Estimación de funcionales de un conjunto.

Alejandro Cholaquidis

CMAT-Facultad de Ciencias, UdelaR
Montevideo Uruguay

Trabajo en conjunto con: Catherine Aaron y Ricardo Fraiman

Coloquio informal de estudiantes - Octubre de 2022

1 Introducción: estimación de conjuntos

2 Estimación de longitudes y superficies

3 Fórmula de Crofton

1 Introducción: estimación de conjuntos

2 Estimación de longitudes y superficies

3 Fórmula de Crofton

Estimación de conjuntos, restricciones de forma

Objetivo general

Dada una muestra $\aleph_n = \{X_1, \ldots, X_n\}$ de X, cuyo suporte es S, estimar S, o ∂S, o $\mu_{d-1}(\partial S)$, o $dim(S)$ si S es una variedad, o determinar si $int(S) = \emptyset$....ETC...

Inicios

Encontrar $\hat{S}_n$ construido a partir de $\aleph_n$ iid de X, con suporte S, que bajo ciertas restricciones de forma en S esté cerca de S

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

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

Estimación de conjuntos, restricciones de forma

Objetivo general

Dada una muestra $\aleph_n = \{X_1, \ldots, X_n\}$ de X, cuyo suporte es S, estimar S, o ∂S, o $\mu_{d-1}(\partial S)$, o $dim(S)$ si S es una variedad, o determinar si $int(S) = \emptyset$....ETC...

Inicios

Encontrar $\hat{S}_n$ construido a partir de $\aleph_n$ iid de X, con suporte S, que bajo ciertas restricciones de forma en S esté cerca de S

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 $\mathcal{B}(\mathbb{R}^d)$,

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

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

Restricciones de forma

Un segundo paso es obtener tasas de convergencia para los estimadores, para eso hay que imponer restricciones de forma en S.

Convexidad

Si S es convexo, el estimador natural es $\mathcal{H}(\aleph_n)$, el cierre convexo de la muestra.

Un poco mas general: r-convexidad

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

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

Si S es r-convexo, el estimador natural es $C_r(\aleph_n)$, ¿y si no conocemos r?...

Alcance Positivo

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

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

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

1 Introducción: estimación de conjuntos

2 Estimación de longitudes y superficies

3 Fórmula de Crofton

Estimación de longitudes y superficies

Objetivo

Dado $S \subset \mathbb{R}^d$, estimar $L_0 = \mu_{d-1}(\partial S)$ a partir de $X_1, \ldots, X_n$ iid de X.

Estimación de longitudes y superficies

Objetivo

Dado $S \subset \mathbb{R}^d$, estimar $L_0 = \mu_{d-1}(\partial S)$ a partir de $X_1, \ldots, X_n$ iid de X.

Modelo de puntos afuera y adentro de S:

Si $X = (Z, Y)$

con $Y = \begin{cases} 1 & \text{si } Z \in S \\ 0 & \text{si } Z \in S^c \end{cases}$

y $\aleph_n = \{X_1, \ldots, X_n\}$ iid de X.

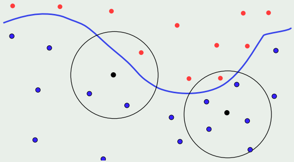


Figura: En rojo los puntos de S^c y en azul los de S.

Estimación de longitudes y superficies

Objetivo

Dado $S \subset \mathbb{R}^d$, estimar $L_0 = \mu_{d-1}(\partial S)$ a partir de $X_1, \ldots, X_n$ iid de X.

Modelo de puntos afuera y adentro de S:

Si $X = (Z, Y)$

con $Y = \begin{cases} 1 & \text{si } Z \in S \\ 0 & \text{si } Z \in S^c \end{cases}$

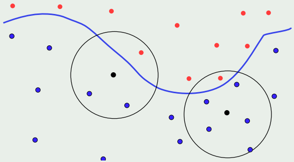


Figura: En rojo los puntos de S^c y en azul los de S.

y $\aleph_n = \{X_1, \ldots, X_n\}$ iid de X.En Cuevas, Fraiman, and Rodríguez-Casal (2007) se propone

$$L_n = \frac{\mu_d\Big\{x : \#(B(x,\varepsilon_n) \cap \mathcal{Z}^0) \geq 1 \quad \text{y} \quad \#(B(x,\varepsilon_n) \cap \mathcal{Z}^1) \geq 1\Big\}}{2\varepsilon_n},$$

donde $\mathcal{Z}^1 = \{X \in \aleph_n : X = (Z,1)\}$ y $\mathcal{Z}^0 = \{X \in \aleph_n : X = (Z,0)\}$.

Estimación de longitudes y superficies

Objetivo

Dado $S \subset \mathbb{R}^d$, estimar $L_0 = \mu_{d-1}(\partial S)$ a partir de $X_1, \dots, X_n$ iid de X.

Modelo de puntos afuera y adentro de S:

Si $X = (Z, Y)$

con $Y = \begin{cases} 1 & \text{si } Z \in S \\ 0 & \text{si } Z \in S^c \end{cases}$

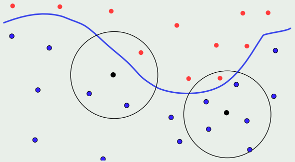


Figura: En rojo los puntos de S^c y en azul los de S.

y $\aleph_n = \{X_1, \dots, X_n\}$ iid de X.En Cuevas, Fraiman, and Rodríguez-Casal (2007) se propone

$$L_n = \frac{\mu_d\left\{x : \#(B(x,\varepsilon_n) \cap \mathcal{Z}^0) \geq 1 \quad \text{y} \quad \#(B(x,\varepsilon_n) \cap \mathcal{Z}^1) \geq 1\right\}}{2\varepsilon_n},$$

donde $\mathcal{Z}^1 = \{X \in \aleph_n : X = (Z,1)\}$ $\quad y \quad$ $\mathcal{Z}^0 = \{X \in \aleph_n : X = (Z,0)\}$.

