Los alcances y sus estimaciones

Alejandro Cholaquidis

CMAT-Facultad de Ciencias, UdelaR
Montevideo Uruguay

Trabajos en conjunto con Antonio Cuevas, Leonardo Moreno y Ricardo Fraiman

Seminario de Probabilidad y Estadística

1. El alcance clásico
   - Estimación de d_S

2. Alcance polinomial

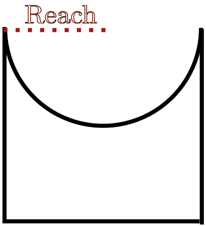
Alcance

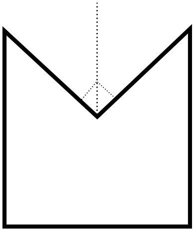
Alcance (de Federer) de un conjunto

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

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

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





Alcance

Algunas conexiones con otras restricciones de forma

- $\text{Alcance}(S) = r > 0 \Rightarrow S$ es r-convexo. El recíproco no es cierto en general, ver Cuevas, Fraiman and Pateiro-López (2012).

Alcance

Algunas conexiones con otras restricciones de forma

- $\text{Alcance}(S) = r > 0 \Rightarrow S$ es r-convexo. El recíproco no es cierto en general, ver Cuevas, Fraiman and Pateiro-López (2012).
- Si ∂S es una variedad $\mathcal{C}^2$ con borde vacío o $\mathcal{C}^2 \Rightarrow \text{Alcance}(S) > 0$, Thäle (2008).

Alcance

Algunas conexiones con otras restricciones de forma

- Alcance$(S) = r > 0 \Rightarrow S$ es r-convexo. El recíproco no es cierto en general, ver Cuevas, Fraiman and Pateiro-López (2012).
- Si ∂S es una variedad $\mathcal{C}^2$ con borde vacío o $\mathcal{C}^2 \Rightarrow$ Alcance$(S) > 0$, Thäle (2008).
- Alcance(S) es cota superior del inverso de la curvatura de una geodésica parametrizada por longitud de arco, Niyogi et al. (2008).

Alcance

Algunas conexiones con otras restricciones de forma

- $\text{Alcance}(S) = r > 0 \Rightarrow S$ es r-convexo. El recíproco no es cierto en general, ver Cuevas, Fraiman and Pateiro-López (2012).
- Si ∂S es una variedad $\mathcal{C}^2$ con borde vacío o $\mathcal{C}^2 \Rightarrow \text{Alcance}(S) > 0$, Thäle (2008).
- $\text{Alcance}(S)$ es cota superior del inverso de la curvatura de una geodésica parametrizada por longitud de arco, Niyogi et al. (2008).

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

$$\mu_d(B(S,\varepsilon)) = \sum_{k=0}^{d} \Phi_k(S)\varepsilon^k, \qquad 0 \leq \varepsilon \leq r,$$

donde μ_d es la medida de Lebesgue d-dimensional.

- $\Phi_0(S) = \mu_d(S)$

Alcance

Algunas conexiones con otras restricciones de forma

- Alcance$(S) = r > 0 \Rightarrow S$ es r-convexo. El recíproco no es cierto en general, ver Cuevas, Fraiman and Pateiro-López (2012).
- Si ∂S es una variedad $\mathcal{C}^2$ con borde vacío o $\mathcal{C}^2 \Rightarrow$ Alcance$(S) > 0$, Thäle (2008).
- Alcance(S) es cota superior del inverso de la curvatura de una geodésica parametrizada por longitud de arco, Niyogi et al. (2008).

Federer, 1959. Si Alcance$(S) = r$

$$\mu_d(B(S,\varepsilon)) = \sum_{k=0}^{d} \Phi_k(S)\varepsilon^k, \qquad 0 \leq \varepsilon \leq r,$$

donde μ_d es la medida de Lebesgue d-dimensional.

- $\Phi_0(S) = \mu_d(S)$
- $\Phi_1(S) = \mu_{d-1}(\partial S)$

Alcance

Algunas conexiones con otras restricciones de forma

- Alcance$(S) = r > 0 \Rightarrow S$ es r-convexo. El recíproco no es cierto en general, ver Cuevas, Fraiman and Pateiro-López (2012).
- Si ∂S es una variedad $\mathcal{C}^2$ con borde vacío o $\mathcal{C}^2 \Rightarrow$ Alcance$(S) > 0$, Thäle (2008).
- Alcance(S) es cota superior del inverso de la curvatura de una geodésica parametrizada por longitud de arco, Niyogi et al. (2008).

Federer, 1959. Si Alcance$(S) = r$

$$\mu_d(B(S,\varepsilon)) = \sum_{k=0}^{d} \Phi_k(S)\varepsilon^k, \qquad 0 \leq \varepsilon \leq r,$$

donde μ_d es la medida de Lebesgue d-dimensional.

