Large sample properties of the regression depth induced median

Yijun Zuo

Department of Statistics and Probability, Michigan State University
East Lansing, MI 48824, USA
zuo@msu.edu

July 12, 2020

Abstract

Notions of depth in regression have been introduced and studied in the literature. Regression depth (RD) of Rousseeuw and Hubert (1999), the most famous one, is a direct extension of Tukey location depth (Tukey, 1975) to regression.

Like its location counterpart, the most remarkable advantage of the notion of depth in regression is to directly introduce the maximum (or deepest) regression depth estimator (aka depth induced median) for regression parameters in a multi-dimensional setting. Classical questions for the regression depth induced median include (i) is it a consistent estimator (or rather under what sufficient conditions, it is consistent)? and (ii) is there any limiting distribution?

Bai and He (1999) (BH99) pioneered an attempt to answer these questions. Under some stringent conditions on (i) the design points, (ii) the conditional distributions of y given $\boldsymbol{x}_i$, and (iii) the error distributions, BH99 proved the strong consistency of the depth induced median. Under another set of conditions, BH99 showed the existence of the limiting distribution of the estimator.

This article establishes the strong consistency of the depth induced median without any of the stringent conditions in BH99, and proves the existence of the limiting distribution of the estimator by sufficient conditions and an approach different from BH99.

AMS 2000 Classification: Primary 62G09; Secondary 62G05, 62G15 62G20.

Key words and phrase: regression depth, maximum depth estimator, depth induced median, consistency, limiting distribution.

Running title: Asymptotic theorems for the regression depth median.

1 Introduction

Depth notions in location have received much attention in the literature. In fact, data depth and its applications remain one of the most active research topics in statistics in the last three decades. Most favored notions of depth in location include (i) halfspace depth (HD)(Tukey, 1975, popularized by Donoho and Gasko, 1992), (ii) simplicial depth (Liu, 1990), and (iii) projection depth (PD) (Liu, 1992 and Zuo and Serfling, 2000, promoted by Zuo, 2003), among others.

Depth notions in regression have also been sporadically proposed. Regression depth (RD) (Rousseeuw and Hubert, 1999) (RH99), the most famous one, is a direct extension of Tukey HD to regression. Others include Carrizosa depth (Carrizosa, 1996) and the projection regression depth induced from Marrona and Yohai, 1993 (MY93) and proposed in Zuo, 2018 (Z18). The latter turns out to be the extension of PD to regression.

One of the prominent advantages of depth notions is that they can be directly employed to introduce median-type deepest estimating functionals (or estimators in the empirical case) for the location or regression parameters in a multi-dimensional setting based on a general min-max stratagem. The maximum (or deepest) regression depth estimator serves as a *robust* alternative to the classical least squares or least absolute deviations estimator of the unknown parameters in a general linear regression model. The latter can be expressed as

$$y = \mathbf{x}'\boldsymbol{\beta} + e, \tag{1}$$

where $'$ denotes the transpose of a vector, and random vector $\mathbf{x} = (x_1, \cdots, x_p)'$ and parameter vector $\boldsymbol{\beta}$ are in $\mathbb{R}^p$ ($p \geq 2$) and random variables y and e are in $\mathbb{R}^1$. If $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1')'$ and $x_1 = 1$, then one has $y = \beta_0 + \mathbf{x}_1'\boldsymbol{\beta}_1 + e$, where $\mathbf{x}_1' = (x_2, \cdots, x_p) \in \mathbb{R}^{p-1}$. Let $\mathbf{w}' = (1, \mathbf{x}_1')$. Then $y = \mathbf{w}'\boldsymbol{\beta} + e$. We use this model or (1) interchangeably depending on the context.

The maximum regression depth estimator possesses the outstanding robustness feature similar to the univariate location counterpart. Indeed, the maximum depth estimator induced from RD, could, asymptotically, resist up to 33% contamination without breakdown, in contrast to 0% for the classical estimators (see Van Aelst and Rousseeuw, 2000) (VAR00).

The asymptotics of the maximum regression depth estimator (denoted by T^*_{RD}, or $\boldsymbol{\beta}^*$) have been considered and established in Bai and He, 1999 (BH99). Under some stringent conditions on (i) the design points, (ii) the conditional distributions of y given $\boldsymbol{x}_i$, and (iii) the error distributions, BH99 proved the strong consistency of the maximum depth estimator. Under another set of conditions, BH99 showed the existence of the limiting distribution of the estimator. This article establishes the strong consistency of the maximum depth estimator without any of the stringent conditions in BH99 and proves the existence of the limiting distribution of the estimator with conditions and an approach very different from BH99.

The rest of article is organized as follows: Section 2 introduces the RD of RH99 and presents pioneering examples of the computation of the RD for population distributions. Section 3 summarizes the important results from Z18 on the RD which are used in later sections. Section 4 establishes the strong and root-n consistency of the T^*_{RD}. Section 5 is devoted to the establishment of limiting distribution of the T^*_{RD}, where the main tool is the Argmax continuous mapping theorem. Assumptions for the theorem to hold are verified via empirical process theory and especially stochastic equicontinuity and VC-classes of functions. The limiting distribution is characterized through an Argmax operation over the infimum of a function involving a Gaussian process.

2 Regression depth of Rousseeuw and Hubert (1999)

Definition 2.1 For any $\boldsymbol{\beta}$ and joint distribution P of $(\mathbf{x}, y)$ in (1), RH99 defined the regression depth of $\boldsymbol{\beta}$, denoted by $\mathrm{RD}(\boldsymbol{\beta}; P)$, to be the minimum probability mass that needs to be passed when titling (the hyperplane induced from) $\boldsymbol{\beta}$ in any way until it is vertical. The maximum regression depth estimating functional T^*_{RD} (also denoted by $\boldsymbol{\beta}^*$) is defined as

$$T^*_{RD}(P) = \operatorname*{argmax}_{\boldsymbol{\beta}\in\mathbb{R}^p} \mathrm{RD}(\boldsymbol{\beta}; P). \tag{2}$$

If there are several $\boldsymbol{\beta}$ that attain the maximum depth value on the right hand side (RHS) of (2), then the average of all those $\boldsymbol{\beta}$ is taken.

The $\mathrm{RD}(\boldsymbol{\beta}; P)$ definition above is rather abstract and not easy to comprehend. Some characterizations, or equivalent definitions of $\mathrm{RD}(\boldsymbol{\beta}; P)$ are summarized below. *In the empirical case, the RD defined originally in RH99, divided by n, is identical to the following.*

Lemma 2.1. The following statements for RD are equivalent.

(i) [Z18]

$$\mathrm{RD}(\boldsymbol{\beta}; P) = \inf_{\boldsymbol{\alpha}\in S(\boldsymbol{\beta})} P\left(|r(\boldsymbol{\beta})| \le |r(\boldsymbol{\alpha})|\right), \tag{3}$$

where $S(\boldsymbol{\beta}) := \{\boldsymbol{\alpha} \in \mathbb{R}^p : \ H_{\boldsymbol{\alpha}} \text{ intersects with } H_{\boldsymbol{\beta}}\}$ for a given $\boldsymbol{\beta}$, $H_{\boldsymbol{\gamma}}$ denotes the unique hyperplane determined by $y = \mathbf{w}'\boldsymbol{\gamma}$, and $r(\boldsymbol{\gamma}) := y - (1, \mathbf{x}')\boldsymbol{\gamma} := y - \mathbf{w}'\boldsymbol{\gamma}$.

