

Estimación de conjuntos bajo restricciones de forma.

Alejandro Cholaquidis.

UDELAR, Uruguay.

Seminario de Probabilidad y Estadística.
Setiembre, 2012.

- 1 Introducción.
- 2 Estimación de conjuntos ρ - cono convexos.
- 3 Otras tasas de convergencia

- 1 Introducción.
- 2 Estimación de conjuntos ρ - cono convexos.
- 3 Otras tasas de convergencia

Distancia entre conjuntos

Distancia de Hausdorff

Sea A and C no vacío, compacto $A \subset \mathbb{R}^d$,

$$d_H(A, C) = \max \left\{ \sup_{a \in A} d(a, C), \sup_{c \in C} d(c, A) \right\},$$

donde

$$d(a, C) = \inf \{ \|a - c\| : c \in C \}.$$

Distancia entre conjuntos

Distancia de Hausdorff

Sea A and C no vacío, compacto $A \subset \mathbb{R}^d$,

$$d_H(A, C) = \max \left\{ \sup_{a \in A} d(a, C), \sup_{c \in C} d(c, A) \right\},$$

donde

$$d(a, C) = \inf \{ \|a - c\| : c \in C \}.$$

Distancia en Medida

Sea A y C en $\mathcal{B}(\mathbb{R}^d)$,

$$d_\mu(A, C) = \mu(A \Delta C),$$

donde μ es una medida de Borel, y $A \Delta C = (A \setminus C) \cup (C \setminus A)$.



Restricciones de Forma

Alcance Positivo

Sea $S \subset \mathbb{R}^d$ compacto, U_S denota el conjunto de puntos con una única proyección sobre S .

$$\text{Alcance}(S, x) = \sup \left\{ r \in \mathbb{R} : B(r, x) \subset U_S \right\},$$

$$\text{Alcance}(S) = \inf_{x \in S} \text{Alcance}(S, x).$$

Restricciones de Forma

Alcance Positivo

Sea $S \subset \mathbb{R}^d$ compacto, U_S denota el conjunto de puntos con una única proyección sobre S .

$$\text{Alcance}(S, x) = \sup \left\{ r \in \mathbb{R} : B(r, x) \subset U_S \right\},$$

$$\text{Alcance}(S) = \inf_{x \in S} \text{Alcance}(S, x).$$

Federer, 1959. Si $\text{Alcance}(S)=r$

$$\mu\{S \oplus B(0, \varepsilon)\} = \sum_{k=0}^d \Phi_k(S) \varepsilon^k, \quad \forall 0 \leq \varepsilon \leq r,$$

donde μ es la medida de Lebesgue. $\Phi_0(S) = \mu(S)$, $\Phi_d(S) = b_d \mathcal{X}(S)$.

Restricciones de Forma

r -convexidad

$S \subset \mathbb{R}^d$ es r -convexo ($r > 0$) si $S = C_r(S)$, donde

$$C_r(S) = \bigcap_{\{\overset{\circ}{B}(x,r) : \overset{\circ}{B}(x,r) \cap S = \emptyset\}} \overset{\circ}{B}(x,r)^c.$$

Restricciones de Forma

r -convexidad

$S \subset \mathbb{R}^d$ es r -convexo ($r > 0$) si $S = C_r(S)$, donde

$$C_r(S) = \bigcap_{\{\overset{\circ}{B}(x,r) : \overset{\circ}{B}(x,r) \cap S = \emptyset\}} \overset{\circ}{B}(x,r)^c.$$

Conjuntos ρ cono convexos

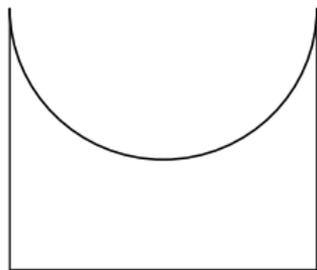
$S \subset \mathbb{R}^d$ compacto es ρ -cono convexo con $\rho \in (0, \pi]$ si :

$$\forall x \in \partial S \exists C_\rho(x) \text{ cono abierto de ángulo } \rho \text{ tal que } C_\rho(x) \subset S^c.$$

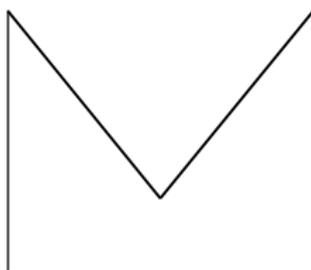
Notación: \mathcal{C}_ρ conjuntos ρ cono convexos.



