Functional depths

Choi Seokjun

11 Dec. 2019

Outline

definition of depth in $\mathbb{R}^p$

Definition (statistical depth in $\mathbb{R}^p$, (Zuo and Serfling, 2000b))

Let $\mathcal{P}$ be some class of of distributions. The bounded and non-negative mapping $D(.,.) : \mathbb{R}^p \times \mathcal{P} \to \mathbb{R}$ is called a statistical depth function if it satisfies the following properties:

- Affine invariance
  $D(Ax + b, P_{AX+b}) = D(X, P_X)$ holds for any $\mathbb{R}^p$-valued random vector X, any $p \times p$ nonsingular matrix A and any $b \in \mathbb{R}^p$.
- Maximality at center
  $D(\theta, P) = sup_{x \in \mathbb{R}^p} D(x, P)$ holds for any $P \in \mathcal{P}$ having a unique center of symmetry θ w.r.t. some notion of symmetry.
- Monotonicity relative to the deepest point
  For any $P \in \mathcal{P}$ having deepest point θ, $D(x, P) \leq D(\theta + \alpha(x - \theta), P)$ holds for all $\alpha \in [0, 1]$.
- Vanishing at infinity $D(x, P) \to 0$ as $||x||_{\mathbb{R}^p} \to \infty$ for each $P \in \mathcal{P}$.

definition of depth in $\mathbb{R}^p$

(Serfling(2006)) Not necessary, but desirable property when setting D:

- Symmetry
  If P is symmetric about θ, then so is $D(x, P)$.
- Continuity of D(x,P) as a function of x
  (or just have upper semi-continuity)
- Continuity of D(x,P) as a function of P
- Quasi-concavicity as a function of x
  The set $\{x : D(x, P) \geq c\}$ is convex for each real c.

Example (on $\mathbb{R}^1$)

If we denote F_P as cdf corresponding distribution measure P, then

- (By Fraiman, Muniz(2001)) $D(x, P) = 1/2 - [1/2 - F_P(x)]$
- (Halfspace depth, By Tukey(1975)) $D(x, P) = min\{F_P(x), lim_{v \to x-} F_P(v)\}$
- (Simplical depth, By Liu(2001)) $D(x, P) = F_P(x)\{1 - lim_{v \to x-} F_P(v)\}$
- (Modified band depth, By Cuevas, Fraiman(2009))
  $D(x, P) = \frac{1}{J-1} \sum_{j=2}^{J} P(x \in [min(X_1, ..., X_j), max(X_1, ..., X_j)])$

definition of depth in $\mathcal{F}$

Definition (statistical depth in $\mathcal{F}$, (Nieto-Reyes and Battey, 2016))

Let $(\mathcal{F}, A, P)$ be probability space and $\mathcal{P}$ be class of all distribution measures on $\mathcal{F}$. The bounded and non-negative mapping $D(.,.) : \mathcal{F} \times \mathcal{P} \to \mathbb{R}$ is called a statistical functional depth function if it satisfies the following properties:

- distance invariance
  $D(f(x), P_{f(X)}) = D(X, P_X)$ for any $x \in \mathcal{F}$ and $f : \mathcal{F} \to \mathcal{F}$ such that for any $y \in \mathcal{F}$, $d(f(x), f(y)) = a_f d(x, y)$, $a_f \in \mathbb{R} - \{0\}$.
- Maximality at center
  For any $P \in \mathcal{P}$ with unique center of symmetry θ w.r.t. some notion of symmetry, $D(\theta, P) = sup_{x \in \mathcal{F}} D(x, P)$.
- Monotonicity (strictly decreasing) relative to the deepest point
  For any $P \in \mathcal{P}$ s.t. $D(z, P) = max_{x \in \mathcal{F}} D(x, P)$ exists (:deepest point z), for $x, y \in \mathcal{F}$, $D(x, P) < D(y, P) < D(z, P)$ s.t. $min\{d(y, z), d(y, x)\} > 0$ and $max\{d(y, z), d(y, x)\} < d(x, z)$.

Definition ((continue.))