(ii) [Z18]

$$\mathrm{RD}(\boldsymbol{\beta}; P) = \inf_{\|\mathbf{v}_2\|=1,\ v_1\in\mathbb{R}} E\Big(\mathbf{I}\big(r(\boldsymbol{\beta})(v_1, \mathbf{v}_2')\mathbf{w} \ge 0\big)\Big) = \inf_{\mathbf{v}\in\mathbb{S}^{p-1}} E\Big(\mathbf{I}\big(r(\boldsymbol{\beta})\mathbf{v}'\mathbf{w} \ge 0\big)\Big), \tag{4}$$

where, $\mathbb{S}^{p-1} := \{\mathbf{u} \in \mathbb{R}^p : \|\mathbf{u}\| = 1\}$ and $\mathbf{I}(A)$ (and throughout) stands for the indicator function of the set A. ∎

Other characterizations are also given in the literature, e.g., in VAR00, in Rousseeuw and Struyf, 2004 (RS04), in Adrover, Maronna, and Yohai (2002), in Mizera (2002) (pages 1689-1690) and in BH99. The latter is specifically defined by

$$\mathrm{RD}(\boldsymbol{\beta}; P_n) = \inf_{\|\mathbf{u}\|=1,\ v\in\mathbb{R}} \min\left\{\sum_{i=1}^n \mathbf{I}(r_i(\boldsymbol{\beta})(\mathbf{u}'\mathbf{x}_i - v) > 0), \sum_{i=1}^n \mathbf{I}(r_i(\boldsymbol{\beta})(\mathbf{u}'\mathbf{x}_i - v) < 0)\right\}, \tag{5}$$

Furthermore, BH99 depended solely on the following alternative definition:

$$\mathrm{RD}(\boldsymbol{\beta}; P_n) = \frac{n}{2} + \frac{1}{2} \inf_{\boldsymbol{\gamma}\in\mathbb{S}^{p-1}} \sum_{i=1}^n \mathrm{sgn}(y_i - \mathbf{w}_i\boldsymbol{\beta})\mathrm{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}). \tag{6}$$

Remarks 2.1

(I) If one assumes that $P(\mathbf{x}'\mathbf{u} = v) = 0$ and $P(r(\boldsymbol{\beta}) = 0) = 0$ for any $\mathbf{u}$, v, and $\boldsymbol{\beta}$, then Definition (5) of BH99 above is *identical* (a.s.) to the original definition of RH99.

(II) Generally, definition (6) is neither identical to the RD of RH99, nor to (5). For example, assume that we have four sample points in $\mathbb{R}^2$ which could be regarded as from a continuous or discrete $\mathbf{Z}$, $\mathbf{Z}_1 = (\frac{1}{8}, 1)'; \mathbf{Z}_2 = (\frac{4}{8}, 0)'; \mathbf{Z}_3 = (\frac{6}{8}, -1)'; \mathbf{Z}_4 = (\frac{7}{8}, 2)'$. Then it is readily seen that for $\boldsymbol{\beta} = (0, 0)'$, RH99 gives RD= 2, which divided by $n = 4$ leads to 1/2 (identical to Def. 2.1), (5) gives 1 whereas (6) yields 1.5 (see Figure 1 for the scatterplot and the line). ■

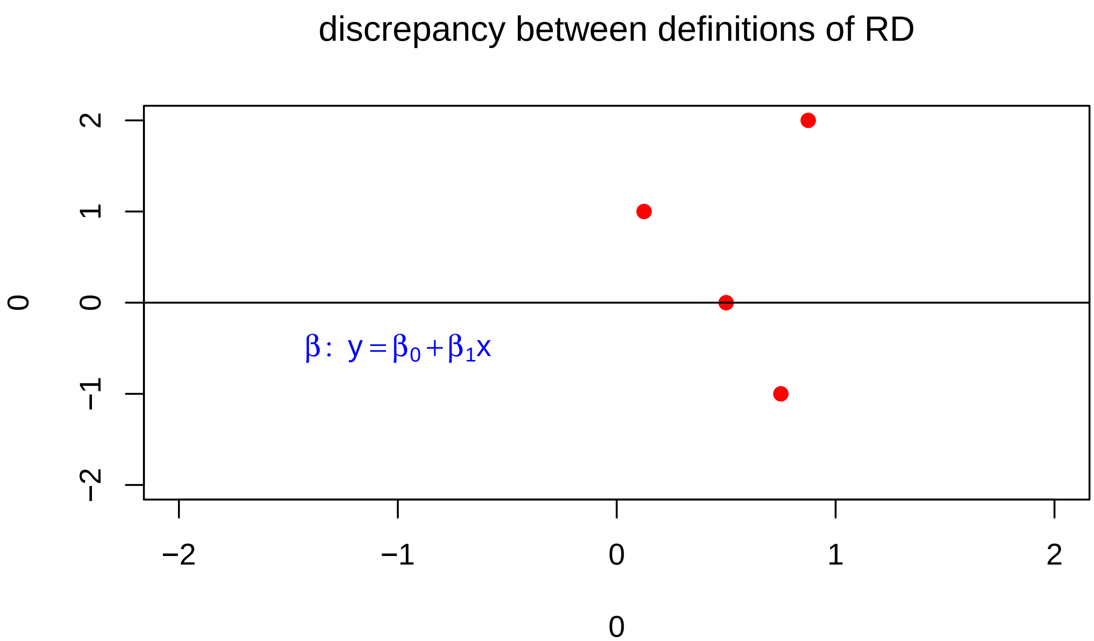

Figure 1: $\boldsymbol{\beta} = (0, 0)'$, the horizontal candidate regression line. RH99 gives its RD=2 while $\text{RD}(\boldsymbol{\beta}, F_n) = 1/2$ (Def. 2.1), RD of (5) gives 1, whereas RD of (6) yields 1.5.

For empirical distributions ($P = P_n$), computing $\text{RD}(\boldsymbol{\beta}, P)$ is quite straightforward and examples have been given in RH99. For a general distribution (probability measure) P, concrete examples of expression of $\text{RD}(\boldsymbol{\beta}, P)$ are not yet given in the literature herebefore. For special classes of distributions, however, one could derive the explicit expression for $\text{RD}(\boldsymbol{\beta}, P)$. In the examples below, for simplicity, we again confine our attention to the case $p = 2$. That is, we have a simple linear regression model $y = \beta_0 + \beta_1 x + e$.

Example 2.1 A random vector $\mathbf{X} \in \mathbb{R}^p$ is said to be *elliptically distributed*, denoted by $\mathbf{X} \sim E(h; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, if its density is of the form

$$f(\mathbf{x}) = c|\boldsymbol{\Sigma}|^{-1/2} h\left((\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right),\ \mathbf{x} \in \mathbb{R}^p, \tag{7}$$

where c is the normalizing constant, $\boldsymbol{\mu}$ is the coordinate-wise median vector (or the mean vector if it exists), $\boldsymbol{\Sigma}$ is a positive definite matrix which is proportional to the covariance matrix if it exists. Generally h is a known function. A straightforward transformation such as $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$ leads to $\mathbf{Z} \sim E(h; \mathbf{0}, \boldsymbol{I_p})$.

To seek concrete expression for $\text{RD}(\boldsymbol{\beta}, P)$ and for the simplicity we restrict to the case $h(x) = \exp\big(-x^2/2\big)$, i.e., the bivariate normal class (laplace, logistic and t classes could be treated similarly). Namely, we have $(x, y) \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. After applying the transformation above, we can assume without loss of generality (w.l.o.g.) that $(x, y) \sim N_2(\mathbf{0}, I_2)$, where I_2 is a 2 by 2 identity matrix. For any $\boldsymbol{\beta} = (\beta_0, \beta_1)'$, by the invariance of RD (see Z18 and Section 3), we can consider the depth of $\boldsymbol{\beta}$ w.r.t. the P that corresponds to the $N_2(\mathbf{0}, I_2)$.

