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

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

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

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

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

Distancia en Medida

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

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

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

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.

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

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

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 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

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

Restricciones de Forma

r-convexidad

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

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

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: $\mathcal{C}_\rho$ conjuntos ρ cono convexos.

Ejemplos

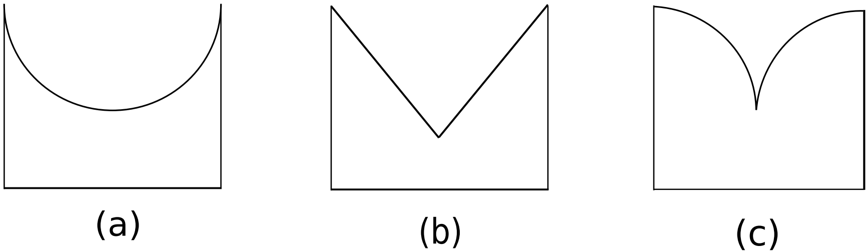

Figure: (a)r-convexo, (b)ρ-cono convexo; (c) No es ρ-c.c.

ρ - Envolvente cono convexa.

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

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

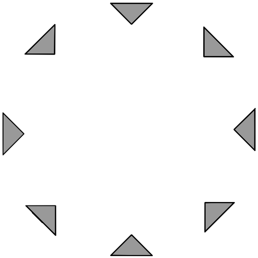

Figure: $\mathscr{C}_\rho(S) \subsetneq \tilde{\mathscr{C}}_\rho(S)$

ρ - Envolvente cono convexa.

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

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

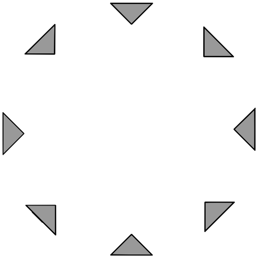

Figure: $\mathscr{C}_\rho(S) \subsetneq \tilde{\mathscr{C}}_\rho(S)$

Observación

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

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

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

siempre que $P_n \to P$ 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

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

Aplicaciones: I) Estimación de conjuntos de nivel

$P \ll \mu$ con densidad f, queremos estimar $S(\lambda) = \{f \geq \lambda\}$.
$S(\lambda) = argmax_{A \in \mathcal{B}(\mathbb{R}^d)} H_\lambda(A)$ siendo

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

Dada $X_1, \ldots, X_n$ i.i.d de P definimos $S_n(\lambda) = argmax_{A \in \mathcal{B}(\mathbb{R}^d)} H_{n,\lambda}(A)$ 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 $S(\lambda) = \{f \geq \lambda\}$.
$S(\lambda) = argmax_{A \in \mathcal{B}(\mathbb{R}^d)} H_\lambda(A)$ siendo

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

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

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

Teorema

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

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

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

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.

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

Vamos a maximizar, para $T \in \Gamma_n = \left\{\mathscr{C}_{\rho,h}(B) : B \subset \aleph_n\right\}$

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

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

Teorema

Si $S \in \mathcal{C}_\rho$, $S_n = argmax_{T\in\Gamma_n} D(T)$ entonces $S_n \to S$ 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

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

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

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

Teorema

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

$$d_H(\mathscr{C}_\rho(\aleph_n), S) = \mathcal{O}\Big(\Big(\frac{\log(n)}{n}\Big)^{1/d}\Big),$$

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

$$d_\mu(\mathscr{C}_\rho(\aleph_n), S) = \mathcal{O}\Big(\Big(\frac{\log(n)}{n}\Big)^{1/d}\Big).$$

Interpretación intuitiva

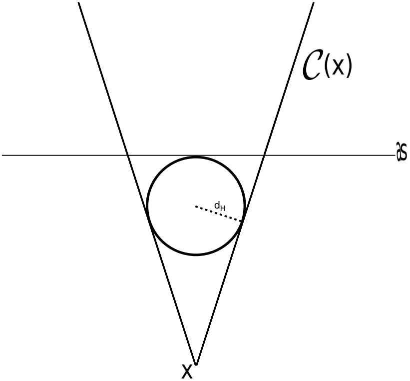

Figure: $x \in \partial\mathscr{C}_\rho(\aleph_n)$

Familias inevitables

Definición

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

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

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

$$\mathcal{O}\Big(E\Big(d_\mu\big(S, \tilde{\mathscr{C}}_{\rho,h}(\aleph_n)\big)\Big)\Big).$$

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

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

Definición

Dado $x \in \mathbb{R}^d$, $\rho \in (0, \pi)$ y $h > 0$, denotemos $\Lambda_{\rho,h}(x) = \{C \in \mathscr{G}_{\rho,h} : x \in C\}$.
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

$$\mathbb{E}\Big(d_\mu\big(S, \tilde{\mathscr{C}}_{\rho,h}(\aleph_n)\big)\Big) = \mathbb{E}\int_S \mathbb{I}_{\{x\in S\setminus \tilde{\mathscr{C}}_{\rho,h}(\aleph_n)\}}dx$$
$$= \int_S \mathbb{P}\big(x \in S \setminus \tilde{\mathscr{C}}_{\rho,h}(\aleph_n)\big)dx$$

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

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

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

Familias inevitables II

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

Segundo sumando:

$\mu(U \cap S) \geq \mu(U \cap B(x, h_2)) = c h_2^d$ entonces

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

$$\int_{\{x: d(x,\partial S) > h_2\}} \sum_{U \in \mathfrak{U}_{x,\rho,h}} \left(1 - \frac{\mu(U \cap S)}{\mu(S)}\right)^n dx \leq k_1 e^{-n c_1}.$$

Familias inevitables II

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

Primer sumando

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

$$\int_{\{x:d(x,\partial S)\leq h_2\}} k\left(1 - \frac{c}{\mu(S)}d(x,\partial S)^d\right)^n dx \leq \int_{\{x:d(x,\partial S)\leq h_2\}} ke^{-c_2 nd(x,\partial S)^d} dx$$

$$= n^{-1/d}\int_0^{c_2 nh_2^d} c_3 e^{-u}u^{d/(d-1)}du \leq n^{-1/d}\int_0^{+\infty} c_3 e^{-u}u^{d/(d-1)}du = \mathcal{O}\left(n^{-1/d}\right).$$

- Cuevas, A., Fraiman, R. Pateiro-López, B. (2011). On statistical properties of sets fullfilling 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*