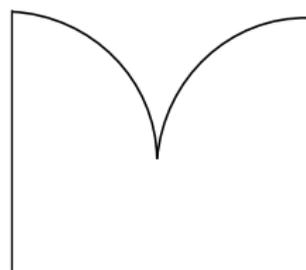
Ejemplos



(a)



(b)



(c)

Figure: (a) r -convexo, (b) ρ -cono convexo; (c) No es ρ -c.c.

ρ - Envolverte como convexa.

Sea $S \subset \mathbb{R}^d$ acotado, definimos

- $\mathcal{C}_\rho(S) = \bigcap \{B: S \subset B \text{ y } B \in \mathcal{C}_\rho\}$ B
- $\tilde{\mathcal{C}}_\rho(S) = \bigcap \{C_\rho(x)^c: C_\rho(x) \subset S^c\}$ $C_\rho(x)^c$

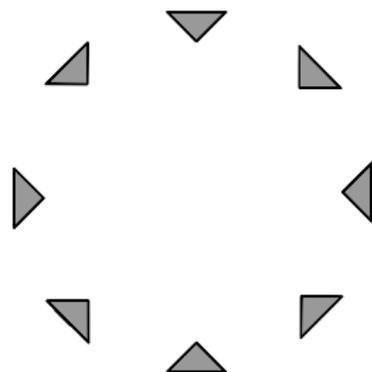


Figure: $\mathcal{C}_\rho(S) \subsetneq \tilde{\mathcal{C}}_\rho(S)$

ρ - Envoltente cono convexa.

Sea $S \subset \mathbb{R}^d$ acotado, definimos

- $\mathcal{C}_\rho(S) = \bigcap \{B: S \subset B \text{ y } B \in \mathcal{C}_\rho\}$ B
- $\tilde{\mathcal{C}}_\rho(S) = \bigcap \{C_\rho(x)^c: C_\rho(x) \subset S^c\}$ $C_\rho(x)^c$

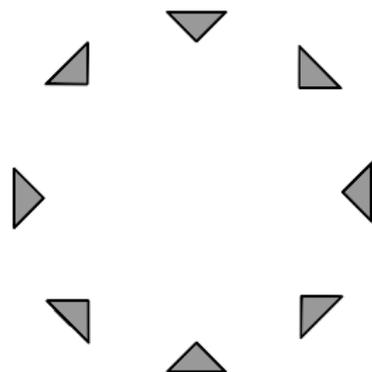


Figure: $\mathcal{C}_\rho(S) \subsetneq \tilde{\mathcal{C}}_\rho(S)$

Observación

- $\mathcal{C}_\rho(S)$ y $\tilde{\mathcal{C}}_\rho(S)$ son ρ -cono convexos.
- $S \subset \mathcal{C}_\rho(S) \subset \tilde{\mathcal{C}}_\rho(S) \subset \text{conv}(S)$.

P-uniformidad

Definición

Sea P una medida de probabilidad en $\mathcal{B}(\mathbb{R}^d)$, $A \in \mathcal{B}(\mathbb{R}^d)$ es P -continuo si $P(\partial A) = 0$.

Definición

$\mathcal{A} \subset \mathcal{B}(\mathbb{R}^d)$ es una clase P -uniforme si

$$\lim_{n \rightarrow \infty} \sup_{A \in \mathcal{A}} |P_n(A) - P(A)| = 0,$$

siempre que $P_n \rightarrow P$ débilmente .

Ejemplo: \mathcal{A} subconjuntos convexos de \mathbb{R}^d .

P-uniformidad, Billingsley Topsøe-1966

Teorema

Sea \mathcal{A} una clase P -continua, si $\partial\mathcal{A} = \{\partial A : A \in \mathcal{A}\}$ es un subconjunto compacto de (\mathcal{M}, d_H) el espacio métrico (completo, loc. compacto) de subconjuntos compactos, entonces \mathcal{A} es P -uniforme.

Teorema

\mathcal{C}_ρ es P -Uniforme, para toda $P \ll \mu$.

Aplicaciones: I) Estimación de conjuntos de nivel

$P \ll \mu$ con densidad f , queremos estimar $S(\lambda) = \{f \geq \lambda\}$.

