Dimensiones y sus estimaciones

Alejandro Cholaquidis

Seminario de Probabilidad y Estadística - Junio 2025.

8 de junio de 2025

Diferentes nociones de dimensión

Estimadores

Resultados de consistencia

Estimación bajo volumen polinomial

Simulaciones

Diferentes nociones de dimensión

Estimadores

Resultados de consistencia

Estimación bajo volumen polinomial

Simulaciones

¿Dimensión? ¿Qué dimensión?

Existen muchas definiciones disponibles:

- Dimensión de Kuratowski - Dimensión de cubrimiento de Lebesgue (para e. topológicos)
- Dimensión de Hausdorff

$$\dim_H(S) = \inf\{s \geq 0 \mid \mathcal{H}^s(S) = 0\} = \sup\{s \geq 0 \mid \mathcal{H}^s(S) = \infty\}$$

$$\mathcal{H}^s(S) = \lim_{\delta \to 0} \inf \left\{ \sum_i (\operatorname{diam} U_i)^s \,\middle|\, \{U_i\} \text{ es un } \delta\text{-cubrimiento } S \right\}$$

- Dimensión de Assouad

$$\dim_A(S) = \inf \left\{ s \geq 0 \,\middle|\, \exists C > 0, \forall 0 < r < R, \forall x \in S, N_r(B(x,R) \cap S) \leq C \left(\frac{R}{r}\right)^s \right\}$$

$N_r(U)$ es el mínimo número de bolas de radio r que cubren U.

¿Dimensión? ¿Qué dimensión?

Existen muchas definiciones disponibles:

- Dimensión de Kuratowski - Dimensión de cubrimiento de Lebesgue (para e. topológicos)
- Dimensión de Hausdorff

$$\dim_H(S) = \text{ínf}\,\{s \geq 0 \,|\, \mathcal{H}^s(S) = 0\} = \sup\,\{s \geq 0 \,|\, \mathcal{H}^s(S) = \infty\}$$

$$\mathcal{H}^s(S) = \lim_{\delta \to 0} \text{ínf} \left\{ \sum_i (\text{diam}\, U_i)^s \,\middle|\, \{U_i\} \text{ es un } \delta\text{-cubrimiento } S \right\}$$

- Dimensión de Assouad

$$\dim_A(S) = \text{ínf} \left\{ s \geq 0 \,\middle|\, \exists C > 0, \forall 0 < r < R, \forall x \in S, N_r(B(x,R) \cap S) \leq C \left(\frac{R}{r}\right)^s \right\}$$

$N_r(U)$ es el mínimo número de bolas de radio r que cubren U.

No vamos a querer estimar ninguna de estas...

¿Dimensión? ¿Qué dimensión?

- Dimensión de Minkowski-Bouligand $\Leftarrow$
- Dimensión puntual $\Leftarrow$
- Dimensión de correlación $\Leftarrow$

¿Por qué Minkowski?: porque es **bastante popular** y **estadísticamente tratable**...

¿Por qué Dimensión Puntual?: queremos enfocarnos también en la **medida de probabilidad subyacente, localmente.**

¿Por qué Dimensión de Correlación?: **buenos resultados empíricos.**

¿Quién se interesa por la dimensión?

- **Dinamistas**: La dimensión es a menudo una característica relevante en el estudio de **atractores** (dimensión no entera de estos conjuntos, a menudo relacionada con comportamientos irregulares o caóticos).
- **Fractalistas**: Una dimensión no entera es una característica típica aquí.
- **Estadísticos**: En relación con la llamada **Hipótesis del Variedad**, la cual está motivada por la observación empírica de que muchos conjuntos de datos multivariados encontrados en la práctica están confinados (o cerca) de un conjunto de menor dimensión.

(*) Fefferman, C., Mitter, S., and Narayanan, H. (2016). Testing the manifold hypothesis. *Journal of the American Mathematical Society*, **29**(4), 983-1049.

