  $g$  son log-cóncavas entonces  $fg$  es log-cóncava.
- (iv) Dada una función log-cóncava  $f : \mathbb{R}^{n+m} \rightarrow \mathbb{R}_+$ ,

$$g(x) = \int_{\mathbb{R}^m} f(x, y) dy$$

es log-cóncava.

- (v) La convolución  $f * g$  es log-cóncava siempre que  $f$  y  $g$  sean log-cóncavas.
- (vi) Dada una función  $f \in C^2(\mathbb{R}^n, \mathbb{R}_+)$  con  $f > 0$ ,  $f$  es log-cóncava si y sólo si  $f \nabla^2 f - \nabla f \nabla f^T$  es negativa semidefinida para cada  $x$ .
- (vii) Una función  $f \in C^2(\mathbb{R}, \mathbb{R}_+)$  con  $f > 0$ ,  $f$  es log-cóncava precisamente cuando  $ff'' \leq (f')^2$ .
- (viii) Para funciones  $f$  y  $g$  log-cóncavas, el producto de Asplus, definido por

$$f \triangle g(x) := \sup_{x_1, x_2 \in \mathbb{R}^n: x = x_1 + x_2} f(x_1)g(x_2),$$

es log-cóncavo.

Demostración: El punto (i) es directo de la desigualdad AM-GM. El punto (ii) es consecuencia inmediata de la igualdad  $\{f \geq t\} = \{\log f \geq \log t\}$  y la hipótesis de log-concavidad de  $f$ . El punto (iii) es evidente. Mientras tanto, el punto (iv) se sigue de la desigualdad de Prekópa-Leindler. Los puntos (iii) y (iv) implican inmediatamente (v). Para verificar el punto (vii) observemos que  $h = \log f$  es cóncava precisamente cuando  $0 \geq h'' = (f''f - (f')^2)/f^2$ . El punto (vi) es una generalización de (vii) que no probaremos. Para finalizar probemos (viii), para lo cual consideraremos  $x, y \in \mathbb{R}^n$  y  $\lambda \in [0, 1]$  dados. Sean  $x_1, x_2, y_1, y_2 \in \mathbb{R}^n$  tales que  $x_1 + x_2 = x$  y  $y_1 + y_2 = y$ . Por definición del producto de Asplus obtenemos la primera desigualdad en

$$\begin{aligned} f \triangle g(\lambda x + (1 - \lambda)y) &\geq f(\lambda x_1 + (1 - \lambda)y_1)g(\lambda x_2 + (1 - \lambda)y_2) \\ &\geq f(x_1)^\lambda f(y_1)^{1-\lambda} g(x_2)^\lambda g(y_2)^{1-\lambda} \\ &= (f(x_1)g(x_2))^\lambda (f(y_1)g(y_2))^{1-\lambda}, \end{aligned}$$

mientras que la segunda desigualdad se da por la log-concavidad de  $f$  y  $g$ . Al

considerar el supremo sobre  $x_1, x_2 \in \mathbb{R}^n$  tales que  $x_1 + x_2 = x$  obtenemos

$$f \triangle g(\lambda x + (1 - \lambda)y) \geq f \triangle g(x)^\lambda (f(y_1)g(y_2))^{1-\lambda},$$

de donde concluimos al considerar el supremo correspondiente sobre  $y_1$  y  $y_2$ . ■

## 2. Definición

Una medida  $\mu$  sobre  $\mathbb{R}^n$  es log-cóncava si para cualquier elección de conjuntos  $A, B \in \mathcal{B}(\mathbb{R}^n)$  y  $\lambda \in [0, 1]$ ,

$$\mu(\lambda A + (1 - \lambda)B) \geq \mu(A)^\lambda \mu(B)^{1-\lambda}.$$

En la definición de una medida log-cóncava aparece una desigualdad reminiscente al caso multiplicativo de la desigualdad de Brunn–Minkowski.

## 2. Proposición

*Cada medida uniforme sobre un convexo compacto es log-cóncava.*

Este resultado es una consecuencia elemental del siguiente resultado.

## 3. Proposición

*Una medida  $\mu$   $\sigma$ -finita cuyo soporte no se encuentre contenido en un hiperplano es log-cóncava si, únicamente si,  $\mu \ll \lambda$  y  $d\mu/d\lambda$  es log-cóncava.*

Demostración de la necesidad de la proposición 3: Comenzamos con la necesidad, poniendo  $f = d\mu/d\lambda$ . Entonces

$$\begin{aligned} \mu(\lambda A + (1 - \lambda)B) &= \int_{\lambda A + (1 - \lambda)B} f(x) dx \\ &\geq \left( \int_A f(x) dx \right)^\lambda \left( \int_B f(x) dx \right)^{1-\lambda} = \mu(A)^\lambda \mu(B)^{1-\lambda}, \end{aligned}$$

donde la desigualdad se da por Prekópa–Leindler. En efecto, para usar tal desigualdad basta considerar  $h = \mathbb{1}_{\lambda A + (1 - \lambda)B} f$ ,  $g = \mathbb{1}_A f$  y  $k = \mathbb{1}_B f$ . ■

Antes de proceder con la demostración de la suficiencia de la proposición 3, veremos algunas propiedades de las normas de Orlicz.

19/04/2023

## 3. Definición

Dada una función  $\Psi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  convexa con  $\Psi(0) = 0$ , definimos la *norma de Orlicz* de una variable aleatoria  $X \in \mathbb{R}^k$  se define

$$\|X\|_\Psi = \inf \left\{ c > 0 : \mathbb{E} \left[ \frac{\|X\|}{c} \right] \leq 1 \right\},$$

donde la consideramos infinita si el conjunto es vacío.

## 4. Proposición

*La norma de Orlicz es una norma, considerando las clases de equivalencia dadas por la igualdad casi segura.*

Demostración: Bastará probar la desigualdad del triángulo y que no es seminorma. Dadas  $X$  y  $Y$  variables aleatorias de norma de Orlicz finita. Consideremos  $c_1 > \|X\|_\Psi$  y  $c_2 > \|Y\|_\Psi$ . Entonces, por ser  $\Psi$  convexa y positiva obtenemos

$$\begin{aligned} \mathbb{E} \left[ \Psi \left( \frac{\|X + Y\|}{c_1 + c_2} \right) \right] &\leq \mathbb{E} \left[ \Psi \left( \frac{\|X\|}{c_1 + c_2} + \frac{\|Y\|}{c_1 + c_2} \right) \right] \\ &\leq \frac{c_1}{c_1 + c_2} \mathbb{E} \left[ \Psi \left( \frac{\|X\|}{c_1} \right) \right] + \frac{c_2}{c_1 + c_2} \mathbb{E} \left[ \Psi \left( \frac{\|Y\|}{c_2} \right) \right] \leq 1, \end{aligned}$$

con la última igualdad dándose por la elección de  $c_1$  y  $c_2$ . Al considerar el ínfimo sobre  $c_1$  y  $c_2$  obtenemos la desigualdad triangular.