$S(\lambda) = \operatorname{argmax}_{A \in \mathcal{B}(\mathbb{R}^d)} H_\lambda(A)$ siendo

$$H_\lambda(A) = P(A) - \lambda\mu(A).$$

Dada X_1, \dots, X_n i.i.d de P definimos $S_n(\lambda) = \operatorname{argmax}_{A \in \mathcal{B}(\mathbb{R}^d)} H_{n,\lambda}(A)$ siendo

$$H_{n,\lambda}(A) = \mathbb{P}_n(A) - \lambda\mu(A).$$

Aplicaciones: I) Estimación de conjuntos de nivel

$P \ll \mu$ con densidad f , queremos estimar $S(\lambda) = \{f \geq \lambda\}$.

$S(\lambda) = \operatorname{argmax}_{A \in \mathcal{B}(\mathbb{R}^d)} H_\lambda(A)$ siendo

$$H_\lambda(A) = P(A) - \lambda\mu(A).$$

Dada X_1, \dots, X_n i.i.d de P definimos $S_n(\lambda) = \operatorname{argmax}_{A \in \mathcal{B}(\mathbb{R}^d)} H_{n,\lambda}(A)$ siendo

$$H_{n,\lambda}(A) = \mathbb{P}_n(A) - \lambda\mu(A).$$

Teorema

Si $S \in \mathcal{C}_\rho$, entonces, tomando $S_n = \operatorname{argmax}_{A \in \mathcal{C}_\rho} H_{n,\lambda}(A)$

- $S_n(\lambda) \rightarrow S(\lambda)$ en d_H y d_μ ,
- $\partial S_n(\lambda) \rightarrow \partial S(\lambda)$ en d_H .

Aplicaciones: II) Conjunto de cambio de distribución

Sea $S \in \mathcal{C}_\rho$, $S \subset [0, 1]^d$, $Z \sim U([0, 1]^d)$ y $X \sim F$, si $Z \in S$, $X \sim G$ si $Z \in S^c$.
Queremos estimar S a partir de $\{(X_i, Z_i)\}_{i=1, \dots, n}$ i.i.d.

$$h_T(x) := \frac{1}{|T|} \sum_{i \in T} \mathbb{I}_{\{X_i \leq x\}}, \quad h_{T^c}(x) := \frac{1}{|T^c|} \sum_{i \in T^c} \mathbb{I}_{\{X_i \leq x\}}, \quad d_i^T := |h_T(X_i) - h_{T^c}(X_i)|$$

Aplicaciones: II) Conjunto de cambio de distribución

Sea $S \in \mathcal{C}_\rho$, $S \subset [0, 1]^d$, $Z \sim U([0, 1]^d)$ y $X \sim F$, si $Z \in S$, $X \sim G$ si $Z \in S^c$.
Queremos estimar S a partir de $\{(X_i, Z_i)\}_{i=1, \dots, n}$ i.i.d.

$$h_T(x) := \frac{1}{|T|} \sum_{i \in T} \mathbb{I}_{\{X_i \leq x\}}, \quad h_{T^c}(x) := \frac{1}{|T^c|} \sum_{i \in T^c} \mathbb{I}_{\{X_i \leq x\}}, \quad d_i^T := |h_T(X_i) - h_{T^c}(X_i)|$$

Vamos a maximizar, para $T \in \Gamma_n = \{\mathcal{C}_{\rho, h}(B) : B \subset \mathbb{R}^n\}$

$$D(T) := \frac{|T|}{n} \frac{|T^c|}{n} N(d_1^T, \dots, d_n^T),$$

$$N_{KS}(d_1, \dots, d_n) := \sup_{1 \leq i \leq n} d_i, \quad N_{Cv}(d_1, \dots, d_n) = (\sum_{i=1}^n d_i^2 / n)^{1/2}$$

Teorema

Si $S \in \mathcal{C}_\rho$, $S_n = \operatorname{argmax}_{T \in \Gamma_n} D(T)$ entonces $S_n \rightarrow S$ en medida.