Estos autores se refieren a la Hipótesis del Variedad (MH, por sus siglas en inglés) en estos términos:

Hipótesis de variedad

"The hypothesis that high dimensional data tend to lie in the vicinity of a low dimensional manifold is the basis of manifold learning".
`https://en.wikipedia.org/wiki/Manifold_hypothesis`

Minkowski-Bouligand (I)

Número de cubrimiento

Sea $S \subset \mathbb{R}^d$ acotado, definimos el *número de cubrimiento*, $N(S, r)$, como el mínimo número de conjuntos de diámetro a lo sumo r que se necesitan para cubrir S.

Minkowski-Bouligand (I)

Número de cubrimiento

Sea $S \subset \mathbb{R}^d$ acotado, definimos el *número de cubrimiento*, $N(S, r)$, como el mínimo número de conjuntos de diámetro a lo sumo r que se necesitan para cubrir S.

Dimension de Minkowski

La dimensión de Minkowski [a] se define como:

$$\dim(S) = \lim_{r \to 0} \frac{\log(N(S, r))}{\log(1/r)}. \tag{1}$$

[a] *Minkowski-Bouligand, Kolmogorov capacity, box-counting, dimensión de entropía*

esto significa que la dimensión es el exponente k tal que $N(S, 1/n) \approx Cn^k$.

Vamos a asumir que el límite (1) existe. Ver **Bishop and Peres (2017)** para un tratamiento formal.

Minkowski-Bouligand (II)

Definimos

- $N_{\text{pack}}(S, r)$: *packing number.* Máxima cantidad de bolas *disjuntas* de radio r centradas en S.

Minkowski-Bouligand (II)

Definimos

- $N_{\text{pack}}(S, r)$: *packing number.* Máxima cantidad de bolas *disjuntas* de radio r centradas en S.
- $N_{\text{sep}}(S, r)$: *separation number.* Máxima cantidad de puntos de un conjunto r-separado: $X \subset S$ es r-separado si $x, y \in X$ implica $|x - y| \geq r$).

En

$$\dim(S) = \lim_{r \to 0} \frac{\log(N(S, r))}{\log(1/r)},$$

es equivalente usar $N(S, r)$, $N_{\text{sep}}(S, r)$, o $N_{\text{pack}}(S, r)$.

Minkowski-Bouligand (II)

Definimos

- $N_{\text{pack}}(S,r)$: *packing number.* Máxima cantidad de bolas *disjuntas* de radio r centradas en S.
- $N_{\text{sep}}(S,r)$: *separation number.* Máxima cantidad de puntos de un conjunto r-separado: $X \subset S$ es r-separado si $x, y \in X$ implica $|x-y| \geq r$).

En

$$\dim(S) = \lim_{r\to 0} \frac{\log(N(S,r))}{\log(1/r)},$$

es equivalente usar $N(S,r)$, $N_{\text{sep}}(S,r)$, o $N_{\text{pack}}(S,r)$.

Una expresión alternativa para $\dim(S)$ es:

$$\dim(S) = d - \lim_{r\to 0} \frac{\log(\mu(B(S,r)))}{\log(r)} = d - \lim_{r\to 0} \frac{\log(V(r))}{\log(r)}.$$

Dimensión Puntual

Dimensión puntual

Si P es una probabilidad con soporte $S \subset \mathbb{R}^d$, la the **dimensión puntual** de P en $x \in S$ es:

$$d_{pw}(x) = \lim_{r \to 0} \frac{\log(P(B(x,r)))}{\log(r)}, \tag{4}$$

siempre que este límite exista. Fue propuesta por **Young (1982)**.

Dimensión Puntual

Dimensión puntual

Si P es una probabilidad con soporte $S \subset \mathbb{R}^d$, la the **dimensión puntual** de P en $x \in S$ es:

$$d_{pw}(x) = \lim_{r \to 0} \frac{\log(P(B(x,r)))}{\log(r)}, \tag{4}$$

siempre que este límite exista. Fue propuesta por **Young (1982)**.

Una noción global de dimensión de S basada en (4), puede ser:

$$D_{pw}(S) = \sup_{x \in S} d_{pw}(x).$$

Se cumple que

$$\dim_H(S) \leq D_{pw}(S).$$

Si P tiene soporte en una variedad y $d_{pw}(x) = q$ c.s., entonces $\dim_H(S) = q$.

Dimensión de Correlación

Introducida por **Grassberger and Procaccia (1983)**; ver también el review **Camastra and Staiano (2016)** y para una definición rigurosa **Pesin (1993)**.