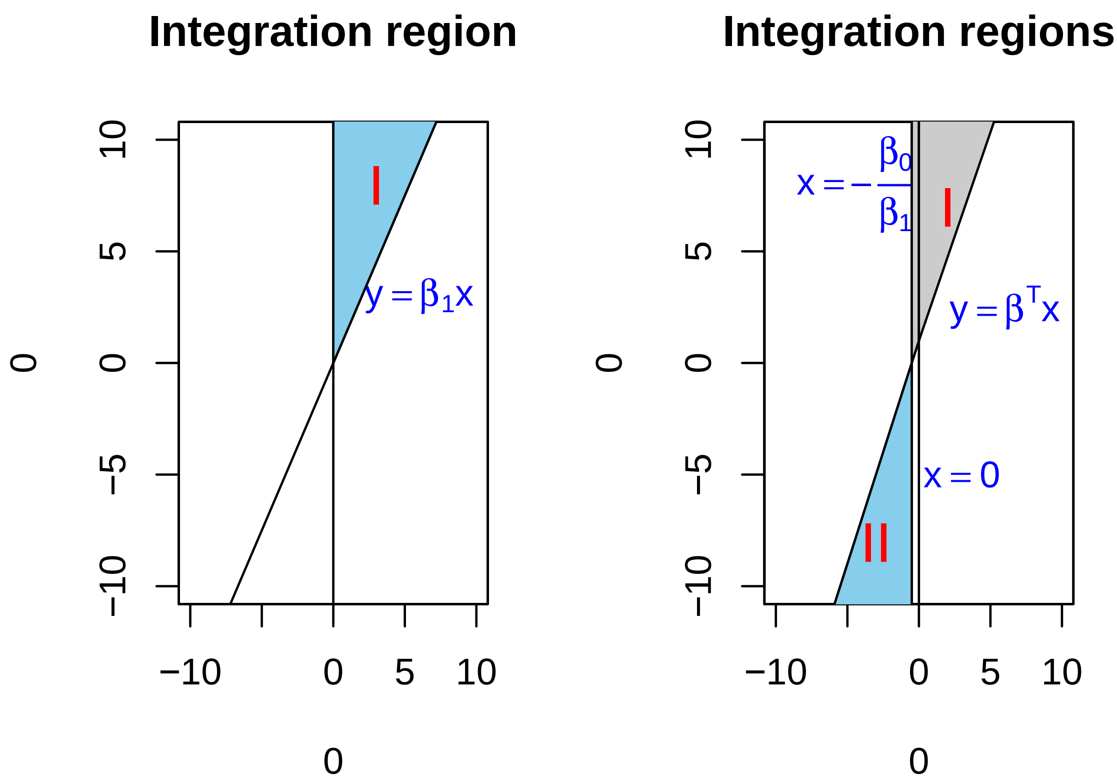

Figure 2: Integration regions. Left: for the region in (ii). Right: for the regions in (iii).

(i) $\boldsymbol{\beta} = (0, 0)'$, then the regression line is $y = 0$, and $\text{RD}(\boldsymbol{\beta}; P) = 1/2$.

(ii) $\beta_0 = 0$ and $\beta_1 > 0$ (the case $\beta_1 < 0$ can be discussed similarly). Denote the region bounded by the regression line $y = \beta_1 x$ and the positive y-axis as I (see the left side of Fig. 2), then it is readily seen that

$$\text{RD}(\boldsymbol{\beta}; N(\mathbf{0}; I_2)) = 2P((y, x) \in \text{I}) = 1 - 2\int_0^\infty \Phi(\beta_1 x) d\Phi(x), \tag{8}$$

where $\Phi(x)$ is the standard normal cumulative distribution function.

(iii) $\beta_0 > 0$ and $\beta_1 > 0$ (the case $\beta_0 > 0$ and $\beta_1 < 0$ and the cases where $\beta_0 < 0$ can be treated similarly). Denote the region formed by the line with positive y part of the vertical line $x = -\beta_0/\beta_1$,$\{x \geq -\beta_0/\beta_1, y \geq \beta_0 + \beta_1 x\}$ as I and with negative y part of the vertical line $x = -\beta_0/\beta_1$ $\{x \le -\beta_0/\beta_1, y \le \beta_0 + \beta_1 x\}$ as II (see the right side of Fig.2), then it is readily seen

$$
\begin{aligned}
&\mathrm{RD}(\boldsymbol{\beta}; N(\mathbf{0}; I_2)) = P((y,x) \in \mathrm{I}) + P((y,x) \in \mathrm{II}) \\
&= 1 - 2\Phi(-\beta_0/\beta_1) + \int_{-\infty}^{-\beta_0/\beta_1} \Phi\left(\beta_0 + \beta_1 x\right) d\Phi(x) \qquad - \int_{-\beta_0/\beta_1}^{\infty} \Phi\left(\beta_0 + \beta_1 x\right) d\Phi(x).
\end{aligned}
$$

(iv) $\beta_0 > 0$ and $\beta_1 = 0$ (the case $\beta_0 < 0$ and $\beta_1 = 0$ can be handled similarly). Denote the region formed by the line $y = \beta_0$ and the part of the positive y-axis $\{y \ge \beta_0\}$ as I then it is readily seen that

$$\mathrm{RD}(\boldsymbol{\beta}; N(\mathbf{0}; I_2)) \quad = \quad P((y,x) \in \mathrm{I}) = 1/2 - \Phi(\beta_0).$$

■

Example 2.2 Assume that (x, y) is uniformly distributed over a unit circle centered at (0,0). By invariance of depth, this will cover a class of distributions of $A(x,y)+b$ for any nonsingular $A \in \mathbb{R}^{2\times 2}$ and $b \in \mathbb{R}^2$.

(i) $\boldsymbol{\beta} = (0,0)'$, then the regression line is $y = 0$, and $\mathrm{RD}(\boldsymbol{\beta}; P) = 1/2$.

(ii) $\beta_0 = 0$ and $\beta_1 > 0$ ($\beta_1 < 0$ can be treated similarly). Denote the region bounded by the regression line $y = \beta_1 x$ and the positive y axis as I, then it is readily seen that

$$\mathrm{RD}(\boldsymbol{\beta}; P) = 2P((x,y) \in \mathrm{I}) = 1/2 - \frac{|\arctan(\beta_1)|}{\pi}. \tag{9}$$

(iii) $\beta_0 > 0$ and $\beta_1 \ge 0$ (the cases where ($\beta_0 > 0, \beta_1 < 0$) or ($\beta_0 < 0$, $\beta_1 \ge$ (or $<$) 0) can be dealt with similarly) and $\Delta = 1 + \beta_1^2 - \beta_0^2 > 0$. That is, the regression line intercepts the unit circle at two points $x_\pm$, where $x_\pm = \frac{-\beta_0\beta_1 \pm \sqrt{1+\beta_1^2-\beta_0^2}}{1+\beta_1^2}$.

(a) Assume that both intersection points have positive y coordinate. Denote the region formed by the regression line and the circle between the vertical lines $x = x_-$ and $x = x_+$, ($\beta_0 + \beta_1 x \le y \le \sqrt{1-x^2}$) as I. Then it is readily seen that

$$\mathrm{RD}(\boldsymbol{\beta}; P) = P((x,y) \in \mathrm{I}) = \int_{x_-}^{x_+} \left(\sqrt{1-x^2} - (\beta_0 + \beta_1 x)\right) dx = g(x_+) - g(x_-) \qquad ,$$

where $g(x) = \frac{1}{2}\left(x\sqrt{1-x^2} + \arcsin(x)\right) - \left(\beta_0 x + \frac{1}{2}\beta_1 x^2\right) := g_1(x) - g_2(x)$,

(b) Assume that the y coordinates of the two intersection points have different signs. The latter implies that $\beta_1 \ne 0$. Denote the region formed by the regression line and the circle and the positive (negative) y-part of vertical line $x = -\beta_0/\beta_1$ as I (II). Then it is readily seen that

$$\mathrm{RD}(\boldsymbol{\beta}; P) = P((x,y) \in \mathrm{I}) + P((x,y) \in \mathrm{II}) = g(x_+) - g(x_-) + 2g_2(-\beta_0/\beta_1) - 2g_2(x_-).$$

(iv) In all other cases, $\mathrm{RD}(\boldsymbol{\beta}; P) = 0$. ■

Remarks 2.1

(I) From the examples, it is readily seen that maximum value of RD is 1/2 (in fact, 1/2 is the maximum possible depth value in many cases, see RH99). Furthermore, the point $\boldsymbol{\beta} = (0,0)'$ is the unique point that attains the maximum depth value in both examples.