Ahora sea  $X$  con  $\|X\|_\Psi = 0$  pero  $X \neq 0$ . Así existe  $\delta > 0$  tal que  $p = \mathbb{P}(\|X\| > \delta) > 0$ . Como  $\|X\|_\Psi = 0$ , para toda  $c > 0$  tendremos que  $\mathbb{E}[\Psi(\|X\|/c)] \leq 1$  lo cual nos dice que

$$\int_0^\infty \mathbb{P}(\|X\| > c\Psi^{-1}(t)) dt \leq 1.$$

No obstante, al tomar  $c$  tal que  $c\Psi^{-1}(t) \leq \delta$  obtenemos que la integral previa es infinita, lo cual es absurdo. ■

Algunas de las elecciones de  $\Psi$  de interés son

$$\begin{aligned} \Psi_1(x) &= e^x - 1, & \Psi_2(x) &= e^{x^2} - 1, & \Psi_p(x) &= e^{x^p} - 1, \\ \Psi_{\log}(x) &= e^{x \log x} - 1, & \tilde{\Psi}_{\log}(x) &= e^{x \log(x+1)} - 1. \end{aligned}$$

Una observación es que  $\|X\|_p \leq \|X\|_{\Psi_p}$ .

### 5. Proposición (Relaciones entre Orlicz y concentración)

Sea  $X$  una variable aleatoria y  $p \in [1, \infty)$ . Tenemos que son equivalentes los siguientes enunciados:

- (i)  $\|X\|_{\Psi_p} < \infty$ ;
- (ii) existen  $C, K > 0$  para las cuales  $\mathbb{P}(\|X\| > t) \leq K e^{-Ct^p}$ .

En particular, si (i) se cumple, (ii) se cumplirá con  $C = \|X\|_{\Psi_p}^{-p}$  y  $K = 2$ . Recíprocamente si (ii) se cumple, entonces

$$\|X\|_{\Psi_p} \leq \left( \frac{1 + K}{C} \right)^{1/p}.$$

Demostración: Probemos que (i) implica (ii). En este caso notemos que

$$\begin{aligned} \mathbb{P}(\|X\| > t) &= \mathbb{P} \left( \Psi \left( \frac{\|X\|}{\|X\|_{\Psi_p}} \right) \geq \Psi_p \left( \frac{t}{\|X\|_{\Psi_p}} \right) \right) \leq 1 \wedge \frac{1}{\Psi_p(t/\|X\|_{\Psi_p})} \\ &\leq 2 \exp \left\{ -\frac{t^p}{\|X\|_{\Psi_p}^p} \right\}, \end{aligned}$$

donde para la última desigualdad hemos usado que  $1 \wedge (e^u - 1)^{-1} \leq 2e^{-u}$  para cada  $u > 0$ .

Ahora probemos que (ii) implica (i). Bajo el supuesto dado tendremos que

$$\begin{aligned}\mathbb{E}\left[e^{\|X\|^p/c^p} - 1\right] &= \mathbb{E}\left[\int_0^{\|X\|^p} c^{-p} e^{s/c^p} ds\right] = \int_0^\infty \mathbb{P}(\|X\| \geq s^{1/p}) c^{-p} e^{s/c^p} ds \\ &\leq \int_0^\infty K e^{-Cs} c^{-p} e^{s/c^p} ds = \frac{K}{c^p} \frac{1}{C - c^{-p}},\end{aligned}$$

de donde vemos que basta tomar  $C$  suficientemente grande. ■

## 6. Proposición

Supongamos que  $\Psi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  es convexa con  $\Psi(0) = 0$  para la cual existe  $c > 0$  tal que

$$\limsup_{x, y \rightarrow \infty} \frac{\Psi(x)\Psi(y)}{\Psi(cxy)} \leq 1.$$

Entonces, para  $X_1, \dots, X_N$  variables aleatorias reales se tendrá

$$\left\| \max_i X_i \right\|_\Psi \leq K \Psi^{-1}(N) \max_i \|X_i\|_\Psi,$$

donde  $K$  es una constante que únicamente depende de  $\Psi$ .

24/04/2023

Demostración de la suficiencia de la proposición 3: Si  $\mu$  tiene dos o más átomos, existen  $a$  y  $b$  tales que  $a \neq b$  y  $\mu(\{a\}), \mu(\{b\}) > 0$ , lo cual implica  $\mu(\{\lambda a + (1 - \lambda)b\}) > 0$ . Esta situación implicaría que todo elemento en  $[a, b]$  es átomo, lo cual contradice el supuesto de  $\sigma$ -finitud.

En caso que  $\mu$  tenga un único átomo, digamos  $a$ , entonces por la hipótesis del soporte de  $\mu$ , existirá un cubo  $C$  que cumple  $\{a\} \cap C$  y  $\mu(C) > 0$ . Por lo tanto, existe una sucesión  $\{\lambda_i\}$  tal que la colección de cubos  $C_n = \lambda_n a + (1 - \lambda_n)C$  es disjunta y

$$\mu(C_n) \geq \mu(\{a\}) \wedge \mu(C).$$

Pero entonces

$$\mu\left(\bigcup_{n \in \mathbb{N}} C_n\right) = \sum_{n \in \mathbb{N}} \mu(C_n) = \infty,$$

a la vez que

$$\mu\left(\bigcup_{n \in \mathbb{N}} C_n\right) \leq \mu(\text{conv}(\{a\}, C)) < \infty$$

por ser la envolvente convexa compacta. Esto implica que no puede haber un átomo.