Dimensión de Correlación

Introducida por **Grassberger and Procaccia (1983)**; ver también el review **Camastra and Staiano (2016)** y para una definición rigurosa **Pesin (1993)**.

Correlation Dimension

Sean X_1, X_2 i.i.d. de X a valores en un espacio métrico (S, d). Definimos

$$p(r) = \mathbb{P}(d(X_1, X_2) < r).$$

Dimensión de Correlación

Introducida por **Grassberger and Procaccia (1983)**; ver también el review **Camastra and Staiano (2016)** y para una definición rigurosa **Pesin (1993)**.

Correlation Dimension

Sean X_1, X_2 i.i.d. de X a valores en un espacio métrico (S, d). Definimos

$$p(r) = \mathbb{P}(d(X_1, X_2) < r).$$

La dimensión de correlación de X con distribución P_X es

$$\dim_{\mathrm{cd}}(S) := \dim_{\mathrm{cd}}(S; P_X) = \lim_{r \to 0} \frac{\log(p(r))}{\log(r)} \tag{5}$$

siempre que este límite exista.

Dimensión de Correlación

Introducida por **Grassberger and Procaccia (1983)**; ver también el review **Camastra and Staiano (2016)** y para una definición rigurosa **Pesin (1993)**.

Correlation Dimension

Sean X_1, X_2 i.i.d. de X a valores en un espacio métrico (S, d). Definimos

$$p(r) = \mathbb{P}(d(X_1, X_2) < r).$$

La dimensión de correlación de X con distribución P_X es

$$\mathrm{dim}_{\mathrm{cd}}(S) := \mathrm{dim}_{\mathrm{cd}}(S; P_X) = \lim_{r \to 0} \frac{\log(p(r))}{\log(r)} \tag{5}$$

siempre que este límite exista.

Lema

Supongamos que existe una medida ν y constantes $0 < c_1 < C_2 < \infty$ tal que 1) para todo $x \in S$:

$$c_1 r^{\ell} \leq \nu(B(x, r)) \leq C_2 r^{\ell}, \quad \textit{for all } r < r_0,$$

2) P_X es estandard respecto de ν y tiene densidad acotada respecto de ν. Entonces, $\mathrm{dim}_{cd}(S) = \ell$.

Diferentes nociones de dimensión

Estimadores

Resultados de consistencia

Estimación bajo volumen polinomial

Simulaciones

Estimador de la dimensión de Minkowski

$\widehat{\dim}_{\text{cap}}$

Sea $r_n \downarrow 0$, entonces

$$\widehat{\dim}_{\text{cap}} = -\frac{\log(N_{\text{sep}}(\aleph_n, r_n))}{\log(r_n)}, \tag{6}$$

donde $N_{\text{sep}}(\aleph_n, r_n)$ es el máximo cardinal de un conjunto r_n-separado de la muestra $\aleph_n$.

Fue considerado por **Kégl (2002)**, y propone un algoritmo para calcularlo.

Estimador de la dimensión de Minkowski

$\widehat{\dim}_{\text{cap}}$

Sea $r_n \downarrow 0$, entonces

$$\widehat{\dim}_{\text{cap}} = -\frac{\log(N_{\text{sep}}(\aleph_n, r_n))}{\log(r_n)}, \tag{6}$$

donde $N_{\text{sep}}(\aleph_n, r_n)$ es el máximo cardinal de un conjunto r_n-separado de la muestra $\aleph_n$.

Fue considerado por **Kégl (2002)**, y propone un algoritmo para calcularlo.

$\widehat{\dim}_{\text{vol}}$: Basado en la función de volumen

Sea $r_n \downarrow 0$, definimos $V_n(r_n) = \mu(B(\aleph_n, r_n))$, y

$$\widehat{\dim}_{\text{vol}} = d - \frac{\log(V_n(r_n))}{\log(r_n)}.$$

Estimador de la dimensión de Minkowski

$\widehat{\dim}_{\text{cap}}$

Sea $r_n \downarrow 0$, entonces

$$\widehat{\dim}_{\text{cap}} = -\frac{\log(N_{\text{sep}}(\aleph_n, r_n))}{\log(r_n)}, \tag{6}$$

donde $N_{\text{sep}}(\aleph_n, r_n)$ es el máximo cardinal de un conjunto r_n-separado de la muestra $\aleph_n$.

