Functional depths

Choi Seokjun

18 Dec. 2019

Contents

1 Definition of the depth

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

As preliminary, I firstly start with introducing a definition of statistical depth on $\mathbb{R}^p$.

Definition 1.1 (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 $P \in \mathcal{P}$.

Any function $D(.,.)$ that satisfies above properties can be used as statistical depth in $\mathbb{R}^p$. I note that, although the definition is made by using population

1

distribution measure but in practice we do not know about true one, we can replace it by empirical distribution measure when n curves observed, (denote P_n) and can calculate sample depth from that empirical distribution. For validity to this replacement, see below part dealing with consistency.

Next, belows are not necessary, but desirable property when setting statistical depth function D: suggested by Serfling(2006). It is notable because it shows some properties of statistical depth more intuitively than definition does.

- 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-concavity as a function of x The set $\{x : D(x, P) \geq c\}$ is convex for each real c.

I give some examples about some 'toy' depth in $\mathbb{R}^1$.

Example 1.1 (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 $\mathcal{F}$

Now, I will introduce the definition of the functional depth, as one of the latest expansions of ordinary depth suggested by Nieto-Reyes and Battey.

Definition 1.2 (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 $\mathcal{F}$, d be metric 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:

1. 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 $y \in \mathcal{F}$, $d(f(x), f(y)) = a_f d(x, y)$, $a_f \in \mathbb{R} - \{0\}$.

2. Maximality at center For any $P \in \mathcal{P}$ with unique center of symmetry θ w.r.t. some notion of symmetry, $D(\theta, P) = sup_{x \in \mathcal{F}} D(x, P)$.

3. 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 $max\{d(y, z), d(y, x)\} < d(x, z)$.

4. Upper semi-continuity in x $D(x, P)$ is upper semi-continuous as a function of x.

5. 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 $\lambda(L_\delta^c) > 0$.

6. 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 $\mathcal{P}$.

I should give an caveat. The definition of functional depth is being studied for now, so there are still different views about it and many researchers are trying to suggest better property bundle for it. So, above definition are not concrete, not absolute-unique thing.

For other suggestion, next ones are often viewed :

- '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)

- 'maximal value at 0 on some Gaussian-type processes (or curves).

Additionally, I comment for each property of above definition.

1. It is direct expansion of 'Affine invariance' from the definition of depth on $\mathbb{R}^p$. It means that depth should be invariant with transformation without changing centrality. Especially, with just rescaled distance metric.

2. It is also direct expansion of 'Maximality at center' the definition on $\mathbb{R}^p$'s. Since we are working in metric space, the condition is rewritten like some form of triangle inequality using distance. And I note that, it implies 'Maximality at 0 with gaussian process' of one item of other suggestion.

   But, if strictly speaking, in general the notion - the center of the distribution on function space - has much ambiguousness itself, especially if we consider other distributions except gaussian or trivial ones with properties like stationary, having obvious mean, symmetric, continuous, etc. So, We can say that the definition above has similar problem.

3. It is expansion of 'monotonicity relative to the deepest point' and 'vanishing at infinity'. In fact, the strictness is too restrictive. If there isn't, it does not violation of the notion of the depth.

   But, in practice, because some suggested depths in past have some 'degeneracy problem': all curves have same depth despite having different level of centrality at some specific distributions, this strictness may help to exclude them that make a situation like degeneracy.

   And, the strictness can be used for more efficient application like classification. Moreover, with second property, this property implies 'null at boundary'.

4. This property is for relation between the depth and the distribution measure, especially cumulative distribution function. Since cdf has upper-semi continuity property, for using the depth to estimate the cdf, the depth have better to have this property, too.

5. This is property that Nieto-Reyes and Battey, who suggest this definition, add newly. Their intention is, they want for the depth to have some robustness that having low influence level in their depth of curve with 'small difference', when some curves are varying in part of domain which data exhibit little variability,

But I think the mathematical expression of the condition has too restrictive part, when think the validity of strict inequality, considering to their intention.

6. It is for consistency of the depth. When empirical distribution to true distribution, if their depth is not converge, it may be ridiculous thing. For excluding these situation, we set proper depth function having property that converges well when both cases that the number of the curves is larger and the number of data points grows up.

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

Below are some 'classical' functional depths. (They are frequently cited at other authors' papers for comparing suggested depth by author of itself.) Bold letters indicates whether each depth satisfying above 1-6 properties respectively.

For notation, $\mathcal{L}^2$ means the space of square integrable function with its norm, and $\mathcal{C}$ means the space of continuous function. Note that, $\mathcal{C}$ equipping sup norm becomes a complete normed vector space (Banach space) without inner product, but it is not problem because above definition is on metric space which is more general setting than inner product space or normed space.

- 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

2 Consistency of functional depth

For showing consistency, I firstly follow the classification of depths to 3 groups by Stanislav Nagy(2018).

From this line, let D be 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.)

These classification do NOT embrace all depths suggested until now, but I'll only focus above three-type-case for simple discussion I do below.

Consistency of functional depth

There are some notions about consistency for statistical depths.

Definition 2.1. 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 2.2. If D is uniformly consistent for any $P \in \mathcal{P}$, then we say D is universally consistent over $\mathcal{F}$.

The consistency theorems depend on below theorem, guaranteeing that empirical distribution from samples converges to true population distribution.

Theorem 2.1 (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:*

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

The proof can be found by many books about probability theory, like 'Real analysis and probability' by Dudley.

I'll introduce only the statements of consistency theorems without proof, since they are too long to shift all thing to here. If need, see of each paper's.

Theorem 2.2 (Consistency of functional band depth (Gijbels, Nagy(2015))). $BD(x, P)$ *is not uniformly consistent over compact subset of* $\mathcal{C}$.