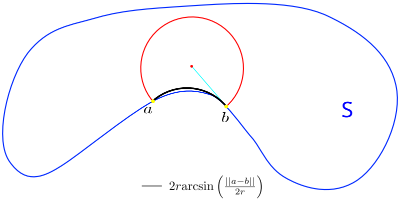
- $\Phi_0(S) = \mu_d(S)$
- $\Phi_1(S) = \mu_{d-1}(\partial S)$
- $\Phi_d(S) = \omega_d \mathcal{X}(S)$, donde $\omega_d = \mu_d(B(0,1))$

ver Federer (1956, 1969)

Alcance, una definición equivalente

El Teorema 1 en Boissonnat et al (2019) establece que cuando S es cerrado, alcance(S) es igual a

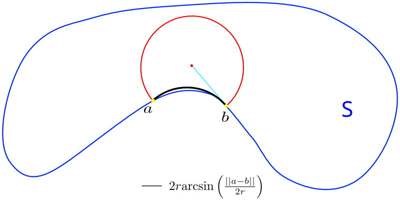
$$\sup \left\{ r > 0,\ \forall a, b \in S, \|a-b\| < 2r \Rightarrow d_S(a,b) \leq 2r\arcsin\left(\frac{\|a-b\|}{2r}\right) \right\}. \quad (1)$$



Alcance, una definición equivalente

El Teorema 1 en Boissonnat et al (2019) establece que cuando S es cerrado, alcance(S) es igual a

$$\sup\left\{ r > 0,\ \forall a, b \in S, \|a-b\| < 2r \Rightarrow d_S(a,b) \leq 2r\arcsin\left(\frac{\|a-b\|}{2r}\right)\right\}. \qquad (1)$$



Si tenemos $\mathcal{X}_n = X_1, \dots, X_n$, (1) sugiere estimar alcance(S) mediante

$$\widehat{\text{alcance}}(S) = \sup\left\{ r > 0,\ \forall X_i, X_j \in \mathcal{X}_n, \|X_i - X_j\| < 2r \Rightarrow \widehat{d_S}(X_i, X_j) \leq 2r\arcsin\left(\frac{\|X_i - X_j\|}{2r}\right)\right\}.$$

donde $\widehat{d_S}$ es una estimación de d_S.

Estimación de d_S

Sea $\mathcal{X}_n = \{X_1, \ldots, X_n\} \subset S$ un conjunto finito y $G_n = (\mathcal{X}_n, E_n)$ un grafo con vértices en $\mathcal{X}_n$ y aristas E_n, definimos

$$d_{G_n}(x, y) = \min_P \sum_{i=1}^{p-1} ||X_{j_{i+1}} - X_{j_i}||,$$

donde $P = (X_{j_1}, \ldots, X_{j_p}) \subset \mathcal{X}_n$ varía sobre todos los caminos en G_n que conectan $x = X_{j_1}$ con $y = X_{j_p}$.

Estimación de d_S

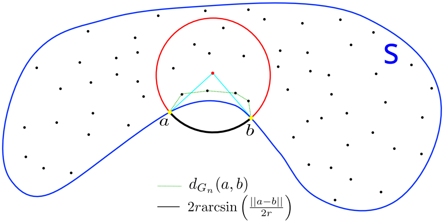
Sea $\mathcal{X}_n = \{X_1, \ldots, X_n\} \subset S$ un conjunto finito y $G_n = (\mathcal{X}_n, E_n)$ un grafo con vértices en $\mathcal{X}_n$ y aristas E_n, definimos

$$d_{G_n}(x, y) = \min_{P} \sum_{i=1}^{p-1} ||X_{j_{i+1}} - X_{j_i}||,$$

donde $P = (X_{j_1}, \ldots, X_{j_p}) \subset \mathcal{X}_n$ varía sobre todos los caminos en G_n que conectan $x = X_{j_1}$ con $y = X_{j_p}$.

Definición

Sea $\{\varepsilon_n\}_n$ tal que $\varepsilon_n > 0$ para todo n, y $G_n = (\mathcal{X}_n, E_n)$ tal que $(X_i, X_j) \in E_n$ si y sólo si $||X_i - X_j|| \leq \varepsilon_n$.



Estimación de d_S

Teorema

Sea $S \subset \mathbb{R}^d$

- compacto
- geodesicamente convexo,
- $r_0 := \text{alcance}(S) > 0$.

Sea $\varepsilon_n \to 0$ tal que $\varepsilon_n > 0$ para todo n y $\tau_n := d_H(\mathcal{X}_n, S)/\varepsilon_n \to 0$.

Estimación de d_S

Teorema

Sea $S \subset \mathbb{R}^d$

- compacto
- geodesicamente convexo,
- $r_0 := \text{alcance}(S) > 0$.