Fue considerado por **Kégl (2002)**, y propone un algoritmo para calcularlo.

$\widehat{\dim}_{\text{vol}}$: Basado en la función de volumen

Sea $r_n \downarrow 0$, definimos $V_n(r_n) = \mu(B(\aleph_n, r_n))$, y

$$\widehat{\dim}_{\text{vol}} = d - \frac{\log(V_n(r_n))}{\log(r_n)}.$$

$$\left| \widehat{\dim}_{\text{vol}} - \widehat{\dim}_{\text{cap}} + \frac{\log(\omega_d)}{\log(r_n)} \right| \leq -d\frac{\log(2)}{\log(r_n)} \qquad \text{c.s.}$$

Estimators of Correlation and Pointwise Dimensions

An Estimator of the Correlation Dimension

Let $X_1, \ldots, X_n$ be an i.i.d. sample from X. Define:

$$\widehat{p}_n(r) = \binom{n}{2}^{-1} \sum_{i \neq j} \mathbb{I}_{\{d(x_i,x_j)<r\}}. \qquad \text{The estimator is:} \qquad \widehat{\dim}_{\mathrm{cd}} = \frac{\log(\widehat{p}_n(r_n))}{\log(r_n)}.$$

Estimators of Correlation and Pointwise Dimensions

An Estimator of the Correlation Dimension

Let $X_1, \dots, X_n$ be an i.i.d. sample from X. Define:

$$\widehat{p}_n(r) = \binom{n}{2}^{-1} \sum_{i \neq j} \mathbb{I}_{\{d(x_i,x_j)<r\}}.$$

The estimator is:

$$\widehat{\dim}_{\mathrm{cd}} = \frac{\log(\widehat{p}_n(r_n))}{\log(r_n)}.$$

A Plug-in Estimator of the Pointwise Dimension

We replace P by its empirical counterpart P_n:

$$\widehat{d}_{pw}(x) = \frac{\log(P_n(B(x, r_n)))}{\log(r_n)}.$$

Diferentes nociones de dimensión

Estimadores

Resultados de consistencia

Estimación bajo volumen polinomial

Simulaciones

(I) - para Minkowski

Sean

- $S \subset \mathbb{R}^d$ compacto, $V(r)$ Lipschitz en $[0, \lambda]$ con $\lambda > 0$.

(I) - para Minkowski

Sean

- $S \subset \mathbb{R}^d$ compacto, $V(r)$ Lipschitz en $[0, \lambda]$ con $\lambda > 0$.
- $\aleph_n = \{X_1, \ldots, X_n\}$ i.i.d. de P_X con soporte S.

(I) - para Minkowski

Sean

- $S \subset \mathbb{R}^d$ compacto, $V(r)$ Lipschitz en $[0, \lambda]$ con $\lambda > 0$.
- $\aleph_n = \{X_1, \ldots, X_n\}$ i.i.d. de P_X con soporte S.
- $\gamma_n = \rho_H(\aleph_n, S)$ distancia de Hausdorff entre $\aleph_n$ y S.

(I) - para Minkowski

Sean

- $S \subset \mathbb{R}^d$ compacto, $V(r)$ Lipschitz en $[0, \lambda]$ con $\lambda > 0$.
- $\aleph_n = \{X_1, \ldots, X_n\}$ i.i.d. de P_X con soporte S.
- $\gamma_n = \rho_H(\aleph_n, S)$ distancia de Hausdorff entre $\aleph_n$ y S.
- $r_n \to 0$ tal que $\gamma_n < r_n$ y

$$\frac{\gamma_n}{V(r_n - \gamma_n)\log(r_n)} \to 0 \qquad \text{c.s.}$$

(I) - para Minkowski

Sean

- $S \subset \mathbb{R}^d$ compacto, $V(r)$ Lipschitz en $[0, \lambda]$ con $\lambda > 0$.
- $\aleph_n = \{X_1, \ldots, X_n\}$ i.i.d. de P_X con soporte S.
- $\gamma_n = \rho_H(\aleph_n, S)$ distancia de Hausdorff entre $\aleph_n$ y S.
- $r_n \to 0$ tal que $\gamma_n < r_n$ y

$$\frac{\gamma_n}{V(r_n - \gamma_n)\log(r_n)} \to 0 \qquad \text{c.s.}$$

Entonces

(a) $\dim(S) = d - \lim_{n\to\infty} \dfrac{\log(V_n(r_n))}{\log(r_n)}$ c.s., donde $V_n(r_n) = \mu(B(\aleph_n, r_n))$.