(II) According to RS04, we say $F_{\mathbf{Z}}$ is *regression symmetric* about $\boldsymbol{\beta}^* = (0,0)'$ in these examples, where $\mathbf{Z} := (x, y)$. We also have a unique T^*_{RD} or $\boldsymbol{\beta}^*$ in both cases. ■

3 Preliminary results

A regression depth functional D is said to be regression, scale and affine *invariant* w.r.t. a given $F_{(\mathbf{x},y)}$ if and only if (iff), respectively, $D(\boldsymbol{\beta}+\mathbf{b};\ F_{(\mathbf{x},\ y+\mathbf{x}'\mathbf{b})}) = D(\boldsymbol{\beta};\ F_{(\mathbf{x},\ y)}),\ \forall\ \mathbf{b} \in \mathbb{R}^p$; $D(s\boldsymbol{\beta};\ F_{(\mathbf{x},\ sy)}) = D(\boldsymbol{\beta};\ F_{(\mathbf{x},\ y)}),\ \forall s(\neq 0) \in \mathbb{R};\ D(A^{-1}\boldsymbol{\beta}; F_{(A'\mathbf{x},\ y)}) = D(\boldsymbol{\beta}; F_{(\mathbf{x},\ y)}), \forall A_{p\times p}$, a nonsingular matrix.

A regression estimating functional $T(\cdot)$ is said to be regression, scale, and affine *equivariant* iff, respectively, $T(F_{(\mathbf{x},\ y+\mathbf{x}'b)}) = T(F_{(\mathbf{x},\ y)}) + b,\ \forall\ b \in \mathbb{R}^p$; $T(F_{(\mathbf{x},\ sy)}) = sT(F_{(\mathbf{x},\ y)}),\ \forall\ s \in \mathbb{R}$; $T(F_{(A'\mathbf{x},\ y)}) = A^{-1}T(F_{(\mathbf{x},\ y)}),\ \ \forall\ $ nonsingular $A \in \mathbb{R}^{p\times p}$.

We now summarize some preliminary results on the RD and its induced maximum depth estimating functional. $F_{\mathbf{Z}}$ and P are used interchangeably and $\mathbf{Z} := (\mathbf{x},\ y)$.

Lemma 3.1 [Z18]

(i) $\mathrm{RD}(\boldsymbol{\beta}; F_{\mathbf{Z}})$ is regression, scale and affine invariant and hence $T^*_{RD}(F_{\mathbf{Z}})$ is regression, scale and affine equivariant. Furthermore, $\mathrm{RD}(\boldsymbol{\beta}; F_{\mathbf{Z}}) \to 0$ as $\|\boldsymbol{\beta}\| \to \infty$, if $P(H_v) = 0$ for any vertical hyperplane H_v.

(ii) $\mathrm{RD}(\boldsymbol{\beta}; P)$ is upper-semicontinuous and concave (in $\boldsymbol{\beta}$), and continuous in $\boldsymbol{\beta}$ if P has a density.

(iii) $\sup_{\boldsymbol{\beta}\in\mathbb{R}^p} |\mathrm{RD}(\boldsymbol{\beta}; F^n_{\mathbf{Z}}) - \mathrm{RD}(\boldsymbol{\beta}; F_{\mathbf{Z}})| \to 0$ almost surely (a.s.) as $n \to \infty$, where $F^n_{\mathbf{Z}}$ is the empirical version of the distribution $F_{\mathbf{Z}}$. ■

In the sequel, we assume that there exists a unique point $T^*_{RD}(F_{\mathbf{Z}})$ (or $\boldsymbol{\beta^*}$, a generic notation for the maximum regression depth point, whereas $\boldsymbol{\eta}$ is also used for the maximum (location or regression) depth point later) that maximizes the underlying regression depth RD. In virtue of (i) of Lemma 3.1, one can assume, w.l.o.g., that $T^*_{RD}(F_{\mathbf{Z}}) = \mathbf{0}$.

Uniqueness is guaranteed if $F_{\mathbf{Z}}$ has a strictly positive density and is regression symmetric about a point $\boldsymbol{\beta}$ ($F_{\mathbf{Z}}$ is *regression symmetric* about $\boldsymbol{\theta}$ if $P(\mathbf{x} \in B, r(\boldsymbol{\theta}) > 0) = P(\mathbf{x} \in B, r(\boldsymbol{\theta}) < 0)$ for any Borel set $B \in \mathbb{R}^{p-1}$, see RS04).

4 Consistency

For a general regression depth functional $D(\boldsymbol{\beta}; F_{\mathbf{Z}})$, let $\boldsymbol{\beta^*}(F_{\mathbf{Z}}) = \arg\max_{\boldsymbol{\beta}\in\mathbb{R}^p} D(\boldsymbol{\beta}; F_{\mathbf{Z}})$, then $\boldsymbol{\beta^*_n} := \boldsymbol{\beta^*}(F^n_{\mathbf{Z}})$ is a natural estimator of $\boldsymbol{\beta^*}$, the maximum regression depth functional.

Is $\boldsymbol{\beta}^*_{\boldsymbol{n}}$ a consistent estimator? This is a very typical question asked in statistics and the argument (or answer) for it is also very standard, almost to the point of cliché as Kim and Pollard (1990) (KP90) have commented.

Let us first deal with the problem in a more general setting. Let M_n be stochastic processes indexed by a metric space Θ of $\boldsymbol{\theta}$, and M: $\Theta \to \mathbb{R}$ be a deterministic function of $\boldsymbol{\theta}$ which attains its maximum at a point $\boldsymbol{\theta}_0$.

The sufficient conditions for the consistency of this type of problem were given in Van Der Vaart (1998) (VDV98) and Van Der Vaart and Wellner (1996) (VW96) and are listed below:

C1: $\sup_{\boldsymbol{\theta}\in\Theta} |M_n(\boldsymbol{\theta}) - M(\boldsymbol{\theta})| = o_p(1)$;

C2: $\sup_{\{\boldsymbol{\theta}\colon\, d(\boldsymbol{\theta},\boldsymbol{\theta_0})\geq\delta\}} M(\boldsymbol{\theta}) < M(\boldsymbol{\theta_0})$, for any $\delta > 0$ and the metric d on Θ;

Then any sequence $\boldsymbol{\theta}_n$ is consistent for $\boldsymbol{\theta}_0$ providing that it satisfies

C3: $M_n(\boldsymbol{\theta_n}) \geq M_n(\boldsymbol{\theta_0}) - o_p(1)$.

Lemma 4.1 [Th. 5.7, VDV98] If **C1** and **C2** hold, then any $\boldsymbol{\theta}_n$ satisfying **C3** is consistent for $\boldsymbol{\theta}_0$.

Remarks 4.1

(I) **C1** basically requires that the $M_n(\boldsymbol{\theta})$ converges to $M(\boldsymbol{\theta})$ in probability uniformly in $\boldsymbol{\theta}$. For the depth process $\mathrm{RD}(\boldsymbol{\beta}; F^n_{\mathbf{Z}})$ and $\mathrm{RD}(\boldsymbol{\beta}; F_{\mathbf{Z}})$, it holds true (the convergence here is actually almost surely (a.s.) and uniformly in $\boldsymbol{\beta}$).

(II) **C2** essentially demands that the unique maximizer $\boldsymbol{\theta_0}$ is well separated. This holds true as long as $D(\boldsymbol{\beta}; F_{\mathbf{Z}})$ is upper semi-continuous and vanishing at infinity, and $\boldsymbol{\theta_0}$ is unique (see, Lemma 4.2 below). Therefore, it holds for RD in light of Lemma 3.1.

(III) **C3** asks that $\boldsymbol{\theta_n}$ is very close to $\boldsymbol{\theta}_0$ in the sense that the difference of images of the two at M_n is within $o_p(1)$. In KP90 and VW96 a stronger version of **C3** is required:

C3' : $\quad M_n(\boldsymbol{\theta_n}) \geq \sup_{\theta\in\Theta} M_n(\boldsymbol{\theta}) - o_p(1),$

which implies **C3**. This strong version mandates that $\boldsymbol{\theta_n}$ *nearly* maximizes $M_n(\boldsymbol{\theta})$. The maximum regression depth estimator $\boldsymbol{\beta}^*_{\boldsymbol{n}}(:= \boldsymbol{\theta_n})$ is defined to be the maximizer of $M_n(\boldsymbol{\theta}) := D(\boldsymbol{\beta}; F^n_{\mathbf{Z}})$, hence **C3**' (and thus **C3**) holds automatically. ■

In light of above, $\boldsymbol{\beta}^*_{\boldsymbol{n}}$ induced from RD is consistent for $\boldsymbol{\beta}^*$. But, we have more.