Because the band type depths are naturally expanded from multi-finite dimensional one and use very intuitive notion, Many researchers try to adjust it to satisfy consistency.

One possible remedy is here: smoothing with integration and decreasing function $w : [0, \infty) \to [0, 1], w(0) = 1, w(\infty) \to 0$ Adjusted band depth: $aBD(x, P) = Ew(inf_{y \in Band(X_1,...,X_k)} ||x - y||)$ for all $x \in \mathcal{C}, P \in \mathcal{P}$. Then, aBD is universally consistent over $\mathcal{C}$. But as cost, lose some simplicity which original one has.

Next, here is the consistency theorem for infimal type depths.

Theorem 2.3 (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. We can observe that $ID(x, P)$ is not universally consistent over $\mathcal{C}$. What is worse, even Wiener measure (Gaussian measure) fails to satisfy them. To consider intuitively, see Brownian motion processes start at 0. Then they are same depth (same boundary level!) regardless of each of behavior at rear. And this degeneracy cannot be solved even if the number of curves become larger.

Finally, this is the consistency theorem for integrated type depths.

Theorem 2.4 (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

Since in practice we can get only finite data points for each curves, we need two direction's convergence property. The first is of the direction of the number of curves grows, and the second is of the direction of the number of data points grows for each curve. It is so complicate to deal with, so there are a little proofs of consistency for only special types of depths, but not for general depths. Below theorem is only for adjust band depth type, h-depth type and integrated type depth.

Theorem 2.5 (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. (either approximated by using kernel smoothing or linear combination of basis on separable spaces.) 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.

The 'some conditions' are too technical for purpose to get intuition, so I intentionally skip them. But I should note that they include some conditions on smoothing kernel which is used for fit to curve from data points, and they make the depth value calculated on approximated curves converge true depth value on population's distribution. If want, see 'Nagy, S., & Ferraty, F. (2019). Data depth for measurable noisy random functions. Journal of Multivariate Analysis, 170, 95-114.'. The proof is proceeded 2 steps. I briefly preview these.

1. show $\tilde{P}_n \to P$ weakly almost every $\omega \in \Omega$ using Varadarajan theorem and some good properties of fitting kernel.

2. using the convergence property of inner D (in $\mathbb{R}^p$), show that expanded D to $\mathcal{F}$ also converges weakly to true P together with the specific form of the depth.

Lastly, I give some convergence rates that are proved lately by some researchers.

Theorem 2.6 (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,

$$sup_{x \in \mathcal{L}^2[0,1]}|FD(x, \tilde{P}_n) - FD(x, P)|$$

$$= O_p(n^{-r\alpha\beta/\{(1+\alpha)(2\beta+1)\}}) \text{ if } r < (2\beta+1)/\beta$$

$$= O_p(\{ln(n)/n\}^{\alpha/(1+\alpha)}) \text{ if } r = (2\beta+1)/\beta$$

$$= O_p(n^{-\alpha/(1+\alpha)}) \text{ if } r > (2\beta+1)/\beta$$

Note that last case means some similar situation of the dense setting case. And we can observe that the rate is similar to full observing case. (think $\alpha = 1$ case: the marginal distribution's convergence is fast enough.) In other cases such as case of the sparse data points, the convergence becomes slower.

3 Application

Application

- Median estimation

- Robust and Nonparametric functional statistics   procedure of with rank (eg. rank test on functional data), 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...)

Now there are many researchers that broaden or newly opening analysis of functional data using depth. And right now, another applications are being developed in real time. So, I expect that the notion of the functional depth will be used wider at even near future.

Finally note that, the Usability in application yields some other criteria about defining and comparing depth such as computational advantage or the validity of central region or value (besides order) of specific depth. For example, a classification problem as a crucial application of depth, the latter is so important. For this reason, there is the trend when suggesting a new depth, not only a theoretical property, but also an applicability also become more to refer.

4 Reference

main reference

- Nieto-Reyes, A., & Battey, H. (2016). A topologically valid definition of depth for functional data. Statistical Science, 31(1), 61-79.

- Nagy, S., & Ferraty, F. (2019). Data depth for measurable noisy random functions. Journal of Multivariate Analysis, 170, 95-114.

other reference

- Fraiman, R., & Muniz, G. (2001). Trimmed means for functional data. Test, 10(2), 419-440.

- Cuevas, A., Febrero, M., & Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. Computational Statistics, 22(3), 481-496.

- Cuevas, A., & Fraiman, R. (2009). On depth measures and dual statistics. A methodology for dealing with general data. Journal of Multivariate Analysis, 100(4), 753-766.

- Lopez-Pintado, S., & Romo, J. (2009). On the concept of depth for functional data. Journal of the American Statistical Association, 104(486), 718-734.

- Lopez-Pintado, S., & Romo, J. (2011). A half-region depth for functional data. Computational Statistics & Data Analysis, 55(4), 1679-1695.

- Long, J. P., & Huang, J. Z. (2015). A study of functional depths. arXiv preprint arXiv:1506.01332.

- Gijbels, I., & Nagy, S. (2015). Consistency of non-integrated depths for functional data. Journal of Multivariate Analysis, 140, 259-282.

- Narisetty, N. N., & Nair, V. N. (2016). Extremal depth for functional data and applications. Journal of the American Statistical Association, 111(516), 1705-1714.

- Hubert, M., Rousseeuw, P., & Segaert, P. (2017). Multivariate and functional classification using depth and distance. Advances in Data Analysis and Classification, 11(3), 445-466.

book reference

- Dudley, R.M., Real analysis and Probability, (1989). Cambridge. ISBN 978-0521007542

- Aneiros, G., Bongiorno, E.G., Cao, R., & Vieu, P., Functional statistics and related field (2017). Springer. ISBN 978-3-319-55846-2