Se obtiene $\mathbb{E}|L_n - L_0| = O(n^{-1/(2d)})$, si ∂S es C^2.

Estimación de longitudes y superficies

Objetivo

Dado $S \subset \mathbb{R}^d$, estimar $L_0 = \mu_{d-1}(\partial S)$ a partir de $X_1, \ldots, X_n$ iid de X.

Estimación de longitudes y superficies

Objetivo

Dado $S \subset \mathbb{R}^d$, estimar $L_0 = \mu_{d-1}(\partial S)$ a partir de $X_1, \ldots, X_n$ iid de X.

Modelo de puntos afuera y adentro de S:

Con el mismo esquema muestral de antes, en Pateiro-López and Rodríguez-Casal (2008) se propone el estimador

$$L_n = \frac{\mu_d\Big(B(C_r(\mathcal{Z}^0), \varepsilon_n) \cap B(C_r(\mathcal{Z}^1), \varepsilon_n)\Big)}{2\varepsilon_n},$$

donde

$$\mathcal{Z}^1 = \{X \in \aleph_n : X = (Z, 1)\} \quad y \quad \mathcal{Z}^0 = \{X \in \aleph_n : X = (Z, 0)\}.$$

Estimación de longitudes y superficies

Objetivo

Dado $S \subset \mathbb{R}^d$, estimar $L_0 = \mu_{d-1}(\partial S)$ a partir de $X_1, \ldots, X_n$ iid de X.

Modelo de puntos afuera y adentro de S:

Con el mismo esquema muestral de antes, en Pateiro-López and Rodríguez-Casal (2008) se propone el estimador

$$L_n = \frac{\mu_d\Big(B(C_r(\mathcal{Z}^0), \varepsilon_n) \cap B(C_r(\mathcal{Z}^1), \varepsilon_n)\Big)}{2\varepsilon_n},$$

donde

$$\mathcal{Z}^1 = \{X \in \aleph_n : X = (Z,1)\} \quad y \quad \mathcal{Z}^0 = \{X \in \aleph_n : X = (Z,0)\}.$$

Tasa

Se obtiene $\mathbb{E}|L_n - L_0| = O(n^{-1/(d+1)})$, si ∂S es C^2.

Estimación de longitudes y superficies

Objetivo

Dado $S \subset \mathbb{R}^d$, estimar $L_0 = \mu_{d-1}(\partial S)$ a partir de $X_1, \ldots, X_n$ iid de X.

Únicamente con puntos dentro de S, y volumen polinomial

Supongamos existe $R > 0$ tal que

$$\mu_d(B(S,\epsilon)) = \mu_d(S) + L_0\epsilon + a_2\epsilon^2 + \cdots + a_{d+1}\epsilon^d \quad \text{para todo } \epsilon \in [0, R].$$

Estimación de longitudes y superficies

Objetivo

Dado $S \subset \mathbb{R}^d$, estimar $L_0 = \mu_{d-1}(\partial S)$ a partir de $X_1, \ldots, X_n$ iid de X.

Únicamente con puntos dentro de S, y volumen polinomial

Supongamos existe $R > 0$ tal que

$$\mu_d(B(S,\epsilon)) = \mu_d(S) + L_0\epsilon + a_2\epsilon^2 + \cdots + a_{d+1}\epsilon^d \quad \text{para todo } \epsilon \in [0,R].$$

En Cuevas and Pateiro-López (2018) se propone un estimador de los coeficientes que consiste en estimar $B(S,\epsilon)$ por $B(\aleph_n,\epsilon)$ y "ajustar un polinomio" a $B(\aleph_n,\epsilon)$ de grado d en $[0,R]$, asumiendo R conocido.

Estimación de longitudes y superficies

Objetivo

Dado $S \subset \mathbb{R}^d$, estimar $L_0 = \mu_{d-1}(\partial S)$ a partir de $X_1, \ldots, X_n$ iid de X.

Únicamente con puntos dentro de S, y volumen polinomial

Supongamos existe $R > 0$ tal que

$$\mu_d(B(S,\epsilon)) = \mu_d(S) + L_0\epsilon + a_2\epsilon^2 + \cdots + a_{d+1}\epsilon^d \quad \text{para todo } \epsilon \in [0,R].$$

En Cuevas and Pateiro-López (2018) se propone un estimador de los coeficientes que consiste en estimar $B(S,\epsilon)$ por $B(\aleph_n,\epsilon)$ y "ajustar un polinomio" a $B(\aleph_n,\epsilon)$ de grado d en $[0,R]$, asumiendo R conocido. No se dan tasas de convergencia. Es el único estimador consistente de

L_0 existente en la literatura, excluyendo el caso S convexo, para este esquema muestral.