Sea $\varepsilon_n \to 0$ tal que $\varepsilon_n > 0$ para todo n y $\tau_n := d_H(\mathcal{X}_n, S)/\varepsilon_n \to 0$.
Entonces, para todo $x, y \in S$,

$$\left(1 - \frac{1}{24}\left(\frac{\pi\varepsilon_n}{2r_0}\right)^2\right) d_S(x,y) \leq d_{G_n}(x,y) \leq (1 + 4\tau_n) d_S(x,y),$$

donde $d_{G_n}(x,y)$ **es la distancia en el grafo construido previamente**, pero incluyendo x e y como vértices, y n es lo suficientemente grande como para garantizar que $\varepsilon_n < 2r_0$.

The estimator

Sea ε_n tal que $\varepsilon_n \to 0$ y $\delta_n/\varepsilon_n^3 \to 0$, donde $\delta_n = d_H(\mathcal{X}_n, S)$.

$$\hat{r}_n = \sup\Big\{ r > 0,\ \forall X_i \neq X_j \in \mathcal{X}_n, \|X_i - X_j\| < 2r \Rightarrow d_{G_n}(X_i, X_j) \leq 2r(1+\varepsilon_n^2)\arcsin\left(\frac{\|X_i - X_j\|}{2r}\right)\Big\}.$$

The estimator

Sea ε_n tal que $\varepsilon_n \to 0$ y $\delta_n/\varepsilon_n^3 \to 0$, donde $\delta_n = d_H(\mathcal{X}_n, S)$.

$$\hat{r}_n = \sup\Big\{r > 0,\ \forall X_i \neq X_j \in \mathcal{X}_n, \|X_i - X_j\| < 2r \Rightarrow d_{G_n}(X_i, X_j) \leq 2r(1+\varepsilon_n^2)\arcsin\Big(\frac{\|X_i - X_j\|}{2r}\Big)\Big\}.$$

Theorem

Sea $S \subset \mathbb{R}^d$ compacto, geodésicamente convexo, y $r_0 = reach(S) > 0$, para todo n suficientemente grande, tenemos que

$$r_0 \leq \hat{r}_n \leq \frac{r_0}{1-\varepsilon_n}. \tag{2}$$

The estimator

Sea ε_n tal que $\varepsilon_n \to 0$ y $\delta_n/\varepsilon_n^3 \to 0$, donde $\delta_n = d_H(\mathcal{X}_n, S)$.

$$\hat{r}_n = \sup\Big\{r > 0,\ \forall X_i \neq X_j \in \mathcal{X}_n, \|X_i - X_j\| < 2r \Rightarrow d_{G_n}(X_i, X_j) \leq 2r(1+\varepsilon_n^2)\arcsin\Big(\frac{\|X_i - X_j\|}{2r}\Big)\Big\}.$$

Theorem

Sea $S \subset \mathbb{R}^d$ compacto, geodésicamente convexo, y $r_0 = reach(S) > 0$, para todo n suficientemente grande, tenemos que

$$r_0 \leq \hat{r}_n \leq \frac{r_0}{1-\varepsilon_n}. \tag{2}$$

Si $\mathcal{X}_n = \{X_1, \dots, X_n\}$ es iid de P_X con soporte S y P_X es estandard [a]*, entonces con probabilidad 1 para n suf. grande*

$$r_0 \leq \hat{r}_n \leq r_0 + c\left(\frac{\log(n)}{n}\right)^{\frac{1}{3d}}\beta_n, \tag{3}$$

donde $\varepsilon_n = c(\log(n)/n)^{\frac{1}{3d}}\beta_n$, $c > (2/(\eta\omega_d))^{1/d}$, η es la constante de estandaridad y $\beta_n \to +\infty$.

[a] existe $\eta > 0$ y $\varepsilon > 0$, tal que para todo $r < \varepsilon$ $P_X(B(x,r)) > \eta\mu(B(x,r) \cap S)$ para todo $x \in S$

Alcance polinomial

Recordemos que

Federer, 1959. si Alcance$(S) = r$

$$\mu_d(B(S,\varepsilon)) = \sum_{k=0}^{d} \Phi_k(S)\varepsilon^k, \qquad 0 \le \varepsilon \le r,$$

Alcance polinomial

Recordemos que

Federer, 1959. si Alcance$(S) = r$

$$\mu_d(B(S,\varepsilon)) = \sum_{k=0}^{d} \Phi_k(S)\varepsilon^k, \qquad 0 \le \varepsilon \le r,$$

Algunas observaciones

- Para $d = 2$, $\mu_d(B(S,\varepsilon))$ es polinomial **para todo** $\varepsilon \ge 0$ si y sólo si S es convexo. No es cierto en $d > 2$.

Alcance polinomial

Recordemos que

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

$$\mu_d(B(S,\varepsilon)) = \sum_{k=0}^{d} \Phi_k(S)\varepsilon^k, \qquad 0 \le \varepsilon \le r,$$

Algunas observaciones

- Para $d = 2$, $\mu_d(B(S,\varepsilon))$ es polinomial **para todo** $\varepsilon \ge 0$ si y sólo si S es convexo. No es cierto en $d > 2$.
- Existen conjuntos S tal que $\mu_d(B(S,\varepsilon))$ es polinomial para todo $r > 0$ pero $\text{reach}(S) = 0$. Ejemplo dos bolas tangentes en $\mathbb{R}^3$.