Theorem 4.1 The maximum regression depth estimator $\boldsymbol{\beta}^*_{\boldsymbol{n}}$ induced from RD is *strongly* consistent for $\boldsymbol{\beta}^*$ (i.e., $\boldsymbol{\beta}^*_{\boldsymbol{n}} - \boldsymbol{\beta}^* = o(1)$ a.s.) provided that $\boldsymbol{\beta}^*$ is unique.

Proof: The proof for the consistency of Lemma 4.1 could be easily extended to the strong consistency with a strengthened version of **C1**

> **C1**': $\sup_{\boldsymbol{\theta}\in\Theta} |M_n(\boldsymbol{\theta}) - M(\boldsymbol{\theta})| = o(1)$ a.s.

In the light of the proof of Lemma 4.1, we need only verify the sufficient conditions **C1**' and **C2-C3**. By (III) of Remark 4.1, **C3** holds automatically, so we need to verify **C1**' and **C2**. **C1**' has been given in Lemma 3.1 for RD. So the only item left is to verify **C2** for RD which is guaranteed by Lemma 4.2 below. ■

Lemma 4.2 Assume that a general (location or regression) depth $D(\boldsymbol{\beta}; F_{\mathbf{Z}})$ is upper semi-continuous in $\boldsymbol{\beta}$ and vanishing when $\|\boldsymbol{\beta}\| \to \infty$. Let $\boldsymbol{\eta} \in \mathbb{R}^p$ be the unique point with $\boldsymbol{\eta} = \arg\max_{\boldsymbol{\beta}\in\mathbb{R}^p} D(\boldsymbol{\beta}; F_{\mathbf{Z}})$ and $D(\boldsymbol{\eta}; F_{\mathbf{Z}}) > 0$. Then for any $\varepsilon > 0$, $\sup_{\boldsymbol{\beta}\in N_\varepsilon^c(\boldsymbol{\eta})} D(\boldsymbol{\beta}; F_{\mathbf{Z}}) < D(\boldsymbol{\eta}; F_{\mathbf{Z}})$, where $N_\varepsilon^c(\boldsymbol{\eta}) = \{\boldsymbol{\beta} \in \mathbb{R}^P : \|\boldsymbol{\beta} - \boldsymbol{\eta}\| \geq \varepsilon\}$ and "c" stands for "complement" of a set.

Proof: Assume conversely that $\sup_{\boldsymbol{\beta}\in N_\varepsilon^c(\boldsymbol{\eta})} D(\boldsymbol{\beta}; F_{\mathbf{Z}}) = D(\boldsymbol{\eta}; F_{\mathbf{Z}})$. Then by the given conditions, there is a sequence of *bounded* $\boldsymbol{\beta_j}$ $(j = 0, 1, \cdots)$ in $N_\varepsilon^c(\boldsymbol{\eta})$ such that $\boldsymbol{\beta_j} \to \boldsymbol{\beta_0} \in N_\varepsilon^c(\boldsymbol{\eta})$ and $D(\boldsymbol{\beta_j}; F_{\mathbf{Z}}) \to D(\boldsymbol{\eta}; F_{\mathbf{Z}})$ as $j \to \infty$. Note that $D(\boldsymbol{\eta}; F_{\mathbf{Z}}) > D(\boldsymbol{\beta_0}; F_{\mathbf{Z}})$. The upper-semicontinuity of $D(\cdot; F_{\mathbf{Z}})$ now leads to a contradiction: for sufficiently large j, $D(\boldsymbol{\beta_j}; F_{\mathbf{Z}}) \leq (D(\boldsymbol{\eta}; F_{\mathbf{Z}}) + D(\boldsymbol{\beta_0}; F_{\mathbf{Z}}))/2 < D(\boldsymbol{\eta}; F_{\mathbf{Z}})$. This completes the proof. ■

Remarks 4.2

> (I) For RD, the sufficient conditions in the Lemma are all satisfied in virtue of Lemma 3.1 and the uniqueness of $\boldsymbol{\eta} = \boldsymbol{\beta}^*$ is guaranteed for special $F_{\mathbf{Z}}$ (see Section 3).
>
> (II) Besides the necessary uniqueness assumption of $\boldsymbol{\beta}^*$ here, BH99 under additional stringent conditions (see their D1-D4) on (i) design points $\mathbf{x}_i$, (ii) the conditional distributions of y given $\mathbf{x_i}$, and (iii) the distributions of error e_i, proved the strong consistency with a very different and unnecessarily complicated approach.
>
> (III) However, if i.i.d x_i is from a univariate Cauchy distribution or any other heavy-tailed ones, then all D1-D3 do not hold, Theorem 2.1 of BH99 is not applicable and no strong consistency result can be obtained meanwhile one can get the result via Theorem 4.1 above, nevertheless. This is the merit and necessity of Theorem 4.1. ■

With the establishment of strong consistency, one naturally wonders about the rate of convergence of the maximum regression depth estimator. Does it possess root-n consistency? To answer the question, we need a stronger version of **C2** for a general depth notion D.

> **C2**': For each small enough positive δ, there exists a positive constant κ such that $\sup_{\|\boldsymbol{\beta}-\boldsymbol{\eta}\|\geq\delta} D(\boldsymbol{\beta}; P) < \alpha^* - \kappa\delta$, where $\alpha^* := D(\boldsymbol{\eta}; P) = \sup_{\boldsymbol{\beta}\in\mathbb{R}^p} D(\boldsymbol{\beta}; P)$.

Remarks 4.3:

(I) When D in **C2**' is RD, Lemma 4.2 provides a choice for the individual κ for every δ. But **C2**' requires more. In the following we provide sufficient conditions for **C2**' to hold.

(II) (i) P has a density; (ii) $h(\boldsymbol{\beta}, \mathbf{v}) := E\left(\mathbf{I}(r(\boldsymbol{\beta})\mathbf{v}'\mathbf{w} \geq 0)\right)$ is differentiable in $\boldsymbol{\beta} \in N_{\boldsymbol{\eta}}$ for a given $\mathbf{v} \in \mathbb{S}^{p-1}$, where $N_{\boldsymbol{\eta}}$ is a small neighborhood of $\boldsymbol{\eta}$; (iii) the directional derivative of h along $\mathbf{u} \in \mathbb{S}^{p-1}$ at $\boldsymbol{\beta} \in N_{\boldsymbol{\eta}}$: $D_{\mathbf{u}}h(\boldsymbol{\beta}, \mathbf{v}) := \nabla h(\boldsymbol{\beta}, \mathbf{v}) \cdot \mathbf{u}$ is continuous in $\mathbf{u}$ and positive uniformly in $\mathbf{u}$, where ∇ is the vector differential operator and "$\cdot$" stands for the inner product; (iv) the Hessian matrix of h has a positive eigenvalues uniformly for $\boldsymbol{\beta}$ over $N_{\boldsymbol{\eta}}$. ■

Let $D(\boldsymbol{\beta}; P)$ be any *regression* (or even *location*) depth functional for $\boldsymbol{\beta} \in \mathbb{R}^p$. We have the following general result for $\boldsymbol{\beta}^*_{\boldsymbol{n}} = \arg\max_{\boldsymbol{\beta} \in \mathbb{R}^p} D(\boldsymbol{\beta}; P_n)$ and $\boldsymbol{\beta}^* = \arg\max_{\boldsymbol{\beta} \in \mathbb{R}^p} D(\boldsymbol{\beta}; P)$:

Lemma 4.3 Let $D(\boldsymbol{\beta}; P)$ be a general depth notion. If (i) $\sup_{\boldsymbol{\beta} \in \mathbb{R}^p} |D(\boldsymbol{\beta}; P_n) - D(\boldsymbol{\beta}; P)| = O_p(n^{-1/2})$ and (ii) **C2**' holds, then $\boldsymbol{\beta}^*_n - \boldsymbol{\beta}^* = O_p(n^{-1/2})$.