1 Introducción: estimación de conjuntos

2 Estimación de longitudes y superficies

3 Fórmula de Crofton

Fórmula de Crofton

En Crofton (1868) se prueba que dado un subconjunto convexo del plano, cuyo borde es γ, la longitud de su borde, $|\gamma|_1$, es

$$|\gamma|_1 = \frac{1}{2}\int_{\theta=0}^{\pi}\int_{p=-\infty}^{+\infty} n_\gamma(\theta,p)dpd\theta, \tag{1}$$

donde $n_\gamma(\theta,p)$ es el número de intersecciones de γ con la recta

$$r_{\theta^*,\theta p}$$

donde $\theta^* \in (\mathcal{S}^+)^1$ es ortogonal a θ, y $dpd\theta$ es la medida de Lebesgue en $\mathbb{R}^2$

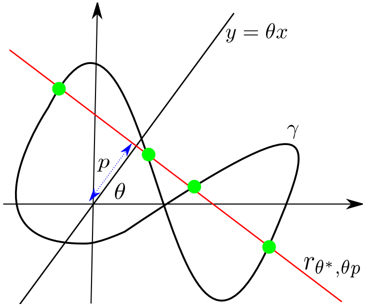


Figura: n_γ cuenta las intersecciones de γ con la recta $r_{\theta^*,\theta p}$ determinada por θ y p.

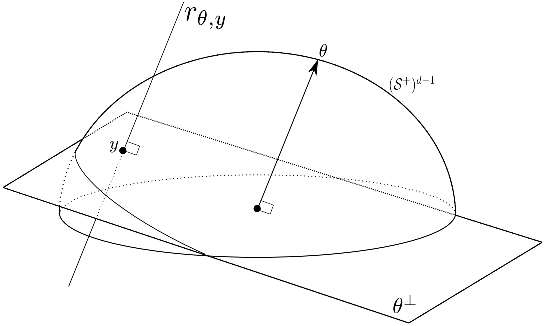
Fórmula de Crofton

En general, si M es una variedad en $\mathbb{R}^d$, compacta de dimensión $d-1$,

$$|M|_{d-1} = \frac{1}{\beta(d)} \int_{\theta \in (\mathcal{S}^+)^{d-1}} \int_{y \in \theta^\perp} n_M(\theta, y) d\mu_{d-1}(y) d\theta. \tag{2}$$

donde

$$\beta(d) = \Gamma(d/2)\Gamma((d+1)/2)^{-1}\pi^{-1/2}.$$



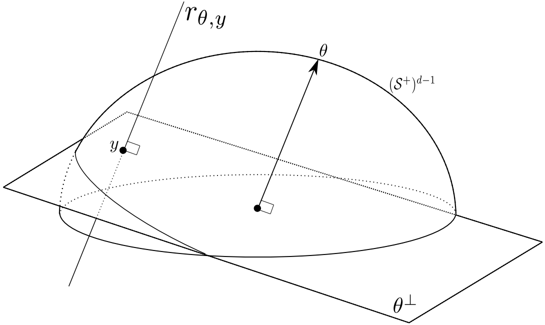
Fórmula de Crofton

En general, si M es una variedad en $\mathbb{R}^d$, compacta de dimensión $d-1$,

$$|M|_{d-1} = \frac{1}{\beta(d)} \int_{\theta \in (\mathcal{S}^+)^{d-1}} \int_{y \in \theta^\perp} n_M(\theta, y) d\mu_{d-1}(y) d\theta. \qquad (2)$$

donde

$$\beta(d) = \Gamma(d/2)\Gamma((d+1)/2)^{-1}\pi^{-1/2}.$$



Estimación

Supongamos que $M = \partial S$, y se tiene $X_1, \ldots, X_n$ iid en S, la fórmula (2) sugiere cambiar M por un estimador $\hat{\partial S}$ y aproximar la integral “tirando rectas al azar”.... ¿qué estimador tomar?

¿Qué estimador tomar?

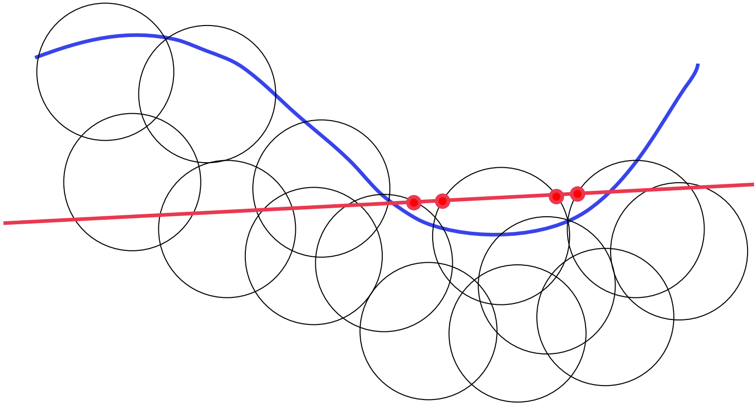
Queremos un estimador $\hat{S}_n$ de S de modo que $n_{\partial S}(\theta, y)$ se parezca, para n "grande", a $n_{\partial \hat{S}_n}(\theta, y)$.

¿Qué estimador tomar?

Queremos un estimador $\hat{S}_n$ de S de modo que $n_{\partial S}(\theta, y)$ se parezca, para n "grande", a $n_{\partial \hat{S}_n}(\theta, y)$. Si tomamos el estimador de Devroye-Wise

$$\hat{S}_{\varepsilon_n}(\aleph_n) = \cup_{i=1}^n \mathcal{B}(X_i, \varepsilon_n) = B(\aleph_n, \epsilon_n)$$

puede pasar que $n_{\partial \hat{S}_n}(\theta, n)$ sobre estime mucho $n_M(\theta, y)$.