- Upper semi-continuity in x
  $D(x, P)$ is upper semi-continuous as a function of x.
- Receptivity to convex hull width across the domain.
  Let $C(\mathcal{F}, P)$ be convex hull in $(\mathcal{F}, A, P)$ defined as
  $C(\mathcal{F}, P) = \{x \in \mathcal{F} : x(v) = \alpha L(v) + (1-\alpha)U(v), v \in V, \alpha \in [0,1]\}$ where $U = \{sup_{x \in E} x(v) : v \in V\}$,
  $L = \{inf_{x \in E} x(v) : v \in V\}$ and E is smallest set in A s.t. $P(E) = P(\mathcal{F})$.
  Then, D has a property that $D(x, P_X) < D(f(x), P_{f(X)})$ for any $x \in C(\mathcal{F}, P)$ with
  $D(x, P) < sup_{y \in \mathcal{F}} D(y, P)$ and $f : \mathcal{F} \to \mathcal{F}$ s.t. $f(y(v)) = \alpha(v) y(v)$ with $\alpha(v) \in (0, 1)$ for all $v \in L_\delta$ and
  $\alpha(v) = 1$ otherwise where $L_\delta = argsup_{H \in V}\{sup_{x,y \in C(\mathcal{F},P)} d(x(H), y(H)) \leq \delta\}$ for any
  $\delta \in inf_{v \in V} d(L(v), U(v)), d(L, U)$ s.t. $\lambda(L_\delta) > 0$ and $\lambda(L_\delta^c) > 0$.
- Continuity in P
  For all $x \in \mathcal{F}$, for all $P \in \mathcal{P}$ and for every $\epsilon > 0$, there exists a $\delta(\epsilon) > 0$ s.t. $|D(x, Q) - D(x - P)| \leq \epsilon$
  P-almost surely for all $Q \in \mathcal{P}$ with $d_P(Q, P) < \delta$ P-almost surely, where d_P is metric on $\mathcal{P}$.

Other requirement?
'convex depth level set'(ex. Narisetty and Nair, 2015), 'null at the boundary'(or, similarly 'Vanishing at infinity')(Mosler and Polyakov, 2012), 'non-degeneracy with gaussian process class' (Chakraborty and Chaudhuri, 2014b)

Check the validity of existing depth on $\mathcal{F}$

- h-depth (Cuevas, Febrero and Fraiman(2007)) : **FTTTTT**
  $D_h(x,P) = E_X(K_h(||x - X||_{\mathcal{L}^2[0,1]}))$ on $\mathcal{L}^2[0,1]^p$
- random-tukey depth (Cuesta-Albertos and Nieto-Reyes(2008)) : **TTFTFT**
  $D_{RT}(x,P) = min_{u \in \{u_j\}_{j=1}^k} min(P_{(u)}(-\infty, \langle u, x \rangle], P_{(u)}[\langle u, x \rangle, \infty))$
  where $P_{(u)}$: marginal distribution measure of u, on $\mathcal{L}^2[0,1]^p$
- band depth(Lopez-Pintado and Romo(2009)) : **TTFTFT**
  $D_J(x,P) = \sum_{j=2}^J P_{S_j}(x \in S_j(P))$ where $S_j(P) = \{y \in \mathcal{F} : y(v) = \alpha_1 X_1(v) + ... + \alpha_j X_j(v), \alpha_k \in$(j-th dim simplex)$, v \in V, X_i \sim P\}$ on $\mathcal{C}$ with sup norm
- modified band depth (Lopez-Pintado and Romo(2009)) : **TTFTFT**
  $D_{MJ}(x,P) = \sum_{j=2}^J E(\lambda\{v \in V : x(v) \in S_j(P)\})$ with above notation, on $\mathcal{C}$ with sup norm
- half-region depth (Lopez-Pintado and Romo(2011)) : **TFFTFT**
  $D_{HR}(x,P) = min\{P(X \in H_x), P(X \in E_x)\}$ where $H_x = \{y \in \mathcal{F} : y(v) \leq x(v)$ for all $v \in V\}$ and $E_x = \{y \in \mathcal{F} : y(v) \geq x(v)$ for all $v \in V\}$ on $\mathcal{C}$ with sup norm
- modified half-region depth (Lopez-Pintado and Romo(2011)) : **TTFTFT**
  $D_{MHR}(x,P) = min\{E(\lambda\{v \in V, X(v) \leq x(v)\}), E(\lambda\{v \in V, X(v) \geq x(v)\})\}/\lambda(V)$ on $\mathcal{C}$ with sup norm

Consistency of functional depth: classification of existing functional depth

For showing consistency, classify depths to 3 groups (Stanislav Nagy(2018)) Let D: some depth in $\mathbb{R}^p$. then

- integrated depth (Fraiman, Muniz(2001) and Cuevas, Fraiman(2009))
  form of $FD(x, P) = \int D(f(x), f(P) d\lambda(f))$
- infimal depth (Mosler(2013))
  form of $ID(x, P) = inf_f D(f(x), f(P))$
- band depth (Lopez-Pintado, Romo(2009))
  form of $BD(x, P) = P(x \in Band(X_1, ..., X_K))$ on $\mathcal{C}$ where
  $Band(x_1, x_2) = \{y \in \mathcal{C} : min\{x_1(v), x_2(v)\} \leq y(v) \leq max\{x_1(v), x_2(v)\}, v \in V\}$
  (extend to convex hull with many X_is.)

Consistency of functional depth

Definition

For given $P \in \mathcal{P}$, let $P_n \to P$ weakly. A functional depth $D(x, P)$ is uniformly consistent for P over $\mathcal{F}$, if

$$sup_{x \in \mathcal{F}} |D(x, P_n) - D(x, P)| \to 0$$

for almost every x as $n \to \infty$.