Alcance polinomial

Recordemos que

Federer, 1959. si Alcance$(S) = r$

$$\mu_d(B(S,\varepsilon)) = \sum_{k=0}^{d} \Phi_k(S)\varepsilon^k, \qquad 0 \le \varepsilon \le r,$$

Algunas observaciones

- Para $d = 2$, $\mu_d(B(S,\varepsilon))$ es polinomial **para todo** $\varepsilon \ge 0$ si y sólo si S es convexo. No es cierto en $d > 2$.
- Existen conjuntos S tal que $\mu_d(B(S,\varepsilon))$ es polinomial para todo $r > 0$ pero reach$(S) = 0$. Ejemplo dos bolas tangentes en $\mathbb{R}^3$.
- La cono convexidad no garantiza volumen polinomial.

Alcance polinomial

Recordemos que

Federer, 1959. si Alcance$(S) = r$

$$\mu_d(B(S,\varepsilon)) = \sum_{k=0}^{d} \Phi_k(S)\varepsilon^k, \qquad 0 \leq \varepsilon \leq r,$$

Algunas observaciones

- Para $d = 2$, $\mu_d(B(S,\varepsilon))$ es polinomial **para todo** $\varepsilon \geq 0$ si y sólo si S es convexo. No es cierto en $d > 2$.
- Existen conjuntos S tal que $\mu_d(B(S,\varepsilon))$ es polinomial para todo $r > 0$ pero reach$(S) = 0$. Ejemplo dos bolas tangentes en $\mathbb{R}^3$.
- La cono convexidad no garantiza volumen polinomial.
- Hay conjuntos cuya función de volumen no es polinomial.

Alcance polinomial

Recordemos que

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

$$\mu_d(B(S,\varepsilon)) = \sum_{k=0}^{d} \Phi_k(S)\varepsilon^k, \qquad 0 \leq \varepsilon \leq r,$$

Algunas observaciones

- Para $d = 2$, $\mu_d(B(S,\varepsilon))$ es polinomial **para todo** $\varepsilon \geq 0$ si y sólo si S es convexo. No es cierto en $d > 2$.
- Existen conjuntos S tal que $\mu_d(B(S,\varepsilon))$ es polinomial para todo $r > 0$ pero $\text{reach}(S) = 0$. Ejemplo dos bolas tangentes en $\mathbb{R}^3$.
- La cono convexidad no garantiza volumen polinomial.
- Hay conjuntos cuya función de volumen no es polinomial.
- Si $\mu(B(S,\varepsilon))$ es polinomial $\Phi_0(S) = \mu(S)$, $\Phi_1(S) = \mu_{d-1}(\partial S)$, ¿qué son los otros coeficientes?

Alcance polinomial

Recordemos que

Federer, 1959. si Alcance$(S) = r$

$$\mu_d(B(S,\varepsilon)) = \sum_{k=0}^{d} \Phi_k(S)\varepsilon^k, \qquad 0 \le \varepsilon \le r,$$

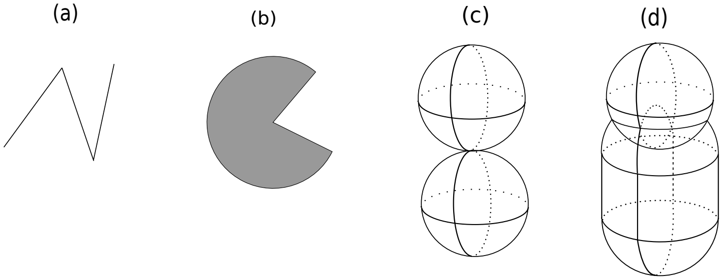
Algunas observaciones

- Para $d = 2$, $\mu_d(B(S,\varepsilon))$ es polinomial **para todo** $\varepsilon \ge 0$ si y sólo si S es convexo. No es cierto en $d > 2$.
- Existen conjuntos S tal que $\mu_d(B(S,\varepsilon))$ es polinomial para todo $r > 0$ pero reach$(S) = 0$. Ejemplo dos bolas tangentes en $\mathbb{R}^3$.
- La cono convexidad no garantiza volumen polinomial.
- Hay conjuntos cuya función de volumen no es polinomial.
- Si $\mu(B(S,\varepsilon))$ es polinomial $\Phi_0(S) = \mu(S)$, $\Phi_1(S) = \mu_{d-1}(\partial S)$, ¿qué son los otros coeficientes?
- **Problema:** estimar el mayor r tal que $\mu_d(B(S,\varepsilon))$ es un polinomio de grado a lo sumo d para todo $\varepsilon \in [0,r]$ (**polynomial reach**)

Ejemplos

Sea $S \in \mathbb{R}^d$ y $V(\epsilon) := \mu_d(B(S,\epsilon))$, definimos el alcance polinomial de S, denotado, **R**, como

$$\mathbf{R} = \sup\{r \geq 0 : \; V \text{ es un polinomio de grado a lo sumo } d \text{ en } [0,r]\}. \tag{4}$$