(I) - para Minkowski

Sean

- $S \subset \mathbb{R}^d$ compacto, $V(r)$ Lipschitz en $[0, \lambda]$ con $\lambda > 0$.
- $\aleph_n = \{X_1, \ldots, X_n\}$ i.i.d. de P_X con soporte S.
- $\gamma_n = \rho_H(\aleph_n, S)$ distancia de Hausdorff entre $\aleph_n$ y S.
- $r_n \to 0$ tal que $\gamma_n < r_n$ y

$$\frac{\gamma_n}{V(r_n - \gamma_n)\log(r_n)} \to 0 \qquad \text{c.s.}$$

Entonces

(a) $\dim(S) = d - \lim_{n\to\infty} \dfrac{\log(V_n(r_n))}{\log(r_n)}$ c.s., donde $V_n(r_n) = \mu(B(\aleph_n, r_n))$.

(b) $\widehat{\dim}_{\text{cap}} = -\frac{\log(N_{\text{sep}}(\aleph_n, r_n))}{\log(r_n)}$ es c.s. consistente bajo las mismas hipótesis en r_n.

(II) - Dimensión puntual

Sean

- $S \subset \mathbb{R}^d$ compacto,
- $\aleph_n = \{X_1, \ldots, X_n\}$ iid de P con soporte S,
- $x \in S$ tal que $\dim_{\mathrm{pw}}(x)$ existe.

Supongamos que P es estandar en x, i.e.,

$$P(B(x,r)) > \delta r^{d'} \qquad \text{for all } r < r_0. \tag{1}$$

Entonces,

$$\dim_{\mathrm{pw}}(x) = \lim_{n\to\infty} \frac{\log\Big(\mathbb{P}_n(B(x,r_n))\Big)}{\log(r_n)} \qquad \text{c.s.},$$

donde $r_n = (\log(n)/n)^{1/d'}$.

(II) - Dimensión puntual

Sean

- $S \subset \mathbb{R}^d$ compacto,
- $\aleph_n = \{X_1, \ldots, X_n\}$ iid de P con soporte S,
- $x \in S$ tal que $\dim_{\mathrm{pw}}(x)$ existe.

Supongamos que P es estandar en x, i.e.,

$$P(B(x,r)) > \delta r^{d'} \qquad \text{for all } r < r_0. \tag{1}$$

Entonces,

$$\dim_{\mathrm{pw}}(x) = \lim_{n\to\infty} \frac{\log\left(\mathbb{P}_n(B(x,r_n))\right)}{\log(r_n)} \qquad \text{c.s.,}$$

donde $r_n = (\log(n)/n)^{1/d'}$.

Uniforme en x

Si $P(B(x,r)) > \delta r^{d'}$ $\forall x \in S$, la convergencia es uniforme en x si

$$r_n = \mathcal{O}\left(\frac{\log(n)}{n}\right)^{1/(2d')}.$$

(II) Dimensión de correlación

Supongamos que X es tal que

$$\dim_{cd}(S) := \lim_{r \to 0} \frac{\log(p(r))}{\log(r)} \tag{5}$$

existe y es finito.

(II) Dimensión de correlación

Supongamos que X es tal que

$$\dim_{cd}(S) := \lim_{r\to 0} \frac{\log(p(r))}{\log(r)} \tag{5}$$

existe y es finito.

Entonces, el estimador

$$\widehat{d}_{pw}(x) = \frac{\log(P_n(B(x, r_n)))}{\log(r_n)} \qquad \text{where} \qquad \widehat{p}_n(r) = \binom{n}{2}^{-1} \sum_{i\neq j} \mathbb{I}_{\{d(x_i,x_j)<r\}}.$$

verifica

$$\dim_{cd}(P) = \lim_{n\to\infty} \frac{\log(\widehat{p}_n(r_n))}{\log(r_n)} \quad c.s.,$$

donde

$$r_n = \left(\frac{\log n}{n}\right)^{\frac{1}{(1+\beta)\dim_{cd}(P)}},$$

y $\beta > 0$.