Lo anterior y la hipótesis del soporte nos dice que  $\mu \ll \lambda$  y en particular, si  $f = d\mu/d\lambda$ ,

$$f(x) = \lim_{\varepsilon \rightarrow 0} \frac{\mu(B(x, \varepsilon))}{\lambda(B(0, \varepsilon))}.$$

Ahora bien, por la log-concavidad de  $\mu$ , tendremos que

$$\begin{aligned}\frac{\mu(B(\lambda x + (1 - \lambda)y, \varepsilon))}{\lambda(B(0, \varepsilon))} &= \frac{\mu(\lambda B(x, \varepsilon) + (1 - \lambda)B(y, \varepsilon))}{\lambda(B(0, \varepsilon))} \\ &\geq \left(\frac{\mu(B(x, \varepsilon))}{\lambda(B(0, \varepsilon))}\right)^\lambda \left(\frac{\mu(B(y, \varepsilon))}{\lambda(B(0, \varepsilon))}\right)^{1-\lambda}.\end{aligned}$$

Al tomar el límite conforme  $\varepsilon \rightarrow 0$  obtenemos el resultado. ■

**1. Corolario**

Las marginales de medidas uniformes sobre convexos son log-cóncavas. En particular, tomando  $K$  un convexo y  $\theta \in \mathbb{S}^{n-1}$ ,

$$\rho(t) = \text{Vol}_{n-1}(C \cap (t\theta + \theta^\perp))$$

nos da una función log-cóncava.

Este resultado nos da la herramienta que se usa principalmente para generar funciones log-cóncavas.

**4. Definición**

Sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$  log-cóncava. Para cada  $x \in \mathbb{R}^n$  definimos

$$\|x\|_f = \left( \int_0^\infty f(rx) r^{n-1} dr \right)^{-1/n}.$$

**7. Proposición**

La función  $\|\cdot\|_f$  es una norma sobre  $\mathbb{R}^n$ . La bola unitaria  $K_f$  es compacta, convexa, simétrica y

$$\text{Vol}_n(K_f) = \frac{1}{n} \int_{\mathbb{S}^{n-1}} \int_0^\infty f(rx) r^{n-1} dr dx = \frac{1}{n} \int f.$$

Demostración: Dados  $x \in \mathbb{R}^n$  y  $\lambda > 0$  notamos que

$$\|\lambda x\|_f = \left( \int_0^\infty f(\lambda r x) r^{n-1} dr \right)^{-1/n} = \left( \int_0^\infty f(r x) \lambda^{-n} dr \right)^{-1/n} = \lambda \|x\|_f,$$

donde la segunda igualdad se debe a un cambio de variable.

La desigualdad triángular la derivaremos en dos pasos. Primeramente consideraremos las funciones

$$m(u) = f\left(\frac{x+y}{2u}\right) u^{-n-1},$$

$$h(u) = f\left(\frac{x}{u}\right) u^{-n-1}$$

y

$$g(u) = f\left(\frac{y}{u}\right) u^{-n-1}.$$

Así,

$$m\left(\frac{s+t}{2}\right) = f\left(\frac{x+y}{s+t}\right) \left(\frac{s+t}{2}\right)^{-n-1} \geq f\left(\frac{x}{s}\right)^{s/(s+t)} f\left(\frac{y}{t}\right)^{t/(s+t)} \left(\frac{s+t}{2}\right)^{-n-1}.$$

De esta expresión deducimos que

$$m\left(\frac{s+t}{2}\right) \geq \underbrace{h(s)g(t) \left(\frac{s+t}{2}\right)^{-n-1}}_{A^{n+1}} s^{(n+1)s/(s+t)} t^{(n+1)t/(s+t)}.$$



Vemos entonces que

$$\begin{aligned}\log A &= \log \frac{2}{t+s} - \frac{s}{t+s} \log \frac{1}{s} - \frac{t}{t+s} \log \frac{1}{t} \\ &\geq \log \frac{2}{t+s} - \log \left( \frac{s}{t+s} \frac{1}{s} + \frac{t}{t+s} \frac{1}{t} \right) = 0,\end{aligned}$$

lo cual implica  $A \geq 1$ . Con ello obtenemos

$$m\left(\frac{s+t}{2}\right) \geq h(s)^{s/(s+t)} g(t)^{t/(s+t)},$$

y en consecuencia

$$\frac{1}{2}\{r : h(r) \geq u\} + \frac{1}{2}\{r : g(r) \geq u\} \subset \{r : m(r) \geq u\}.$$

Por consiguiente, al usar Brunn–Minkowski en dimensión uno obtenemos

$$\int_0^\infty m(r) dr \geq \frac{1}{2} \int_0^\infty h(r) dr + \frac{1}{2} \int_0^\infty g(r) dr,$$

lo cual nos da

$$\frac{1}{\|(x+y)/2\|_f^n} \geq \frac{1}{2} \frac{1}{\|x\|_f} + \frac{1}{2} \frac{1}{\|y\|_f}.$$

Que  $f$  es norma se sigue entonces de que  $t \mapsto t^{-1/n}$  es decreciente y convexa. Observemos ahora que

$$\text{Vol}_n(K_f) = \int_{\mathbb{R}^n} \mathbb{1}_{\{\|y\|_f \leq 1\}} dy.$$

Debido a que podemos hacer una biyección  $x \mapsto (u, \theta) \in \mathbb{R}_+ \times \mathbb{S}^{n-1}$  dada por las descomposición de  $x$  es coordenadas polares, obtenemos

$$\text{Vol}_n(K_f) = \int_{\mathbb{R}_+ \times \mathbb{S}^{n-1}} \mathbb{1}_{\{\|ux\|_f \leq 1\}} u^{-n-1} dx du.$$

Al considerar el cambio de variable  $u = t^{1/n}$  deducimos

$$\begin{aligned}\text{Vol}_n(K_f) &= \int_{\mathbb{S}^{n-1}} \int_0^\infty \mathbb{1}_{\{\|t^{1/n}x\|_f \leq 1\}} \frac{1}{n} dt dx \\ &= \frac{1}{n} \int_{\mathbb{S}^{n-1}} \int_0^\infty \mathbb{1}_{\{t \leq \|x\|_f^{-n}\}} dt dx \\ &= \frac{1}{n} \int_{\mathbb{S}^{n-1}} \|x\|_f^{-n} dx.\end{aligned}$$

### 1. Observación

Cuando  $K$  es un convexo que contiene a 0,

$$\text{Vol}_n(K) = n v_n \int_{\mathbb{S}^{n-1}} \int_0^{r_\theta} s^{n-1} ds d\sigma(\theta),$$

donde  $\sigma(\mathbb{S}^{n-1}) = 1$ ,  $v_n = \text{Vol}_n(B(0, 1))$  y  $r_\theta$  es el radio de  $K$  en la dirección  $\theta$ . Más aún,

$$\text{Vol}_n(K) = v_n \int_{\mathbb{S}^{n-1}} r_\theta d\sigma(\theta).$$



**8. Proposición**

Si  $K$  es simétrico, entonces existe una norma  $\|\cdot\|_K$  de  $\mathbb{R}^n$  tal que  $K = \{x \in \mathbb{R}^n : \|x\|_K \leq 1\}$ . En adición se cumple la relación  $r_\theta = \|\theta\|_K^{-1}$ .