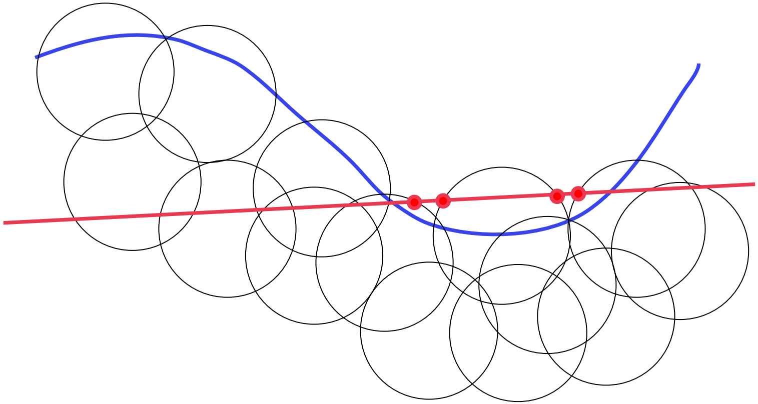


¿Qué estimador tomar?

Queremos un estimador $\hat{S}_n$ de S de modo que $n_{\partial S}(\theta, y)$ se parezca, para n "grande", a $n_{\partial \hat{S}_n}(\theta, y)$. Si tomamos el estimador de Devroye-Wise

$$\hat{S}_{\varepsilon_n}(\aleph_n) = \cup_{i=1}^n \mathcal{B}(X_i, \varepsilon_n) = B(\aleph_n, \epsilon_n)$$

puede pasar que $n_{\partial \hat{S}_n}(\theta, n)$ sobre estime mucho $n_M(\theta, y)$.



¿y si tomamos $C_r(\aleph_n)$?...

Primera propuesta: Devroye-Wise tuneado

Supongamos que $\mathcal{X}$ es un subconjunto de S (no necesariamente finito). Sea $\varepsilon > 0$, denotamos

$$\hat{S}_\varepsilon(\mathcal{X}) = B(\mathcal{X}, \varepsilon).$$

[1] con medida finita $d\theta dp$ son una cantidad finita porque $|\partial S|_{d-1} < \infty$.

Primera propuesta: Devroye-Wise tuneado

Supongamos que $\mathcal{X}$ es un subconjunto de S (no necesariamente finito). Sea $\varepsilon > 0$, denotamos

$$\hat{S}_{\varepsilon}(\mathcal{X}) = B(\mathcal{X}, \varepsilon).$$

Consideramos $r_{\theta,y} = y + \mathbb{R}\theta$. Vamos a definir $\hat{n}_{\varepsilon,\mathcal{X}}(\theta, y)$.

[1] con medida finita $d\theta dp$ son una cantidad finita porque $|\partial S|_{d-1} < \infty$.

Primera propuesta: Devroye-Wise tuneado

Supongamos que $\mathcal{X}$ es un subconjunto de S (no necesariamente finito). Sea $\varepsilon > 0$, denotamos

$$\hat{S}_\varepsilon(\mathcal{X}) = B(\mathcal{X}, \varepsilon).$$

Consideramos $r_{\theta,y} = y + \mathbb{R}\theta$. Vamos a definir $\hat{n}_{\varepsilon,\mathcal{X}}(\theta, y)$. Si $\hat{S}_\varepsilon(\mathcal{X}) \cap r_{\theta,y} = \emptyset$, **definimos**

$\hat{n}_{\varepsilon,\mathcal{X}}(\theta, y) = 0$.

[1] con medida finita $d\theta dp$ son una cantidad finita porque $|\partial S|_{d-1} < \infty$.

Primera propuesta: Devroye-Wise tuneado

Supongamos que $\mathcal{X}$ es un subconjunto de S (no necesariamente finito). Sea $\varepsilon > 0$, denotamos

$$\hat{S}_\varepsilon(\mathcal{X}) = B(\mathcal{X}, \varepsilon).$$

Consideramos $r_{\theta,y} = y + \mathbb{R}\theta$. Vamos a definir $\hat{n}_{\varepsilon,\mathcal{X}}(\theta, y)$. Si $\hat{S}_\varepsilon(\mathcal{X}) \cap r_{\theta,y} = \emptyset$, **definimos**

$\hat{n}_{\varepsilon,\mathcal{X}}(\theta, y) = 0.$

En caso contrario

- Sean $I_1, \ldots, I_m$ las componentes conexas de $\hat{S}_\varepsilon(\mathcal{X}) \cap r_{\theta,y}$. Las ordenamos de modo que $I_i = (a_i, b_i)$, con $-\infty \leq a_1 < b_1 < \cdots < a_m < b_m \leq +\infty$.[1]

[1] con medida finita $d\theta dp$ son una cantidad finita porque $|\partial S|_{d-1} < \infty$.

Primera propuesta: Devroye-Wise tuneado

Supongamos que $\mathcal{X}$ es un subconjunto de S (no necesariamente finito). Sea $\varepsilon > 0$, denotamos

$$\hat{S}_\varepsilon(\mathcal{X}) = B(\mathcal{X}, \varepsilon).$$

Consideramos $r_{\theta,y} = y + \mathbb{R}\theta$. Vamos a definir $\hat{n}_{\varepsilon,\mathcal{X}}(\theta, y)$. Si $\hat{S}_\varepsilon(\mathcal{X}) \cap r_{\theta,y} = \emptyset$, **definimos**

$\hat{n}_{\varepsilon,\mathcal{X}}(\theta, y) = 0.$

En caso contrario

- Sean $I_1, \ldots, I_m$ las componentes conexas de $\hat{S}_\varepsilon(\mathcal{X}) \cap r_{\theta,y}$. Las ordenamos de modo que $I_i = (a_i, b_i)$, con $-\infty \leq a_1 < b_1 < \cdots < a_m < b_m \leq +\infty$.[1]
- Si existen $I_i, I_{i+1}, \ldots, I_{i+\ell}$, tal que para todo $a_i < \lambda < b_{i+\ell}$ si $t = y + \lambda\theta \in r_{\theta,y}$ entonces $d(t, \mathcal{X}) \leq 4\varepsilon$, definimos $A_i = (a_i, b_{i+\ell})$:"pegamos intervalos que distan menos de 4ε de $\mathcal{X}$".