Diferentes nociones de dimensión

Estimadores

Resultados de consistencia

Estimación bajo volumen polinomial

Simulaciones

Alcance

Alcance positivo

Denotamos $\mathrm{Unp}(S)$ el conjunto de puntos $x \in \mathbb{R}^d$ que tienen una única proyección sobre S.

Si $x \in S$, definimos

$$\mathrm{reach}(S,x) = \sup\{r > 0 : \mathring{B}(x,r) \subset \mathrm{Unp}(S)\}.$$

y

$$\mathrm{reach}(S) = \mathrm{inf}\{\mathrm{reach}(S,x) : x \in S\}.$$

El conjunto S tiene **alcance positivo** si $\mathbf{r} = \mathrm{reach}(S) > 0$.

Alcance

Alcance positivo

Denotamos $\mathrm{Unp}(S)$ el conjunto de puntos $x \in \mathbb{R}^d$ que tienen una única proyección sobre S.

Si $x \in S$, definimos

$$\mathrm{reach}(S,x) = \sup\{r > 0 : \mathring{B}(x,r) \subset \mathrm{Unp}(S)\}.$$

y

$$\mathrm{reach}(S) = \inf\{\mathrm{reach}(S,x) : x \in S\}.$$

El conjunto S tiene **alcance positivo** si $\mathbf{r} = \mathrm{reach}(S) > 0$.

Si $\mathrm{reach}(S) > 0$, entonces V es polinomial

Si $\mathbf{r} > 0$, $V(r)$ is un polinomio en $[0,\mathbf{r}]$ de la forma:

$$V(r) = \theta_0 + \theta_1 r + \ldots + \theta_d r^d, \qquad r \in [0,\mathbf{r}].$$

Alcance

Alcance positivo

Denotamos Unp(S) el conjunto de puntos $x \in \mathbb{R}^d$ que tienen una única proyección sobre S.

Si $x \in S$, definimos

$$\text{reach}(S,x) = \sup\{r > 0 : \mathring{B}(x,r) \subset \text{Unp}(S)\}.$$

y

$$\text{reach}(S) = \text{ínf}\{\text{reach}(S,x) : x \in S\}.$$

El conjunto S tiene **alcance positivo** si $\mathbf{r} = \text{reach}(S) > 0$.

Si reach(S) > 0, entonces V es polinomial

Si $\mathbf{r} > 0$, $V(r)$ is un polinomio en $[0, \mathbf{r}]$ de la forma:

$$V(r) = \theta_0 + \theta_1 r + \ldots + \theta_d r^d, \qquad r \in [0, \mathbf{r}].$$

- $\theta_0 = \mu(S)$
- $\theta_1 = \mu_{d-1}(\partial S)$
- $\theta_d = \mu(B(0,1))\mathcal{X}(S)$, donde $\mathcal{X}(S)$ es la característica de Euler-Poincaré de S.

Alcance polinomial

Sea $S \subset \mathbb{R}^d$ compacto y $V(r) = \mu(B(S,r))$, definimos el *alcance polinomial* **R** de S as

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

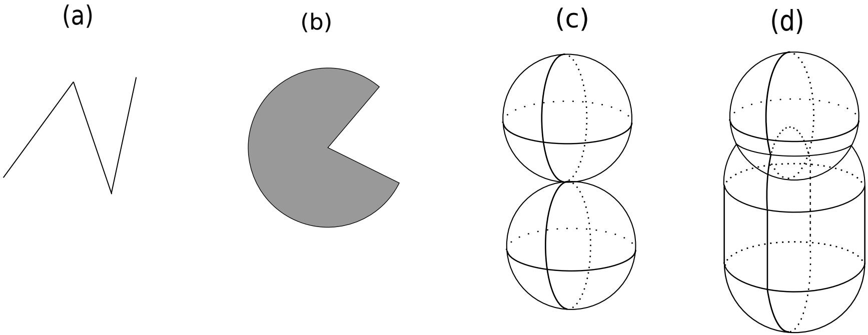


Figura: Conjuntos con $\mathbf{R} > 0$ pero alcance 0.