**5.3. Primeros ejemplos de convexos****5.3.1. Ejemplos fundamentales**

**El cubo en  $\mathbb{R}^n$**  Dado por el convexo  $[-1, 1]^n$ .

**1. Proposición**

- La bola euclidiana de radio máximo contenida en el cubo es de radio uno.
- La bola euclidiana de radio mínimo que incluye el cubo es de radio  $n^{1/2}$ .

**El  $n$ -simplex regular**

**1. Definición**

El  $n$ -simplex regular es la envolvente convexa de  $n+1$  puntos igualmente espaciados.

**2. Proposición**

Si  $K$  es un  $n$ -simplex y  $r$  y  $R$  son tales que

$$rB_2^n \subset K \subset RB_2^n,$$

entonces  $R/r = n$ .

**Conos**

**2. Definición**

Un cono convexo en  $\mathbb{R}^n$  es la envolvente convexa de un punto  $a \in \mathbb{R}^n$  y un convexo  $K \subset \mathbb{R}^{n-1}$ .

**3. Proposición**

Dado un cono  $C$  con base  $K$ ,

$$\text{Vol}_n(C) = \frac{\text{Vol}_{n-1}(K)h}{n}$$

donde  $h = d(a, \mathbb{R}^{n-1})$ .

**Octaedro de dimensión  $n$**

**3. Definición**

Un octaedro es la envolvente convexa de los  $2n$  puntos  $\pm e_i$ ,  $i \in [n]$ . A él lo denotaremos por  $B_1^n$  pues coincide con la bola unitaria dada por  $\|\cdot\|_1$ .

#### 4. Proposición

Las esferas circunscritas e inscritas a  $B_1^n$  tienen un cociente de radios de  $n^{-1/2}$ .

#### 5. Proposición

Se tiene que

$$\text{Vol}_n(B_1^n) = \frac{2^n}{n!}.$$

### Bola euclidiana

#### 4. Definición

Consideramos la bola euclidiana

$$B_2^n = \left\{ x \in \mathbb{R}^n : \sum_{i=1}^n x_i^2 \leq 1 \right\}.$$

El volumen de  $B_2^n$  lo denotaremos por  $v_n$ . La superficie de  $B_2^n$  es la *esfera euclidiana* es dada por

$$\mathbb{S}_2^{n-1} = \partial B_2^n.$$

#### 6. Proposición

Se cumple que  $\sigma_{n-1}(\mathbb{S}_2^{n-1}) = n v_n$  y

$$v_n = \frac{\pi^{n/2}}{\Gamma(1 + n/2)}.$$

Demostración: Notemos que

$$\begin{aligned} \sigma_{n-1}(\partial B_2^n) &= \liminf_{\varepsilon \rightarrow 0} \frac{\text{Vol}_n(B_2^n + \varepsilon \mathbb{S}_2^{n-1}) - \text{Vol}_n(B_2^n)}{\varepsilon} \\ &= \liminf_{\varepsilon \rightarrow 0} \frac{(1 + \varepsilon)^n - 1}{\varepsilon} v_n = n v_n. \end{aligned}$$

Mientras tanto, si  $(r, \theta) \in \mathbb{R}_+ \times \mathbb{S}_2^{n-1}$  es la descomposición polar de  $x \in \mathbb{R}^n$ , obtenemos

$$\begin{aligned} \int_{\mathbb{R}^n} f &= \int_{\mathbb{R}_+ \times \mathbb{S}_2^{n-1}} f(r\theta) r^{n-1} dr \times d\sigma_{n-1}(\theta) \\ &= n v_n \int_{\mathbb{R}_+ \times \mathbb{S}_2^{n-1}} f(r\theta) r^{n-1} dr \times d\sigma(\theta), \end{aligned}$$

donde  $\sigma$  es una normalización de  $\sigma_{n-1}$  con masa uno. Al considerar  $f(x) = \exp\{-|x|^2/2\}$  vemos por un lado que

$$\int_{\mathbb{R}^n} f = \left( \int_{\mathbb{R}} e^{-x^2/2} dx \right)^n = \sqrt{2\pi}^n$$

y por otro

$$\int_{\mathbb{R}^n} f = n v_n \int_{\mathbb{R}_+} \int_{\mathbb{S}_2^{n-1}} e^{-r^2/2} r^{n-1} d\sigma dr = v_n 2^{n/2} \Gamma\left(\frac{n}{2} + 1\right).$$

**1. Observación**

Conforme  $n \rightarrow \infty$ ,

$$v_n \sim C \left( \frac{2\pi e}{n} \right)^{n/2}.$$

Nos interesa saber en qué parte de la bola euclidiana se concentra la masa del volumen, lo que nos dará un ejemplo de concentración de la medida.

Sean  $B = v_n^{-1/n} B_2^n$  de tal forma que  $\text{Vol}_n(B) = 1$ . La intersección de  $B$  con un hiperplano  $H$  que contiene a  $0$  es una bola euclidiana de  $\mathbb{R}^{n-1}$  de tal forma que

$$\text{Vol}_{n-1}(B \cap H) = v_{n-1} v_n^{-(n-1)/n} \xrightarrow{n \rightarrow \infty} \sqrt{e}.$$

Si ahora consideramos un hiperplano que no contenga al cero, es decir de la forma  $x + H$ , entonces la  $n - 1$  bola euclidiana  $B_2^n \cap (x + H)$  tiene radio  $\sqrt{r_0^2 - x^2}$ , donde  $r_0 = v_n^{-1/n}$ , de modo que su volumen es aproximadamente

$$\left( \frac{\sqrt{r_0^2 - x^2}}{r_0} \right)^{n-1} \sqrt{e} = \sqrt{e} \left( 1 - \frac{x^2}{r_0^2} \right)^{(n-1)/2}.$$

Puesto que  $r_0 \sim (2\pi e/n)^{-1/2}$ , deducimos

$$\text{Vol}_{n-1}(B \cap (x + H)) \sim \sqrt{e} \left( 1 - \frac{2\pi e x^2}{n} \right)^{(n-1)/2} \sim \sqrt{e} \exp\{-\pi e x^2\}.$$

Hemos demostrado el siguiente resultado.