[1] con medida finita $d\theta dp$ son una cantidad finita porque $|\partial S|_{d-1} < \infty$.

Primera propuesta: Devroye-Wise tuneado

Supongamos que $\mathcal{X}$ es un subconjunto de S (no necesariamente finito). Sea $\varepsilon > 0$, denotamos

$$\hat{S}_\varepsilon(\mathcal{X}) = B(\mathcal{X}, \varepsilon).$$

Consideramos $r_{\theta,y} = y + \mathbb{R}\theta$. Vamos a definir $\hat{n}_{\varepsilon,\mathcal{X}}(\theta, y)$. Si $\hat{S}_\varepsilon(\mathcal{X}) \cap r_{\theta,y} = \emptyset$, **definimos**

$\hat{n}_{\varepsilon,\mathcal{X}}(\theta, y) = 0.$

En caso contrario

- Sean $I_1, \ldots, I_m$ las componentes conexas de $\hat{S}_\varepsilon(\mathcal{X}) \cap r_{\theta,y}$. Las ordenamos de modo que $I_i = (a_i, b_i)$, con $-\infty \leq a_1 < b_1 < \cdots < a_m < b_m \leq +\infty$.[1]
- Si existen $I_i, I_{i+1}, \ldots, I_{i+\ell}$, tal que para todo $a_i < \lambda < b_{i+\ell}$ si $t = y + \lambda\theta \in r_{\theta,y}$ entonces $d(t, \mathcal{X}) \leq 4\varepsilon$, definimos $A_i = (a_i, b_{i+\ell})$:"pegamos intervalos que distan menos de 4ε de $\mathcal{X}$".
- Sea j el número de intervalos disjuntos que se obtienen al final: $A_1, \ldots, A_j$, **definimos** $\hat{n}_{\varepsilon,\mathcal{X}}(\theta, y) = 2j.$

[1] con medida finita $d\theta dp$ son una cantidad finita porque $|\partial S|_{d-1} < \infty$.

Restricción geométrica y teorema

H1

Vamos a suponer ∂S es C^2 y existe N_S tal que para todo $\theta \in (\mathcal{S}^+)^{d-1}$ y $y \in \theta^\perp$, $n_{\partial S}(\theta, y) \leq N_S$ ctp $d\mu_{d-1} d\theta$. Además $S = \overline{int(S)}$.

Restricción geométrica y teorema

H1

Vamos a suponer ∂S es C^2 y existe N_S tal que para todo $\theta \in (\mathcal{S}^+)^{d-1}$ y $y \in \theta^\perp$, $n_{\partial S}(\theta, y) \leq N_S$ ctp $d\mu_{d-1} d\theta$. Además $S = \overline{int(S)}$.

Consideremos

$$\hat{I}_{d-1}(\mathcal{X}, \varepsilon) = \frac{1}{\beta(d)} \int_{\theta \in (\mathcal{S}^+)^{d-1}} \int_{y \in \theta^\perp} \hat{n}_{\varepsilon, \mathcal{X}}(\theta, y) d\mu_{d-1}(y) d\theta. \tag{3}$$

Restricción geométrica y teorema

H1

Vamos a suponer ∂S es C^2 y existe N_S tal que para todo $\theta \in (\mathcal{S}^+)^{d-1}$ y $y \in \theta^\perp$, $n_{\partial S}(\theta, y) \leq N_S$ ctp $d\mu_{d-1} d\theta$. Además $S = \overline{int(S)}$.

Consideremos

$$\hat{I}_{d-1}(\mathcal{X}, \varepsilon) = \frac{1}{\beta(d)} \int_{\theta \in (\mathcal{S}^+)^{d-1}} \int_{y \in \theta^\perp} \hat{n}_{\varepsilon, \mathcal{X}}(\theta, y) d\mu_{d-1}(y) d\theta. \tag{3}$$

Teorema

Si $\mathcal{X} = \aleph_n = \{X_1, \ldots, X_n\}$ iid de X con distribución P_X con soporte S, con densidad, supongamos que se cumple **H1**. Si tomamos $\varepsilon_n = 2 \max_i \min_j ||X_i - X_j||$, con probabilidad 1, para n suf. grande

$$\hat{I}_{d-1}(\aleph_n, \varepsilon_n) = |\partial S|_{d-1} + \mathcal{O}\left(\left(\frac{\ln(n)}{n}\right)^{\frac{1}{d}}\right).$$

Restricción geométrica y teorema

H1

Vamos a suponer ∂S es C^2 y existe N_S tal que para todo $\theta \in (\mathcal{S}^+)^{d-1}$ y $y \in \theta^\perp$, $n_{\partial S}(\theta, y) \leq N_S$ ctp $d\mu_{d-1}d\theta$. Además $S = \overline{int(S)}$.