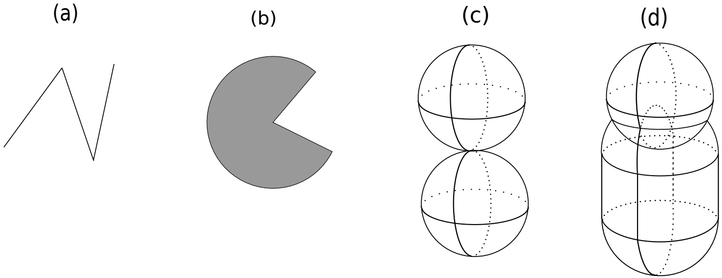


- (a) (b) (c) (d) tienen alcance polinomial > 0.

Ejemplos

Sea $S \in \mathbb{R}^d$ y $V(\epsilon) := \mu_d(B(S,\epsilon))$, definimos el alcance polinomial de S, denotado, $\mathbf{R}$, como

$$\mathbf{R} = \sup\{r \geq 0 : \ V \text{ es un polinomio de grado a lo sumo } d \ \text{ en } \ [0,r]\}. \tag{4}$$

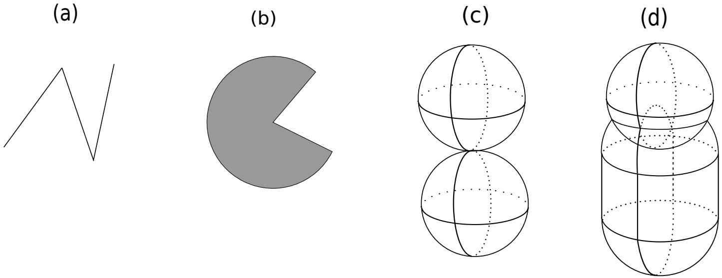


- (a) (b) (c) (d) tienen alcance polinomial > 0.
- Los conjuntos (c) y (d) tienen alcance polinomial para todo $r > 0$.

Ejemplos

Sea $S \in \mathbb{R}^d$ y $V(\epsilon) := \mu_d(B(S,\epsilon))$, definimos el alcance polinomial de S, denotado, $\mathbf{R}$, como

$$\mathbf{R} = \sup\{r \geq 0 : \ V \text{ es un polinomio de grado a lo sumo } d \text{ en } [0,r]\}. \tag{4}$$



- (a) (b) (c) (d) tienen alcance polinomial > 0.
- Los conjuntos (c) y (d) tienen alcance polinomial para todo $r > 0$.

Algunas ideas

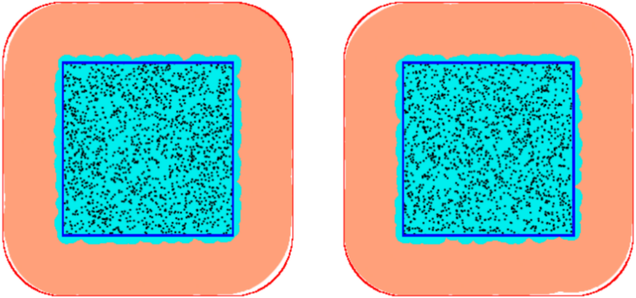
¿Cómo estimar V cuando $\mathbf{R}$ es conocido?

Si $\mathcal{X}_n = \{X_1, \ldots, X_n\} \subset S$ y $V_n(\epsilon) := \mu_d(B(\mathcal{X}_n, \epsilon))$ en Cuevas and Pateiro-López (2018) se prueba que

$$\sup_{x:0<a\leq x\leq b<\infty} |V(x) - V_n(x)| \to 0 \quad a.s.$$

Siempre vale que $V_n \leq V$ y se puede probar que si $I \subset (0, +\infty)$

$$||V_n - V||_{L^2(I)} = \mathcal{O}(d_H(\aleph_n, S)) = \mathcal{O}(\log(n)/n)^{1/d})$$



Algunas ideas

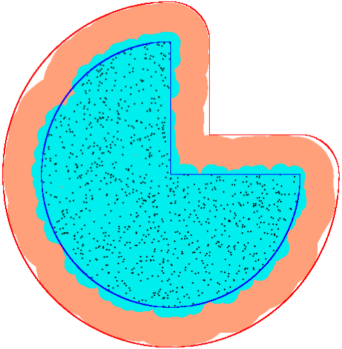
¿Cómo estimar V cuando $\mathbf{R}$ es conocido?