**7. Proposición**

Sea  $\mu_n$  la distribución marginal de la distribución uniforme sobre  $B_2^n$ . Entonces

$$\mu_n \Rightarrow \eta$$

donde  $\eta$  es la distribución normal con media cero y varianza  $1/2\pi e$ .

**NO HAY DEPENDENCIA DE LA VARIANZA CON RESPECTO A  $n$ .** En particular 96 % se ubica en el conjunto

$$\left\{ x \in \mathbb{R}^n : -\frac{1}{2} \leq x_1 \leq \frac{1}{2} \right\}.$$

**5.4. La gran conjetura**

La primera formulación se debe a BOURGAIN con la pregunta:

Sea  $K$  un convexo simétrico compacto de  $\mathbb{R}^n$  con  $\text{Vol}_n(K) = 1$ . ¿Existe  $u \in \text{SL}_n(\mathbb{R})$  de tal forma que todas las secciones  $u(K) \cap H$  con  $0 \in H$  tengan casi el mismo volumen  $n - 1$ -dimensional?

Es posible reinterpretar tal pregunta como:

¿Podemos transformar  $K$  en una casi-bola?

Por su parte, BUSEMANN y PETTY preguntaron:

Sean  $K$  y  $T$  dos convexos simétricos compactos tales que para cada  $\theta \in \mathbb{S}^{n-1}$ ,  $\text{Vol}_{n-1}(K \cap \theta^\perp) \leq \text{Vol}_{n-1}(T \cap \theta^\perp)$ . ¿Entonces  $\text{Vol}_n(K) \leq \text{Vol}_n(T)$ ?

### 1. Proposición

*Si  $K$  es un elipsoide la pregunta de BP tiene respuesta afirmativa.*

### 2. Proposición

*Si  $n \leq 4$  es un elipsoide la pregunta de BP tiene respuesta afirmativa, pero ella no es cierta para  $n \geq 5$ .*

Un contraejemplo cuando  $n$  es grande, es el siguiente: Sea  $K = [-1/2, 1/2]^n$  y  $T = \lambda_n B_2^n$  con  $\text{Vol}(T) = 9/10$ . Puede probarse que  $\text{Vol}_{n-1}(K \cap \theta^\perp) \leq \sqrt{2}$ . Por su parte

$$\text{Vol}_{n-1}(T \cap \theta^\perp) = \left(\frac{9}{10}\right)^{(n-1)/n} \frac{\Gamma(1 + n/2)^{(n-1)/n}}{\Gamma((n+1)/2)} \rightarrow 0.9\sqrt{e}.$$

**Conjetura (BP corregida)** Sean  $K$  y  $T$  convexos simétricos compactos tales que  $\text{Vol}_{n-1}(K \cap \theta^\perp) \leq \text{Vol}_{n-1}(T \cap \theta^\perp)$ . ¿Existe  $c > 1$  universal tal que  $\text{Vol}_n(K) \leq c \text{Vol}_n(T)$ ?

**Bourgain slicing conjecture** Si  $K \subset \mathbb{R}^n$  es un convexo con  $\text{Vol}_n(K) = 1$ , ¿existe  $c > 0$  universal y un hiperplano  $H$  tal que  $\text{Vol}_{n-1}(K \cap H) \geq 1/c$ ?

# Concentración de la medida

# 6

17/05/2023

## 6.1. Concentración clásica

La idea detrás de la cota de Cramer–Chernoff es considerar la desigualdad

$$\mathbb{P}(X \geq t) \leq e^{-\lambda t} \mathbb{E}[e^{\lambda X}]$$

bajo el supuesto de que  $X$  tenga momentos exponenciales. Si definimos  $\Psi_X(\lambda) = \log \mathbb{E}[e^{\lambda X}]$ , lo anterior nos da

$$\mathbb{P}(X \geq t) \leq e^{-(\lambda t - \Psi_X(\lambda))},$$

de modo que nos interesa el encontrar el valor de  $\lambda$  que nos dé la mejor cota posible.

### 1. Definición (Transformada de Fenchel–Legendre)

Ella se define por

$$\Psi_X^*(t) = \sup_{\lambda \geq 0} (\lambda t - \Psi_X(\lambda)).$$

### 1. Proposición

$$\mathbb{P}(X \geq t) \leq e^{-\Psi_X^*(t)}$$

### 1. Ejemplo

Si  $X \sim N(0, \sigma^2)$ ,  $\Psi_X(\lambda) = \lambda^2 \sigma^2 / 2$  y

$$\Psi_X^*(t) = \sup_{\lambda \geq 0} \left( \lambda t - \frac{\lambda^2 \sigma^2}{2} \right) = \frac{t^2}{2\sigma^2}.$$

### 2. Ejemplo

Si  $Y \sim \text{Poisson}(\nu)$  y  $Z = Y - \nu$ ,  $\Psi_Z(\lambda) = \nu(e^\lambda - \lambda - 1)$  y por ende

$$\Psi_Z^*(t) = \sup_{\lambda \geq 0} (\lambda t - \nu(e^\lambda - \lambda - 1)).$$

Vemos que la derivada de lo de adentro del supremo es  $t - \nu e^\lambda + \nu$ , lo cual nos da  $\lambda = \log(1 + t/\nu)$  y en consecuencia

$$\Psi_Z^*(t) = \nu h\left(\frac{t}{\nu}\right),$$

donde  $h(x) = (1 + x) \log(1 + x) - x$ .

## 2. Definición

Si  $\Psi_X(\lambda) \leq \lambda^2 \nu / 2$  para cualquier  $\lambda \geq 0$  diremos que  $X$  es sub-gaussiana de constante  $\nu$ . Esto lo denotaremos por  $X \in \mathcal{G}(\nu)$ . Si la desigualdad solo se cumple para  $\lambda \geq 0$ , lo denotaremos por  $X \in \mathcal{G}_+(\nu)$ .

En adición, diremos que  $X$  es sub-gamma a la derecha si  $\Psi_X(\lambda) \leq \lambda^2 \nu / 2 (1 - c\lambda)$  para  $\lambda \in (0, 1/c)$ , propiedad que denotaremos por  $X \in \Gamma_+(\nu, c)$ .