Proof: Denote $\Delta_n := \sup_{\boldsymbol{\beta} \in \mathbb{R}^p} |D(\boldsymbol{\beta}; P_n) - D(\boldsymbol{\beta}; P)|$. Let $\delta = 2\Delta_n/\kappa$. In light of **C2**', we have for every n

$$\begin{aligned}\sup_{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\beta}^*\|\geq\delta} D(\boldsymbol{\beta}; P_n) \;\leq\; & \sup_{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\beta}^*\|\geq\delta} |D(\boldsymbol{\beta}; P_n) - D(\boldsymbol{\beta}; P)| + \sup_{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\beta}^*\|\geq\delta} D(\boldsymbol{\beta}; P) \\ < \;& \Delta_n + \alpha^* - \kappa\delta = \alpha^* - \Delta_n,\end{aligned}$$

which, in conjunction with the definition of $\boldsymbol{\beta}^*_{\boldsymbol{n}}$, implies that $\|\boldsymbol{\beta}^*_n - \boldsymbol{\beta}^*\| < \delta$. ■

Theorem 4.2 If **(A0):** $E|r(\boldsymbol{\beta})|^2$ and $E|\mathbf{v}'\mathbf{w}|^2$ exist uniformly in $\boldsymbol{\beta} \in \mathbb{R}^p$ and $\mathbf{v} \in \mathbb{S}^{p-1}$, then (i) $\sup_{\boldsymbol{\beta} \in \mathbb{R}^p} |\mathrm{RD}(\boldsymbol{\beta}; P_n) - \mathrm{RD}(\boldsymbol{\beta}; P)| = O_p(n^{-\frac{1}{2}})$ and (ii) $\boldsymbol{\beta}^*_n - \boldsymbol{\beta}^* = O_p(n^{-\frac{1}{2}})$ if **C2**' holds.

Proof: Write $f(y, \mathbf{w}, \boldsymbol{\beta}, \mathbf{v}) := (y - \mathbf{w}'\boldsymbol{\beta})(\mathbf{v}'\mathbf{w})$, $\forall \boldsymbol{\beta} \in \mathbb{R}^p, \mathbf{v} \in \mathbb{S}^{p-1}$. In light of (ii) of Lemma 2.1, $\mathrm{RD}(\boldsymbol{\beta}; P) = \inf_{\mathbf{v} \in \mathbb{S}^{p-1}} P\left(f(y, \mathbf{w}, \boldsymbol{\beta}, \mathbf{v}) \geq 0\right)$. Define a class of functions (for the notation convention, see p140 of Pollard,1984 (P84)) $\mathscr{F} = \{(f(\cdot, \cdot, \boldsymbol{\beta}, \mathbf{v}),\ \forall\ \boldsymbol{\beta} \in \mathbb{R}^p,$ and $\mathbf{v} \in \mathbb{S}^{p-1}\}$. Note that (see Z18)

$$\begin{aligned}\{f(y, \mathbf{w}, \boldsymbol{\beta}, \mathbf{v}) \geq 0, \boldsymbol{\beta} \in \mathbb{R}^p, \mathbf{v} \in \mathbb{S}^{p-1}\} = \Big\{ &\Big(\{(y - \boldsymbol{\beta}'\mathbf{w}) \geq 0\} \cap \{\mathbf{u}'\mathbf{x} < v\}\Big) \\ \cup &\Big(\{(y - \boldsymbol{\beta}'\mathbf{w}) < 0\} \cap \{\mathbf{u}'\mathbf{x} \geq v\}\Big), \mathbf{u} \in \mathbb{S}^{p-2}, v \in \mathbb{R}^1, \boldsymbol{\beta} \in \mathbb{R}^p\Big\}.\end{aligned}$$

The RHS above is built up from sets of the form $\{g \geq 0\}$ with g in the finite-dimensional vector space of functions. By Lemmas II.28 and II.15 of P84, the class of graphs of functions in $\mathscr{F}$ has polynomial discrimination and $\mathscr{F}$ has VC subgraphs with a square integrable envelope (see II.5 of P84 and 2.6 of VW96 for discussions). By Corollary 3.2 of KP90, we have that

$$\sup_{\boldsymbol{\beta} \in \mathbb{R}^p,\ \mathbf{v} \in \mathbb{S}^{p-1}} \big|P_n(f(y, \mathbf{w}, \boldsymbol{\beta}, \mathbf{v}) \geq 0) - P(f(y, \mathbf{w}, \boldsymbol{\beta}, \mathbf{v}) \geq 0)\big| = O_p(n^{-1/2}).$$

Thus we have

$$\begin{aligned}\sup_{\boldsymbol{\beta} \in \mathbb{R}^p} \left| \inf_{\mathbf{v} \in \mathbb{S}^{p-1}} P_n(f(y, \mathbf{w}, \boldsymbol{\beta}, \mathbf{v}) \geq 0) - \inf_{\mathbf{v} \in \mathbb{S}^{p-1}} P(f(y, \mathbf{w}, \boldsymbol{\beta}, \mathbf{v}) \geq 0) \right| &\leq \\ \sup_{\boldsymbol{\beta} \in \mathbb{R}^p,\ \mathbf{v} \in \mathbb{S}^{p-1}} \big|P_n(f(y, \mathbf{w}, \boldsymbol{\beta}, \mathbf{v}) \geq 0) - P(f(y, \mathbf{w}, \boldsymbol{\beta}, \mathbf{v}) \geq 0)\big| &= O_p(n^{-1/2}),\end{aligned}$$

where the inequality follows from the fact that $|\inf_A f - \inf_A g| \leq \sup_A |f - g|$. It follows that $\sup_{\boldsymbol{\beta} \in \mathbb{R}^p} |\mathrm{RD}(\boldsymbol{\beta}; P_n) - \mathrm{RD}(\boldsymbol{\beta}; P)| = O_p(n^{-1/2})$, the first part of the theorem is obtained.

This first part, in conjunction with **C2**' and Lemma 4.3, yields the desired second part of the theorem. That is, $\boldsymbol{\beta}_n^* - \boldsymbol{\beta}^* = O_p(n^{-1/2})$. ■

Remarks 4.4:

(I) The approach of the first part of the proof could be extended for any depth notions that are defined based on sets that form a VC class such as the location counterpart, Tukey halfspace depth (HD), where one has a class of halfspaces, a VC class of sets.

That is, utilizing the approach, one can prove that the maximum Tukey location depth estimator (aka Tukey median) is root-n consistent (uniformly tight) if **C2**' holds for the HD. For the latter, a sufficient condition was given in Nolan (1999) ((ii) of Lemma 2), BH99 ((N2) in Theorem 4.1), and Massé (2002) ((b) of Proposition 3.2 and Theorem 3.5). That is, the approach here covers the uniform tightness result in those papers.

(II) BH99 obtained the root-n consistency for $\boldsymbol{\beta}^*_{\boldsymbol{n}}$ with a very different approach under more assumptions, such as their (D1)-(D4) and (C1), (C2), and (C3), on the random vector $\mathbf{x}$, on the conditional distribution of y given $\mathbf{x}$, and on the error distributions. ■

5 Limiting distribution

With the root-n consistency of the maximum regression depth estimator established, we are now in a position to address the natural question: does it have a limiting distribution?

Since the tool employed for establishing the limiting distributions is the *Argmax* theorem, we first cite it below from VW96 (Theorem of 2.7 of KP90 is an earlier version).

Lemma 5.1 [Th. 3.2.2, VW96, Argmax continuous mapping] Let M_n, M be stochastic processes indexed by a metric space S such that $M_n \xrightarrow{d} M$ in $l^\infty(K)$ for every compact $K \subset S$. Suppose that almost all sample paths $\mathbf{s} \mapsto M(\mathbf{s})$ are upper semicontinuous and possess a unique maximum at a random point $\widehat{\mathbf{s}}$, which, as a random map into S, is tight. If the sequence $\widehat{\mathbf{s}}_n$ is uniformly tight and satisfies $M_n(\widehat{\mathbf{s}}_n) \geq \sup_{\mathbf{s}} M_n(\mathbf{s}) - o_p(1)$, then $\widehat{\mathbf{s}}_n \xrightarrow{d} \widehat{\mathbf{s}}$, where $\xrightarrow{d}$ stands for convergence in distribution. ■

To establish the limiting distribution for $\widehat{\mathbf{s}}_n := \sqrt{n}\boldsymbol{\beta}^*_{\boldsymbol{n}}$, we need (A) to identify the processes M_n and M and show that $M_n \xrightarrow{d} M$ in $l^\infty(K)$ for any compact $K \subset \mathbb{R}^p$. (B) to show that almost all sample paths of $M(\mathbf{s})$ are upper semicontinuous and possess a unique maximum at a random point $\widehat{\mathbf{s}}$, which is tight, and (C) to show that $\widehat{\mathbf{s}}_n$ is uniformly tight and $M_n(\widehat{\mathbf{s}}_n) \geq \sup_s M_n(\mathbf{s}) - o_p(1)$.