Si $\mathcal{X}_n = \{X_1, \ldots, X_n\} \subset S$ y $V_n(\epsilon) := \mu_d(B(\mathcal{X}_n, \epsilon))$ en Cuevas and Pateiro-López (2018) se prueba que

$$\sup_{x:0<a\leq x\leq b<\infty} |V(x) - V_n(x)| \to 0 \quad a.s.$$

Siempre vale que $V_n \leq V$ y se puede probar que si $I \subset (0, +\infty)$

$$||V_n - V||_{L^2(I)} = \mathcal{O}(d_H(\aleph_n, S)) = \mathcal{O}(\log(n)/n)^{1/d})$$



Algunas ideas

$$\sup_{x:0<a\leq x\leq b<\infty} |V(x) - V_n(x)| \to 0 \quad a.s.$$

sugiere que cuando $\mathbf{R}$ es conocido, se puede estimar V por

$$P^I_{n,d} = \text{argmin}_{\pi\in\Pi_d(I)} ||V_n - \pi||_{L^2(I)}$$

donde $\Pi_d(I)$ denota el espacio (cerrado) de polinomios de grado a lo sumo d, $I \subset (0, \mathbf{R}]$.

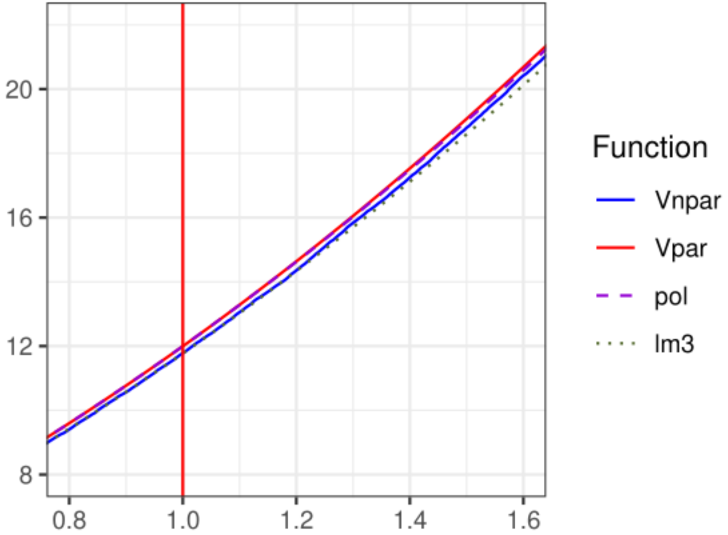
Siempre vale que si $I \subset (0, \mathbf{R}]$

$$||P^I_{n,d} - V_n||_{L^2(I)} \leq ||V - V_n||_{L^2(I)} = \mathcal{O}(\log(n)/n)^{1/d})$$

ya que $P_{n,d}$ es la mejor aproximación a V_n y V es otro polinomio de grado d. Por lo tanto $P_{n,d} \to V$ cuando $n \to \infty$.

Observar que es suficiente tener una cota inferior de $\mathbf{R}$ para estimar el polinomio V

No se ve nada...



Un algoritmo para obtener una cota inferior

Dado un intervalo cerrado $I \subset [0, \infty)$, denotamos $\Pi_D(I)$ al subespacio de todos los polinomios de grado a lo sumo $D \in \mathbb{N}$ en $L^2(I)$.

$$P^I_{n,D} = \text{argmin}_{\pi \in \Pi_D(I)} \|V_n - \pi\|_{L^2(I)},$$

Inputs del algoritmo:

- $\aleph_n = \{X_1 \ldots, X_n\}$

Un algoritmo para obtener una cota inferior

Dado un intervalo cerrado $I \subset [0, \infty)$, denotamos $\Pi_D(I)$ al subespacio de todos los polinomios de grado a lo sumo $D \in \mathbb{N}$ en $L^2(I)$.

$$P^I_{n,D} = \text{argmin}_{\pi \in \Pi_D(I)} \|V_n - \pi\|_{L^2(I)},$$

Inputs del algoritmo:

- $\aleph_n = \{X_1 \ldots, X_n\}$
- Una grilla $0 < r_1 < r_2 < \cdots < r_K$ donde asumimos que $\mathbf{R} < r_K$.
- $U_n = \left(\frac{\log n}{n}\right)^{\frac{1}{2d} - \eta}$ con η cualquier valor > 0

Un algoritmo para obtener una cota inferior

Dado un intervalo cerrado $I \subset [0, \infty)$, denotamos $\Pi_D(I)$ al subespacio de todos los polinomios de grado a lo sumo $D \in \mathbb{N}$ en $L^2(I)$.

$$P^I_{n,D} = \text{argmin}_{\pi \in \Pi_D(I)} \|V_n - \pi\|_{L^2(I)},$$

Inputs del algoritmo:

- $\aleph_n = \{X_1 \ldots, X_n\}$
- Una grilla $0 < r_1 < r_2 < \cdots < r_K$ donde asumimos que $\mathbf{R} < r_K$.
- $U_n = \left(\frac{\log n}{n}\right)^{\frac{1}{2d} - \eta}$ con η cualquier valor > 0
- $\ell \geq d$, $\ell \in \mathbb{N}$, para las aproximaciones polinomiales.