- 1 Introducción.
- 2 Estimación de conjuntos ρ - cono convexos.
- 3 Otras tasas de convergencia

Convergencia en Hausdorff

Conjunto estándar

$S \subset \mathbb{R}^d$ es *estándar* respecto de una medida ν si existe $\lambda > 0$ y $\delta > 0$ tal que

$$\nu(B(x, \varepsilon) \cap S) \geq \delta \mu(B(x, \varepsilon)) \text{ para todo } x \in S \text{ and } 0 < \varepsilon \leq \lambda.$$

Convergencia en Hausdorff

Conjunto estándar

$S \subset \mathbb{R}^d$ es *estándar* respecto de una medida ν si existe $\lambda > 0$ y $\delta > 0$ tal que

$$\nu(B(x, \varepsilon) \cap S) \geq \delta \mu(B(x, \varepsilon)) \text{ para todo } x \in S \text{ and } 0 < \varepsilon \leq \lambda.$$

Teorema

$S \subset \mathbb{R}^d$, $S \in \mathcal{C}_\rho$, *estándar* respecto de la medida de Lebesgue, X_1, X_2, \dots, X_n i.i.d. v.a. con soporte S , entonces

$$d_H(\mathcal{C}_\rho(\mathbb{N}_n), S) = \mathcal{O}\left(\left(\frac{\log(n)}{n}\right)^{1/d}\right),$$

si además $\mu\{x : B(x, t) \subset S\} = \mathcal{O}(t)$

$$d_\mu(\mathcal{C}_\rho(\mathbb{N}_n), S) = \mathcal{O}\left(\left(\frac{\log(n)}{n}\right)^{1/d}\right).$$



- 1 Introducción.
- 2 Estimación de conjuntos ρ - cono convexos.
- 3 Otras tasas de convergencia**

Familias inevitables

Definición

$S \subset \mathbb{R}^d$ compacto satisface la condición de ρ , h -cono convexidad si

$$\forall x \in \partial S \exists \xi = \xi(x) \text{ tal que } C_{\rho, \xi, h}(x) \subset S^\circ.$$

Supongamos que $S = \tilde{\mathcal{C}}_{\rho, h}(S)$, queremos calcular

$$\mathcal{O}\left(E\left(d_\mu(S, \tilde{\mathcal{C}}_{\rho, h}(N_n))\right)\right).$$

Notación: Dado $a \in (0, \pi)$ y $b > 0$

$$\mathcal{G}_{a, b} = \{C_{a, \xi, b}(x) : x \in \mathbb{R}^d, \xi \in S_1^{d-1}\}.$$

Definición

Dado $x \in \mathbb{R}^d$, $\rho \in (0, \pi)$ y $h > 0$, denotemos $\Lambda_{\rho, h}(x) = \{C \in \mathcal{G}_{\rho, h} : x \in C\}$. La familia de conjuntos $\mathfrak{U}_{x, \rho, h}$ es inevitable para $\Lambda_{\rho, h}(x)$ si para todo $C \in \Lambda_{\rho, h}(x)$ existe $U \in \mathfrak{U}_{x, \rho, h}$ tal que $U \subset C$.



Familias inevitables II

$$\begin{aligned} \mathbb{E}\left(d_\mu(\mathcal{S}, \tilde{\mathcal{C}}_{\rho,h}(\mathbb{N}_n))\right) &= \mathbb{E} \int_{\mathcal{S}} \mathbb{I}_{\{x \in \mathcal{S} \setminus \tilde{\mathcal{C}}_{\rho,h}(\mathbb{N}_n)\}} dx \\ &= \int_{\mathcal{S}} \mathbb{P}(x \in \mathcal{S} \setminus \tilde{\mathcal{C}}_{\rho,h}(\mathbb{N}_n)) dx \end{aligned}$$

Supongamos que podemos construir para cada x familias inevitables $\mathfrak{U}_{x,\rho,h}$ con una cantidad finita de elementos, entonces:

$$\mathbb{P}(x \in \mathcal{S} \setminus \tilde{\mathcal{C}}_{\rho,h}(\mathbb{N}_n)) = \mathbb{P}(\exists C \in \Lambda_{\rho,h}(x) : C \cap \mathbb{N}_n = \emptyset) \leq \sum_{U \in \mathfrak{U}_{x,\rho,h}} \mathbb{P}(U \cap \mathbb{N}_n = \emptyset).$$

$$\int_{\mathcal{S}} \sum_{U \in \mathfrak{U}_{x,\rho,h}} \mathbb{P}(U \cap \mathbb{N}_n = \emptyset) dx = \int_{\mathcal{S}} \sum_{U \in \mathfrak{U}_{x,\rho,h}} \left(1 - \frac{\mu(U \cap \mathcal{S})}{\mu(\mathcal{S})}\right)^n dx$$