Consideremos

$$\hat{I}_{d-1}(\mathcal{X}, \varepsilon) = \frac{1}{\beta(d)} \int_{\theta \in (\mathcal{S}^+)^{d-1}} \int_{y \in \theta^\perp} \hat{n}_{\varepsilon,\mathcal{X}}(\theta, y) d\mu_{d-1}(y) d\theta. \tag{3}$$

Teorema

Si $\mathcal{X} = \aleph_n = \{X_1, \ldots, X_n\}$ iid de X con distribución P_X con soporte S, con densidad, supongamos que se cumple **H1**. Si tomamos $\varepsilon_n = 2\,\text{máx}_i\,\text{mín}_j\,||X_i - X_j||$, con probabilidad 1, para n suf. grande

$$\hat{I}_{d-1}(\aleph_n, \varepsilon_n) = |\partial S|_{d-1} + \mathcal{O}\left(\left(\frac{\ln(n)}{n}\right)^{\frac{1}{d}}\right).$$

Como no tenemos una fórmula cerrada para $\hat{n}_{\varepsilon,\mathcal{X}}(\theta, y)$ usamos Montecarlo para estimar (3)

Montecarlo

Generamos $\theta_1, \ldots, \theta_k$ uniformes en $(\mathcal{S}^+)^{d-1}$.

Montecarlo

Generamos $\theta_1, \dots, \theta_k$ uniformes en $(\mathcal{S}^+)^{d-1}$. Para cada $i = 1, \dots, k$, generamos

$\aleph_i = \{y_1^i, \dots, y_\ell^i\}$ uniformes en $[-L, L]^{d-1} \subset \theta_i^\perp$, independientes de $\theta_1, \dots, \theta_k$, donde

$$L = \max_{j=1,\dots,n} ||X_j||.$$

Montecarlo

Generamos $\theta_1, \ldots, \theta_k$ uniformes en $(\mathcal{S}^+)^{d-1}$. Para cada $i = 1, \ldots, k$, generamos

$\aleph_i = \{y_1^i, \ldots, y_\ell^i\}$ uniformes en $[-L, L]^{d-1} \subset \theta_i^\perp$, independientes de $\theta_1, \ldots, \theta_k$, donde

$$L = \max_{j=1,\ldots,n} ||X_j||.$$

El estimador por Montecarlo de $\hat{I}_{d-1}(\mathcal{X}, \varepsilon_n)$ es

$$\frac{(2L)^{d-1}}{\beta(d)} \frac{1}{\ell k} \sum_{i=1}^{k} \sum_{j=1}^{\ell} \hat{n}_{\varepsilon_n, \mathcal{X}}(\theta_i, y_j^i)$$

Una conjetura sobre el estimador de Devroye-Wise

Recordemos que

$$\hat{S}_{\varepsilon_n}(\aleph_n) = \cup_{i=1}^{n} \mathcal{B}(X_i, \varepsilon_n) = B(\aleph_n, \epsilon_n)$$

Una conjetura sobre el estimador de Devroye-Wise

Recordemos que

$$\hat{S}_{\varepsilon_n}(\aleph_n) = \cup_{i=1}^{n} \mathcal{B}(X_i, \varepsilon_n) = B(\aleph_n, \epsilon_n)$$

Estimación de ∂S

En Cuevas and Rodriguez-Casal (2004) se prueba que $d_H(\partial \hat{S}_{\varepsilon_n}, \partial S) \to 0$ c.s, si $\varepsilon_n = d_H(\aleph_n, S)$. Pero, como vimos, $|\partial \hat{S}_{\varepsilon_n}|_{d-1}$ sobre estima $|\partial S|_{d-1}$.

Una conjetura sobre el estimador de Devroye-Wise

Recordemos que

$$\hat{S}_{\varepsilon_n}(\aleph_n) = \cup_{i=1}^{n} \mathcal{B}(X_i, \varepsilon_n) = B(\aleph_n, \epsilon_n)$$

Estimación de ∂S

En Cuevas and Rodriguez-Casal (2004) se prueba que $d_H(\partial \hat{S}_{\varepsilon_n}, \partial S) \to 0$ c.s, si $\varepsilon_n = d_H(\aleph_n, S)$. Pero, como vimos, $|\partial \hat{S}_{\varepsilon_n}|_{d-1}$ sobre estima $|\partial S|_{d-1}$.

Conjetura

Si tomamos $\varepsilon_n \gg d_H(\aleph_n, S)$,

- ¿ se cumple que $|\partial \hat{S}_{\varepsilon_n}|_{d-1} \to |\partial S|_{d-1}$?

Una conjetura sobre el estimador de Devroye-Wise

Recordemos que

$$\hat{S}_{\varepsilon_n}(\aleph_n) = \cup_{i=1}^n \mathcal{B}(X_i, \varepsilon_n) = B(\aleph_n, \epsilon_n)$$

Estimación de ∂S

En Cuevas and Rodriguez-Casal (2004) se prueba que $d_H(\partial \hat{S}_{\varepsilon_n}, \partial S) \to 0$ c.s, si $\varepsilon_n = d_H(\aleph_n, S)$. Pero, como vimos, $|\partial \hat{S}_{\varepsilon_n}|_{d-1}$ sobre estima $|\partial S|_{d-1}$.

Conjetura

Si tomamos $\varepsilon_n \gg d_H(\aleph_n, S)$,

- ¿ se cumple que $|\partial \hat{S}_{\varepsilon_n}|_{d-1} \to |\partial S|_{d-1}$?
- ¿cómo tomar ϵ_n?, y ¿que tasa de convergencia se obtiene?