Un algoritmo para obtener una cota inferior

Dado un intervalo cerrado $I \subset [0, \infty)$, denotamos $\Pi_D(I)$ al subespacio de todos los polinomios de grado a lo sumo $D \in \mathbb{N}$ en $L^2(I)$.

$$P^I_{n,D} = \operatorname{argmin}_{\pi \in \Pi_D(I)} \|V_n - \pi\|_{L^2(I)},$$

Inputs del algoritmo:

- $\aleph_n = \{X_1 \ldots, X_n\}$
- Una grilla $0 < r_1 < r_2 < \cdots < r_K$ donde asumimos que $\mathbf{R} < r_K$.
- $U_n = \left(\frac{\log n}{n}\right)^{\frac{1}{2d} - \eta}$ con η cualquier valor > 0
- $\ell \geq d$, $\ell \in \mathbb{N}$, para las aproximaciones polinomiales.
- Definimos $I_i = [0, r_i]$ y $J_i = [r_i, r_K]$ para todo $i = 1, \ldots, K - 1$

Un algoritmo para obtener una cota inferior

Dado un intervalo cerrado $I \subset [0, \infty)$, denotamos $\Pi_D(I)$ al subespacio de todos los polinomios de grado a lo sumo $D \in \mathbb{N}$ en $L^2(I)$.

$$P^I_{n,D} = \text{argmin}_{\pi \in \Pi_D(I)} \|V_n - \pi\|_{L^2(I)},$$

Inputs del algoritmo:

- $\aleph_n = \{X_1 \ldots, X_n\}$
- Una grilla $0 < r_1 < r_2 < \cdots < r_K$ donde asumimos que $\mathbf{R} < r_K$.
- $U_n = \left(\frac{\log n}{n}\right)^{\frac{1}{2d} - \eta}$ con η cualquier valor > 0
- $\ell \geq d$, $\ell \in \mathbb{N}$, para las aproximaciones polinomiales.
- Definimos $I_i = [0, r_i]$ y $J_i = [r_i, r_K]$ para todo $i = 1, \ldots, K - 1$

Paso 0. Si $\|V_n - P^{I_1}_{n,d}\|_{L^2(I_1)} > U_n$ entonces el output es $\hat{\mathbf{R}} = 0$ y **el algoritmo termina**.

Un algoritmo para obtener una cota inferior

Dado un intervalo cerrado $I \subset [0,\infty)$, denotamos $\Pi_D(I)$ al subespacio de todos los polinomios de grado a lo sumo $D \in \mathbb{N}$ en $L^2(I)$.

$$P^I_{n,D} = \text{argmin}_{\pi \in \Pi_D(I)} \|V_n - \pi\|_{L^2(I)},$$

Inputs del algoritmo:

- $\aleph_n = \{X_1 \ldots, X_n\}$
- Una grilla $0 < r_1 < r_2 < \cdots < r_K$ donde asumimos que $\mathbf{R} < r_K$.
- $U_n = \left(\frac{\log n}{n}\right)^{\frac{1}{2d}-\eta}$ con η cualquier valor > 0
- $\ell \geq d$, $\ell \in \mathbb{N}$, para las aproximaciones polinomiales.
- Definimos $I_i = [0, r_i]$ y $J_i = [r_i, r_K]$ para todo $i = 1, \ldots, K-1$

Paso 0. Si $\|V_n - P^{I_1}_{n,d}\|_{L^2(I_1)} > U_n$ entonces el output es $\hat{\mathbf{R}} = 0$ y **el algoritmo termina**.

Paso 1. Si el algoritmo no terminó en el paso 0, definimos para $i = 1, \ldots, K-1$

$$c_i = \frac{\|V_n - P^{I_i}_{n,d}\|_{L^2(I_i)}}{\|V_n - P^{J_i}_{n,\ell}\|_{L^2(J_i)}}.$$

Un algoritmo para obtener una cota inferior

Dado un intervalo cerrado $I \subset [0, \infty)$, denotamos $\Pi_D(I)$ al subespacio de todos los polinomios de grado a lo sumo $D \in \mathbb{N}$ en $L^2(I)$.

$$P^I_{n,D} = \text{argmin}_{\pi \in \Pi_D(I)} \|V_n - \pi\|_{L^2(I)},$$

Inputs del algoritmo:

- $\aleph_n = \{X_1 \ldots, X_n\}$
- Una grilla $0 < r_1 < r_2 < \cdots < r_K$ donde asumimos que $\mathbf{R} < r_K$.
- $U_n = \left(\frac{\log n}{n}\right)^{\frac{1}{2d} - \eta}$ con η cualquier valor > 0
- $\ell \geq d$, $\ell \in \mathbb{N}$, para las aproximaciones polinomiales.
- Definimos $I_i = [0, r_i]$ y $J_i = [r_i, r_K]$ para todo $i = 1, \ldots, K-1$

Paso 0. Si $\|V_n - P^{I_1}_{n,d}\|_{L^2(I_1)} > U_n$ entonces el output es $\hat{\mathbf{R}} = 0$ y **el algoritmo termina**.