## 3. Ejemplo

Afirmamos que una variable aleatoria con distribución  $\Gamma(a, b)$  es sub-gamma. En efecto, si  $Y$  es una variable aleatoria cuya distribución tiene la densidad

$$f(x) = \frac{x^{a-1} e^{-x/b}}{\Gamma(a) b^a} \mathbb{1}_{x>0},$$

$\mathbb{E}[Y] = ab$  y  $\text{Var}(Y) = ab^2$  y entonces, para  $\lambda < 1/b$ , al definir  $X = Y - \mathbb{E}[Y]$ ,

$$\mathbb{E}[e^{\lambda X}] = \exp\{-\lambda ab - a \log(1 - \lambda b)\}.$$

Puesto que para cada  $u \in (0, 1)$  se cumple la desigualdad  $-\log(1 - u) - u \leq u^2 / 2(1 - u)$ , obtenemos

$$\Psi_X(\lambda) \leq \frac{a(\lambda b)^2}{2(1 - \lambda b)},$$

con lo cual obtenemos la afirmación con  $\nu = ab^2$  y  $c = b$ .

## 2. Proposición

Sea  $\Psi$  convexa, derivable en  $[0, b]$  con  $b \in (0, \infty]$ , y tal que  $\Psi(0) = \Psi'(0) = 0$ . Para  $t \geq 0$  definimos

$$\Psi^*(t) = \sup_{\lambda \in [0, b]} (\lambda t - \Psi(\lambda)).$$

Entonces tendremos que  $\Psi^* \geq 0$ ,  $\Psi^*$  es creciente y convexa y

$$(\Psi^*)^{-1}(y) = \inf_{\lambda \in (0, b)} \frac{y + \Psi(\lambda)}{\lambda}.$$

Demostración: Es claro que  $\Psi^*$  es convexa y creciente, de modo que la primera propiedad se sigue al observar que  $\Psi^*(0) = 0$ . Verifiquemos la fórmula de inversión. Sea  $u = \inf_{\lambda \in (0, b)} \lambda^{-1}(y + \Psi(\lambda))$ . Entonces se cumple que  $u \geq t$  precisamente cuando

$$\frac{y + \Psi(\lambda)}{\lambda} \geq t \text{ para todo } \lambda \in (0, b),$$

lo cual es equivalente a  $y \geq \lambda t - \Psi(\lambda)$  para toda  $\lambda \in (0, b)$ . Así,  $u \geq t$  sucede si y sólo si  $y \geq \Psi^*(t)$ , lo cual nos da el resultado deseado. ■

## 3. Proposición

Sean  $Z_1, \dots, Z_n$  variables aleatorias reales tales que para toda  $\lambda \in (0, b)$  se satisface

$\Psi_{Z_i}(\lambda) \leq \Psi(\lambda)$  con  $\Psi$  una función convexa, derivable que satisface  $\Psi(0) = \Psi'(0) = 0$ . Entonces

$$\mathbb{E}\left[\max_{i \leq n} Z_i\right] \leq (\Psi^*)^{-1}(\log n).$$

En el caso sub-gamma se tiene que

$$(\Psi^*)^{-1}(x) \leq \sqrt{2vx} + cx,$$

lo cual permite dar una cota del máximo de variables aleatorias sub-gamma.

### 1. Teorema (Bennett)

Sean  $X_1, \dots, X_n$  variables aleatorias independientes con  $X_i \leq b$  para toda  $i$ . Sea  $S = \sum_{i=1}^n (X_i - \mathbb{E}[X_i])$  y  $v = \sum_{i=1}^n \mathbb{E}[X_i^2]$ . Consideremos también  $\phi(u) = e^u - u - 1$  para toda  $u \in \mathbb{R}$ . Entonces

$$\Psi_S(\lambda) \leq \frac{v}{b^2} \phi(b\lambda).$$

Es decir que  $S$  tiene concentración sub-Poisson.

### 2. Teorema (Bernstein)

Sean  $X_1, \dots, X_n$  independientes. Supongamos que existen  $c, v > 0$  que satisfagan  $\sum_{i=1}^n \mathbb{E}[X_i^2] \leq v$  y  $\sum_{i=1}^n \mathbb{E}[(X_i)_+^q] \leq q! v e^{q-2}/2$  para toda  $q \geq 3$ . Entonces, para cualquier  $\lambda \in (0, 1/c)$ ,

$$\Psi_S(\lambda) \leq \frac{v\lambda^2}{2(1 - c\lambda)},$$

y en particular, para  $u > 0$ ,

$$\Psi^*(t) \geq \frac{v}{c} g\left(\frac{ct}{2}\right),$$

donde  $g(u) = 1 + u - \sqrt{1 + 2u}$ . Es decir que  $S$  es sub-Gamma.

Demostración del teorema de Bernstein: Nótese que si  $\psi(u) = e^u - u - 1$  entonces  $\psi(u) \leq u^2/2$  cuando  $u \leq 0$ . Luego,

$$\begin{aligned} \phi(\lambda X_i) &= \psi(\lambda(X_i)_-) \mathbb{1}_{\{X_i \leq 0\}} + \psi(\lambda(X_i)_+) \mathbb{1}_{\{X_i > 0\}} \\ &\leq \frac{\lambda^2 (X_i)_-^2}{2} + \sum_{q \geq 2} \frac{\lambda^q (X_i)_+^q}{q!}. \end{aligned}$$

Entonces

$$\sum_{i=1}^n \mathbb{E}[\phi(\lambda X_i)] \leq \frac{v}{2} \sum_{q \geq 2} \lambda^q c^{q-2},$$

de donde

$$\begin{aligned}
 \Psi_s(\lambda) &\leq \sum_{i=1}^n \log \mathbb{E}[e^{\lambda X_i}] - \lambda \mathbb{E}[X_i] \\
 &\leq \sum_{i=1}^n \mathbb{E}[e^{\lambda X_i}] - 1 - \lambda \mathbb{E}[X_i] \\
 &= \sum_{i=1}^n \mathbb{E}[\psi(\lambda X_i)] \\
 &\leq \frac{v\lambda^2}{2} \sum_{q \geq 0} (\lambda c)^q = \frac{v\lambda^2}{2(1-\lambda c)}.
 \end{aligned}$$

■

## 6.2. Desigualdades de entropía (teoría de la información)

### 1. Teorema (Efron–Stein)

Sean  $X_1, \dots, X_n$  variables aleatorias independientes y sea  $Z = f(X_1, \dots, X_n)$  tal que  $\mathbb{E}[Z] < \infty$ . Entonces

$$\text{Var}(Z) \leq \sum_{i=1}^n \mathbb{E}[\text{Var}^{(i)}(Z)],$$

donde  $\text{Var}^{(i)}(Z) = \mathbb{E}^{(i)}[(Z - \mathbb{E}^{(i)}[Z])^2]$  y  $\mathbb{E}^{(i)}[\cdot] = \mathbb{E}[\cdot | X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n]$ .