Una conjetura sobre el estimador de Devroye-Wise

Recordemos que

$$\hat{S}_{\varepsilon_n}(\aleph_n) = \cup_{i=1}^{n} \mathcal{B}(X_i, \varepsilon_n) = B(\aleph_n, \epsilon_n)$$

Estimación de ∂S

En Cuevas and Rodriguez-Casal (2004) se prueba que $d_H(\partial \hat{S}_{\varepsilon_n}, \partial S) \to 0$ c.s, si $\varepsilon_n = d_H(\aleph_n, S)$. Pero, como vimos, $|\partial \hat{S}_{\varepsilon_n}|_{d-1}$ sobre estima $|\partial S|_{d-1}$.

Conjetura

Si tomamos $\varepsilon_n \gg d_H(\aleph_n, S)$,

- ¿ se cumple que $|\partial \hat{S}_{\varepsilon_n}|_{d-1} \to |\partial S|_{d-1}$?
- ¿cómo tomar ϵ_n?, y ¿que tasa de convergencia se obtiene?

Usando el cierre r-convexo

En el caso del cierre r-convexo, con la fórmula de Crofton podemos calcular $|\partial C_\alpha(\aleph_n)|_{d-1}$, y se cumple:

$$|\partial C_r(\aleph_n)|_{d-1} \to |\partial S|_{d-1}. :$$

Teorema

Sea $S \subset \mathbb{R}^d$ compacto tal que ∂S es C^2, y $\aleph_n = \{X_1, \ldots, X_n\}$ iid de X, con densidad con soporte S. Sea $\alpha = reach(\partial S)$ y $\alpha' < \alpha$. Con probabilidad 1, para n suf. grande

$$\Big| |\partial S|_{d-1} - |\partial C_{\alpha'}(\mathcal{X})|_{d-1} \Big| = \mathcal{O}\left(\left(\frac{\ln(n)}{n} \right)^{\frac{2}{d+1}} \right).$$

Usando el cierre r-convexo

En el caso del cierre r-convexo, con la fórmula de Crofton podemos calcular $|\partial C_\alpha(\aleph_n)|_{d-1}$, y se cumple:

$$|\partial C_r(\aleph_n)|_{d-1} \to |\partial S|_{d-1}. :$$

Teorema

Sea $S \subset \mathbb{R}^d$ compacto tal que ∂S es C^2, y $\aleph_n = \{X_1, \dots, X_n\}$ iid de X, con densidad con soporte S. Sea $\alpha = reach(\partial S)$ y $\alpha' < \alpha$. Con probabilidad 1, para n suf. grande

$$\left| |\partial S|_{d-1} - |\partial C_{\alpha'}(\mathcal{X})|_{d-1} \right| = \mathcal{O}\left(\left(\frac{\ln(n)}{n} \right)^{\frac{2}{d+1}} \right).$$

Se mejora la tasa pero calcular $C_{\alpha'}(\mathcal{X})$ es costoso computacionalmente.

Usando el cierre r-convexo

En el caso del cierre r-convexo, con la fórmula de Crofton podemos calcular $|\partial C_\alpha(\aleph_n)|_{d-1}$, y se cumple:

$$|\partial C_r(\aleph_n)|_{d-1} \to |\partial S|_{d-1}. :$$

Teorema

Sea $S \subset \mathbb{R}^d$ compacto tal que ∂S es C^2, y $\aleph_n = \{X_1, \ldots, X_n\}$ iid de X, con densidad con soporte S. Sea $\alpha = reach(\partial S)$ y $\alpha' < \alpha$. Con probabilidad 1, para n suf. grande

$$\Big| |\partial S|_{d-1} - |\partial C_{\alpha'}(\mathcal{X})|_{d-1} \Big| = \mathcal{O}\left(\left(\frac{\ln(n)}{n} \right)^{\frac{2}{d+1}} \right).$$

Se mejora la tasa pero calcular $C_{\alpha'}(\mathcal{X})$ es costoso computacionalmente.

Problema abierto:

¿Cuál es la mejor tasa que se puede obtener, con algún estimador, para el problema de estimar L_0?

Usando el cierre r-convexo

En el caso del cierre r-convexo, con la fórmula de Crofton podemos calcular $|\partial C_\alpha(\aleph_n)|_{d-1}$, y se cumple:

$$|\partial C_r(\aleph_n)|_{d-1} \to |\partial S|_{d-1}. :$$

Teorema

Sea $S \subset \mathbb{R}^d$ compacto tal que ∂S es C^2, y $\aleph_n = \{X_1, \ldots, X_n\}$ iid de X, con densidad con soporte S. Sea $\alpha = reach(\partial S)$ y $\alpha' < \alpha$. Con probabilidad 1, para n suf. grande

$$\Big| |\partial S|_{d-1} - |\partial C_{\alpha'}(\mathcal{X})|_{d-1} \Big| = \mathcal{O}\left(\left(\frac{\ln(n)}{n} \right)^{\frac{2}{d+1}} \right).$$

Se mejora la tasa pero calcular $C_{\alpha'}(\mathcal{X})$ es costoso computacionalmente.

Problema abierto:

¿Cuál es la mejor tasa que se puede obtener, con algún estimador, para el problema de estimar L_0?

lo que si se sabe

La tasa óptima para $\mu_d(S)$ es $n^{-(d+3)/(2d+2)}$.

Aaron, C. and Bodart, O. (2016). Local convex hull support and boundary estimation. *Journal of Multivariate Analysis* **147** 82–101.