Paso 1. Si el algoritmo no terminó en el paso 0, definimos para $i = 1, \ldots, K-1$

$$c_i = \frac{\|V_n - P^{I_i}_{n,d}\|_{L^2(I_i)}}{\|V_n - P^{J_i}_{n,\ell}\|_{L^2(J_i)}}.$$

Output Sea i el primer índice tal que $c_i > 1$. El output es $\hat{\mathbf{R}} := r_{i-1}$.

Teorema de sub-estimación

Teorema

Sea $\mathbf{R}$ el alcance polinomial de $S \subset \mathbb{R}^d$. Suponemos $0 \leq \mathbf{R} < r_K$. Consideremos el algoritmo anterior basado en una muestra iid de n datos con soporte S. Asumimos que

- S es estandard
- $L_0(\partial S) < \infty$ donde L_0 es el contenido de Minkowski de ∂S.

Existe un grado ℓ a partir del cual :

1. el output $\hat{\mathbf{R}}$ es un sub estimador de $\mathbf{R}$: con prob. 1 para n suficientemente grande $\hat{\mathbf{R}} \leq \mathbf{R}$.
2. Con probabilidad uno, $\hat{\mathbf{R}}$ es constante para n suficientemente grande.

Simulaciones para el pacman

	$n = 2000$			$n = 3000$			$n = 4000$		
	gr 1	gr 2	gr 3	gr 1	gr 2	gr 3	gr 1	gr 2	gr 3
Mean	0.6212	0.7692	0.7196	0.606	0.7431	0.7104	0.6024	0.738	0.704
Var	0.0083	0.0225	0.0077	0.0027	0.0224	0.0040	0.0009	0.0223	0.0015
	0	2	50	0	0	26	0	0	10
Mean	0.6096	0.7398	0.7096	0.6024	0.7149	0.7044	0.6016	0.6954	0.7024
Var	0.0043	0.0227	0.0050	0.0016	0.0212	0.0023	0.0006	0.0195	0.0009
	0	0	24	0	0	13	0	0	6

Cuadro: Pacman: media y varianza **en 1000 replicas** de $\hat{\mathbf{R}}$, para 3 grillas y $\ell = 8$ en filas 1,2,3. Filas 4,5,6 $\ell = 10$. Filas 3 y 6 cantidad de veces que el algoritmo da un valor mayor que $\mathbf{R}$.

Aamari, Eddie and Levrard, Clément. (2019). Nonasymptotic rates for manifold, tangent space and curvature estimation. *The Annals of Statistics*, **47**, (1), 177–204.

Aamari, Eddie and Kim, Jisu and Chazal, Frédéric and Michel, Bertrand and Rinaldo, Alessandro and Wasserman, Larry. (2019). Estimating the reach of a manifold. *Electron. J. Statist.* , **13**, (5), 1359–1399.

Berenfeld, Clément and Harvey, John and Hoffmann, Marc and Shankar, Krishnan. (2022). Estimating the reach of a manifold via its convexity defect function. *Discrete & Computational Geometry*, **67**, (2), 403–438,

Boissonnat, J. D., Lieutier, A., and Wintraecken, M. (2019). The reach, metric distortion, geodesic convexity and the variation of tangent spaces. *Journal of Applied and Computational Topology*, 3(1), 29–58.

Berrendero, J.R., Cholaquidis, A., Cuevas, A. and Fraiman, R. (2014). A geometrically motivated parametric model in manifold estimation. *Statistics* **48** 983–1004.

Niyogi, P., Smale, S., and Weinberger, S.(2008) Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry,* 39, 419–441.

Colesanti, A., and Manselli, P. (2010). Geometric and isoperimetric properties of sets of positive reach in E^d. *Atti Semin. Mat. Fis. Univ. Modena Reggio Emilia* **57** 97–113.

Cuevas, A., Fraiman, R., and Pateiro-López, B. (2012). On statistical properties of sets fulfilling rolling-type conditions. *Adv. in Appl. Probab.* **44** 311–329.

Cuevas, A. and Pateiro-López, B. (2018). Polynomial volume estimation and its applications. *Journal of Statistical Planning and Inference* **196** 174–184.

Federer, H. (1956). Curvature measures. *Transactions of the American Mathematical Society* **93**(3) 418–491.

Federer, H. (1969). *Geometric Measure Theory* Springer-Verlag, Berlin.

Niyogi, P., Smale, S., and Weinberger, S. (2008). Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.* **39** 419?441.

Pateiro-López, B. (Unpublished). Set estimation under convexity type restrictions. Phd. Thesis. Universidade de Santiago de Compostela. `http://eio.usc.es/pub/pateiro/files/thesis_beatrizpateirolopez.pdf`

Thäle, C. (2008). 50 years sets with positive reach. A survey. *Surveys in Mathematics and its Applications* 3, 123–165.

Walther, G. (1999). On a generalization of Blaschke's rolling theorem and the smoothing of surfaces, *Math. Meth. Appl. Sci.* **22** 301–316.