Definition

If D is uniformly consistent for any $P \in \mathcal{P}$, then we say D is universally consistent over $\mathcal{F}$.

Theorem (Varadarajan(1956))

Let (S, d) be a sparable metric space and μ be any distribution (Borel probability measure) on S. Then the empirical measure μ_n converges to μ almost surely:

$$P(\{w : \mu_n(.)(w) \to \mu\}) = 1$$

Consistency of functional depth

Theorem (Consistency of functional band depth (Gijbels, Nagy(2015)))

$BD(x, P)$ is not uniformly consistent over compact subset of $\mathcal{C}$.

Possible remedy: smoothing with integration and decreasing function $w : [0, \infty) \to [0, 1], w(0) = 1, w(\infty) \to 0$
Adjusted band depth: $aBD(x, P) = Ew(inf_{y \in Band(X_1,\ldots,X_k)}||x - y||)$ for all $x \in \mathcal{C}, P \in \mathcal{P}$. Then, aBD is universally consistent over $\mathcal{C}$.

Theorem (Consistency of functional infimal depth (Gijbels, Nagy(2015)))

$ID(x, P)$ is uniformly consistent over $\mathcal{C}$ for P
when P is mixture of P_1, P_2 s.t.

- *all marginal distribution of P_1 have continuous dist. functions.*
- *P_2 is concentrated in finite-dimensional subspace of $\mathcal{C}$.*

Note that the conditions are too restrictive. (Wiener measure fails to satisfy them.) And it means that $ID(x, P)$ is not universally consistent over $\mathcal{C}$.

Consistency of functional depth

Theorem (Consistency of functional integrated depth (Nagy, Gijbels, Omelka, Hlubinka(2016)))

$FD(x, P)$ is uniformly consistent over $\mathcal{C}$.

Note that, using the definition of integration, $\mathcal{C}$ can be extend to Borel-measurable (may be discontinuous) functions, include $\mathcal{L}^2$.

Consistency of functional depth: In practice

Theorem (Consistency over partial observability, (Nagy,Ferraty(2018))

Let $P \in \mathcal{P}$ on $\mathcal{L}^2[0,1]$ and $\tilde{P}_n$ be empirical distribution of fitted n curves. Then (under some assumptions,)

$$sup_{x \in \mathcal{L}^2} |D(x, \tilde{P}_n) - D(x, P)| \to 0$$

almost every x as $n \to \infty$ when D is adjust band depth type, h-depth type. If all marginal distribution of P is absolutely continuous, then also true for integrated depth type.

Proof:
step1: show $\tilde{P}_n \to P$ weakly almost every $\omega \in \Omega$ using Varadarajan theorem and some good properties of fitting kernel.
step2: using convergence property of inner D, show outer D converges weakly.

Consistency of functional depth: In practice

Theorem (convergence rate of FD (Nagy,Ferraty(2018)))

Let P_n be empirical distribution of (true) n curves, and $\tilde{P}_n$ be one of fitted n curves.
Suppose $P(|X(s) - X(t)| \leq L|s-t|^{\beta}) = 1$ for all $s, t \in [0,1]$.
Then, for any $P \in \mathcal{P}$ on $\mathcal{L}^2[0,1]$, under some conditions,

$$sup_{x \in \mathcal{L}^2[0,1]} |FD(x, P_n) - FD(x, P)| = O_p(n^{-1/2})$$

Moreover, if number of data points of n-th curve is comparable to n^r and $sup_{v \in [0,1]} sup_{|s-s'| \leq \epsilon} |F_{(v)}(s) - F_{(v)}(s')| \leq K\epsilon^{\alpha}$ for some $\alpha \in (0,1]$ where $F_{(v)}$: marginal cdf of P at v, then under some conditions,

$$\begin{aligned} &sup_{x \in \mathcal{L}^2[0,1]} |FD(x, \tilde{P}_n) - FD(x, P)| \\ &= O_p(n^{-r\alpha\beta/\{(1+\alpha)(2\beta+1)\}}) \text{ if } r < (2\beta+1)/\beta \\ &= O_p(\{ln(n)/n\}^{\alpha/(1+\alpha)}) \text{ if } r = (2\beta+1)/\beta \\ &= O_p(n^{-\alpha/(1+\alpha)}) \text{ if } r > (2\beta+1)/\beta \end{aligned}$$

Note that last case is dense setting, and the rate is similar to full observing case. In other cases, become slower.

Application

- Median estimation
- Robust and Nonparametric functional statistics
  procedure of with rank, nonparametric estimation of distribution or summary statistic, ...
- Exploratory Data Analysis (EDA)
  outlier detection, data expression (ex. functional box plot), ...
  (Center? Cluster? Symmetry? range(width)? gap(separation)? other irregularities?)
- classification
  when data can be classified by relation to the center.
  (if needed, after some transformation)
- (and other things...)

Note: Usability in application yields some other criteria about comparing depth.
eg. 1. width(using depth) vs std relation? 2. validity of central region? 3. computational advantage?