Demostración: Usaremos la técnica de martingalas de Doob. Consideremos  $\mathbb{E}_i[\cdot] = \mathbb{E}[\cdot | X_1, \dots, X_i]$  y  $\mathbb{E}_0 = \mathbb{E}$ . Definamos además  $\Delta_i = \mathbb{E}_i[Z] - \mathbb{E}_{i-1}[Z]$ . Con ello  $Z - \mathbb{E}[Z] = \sum_{i=1}^n \Delta_i$ .

Comenzamos notando que las variables  $\Delta_i$  tienen correlación cero. Para verlo observemos que si  $j > i$ ,

$$\mathbb{E}_i[\Delta_j] = \mathbb{E}_i[\mathbb{E}_j[Z]] - \mathbb{E}_i[\mathbb{E}_{j-1}[Z]] = 0$$

por la propiedad torre. Por la misma razón

$$\mathbb{E}[\Delta_i \Delta_j] = \mathbb{E}[\Delta_i \mathbb{E}_i[\Delta_j]] = 0.$$

En tanto, es claro que  $\mathbb{E}_i[\mathbb{E}^{(i)}[Z]] = \mathbb{E}_{i-1}[Z]$ . Entonces la desigualdad de Jensen implica la cota

$$\Delta_i^2 \leq \mathbb{E}_i[(Z - \mathbb{E}^{(i)}[Z])^2].$$

El resultado se sigue de ambas observaciones. ■

### 1. Definición

Consideremos una función  $\psi \in C^2(\mathbb{R}_+, \mathbb{R})$  con  $\psi'' > 0$  y  $1/\psi''$  cóncava. Definimos la  $\psi$ -entropía de  $Z$  como

$$H_\psi(Z) = \mathbb{E}[\psi(Z)] - \psi(\mathbb{E}[Z]).$$

Si  $\psi(x) = x^2$  obtenemos la varianza mientras que al tomar  $\psi(x) = x \log x$  obtenemos la entropía clásica, la cual denotaremos por  $\text{Ent}(\cdot)$ . Más aún, todas las funciones de la forma  $\psi : x \mapsto x^\alpha \log x^\beta$ , con  $\alpha \in [1, 2]$  funciona.



**2. Teorema (Generalización de Efron–Stein)**

Se tiene que

$$H_\phi(Z) \leq \mathbb{E} \left[ \sum_{i=1}^n H_\phi^{(i)}(Z) \right],$$

donde  $H_\phi^{(i)}(Z) = \mathbb{E}^{(i)}[\phi(Z)] - \phi(\mathbb{E}^{(i)}[Z])$ .

**6.3. Desigualdades Log-Sobolev**

Consideraremos funciones de la forma  $f : \{-1, 1\}^n \rightarrow \mathbb{R}^n$ , un vector  $Z$  de variables aleatorias Rademacher de parámetro  $1/2$  con entradas independientes. Definimos

$$\nabla_i f(x) = \frac{f(x) - f(x^{(i)})}{2},$$

donde  $x^{(i)} = (x_1, \dots, x_{i-1}, -x_i, x_{i+1}, \dots, x_n)$ .

**1. Teorema (Log-Sobolev sobre el hipercubo)**

Sean  $f$  y  $Z$  como antes. Entonces tendremos que

$$\text{Ent}(f^2) = \text{Ent}(f(Z)^2) \leq 2\mathbb{E}[\|\nabla f(Z)\|^2].$$

Demostración: Por la generalización de Efron–Stein vemos que

$$\text{Ent}(f^2) \leq \mathbb{E} \left[ \sum_{i=1}^n \text{Ent}^{(i)}(f(Z)^2) \right].$$

Consecuentemente basta probar que se cumple la desigualdad

$$\text{Ent}^{(i)}(f(Z)^2) \leq 2\mathbb{E}^{(i)}[(\nabla_i f(Z))^2].$$

Entonces basta demostrar el resultado en el caso unidimensional, es decir  $\text{Ent}(f^2) \leq 2^{-1}\mathbb{E}[(f(1) - f(-1))^2]$ . Si ponemos  $a = f(1)$  y  $b = f(-1)$ , la desigualdad que debemos demostrar que

$$\frac{a^2}{2} \log a^2 + \frac{b^2}{2} \log b^2 - \frac{a^2 + b^2}{2} \log \frac{a^2 + b^2}{2} \leq \frac{(a - b)^2}{2}.$$

Esto se deja como ejercicio para el lector. ■

**2. Teorema (Log-Sobolev gaussiano)**

Sea  $X$  un vector gaussiano con media cero y matriz de covarianzas igual a la identidad. Dada  $f \in C^1(\mathbb{R}^n)$ ,

$$\text{Ent}(f(X)^2) \leq 2\mathbb{E}[\|\nabla f(X)\|^2].$$

Demostración: Procediendo como en la demostración anterior basta demostrar el caso unidimensional. Sea  $X_n = n^{-1/2} \sum_{i=1}^n \epsilon_i$  donde  $\epsilon_i$  son variables aleatorias Rademacher de parámetro  $1/2$  independientes. Sabemos que  $X_n \Rightarrow X$ , de modo que

$\text{Ent}(f(X_n)^2) \rightarrow \text{Ent}(f(X)^2)$ . Por otra parte, la desigualdad log-Sobolev en el hipercubo nos permite obtener

$$\begin{aligned} \text{Ent}(f(X_n)^2) &\leq 2\mathbb{E}\left[\sum_{i=1}^n \left(\frac{f(X_n) - f(X_n - 2\varepsilon/\sqrt{n})}{2}\right)^2\right] \\ &= 2\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \left(\frac{f(X_n) - f(X_n - 2\varepsilon_i/\sqrt{n})}{2\varepsilon_i/\sqrt{n}}\right)^2\right] \\ &\rightarrow 2\mathbb{E}[f'(X)^2]. \end{aligned}$$

### 3. Teorema (Tsirelson–Ibragimov–Sudakov)

Sean  $X_1, \dots, X_n$  variables aleatorias gaussianas independientes e idénticamente distribuidas. Sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  una función Lipschitz de constante  $L$ . Entonces  $Z = f(X_1, \dots, X_n)$  es sub-gaussiana de constante  $L^2$ . Esto es,

$$\mathbb{E}[e^{\lambda(Z - \mathbb{E}[Z])}] \leq e^{\lambda^2 L^2 / 2},$$

lo que nos dice

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq t) \leq e^{-t^2 / 2L^2},$$

### 1. Lema (de Herbst)

Sea  $Z$  tal que existe  $v > 0$  para la cual se cumple

$$\text{Ent}(e^{\lambda Z}) \leq \frac{\lambda^2 v}{2} \mathbb{E}[e^{\lambda Z}]$$

para toda  $\lambda > 0$ , entonces  $Z \in \mathcal{G}(v)$ .