Consistencia bajo volumen polinomial

Sea $S \subset \mathbb{R}^d$ compacto. Supongamos $V(r) = \sum_{j=k}^{d} \theta_j r^j, r \in [0, \mathbf{R}]$, donde θ_k es el primer coeficinete no nulo.

Entonces:

(a) $\dim(S) = d - k$.

Consistencia bajo volumen polinomial

Sea $S \subset \mathbb{R}^d$ compacto. Supongamos $V(r) = \sum_{j=k}^{d} \theta_j r^j$, $r \in [0, \mathbf{R}]$, donde θ_k es el primer coeficinete no nulo.

Entonces:

(a) $\dim(S) = d - k$.

(b) Sea $\gamma_n = \rho_H(\aleph_n, S)$. Supongamos $r_n \to 0$ $\gamma_n / r_n \to 0$ c.s. Entonces

$$\left| \dim(S) - \widehat{\dim}_{\mathrm{vol}} \right| = \left| \dim(S) - \frac{\log(N_{\mathrm{sep}}(\aleph_n, r_n))}{\log(r_n)} \right| \leq \frac{|\log(2\theta_k)|}{|\log(r_n)|}, \qquad \text{c.s.}$$

Consistencia bajo volumen polinomial

Sea $S \subset \mathbb{R}^d$ compacto. Supongamos $V(r) = \sum_{j=k}^{d} \theta_j r^j$, $r \in [0, \mathbf{R}]$, donde θ_k es el primer coeficinete no nulo.

Entonces:

(a) $\dim(S) = d - k$.

(b) Sea $\gamma_n = \rho_H(\aleph_n, S)$. Supongamos $r_n \to 0$ $\gamma_n/r_n \to 0$ c.s. Entonces

$$\left|\dim(S) - \widehat{\dim}_{\text{vol}}\right| = \left|\dim(S) - \frac{\log(N_{\text{sep}}(\aleph_n, r_n))}{\log(r_n)}\right| \leq \frac{|\log(2\theta_k)|}{|\log(r_n)|}, \qquad \text{c.s.}$$

(c) Además, si $r_n \to 0$ y $\gamma_n/r_n \to 0$, c.s.:

$$\left|\dim(S) - \widehat{\dim}_{\text{cap}}\right| = \left|\dim(S) - \frac{\log(N_{\text{sep}}(\aleph_n, r_n))}{\log(r_n)}\right| \leq \frac{|\log(2\theta_k)| + |\log(\omega_d)| + d\log(2)}{|\log(r_n)|}.$$

Consistencia bajo volumen polinomial

Sea $S \subset \mathbb{R}^d$ compacto. Supongamos $V(r) = \sum_{j=k}^{d} \theta_j r^j$, $r \in [0, \mathbf{R}]$, donde θ_k es el primer coeficinete no nulo.

Entonces:

(a) $\dim(S) = d - k$.

(b) Sea $\gamma_n = \rho_H(\aleph_n, S)$. Supongamos $r_n \to 0$ $\gamma_n / r_n \to 0$ c.s. Entonces

$$\left| \dim(S) - \widehat{\dim}_{\text{vol}} \right| = \left| \dim(S) - -\frac{\log(N_{\text{sep}}(\aleph_n, r_n))}{\log(r_n)} \right| \leq \frac{|\log(2\theta_k)|}{|\log(r_n)|}, \qquad \text{c.s.}$$

(c) Además, si $r_n \to 0$ y $\gamma_n / r_n \to 0$, c.s.:

$$\left| \dim(S) - \widehat{\dim}_{\text{cap}} \right| = \left| \dim(S) - -\frac{\log(N_{\text{sep}}(\aleph_n, r_n))}{\log(r_n)} \right| \leq \frac{|\log(2\theta_k)| + |\log(\omega_d)| + d\log(2)}{|\log(r_n)|}.$$

Si aplicamos $\lfloor x + 1/2 \rfloor$ a $\widehat{\dim}_{\text{cap}}$ o $\widehat{\dim}_{\text{vol}}$, tenemos esitmadores que, para n suficientemente grande son *exactamente* $\dim(S)$.