Familias inevitables II

$$\mathbb{E}\left(d_\mu(\mathcal{S}, \tilde{\mathcal{C}}_{\rho, h}(\mathbb{N}_n))\right) = \int_{\{x: d(x, \partial\mathcal{S}) \leq h_2\}} \sum_{U \in \mathfrak{U}_{x, \rho, h}} \left(1 - \frac{\mu(U \cap \mathcal{S})}{\mu(\mathcal{S})}\right)^n dx +$$

$$\int_{\{x: d(x, \partial\mathcal{S}) > h_2\}} \sum_{U \in \mathfrak{U}_{x, \rho, h}} \left(1 - \frac{\mu(U \cap \mathcal{S})}{\mu(\mathcal{S})}\right)^n dx.$$

Segundo sumando:

$\mu(U \cap \mathcal{S}) \geq \mu(U \cap B(x, h_2)) = ch_2^d$ entonces

$$\int_{\{x: d(x, \partial\mathcal{S}) > h_2\}} \sum_{U \in \mathfrak{U}_{x, \rho, h}} \left(1 - \frac{\mu(U \cap \mathcal{S})}{\mu(\mathcal{S})}\right)^n dx = k \left(1 - \frac{c}{\mu(\mathcal{S})} h_2^d\right)^n \mu(\{x : d(x, \partial\mathcal{S}) > h_2\}).$$

$$\int_{\{x: d(x, \partial\mathcal{S}) > h_2\}} \sum_{U \in \mathfrak{U}_{x, \rho, h}} \left(1 - \frac{\mu(U \cap \mathcal{S})}{\mu(\mathcal{S})}\right)^n dx \leq k_1 e^{-nc_1}.$$

Familias inevitables II

$$\mathbb{E} \left(d_{\mu}(\mathcal{S}, \tilde{\mathcal{C}}_{\rho, h}(\mathbb{N}_n)) \right) = \int_{\{x: d(x, \partial \mathcal{S}) \leq h_2\}} \sum_{U \in \mathcal{U}_{x, \rho, h}} \left(1 - \frac{\mu(U \cap \mathcal{S})}{\mu(\mathcal{S})} \right)^n dx + \\ \int_{\{x: d(x, \partial \mathcal{S}) > h_2\}} \sum_{U \in \mathcal{U}_{x, \rho, h}} \left(1 - \frac{\mu(U \cap \mathcal{S})}{\mu(\mathcal{S})} \right)^n dx.$$

Primer sumando

$U \cap B(x, t) = C_{\gamma/2, \xi_j, t}(x)$ por lo tanto si $d(x, \partial \mathcal{S}) = t$ entonces $\frac{\mu(U \cap \mathcal{S})}{\mu(\mathcal{S})} \geq \frac{\mu(U \cap B(x, t))}{\mu(\mathcal{S})} = \frac{c}{\mu(\mathcal{S})} t^d$.

$$\int_{\{x: d(x, \partial \mathcal{S}) \leq h_2\}} k \left(1 - \frac{c}{\mu(\mathcal{S})} d(x, \partial \mathcal{S})^d \right)^n dx \leq \int_{\{x: d(x, \partial \mathcal{S}) \leq h_2\}} k e^{-c_2 n d(x, \partial \mathcal{S})^d} dx \\ = n^{-1/d} \int_0^{c_2 n h_2^d} c_3 e^{-u} u^{d/(d-1)} du \leq n^{-1/d} \int_0^{+\infty} c_3 e^{-u} u^{d/(d-1)} du = \mathcal{O}(n^{-1/d}).$$

-  Cuevas, A., Fraiman, R. Pateiro-López, B. (2011). On statistical properties of sets fulfilling rolling-type conditions. *Applied probability trust*.
-  Cuevas, A. and Rodriguez-Casal, A. (2004). On boundary estimation. *Adv. in Appl. Probab.* 36 340-354.
-  Rodriguez-Casal, A. (2006). Set estimation under convexity type assumptions. *Annales de l'I.H.P.- Probabilites & Statistiques*, vol. 43, pp. 763-774.
-  Pateiro-López, B. and Rodríguez-Casal, A. (2008). "Length and surface area estimation under smoothness restrictions". *Advances in Applied Probability Vol. 40 (2)*, pp. 348-358