In virtue of Theorem 4.2, part of (C) already holds under certain conditions for $\widehat{\mathbf{s}}_n = \sqrt{n}\boldsymbol{\beta}^*_{\boldsymbol{n}}$. So we need to verify the (A) and (B) and the second part of (C).

By (ii) of Lemma 2.1 and (2), we have that

$$\boldsymbol{\beta}^* = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\max}\, \mathrm{RD}(\boldsymbol{\beta}; P) = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\max} \inf_{\mathbf{v}\in\mathbb{S}^{p-1}} E\Big(\mathbf{I}\big(f(y, \mathbf{w}, \boldsymbol{\beta}, \mathbf{v}) \geq 0\big)\Big),$$

where $f(y, \mathbf{w}, \boldsymbol{\beta}, \mathbf{v}) = (y - \mathbf{w}'\boldsymbol{\beta})\mathbf{v}'\mathbf{w}$ given in the proof of Theorem 4.2. For a given $\boldsymbol{\beta}$ define

$$V(\boldsymbol{\beta}) = \{\mathbf{v} \in \mathbb{S}^{p-1} : \mathrm{RD}(\boldsymbol{\beta}; P) = P(f(y, \mathbf{w}, \boldsymbol{\beta}, \mathbf{v}) \geq 0) = \inf_{\mathbf{u} \in \mathbb{S}^{p-1}} P(f(y, \mathbf{w}, \boldsymbol{\beta}, \mathbf{u}) \geq 0)\},$$

i.e., the collection of $\mathbf{v}$ at which $P(f(y, \mathbf{w}, \boldsymbol{\beta}, \mathbf{v}) \geq 0)$ attains the infimum over $\mathbf{v} \in \mathbb{S}^{p-1}$.

Assume by Lemma 3.1 that $\boldsymbol{\beta^*}$ is $\mathbf{0}$. Hereafter $\boldsymbol{\beta}$ is assumed to be in a small bounded neighborhood Θ of $\mathbf{0}$ by virtue of Theorem 4.1. Assume for $\mathbf{v} \in \mathbb{S}^{p-1}$ and $\boldsymbol{\beta} \in \Theta$ that

$$\mathbf{A1}: \qquad P(f(y, \mathbf{w}, \boldsymbol{\beta}, \mathbf{v}) \geq 0) = P(f(y, \mathbf{w}, \mathbf{0}, \mathbf{v}) \geq 0) + \mathbf{g}(\mathbf{v}) \cdot \boldsymbol{\beta} + o(\|\boldsymbol{\beta}\|),$$

where $\mathbf{g}(\mathbf{v})$ is the $\nabla h(\mathbf{0}, \mathbf{v})$. The latter is defined in (ii) of (II) of Remarks 4.3. That is, the LHS permits a Taylor expansion at $\boldsymbol{\beta^*} = \mathbf{0}$. Furthermore,

$$\mathbf{A2}: \qquad V(\mathbf{0}) = \mathbb{S}^{p-1}.$$

That is, along any direction $\mathbf{v} \in \mathbb{S}^{p-1}$, $P(f(y, \mathbf{w}, \mathbf{0}, \mathbf{v}) \geq 0) = \alpha^* := \mathrm{RD}(\boldsymbol{\beta^*}, P)$. And

$$\mathbf{A3}: \qquad \mathbf{g}(\mathbf{v}) \text{ is continuous in } \mathbf{v} \text{ and } \sup_{\mathbf{v} \in V(\mathbf{0})} \|\mathbf{g}(\mathbf{v})\| < \infty.$$

That is, $\mathbf{g}(\mathbf{v})$ is uniformly bounded over $V(\mathbf{0})$.

Theorem 5.1 If **C2**' and **A0-A3** hold, then for $\boldsymbol{\beta^*_n}$ induced from RD, as $n \to \infty$,

$$\sqrt{n}(\boldsymbol{\beta^*_n} - \boldsymbol{\beta^*}) \xrightarrow{d} \operatorname*{argmax}_{\mathbf{s}} \inf_{\mathbf{v} \in V(\mathbf{0})} \{E_P(f(y, \mathbf{w}, \mathbf{0}, \mathbf{v}) \geq 0) + \mathbf{g}(\mathbf{v}) \cdot \mathbf{s}\},$$

where E_P is the limit of the empirical process $E_n = \sqrt{n}(P_n - P)$ in $l^\infty(\mathscr{F})$, a P-Brownian bridge (see Def. VII. 14 of P84), and $\mathscr{F} = \{f(\cdot, \cdot, \mathbf{0}, \mathbf{v}) \geq 0, \mathbf{v} \in V(\mathbf{0})\}$.

Proof: **C2**' guarantees the uniqueness of $\boldsymbol{\beta}^*(:= \boldsymbol{\eta})$, which can be assumed, w.l.o.g., to be $\mathbf{0}$. (ii) of Lemma 2.1 yields

$$\mathrm{RD}(\boldsymbol{\beta}; P) = \inf_{\mathbf{v} \in \mathbb{S}^{p-1}} E(\mathbf{I}((y - \mathbf{w}'\boldsymbol{\beta})\mathbf{v}'\mathbf{w} \geq 0)) = \inf_{\mathbf{v} \in \mathbb{S}^{p-1}} P(f(y, \mathbf{w}, \boldsymbol{\beta}, \mathbf{v}) \geq 0).$$

Note that

$$n^{1/2}\boldsymbol{\beta}^*_n = n^{1/2} \operatorname*{argmax}_{\boldsymbol{\beta} \in \mathbb{R}^p} \inf_{\mathbf{v} \in \mathbb{S}^{p-1}} P_n\left(f(y, \mathbf{w}, \boldsymbol{\beta}, \mathbf{v}) \geq 0\right). \qquad (10)$$

Hence for any compact K and $\mathbf{s} \in K \subset \mathbb{R}^p$ and sufficiently large n,

$$\begin{aligned} n^{1/2}P_n(f(y, \mathbf{w}, \mathbf{s}/n^{1/2}, \mathbf{v}) \geq 0) &= n^{1/2}P(f(y, \mathbf{w}, \mathbf{s}/n^{1/2}, \mathbf{v}) \geq 0) + E_n(f(y, \mathbf{w}, \mathbf{s}/n^{1/2}, \mathbf{v}) \geq 0) \\ &= n^{1/2}P(f(y, \mathbf{w}, \mathbf{s}/n^{1/2}, \mathbf{v}) \geq 0) + E_n(f(y, \mathbf{w}, \mathbf{0}, \mathbf{v}) \geq 0) + o_p(1) \\ &= n^{1/2}P(f(y, \mathbf{w}, \mathbf{0}, \mathbf{v}) \geq 0) + \mathbf{g}(\mathbf{v}) \cdot \mathbf{s} + o(\|\mathbf{s}\|) + E_n(f(y, \mathbf{w}, \mathbf{0}, \mathbf{v}) \geq 0) + o_p(1), \qquad (11) \end{aligned}$$

where the second equality follows from the stochastic equicontinuity Lemma VII. 15 of P84 (see, Lemma II.18; Example II.26; Lemma II. 28; Lemmas II.25, II.36 and Example VII.18 of P84), the last equality follows from the **A1**. Then we can define that

$$M_n(\mathbf{s}) := n^{1/2} \inf_{\mathbf{v} \in \mathbb{S}^{p-1}} P_n(f(y, \mathbf{w}, \mathbf{s}n^{-1/2}, \mathbf{v})) - n^{1/2}\alpha^*, \qquad (12)$$

where $\alpha^* = \text{RD}(\boldsymbol{\beta^*}; P)$. Note that by (10), it is readily seen that $\widehat{\mathbf{s}}_n := n^{1/2}\boldsymbol{\beta}^*_n$ maximizes $M_n(\mathbf{s})$ and is uniformly tight in virtue of Theorem 4.2, therefore (C) is completely verified.