Diferentes nociones de dimensión

Estimadores

Resultados de consistencia

Estimación bajo volumen polinomial

Simulaciones

(I) - Simulaciones

¿Cómo elegimos r_n?

Consideremos

$$\widehat{\dim}_{\text{cap}} = -\frac{\log(N_{\text{sep}}(\aleph_n, r_n))}{\log(r_n)}, \tag{6}$$

Esto significa que la dimensión es el exponente k tal que $N_{\text{sep}}(\aleph_n, r_n) \approx C r_n^k$.

Esto sugiere elegir una grilla de valores de r_n y hacer una regresión lineal entre entre $\log(r_n)$ y $\log(N_{\text{sep}}(\aleph_n, r_n))$, **y lo mismo se puede hacer para el resto de los estimadores**.

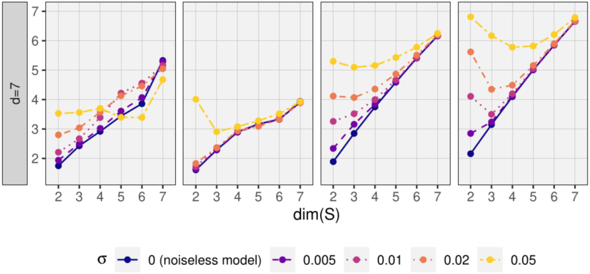
(I) - Sin ruido

Manifold	d	$\dim(\mathcal{M}_i)$	BC	CAP	CD	PW
$\mathcal{M}_1$	11	10	10.48	6.65	9.06	9.99
$\mathcal{M}_2$	5	3	2.30	2.20	2.89	3.15
$\mathcal{M}_3$	6	4	2.68	2.70	3.59	4.36
$\mathcal{M}_4$	8	4	4.11	4.07	3.79	4.36
$\mathcal{M}_5$	3	2	1.85	1.68	1.99	2.15
$\mathcal{M}_6$	36	6	12.15	5.56	5.79	7.27
$\mathcal{M}_7$	3	2	2.10	2.41	1.98	2.15
$\mathcal{M}_8$	72	12	25.03	8.54	11.69	16.02
$\mathcal{M}_9$	20	20	26.88	8.74	14.43	16.51
$\mathcal{M}_{10}$	11	10	2.45	5.43	8.34	9.12
$\mathcal{M}_{11}$	3	2	1.95	2.65	2.02	2.16
$\mathcal{M}_{12}$	20	20	10.71	7.16	14.04	18.95
$\mathcal{M}_{13}$	13	1	0.96	1.26	1.25	1.25

Promedio en 20 réplicas uniformes con soporte en diferentes variedades $n = 2500$.
$BC = \widehat{\dim}_{\text{bc}} = \frac{\log(N_{\text{box}}(\aleph_n, r_n))}{\log(1/r_n)}$, $CAP = \widehat{\dim}_{\text{cap}} = -\frac{\log(N_{\text{sep}}(\aleph_n, r_n))}{\log(r_n)}$ $CD = \widehat{\dim}_{\text{cd}} = \frac{\log(\widehat{p}_n(r_n))}{\log(r_n)}$.
$PW = \widehat{d}_{pw}(x) = \frac{\log(P_n(B(x, r_n)))}{\log(r_n)}$.

(II) - Resultados, modelos con ruido

A cada punto de la muestra se le suma $N(0, \sigma^2 I_d)$.



$S \subset \mathbb{R}^d$ cubo de dimensiones $2, \ldots, 7$, $n = 2500$, promedio en 20 réplicas.

Columnas 1) $\widehat{\dim}_{bc} = \frac{\log(N_{box}(\aleph_n, r_n))}{\log(1/r_n)}$, 2) $\widehat{\dim}_{cap} = -\frac{\log(N_{sep}(\aleph_n, r_n))}{\log(r_n)}$, 3) $\widehat{\dim}_{cd} = \frac{\log(\widehat{p}_n(r_n))}{\log(r_n)}$.
4) $\widehat{d}_{pw}(x) = \frac{\log(P_n(B(x, r_n)))}{\log(r_n)}$.

¡Gracias!