Demostración: Notemos que al definir  $F(\lambda) = \mathbb{E}[e^{\lambda Z}]$ ,  $F'(\lambda) = \mathbb{E}[Ze^{\lambda Z}]$  y puesto que la condición se escribe

$$\lambda F'(\lambda) - F(\lambda) \log F(\lambda) \leq \frac{\lambda^2 v}{2} F(\lambda),$$

lo que se lo mismo que

$$\left(\frac{\log F(\lambda)}{\lambda}\right)' \leq \frac{v}{2}.$$

Pero entonces

$$\frac{\log F(\lambda)}{\lambda} - \frac{F'(0)}{F(0)} = \frac{\log F(\lambda)}{\lambda} - \mathbb{E}[Z] \leq \frac{v\lambda}{2}.$$

Demostración del teorema: Sabemos que es posible aproximar  $f$  por una sucesión de funciones derivables tales que  $\|\nabla f\|_\infty \leq L$ . Supongamos que  $\mathbb{E}[Z] = 0$ . Usando la desigualdad log-Sobolev con  $x \mapsto e^{\lambda f(x)/2}$  deducimos

$$\text{Ent}(e^{\lambda f}) \leq 2\mathbb{E}[\|\nabla e^{\lambda f/2}\|^2] = \frac{\lambda^2}{2} \mathbb{E}[e^{\lambda f} \|\nabla f\|^2] \leq \frac{\lambda^2 L^2}{2} \mathbb{E}[e^{\lambda f}].$$

La conclusión se sigue del lema de Herbst.

# **Parte III.**

## **Tareas**

**Ejercicio 1:**

Sean  $X_1, \dots, X_n$  variables aleatorias reales tales que  $\mathbb{E}[X_i] = 0$  para toda  $i$ . Supongamos que

- $\sum_i \text{Var}(X_i) = o(n^2)$ ,
- existe  $\phi : \mathbb{N} \rightarrow \mathbb{R}_+$  que satisface  $|\text{Cov}(X_i, X_j)| \leq \phi(|i - j|)$  para cualesquiera  $i, j$  distintas y en adición

$$\frac{1}{n} \sum_{i=1}^n \phi(i) \xrightarrow{n \rightarrow \infty} 0.$$

Entonces

$$S_n := \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} 0.$$

Demostración: Fijemos  $\varepsilon > 0$ . Debido a que las variables aleatorias son centradas,  $\mathbb{E}[S_n^2] = \text{Var}(S_n)$ . Por consiguiente, usando la desigualdad de Markov,

$$\mathbb{P}(|S_n| > \varepsilon) = \mathbb{P}(S_n^2 > \varepsilon^2) \leq \frac{1}{\varepsilon^2} \text{Var}(S_n).$$

Ahora bien, observando que

$$\text{Var}(S_n) = \frac{1}{n^2} \sum_{i,j=1}^n \text{Cov}(X_i, X_j) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) + \frac{2}{n^2} \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j),$$

por la primera parte de la segunda hipótesis,

$$(6.1) \quad \mathbb{P}(|S_n| > \varepsilon) \leq \frac{1}{\varepsilon^2 n^2} \sum_{i=1}^n \text{Var}(X_i) + \frac{2}{\varepsilon^2 n^2} \sum_{1 \leq i < j \leq n} \phi(j - i).$$

En vista de la primera hipótesis,  $(\varepsilon n)^{-2} \sum_{i=1}^n \text{Var}(X_i) = o(1)$ , de modo que resta probar que el segundo sumando de la cota en (6.1). Para ello observemos que

$$\sum_{1 \leq i < j \leq n} \phi(j - i) = \sum_{i=1}^{n-1} (n - i) \phi(i)$$

y por consiguiente,

$$\frac{1}{n^2} \sum_{1 \leq i < j \leq n} \phi(j - i) \leq \frac{1}{n} \sum_{i=1}^n \phi(i),$$

donde hemos usado la positividad de  $\phi$  para añadir  $\phi(n)$  a la suma. Por la cota previa y la última parte de la segunda hipótesis concluimos que

$$2(\varepsilon n)^{-2} \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j) = o(1),$$

lo cual nos permite deducir que

$$\lim_{n \rightarrow \infty} \mathbb{P}(|S_n| > \varepsilon) = 0.$$

Puesto que  $\varepsilon > 0$  fue arbitrario, concluimos que  $S_n$  converge a 0 en probabilidad. ■

## **Parte IV.**

# **Referencias**

# Referencias

---

## Principales

- Boucheron, S., Lugosi, G., & Massart, P. (2016). *Concentration Inequalities: A Nonasymptotic Theory of Independence* (First published in paperback). Oxford University Press.
- Devroye, L., & Lugosi, G. (2001). *Combinatorial Methods in Density Estimation*. Springer New York. <https://doi.org/10.1007/978-1-4613-0125-7>
- Feller, W. (1968). *An Introduction to Probability Theory and its Applications* (3.<sup>a</sup> ed.). Wiley.
- Gine, E., & Zinn, J. (1984). Some Limit Theorems for Empirical Processes. *The Annals of Probability*, 12(4). <https://doi.org/10.1214/aop/1176993138>
- Ledoux, M., & Talagrand, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes*. Springer-Verlag.
- Massart, P. (2007). *Concentration Inequalities and Model Selection: Ecole d'Été de Probabilités de Saint-Flour XXXIII - 2003*. Springer-Verlag.
- Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Institute of Mathematical Statistics ; American Statistical Association.
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley.
- Talagrand, M. (2005). *The Generic Chaining*. Springer-Verlag. <https://doi.org/10.1007/3-540-27499-5>
- van de Geer, S. (2006). *Applications of Empirical Process Theory* (Reprint). Cambridge Univ. Pr.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.
- Williams, D. (2018). *Probability with Martingales* (22nd printing). Cambridge Univ. Pr.

## Suplementarias

- Dedecker, J., & Doukhan, P. (2003). A new covariance inequality and applications. *Stochastic Processes and their Applications*, 106(1), 63-80. [https://doi.org/10.1016/S0304-4149\(03\)00040-1](https://doi.org/10.1016/S0304-4149(03)00040-1)