Aaron, C., and Cholaquidis. (2020). On boundary detection. *Ann. Inst. H. Poincaré Probab. Statist.* **56**(3) 2028–2050.

Aaron, C., Cholaquidis, A., and Cuevas, A. (2017). Stochastic detection of low dimensionality and data denoising via set estimation techniques. *Electronic Journal of Statistics* **11**(2) 4596–4628.

Alesker, S. (2018). Some conjectures on intrinsic volumes of Riemannian manifolds and Alexandrov spaces, *Arnold Mathematical Journal* **4**(1) 1–17.

Arias-Castro, E. And Rodríguez-Casal, A. (2017). On estimating the perimeter using the alpha-shape, *Ann. Inst. H. Poincaré Probab. Statist.* **53**(3) 1051–1068.

Arias-Castro, E., Pateiro-López, B., and Rodríguez-Casal, A. (2018). Minimax estimation of the volume of a set under the rolling ball condition. *Journal of the American Statistical Association.*

Baíllo, A. and Chacón, J.E. (2018). A survey and a new selection criterion for statistical home range estimation, preprint: arXiv:1804.05129

Baddeley, A. J., Gundersen, H. J. G., and Cruz-Orive, L. M. (1986). Estimation of surface area from vertical sections. *J. Microsc.* **142**(3) 259–276.

Baddeley, A. J. and Jensen, E.B. V. (2004). *Stereology for Statisticians*. Chapman and Hall, London.

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

Bräker, H. and Hsing, T. (1998). On the area and perimeter of a random convex hull in a bounded convex set. *Probability Theory and Related Fields* **111**(4) 517–550.

Burt, W. H. (1943). Territoriality and home range concepts as applied to mammals. *J. Mammal.* **24**, 346–352.

Cholaquidis, A., Fraiman, R., Lugosi, G., and Pateiro-López, B. (2016). Set estimation from reflected Brownian motion. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78**(5) 1057–1078.

Cholaquidis, A., Fraiman, R., Mordecki, E., Papalardo, C. (2021). Level sets and drift estimation for reflected Brownian motion with drift, *Statistical Sinica* 31(1) 29–51.

Cholaquidis, A., Fraiman, R., and Hernandez, M. (2021). Home range estimation under a restricted sampling scheme, https://arxiv.org/abs/2106.02035

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.

Crofton, M. W. (1868). On the theory of local probability, applied to straight lines drawn at random in a plane: The methods used being also extended to the proof of certain new theorems in the integral calculus, *Philosophical Transactions of the Royal Society of London* **158** 181–199.

Cuevas, A. and Rodriguez-Casal, A. (2004). On boundary estimation. *Adv. in Appl. Probab.* **36** 340–354.

Cuevas, A., Fraiman, R., and Rodríguez-Casal, A. (2007). A nonparametric approach to the estimation of lengths and surface areas. *Ann. Statist.* **35**(3) 1031–1051.

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., Fraiman, R., and Györfi, L. (2013). Towards a universally consistent estimator of the Minkowski content. *ESAIM: Probability and Statistics* **17** 359–369.

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

Devroye, L. and G. L. Wise (1980). Detection of abnormal behavior via nonparametric estimation of the support. *SIAM J. Appl. Math.* **3** 480–489.

Edelsbrunner, E., Kirkpatrick, D., and Seidel, R. (1983). On the shape of points in the plane. *IEEE Transaction on Information Theory* **29** 551–559.

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

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

Gokhale, A.M. (1990) Unbiased estimation of curve length in 3D using vertical slices. *J. Microsc.* **195** 133–141.

Güney, M., Kaplan, S., Onger, M., and Marangoz, A. (2012) Surface area estimation: A brief review. *NeuroQuantology.* **12**.

Jiménez, R. and Yukich, J.E. (2011). Nonparametric estimation of surface integrals. *Ann. Statist.* **39** 232–260.

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`

Pateiro-López, B. and Rodríguez-Casal, A. (2008). Length and surface area estimation under smoothness restrictions. *Advances in Applied Probability* **40**(2) 348–358.

Pateiro-López, B. and Rodríguez-Casal, A. (2009). Surface area estimation under convexity type assumptions, *Journal of Nonparametric Statistics* **21**(6) 729–741.

Kim, J.C. and Korostelëv, A. (2000). Estimation of smooth functionals in image models. *Math. Methods Statist.* **9**(2) 140–159.

Korostelëv, A.P. and Tsybakov, A.B. (1993). *Minimax Theory of Image Reconstruction*.Springer-Verlag, Berlin.

Penrose M.D. (2021). Random Euclidean coverage from within. preprint: arXiv:2101.06306

Rodríguez-Casal, A. (2007) Set estimation under convexity type assumptions, *Annales de l'Institut Henri Poincaré (B): Probability and Statistics* **43** 763–774.

Rodríguez-Casal, A. and Saavedra-Nieves, P. (2019). Extent of occurrence reconstruction using a new data-driven support estimator. preprint: arXiv:1907.08627

Santaló, L. A. (2004). Integral Geometry and Geometric Probability. Cambridge University Press.

Sarabia-Vallejos, Mauricio A. and Ayala-Jeria, Pedro and Hurtado, Daniel E. Three-dimensional whole-organ characterization of the regional alveolar morphology in normal murine lungs *Frontiers in Physiology.* **21**.

Thäle, C. And Yukich J.E. (2016). Asymptotic theory for statistics of the Poisson?Voronoi approximation, *Bernouilli* **22**(4) 2372–2400.

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