Now we need to verify (A) and (B) for

$$M(\mathbf{s}) := \inf_{\mathbf{v} \in V(\mathbf{0})} \{E_P(f(y, \mathbf{w}, \mathbf{0}, \mathbf{v}) \geq 0) + \mathbf{g}(\mathbf{v}) \cdot \mathbf{s}\}. \tag{13}$$

Following lemmas are needed to fulfil the task above (detailed proofs are given in the supplement provided).

Lemma 5.2 In light of **A2** and **A3**,

> **R1:** The sample path of $M(\mathbf{s})$ is continuous in $\mathbf{s}$ a.s., and furthermore $M(\mathbf{s}) \to -\infty$ as $\|\mathbf{s}\| \to \infty$ a.s.; **R2**: $M(\mathbf{s})$ is concave in $\mathbf{s}$ a.s. ■

Let $\widehat{\mathbf{s}}$ be a maximizer of $M(\mathbf{s})$. The existence of a $\widehat{\mathbf{s}}$ is guaranteed by **R1** and **R2**. To show the tightness of $\widehat{\mathbf{s}}$, it suffices to show its measurability (see page 8 of VDV98). The latter is straightforward (see page 197 of P84, or pages 295-296 of Massé, 2002, for example). Now we have to show that $\widehat{\mathbf{s}}$ is unique. Define that

$$M(\mathbf{s}, \mathbf{v}) = E_P(f(y, \mathbf{w}, \mathbf{0}, \mathbf{v}) \geq 0) + \mathbf{g}(\mathbf{v}) \cdot \mathbf{s}, \quad \mathcal{V}(\widehat{\mathbf{s}}) := \{\mathbf{v} \in V(\mathbf{0}), M(\widehat{\mathbf{s}}) = M(\widehat{\mathbf{s}}, \mathbf{v})\}, \tag{14}$$

the latter is clearly non-empty. Suppose that $\widehat{\mathbf{t}}$ is another maximizer of $M(\mathbf{s})$, then by **R2**, $\alpha\widehat{\mathbf{s}} + (1-\alpha)\widehat{\mathbf{t}}$ is also a maximum point for every $\alpha \in [0, 1]$. Following Nolan, 1999, one has

Lemma 5.3 If **A2** and **A3** hold, then

R3: $\inf_{\mathbf{v} \in \mathcal{V}(\widehat{\mathbf{s}})} \mathbf{v}'x \leq 0$, $\forall\, x \in \mathbb{R}^p$; **R4**: $\mathcal{V}(\alpha\widehat{\mathbf{s}} + (1-\alpha)\widehat{\mathbf{t}}) = \mathcal{V}(\widehat{\mathbf{s}}) \cap \mathcal{V}(\widehat{\mathbf{t}})$, $\forall\, \alpha \in (0, 1)$. ■

Lemma 5.4 If **A2** and **A3** hold, then $\widehat{\mathbf{s}}$ is unique. ■

Recall $M(\mathbf{s}) := \inf_{\mathbf{v} \in V(\mathbf{0})} \{E_P(f(y, \mathbf{w}, \mathbf{0}, \mathbf{v}) \geq 0) + \mathbf{g}(\mathbf{v}) \cdot \mathbf{s}\}$. Note that by (11)

$$\begin{aligned} M_n(\mathbf{s}) = \inf_{\mathbf{v} \in \mathbb{S}^{p-1}} \Big(& n^{1/2}\big(P(f(y, \mathbf{w}, \mathbf{0}, \mathbf{v}) \geq 0) - \alpha^*\big) + \mathbf{g}(\mathbf{v}) \cdot \mathbf{s} + o(\|\mathbf{s}\|) \\ & + E_n(f(y, \mathbf{w}, \mathbf{0}, \mathbf{v}) \geq 0) + o_p(1)\Big), \end{aligned} \tag{15}$$

Lemma 5.5 If **A1-A3** hold, then $M_n(\mathbf{s}) \xrightarrow{d} M(\mathbf{s})$ uniformly over $\mathbf{s} \in K$. ■

Remarks 5.1

(I) Theorem 5.1 could be adapted to cover the location counterpart (maximum halfspace depth estimator (aka Tukey median)). The assumptions **A1-A3** and **C2**' hold under the conditions given in Nolan, 1999 and BH99.

(II) Utilizing a different approach, BH99 treated the limit distribution of $\boldsymbol{\beta^*_n}$. BH99 skipped the verification of the two key conditions ((W1) and (W3)) in their uniqueness lemma 3.3 though. Their result does not cover the result here. ■

Acknowledgments

The author thanks Hanshi Zuo, Wei Shao, and Professor Emeritus James Stapleton for their careful proofreading and an anonymous referee and the Co-Editor-in-Chief Yimin Xiao for their insightful comments and suggestions, all of which have led to improvements.

References

[1] Adrover, J., Maronna, R. and Yohai, V. (2002), "Relationships between maximum depth and projection regression estimates", *J. Statist. Plann. Inference*, 105, 363-375

[2] Bai, Z. D. and He, X. (1999), "Asymptotic distributions of the maximal depth estimators for regression and multivariate location", *Ann. Statist.*, Vol. 27, No. 5, 1616–1637.

[3] Dohono, D. L. and Gasko, M. (1992), "Breakdown properties of location estimates based on halfspace depth and projected outlyingness", *Ann. Statist.*, 20 1803-1827.

[4] Kim, J. and Pollard, D. (1990), "Cube root asymptotics". *Ann. Statist.*, 18 191-219.

[5] Liu, R. Y. (1990), "On a notion of data depth based on random simplices", *Ann. Statist.*, 18, 405-414, 1990.

[6] Liu, R. Y. (1992), "Data depth and multivariate rank tests", *In Y. Dodge (ed.), L1-Statistical Analysis and Related Methods*, North-Holland, Amsterdam, 279-294.

[7] Massé, J.C. (2002), "Asymptotic for the Tukey Median", *J. Multivariate Anal.* 81, 286-300.

[8] Maronna, R. A., and Yohai, V. J. (1993), "Bias-Robust Estimates of Regression Based on Projections", *Ann. Statist.*, 21(2), 965-990.

[9] Mizera, I. (2002), "On depth and deep points: a calculus", *Ann. Statist.*, 30(6), 1681–1736.

[10] Nolan, D. (1999), "On min-max majority and deepest points", *Statist. Probab. Let.* 43, 325–333.

[11] Pollard, D. (1984), *Convergence of Stochastic Processes*, Springer, Berlin.

[12] Rousseeuw, P. J., and Hubert, M. (1999), "Regression depth" (with discussion and a reply by the authors and Stefan Van Aelst), *J. Amer. Statist. Assoc.*, 94: 388–433.

[13] Rousseeuw, P. J., and Struyf, A. (2004), "Characterizing angular symmetry and regression symmetry", *J. Statist. Plann. Inference*, 122, 161-173.

[14] Tukey, J. W. (1975), "Mathematics and the picturing of data", *In: James, R.D. (ed.), Proceeding of the International Congress of Mathematicians*, Vancouver 1974 (Volume 2), Canadian Mathematical Congress, Montreal, 523-531.

[15] Van Aelst, S., and Rousseeuw, P. J. (2000), "Robustness of Deepest Regression", *J. Multivariate Anal.*, 73, 82–106.

[16] Van Der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambriage University Press,

[17] Van Der Vaart, A. W. and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes with Applications to Statistics*, Springer, New York.

[18] Zuo, Y. (2003) "Projection-based depth functions and associated medians", *Ann. Statist.*, 31, 1460-1490.

[19] Zuo, Y. (2018), "On general notions of depth in regression", *Statistical Science* (in press), arXiv:1805.02046.

[20] Zuo, Y., and Serfling, R. (2000), "General notions of statistical depth function", *Ann. Statist.*, 28, 461-482.