# **Supplement** (a slightly modified version of the original one) to

# 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

December 13, 2021

# 1 Detailed Proofs of the Lemmas in the Proof of Theorem 5.1

Recall that

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

where $\alpha^* = \mathrm{RD}(\boldsymbol{\beta}^*; P)$, $\mathbf{s} \in K$, a compact set in $\mathbb{R}^p$. Now we need to verify (A) and (B) for

$$M(\mathbf{s}) := \inf_{\mathbf{v}\in V(\mathbf{0})} \{E_P(f(y,\mathbf{w},\mathbf{0},\mathbf{v}) \geq 0) + \mathbf{g}(\mathbf{v})\cdot\mathbf{s}\}. \tag{2}$$

**Lemma 5.2** In light of **A2** and **A3** (defined before Theorem 5.1),

> **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..

**Proof**: Write $M(\mathbf{s},\mathbf{v}) = E_P(f(y,\mathbf{w},\mathbf{0},\mathbf{v}) \geq 0) + \mathbf{g}(\mathbf{v})\cdot\mathbf{s}$. The continuity and concavity of $M(\mathbf{s},\mathbf{v})$ in $\mathbf{s}$ is obvious. The assertion on $M(\mathbf{s})$ follows since the infimum preserves these properties. We need to show the second part of **R1**.

First by the compactness of $V(\mathbf{0})$, the continuity and boundedness of $g(\mathbf{v})$ over $V(\mathbf{0})$, for an arbitrary $\mathbf{s}$, there is a $\mathbf{v}_0 \in V(\mathbf{0})$ such that

$$\inf_{\mathbf{v}\in V(\mathbf{0})} \mathbf{g}(\mathbf{v})\cdot\mathbf{s} = \mathbf{g}(\mathbf{v}_0)\cdot\mathbf{s}. \tag{3}$$

By the oddness of $\mathbf{g}(\mathbf{v})$ in $\mathbf{v}$, it can be shown that the $\mathbf{g}(\mathbf{v}_0)\cdot\mathbf{s} < 0$ (see the related result **R3** in Lemma 5.3). Now we have that

$$-\infty \ \leq\ M(\mathbf{s}) \leq E_P(f(y,\mathbf{w},\mathbf{0},\mathbf{v_0}) \geq 0) + \mathbf{g}(\mathbf{v_0})\cdot\mathbf{s} \to -\infty\ (a.s.),\ \text{as } \|\mathbf{s}\| \to \infty, \tag{4}$$

where the second inequality follows from the definition of infimum in $M(\mathbf{s})$. ■

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 Van Der Vaart (1998) (VDV98)). The latter is straightforward (see page 197 of Pollard, 1984 (P84), or pages 295-296 of Massé, 2002, for example). Now we have to show that $\widehat{\mathbf{s}}$ is unique. Recall that $M(\mathbf{s},\mathbf{v}) = E_P(f(y,\mathbf{w},\mathbf{0},\mathbf{v}) \geq 0) + \mathbf{g}(\mathbf{v})\cdot\mathbf{s}$. Define

$$\mathscr{V}(\widehat{\mathbf{s}}) := \{\mathbf{v} \in V(\mathbf{0}), M(\widehat{\mathbf{s}}) = M(\widehat{\mathbf{s}},\mathbf{v})\},$$

which 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 can show that

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

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

Equipped with the results above, we now are in the position to show that

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

**Proof**: Define $\mathscr{G} := \text{span}\Big(\{\mathbf{g}(\mathbf{v}) : \mathbf{v} \in \mathscr{V}(\alpha\widehat{\mathbf{s}} + (1-\alpha)\widehat{\mathbf{t}}, \text{ for an } \alpha \in (0,1)\}\Big)$. Let $r$ be the dimension of $\mathscr{G}$. In the sequel, consider different cases of $r$.

If $r = 1$, then there exists a $\mathbf{v} \in \mathscr{V}(\alpha\widehat{\mathbf{s}} + (1-\alpha)\widehat{\mathbf{t}})$ such that $\mathscr{G}$ is spanned by $\mathbf{g}(\mathbf{v})$.

Note that $E_P(f(y, \mathbf{w}, \mathbf{0}, \mathbf{v}) \geq 0) = -E_P(f(y, \mathbf{w}, \mathbf{0}, -\mathbf{v}) \geq 0)$ and $\mathbf{g}(-\mathbf{v}) = -\mathbf{g}(\mathbf{v})$. Now, by the definitions of $M(\mathbf{s})$, $M(\mathbf{s}, \mathbf{v})$ and $\mathscr{V}(\widehat{\mathbf{s}})$ and Lemma 5.3, we have

$$M(\widehat{\mathbf{s}}) = M(\widehat{\mathbf{s}}, \mathbf{v}) = \inf_{\mathbf{v} \in \mathcal{V}(\mathbf{0})} M(\widehat{\mathbf{s}}, \mathbf{v}) = -\sup_{\mathbf{v} \in \mathcal{V}(\mathbf{0})} (-M(\widehat{\mathbf{s}}, \mathbf{v}))$$

$$= -\sup_{\mathbf{v} \in \mathcal{V}(\mathbf{0})} M(\widehat{\mathbf{s}}, -\mathbf{v}) \leq -\sup_{\mathbf{v} \in \mathcal{V}(\mathbf{0})} M(\widehat{\mathbf{s}}) = -M(\widehat{\mathbf{s}}),$$

which implies that $M(\widehat{\mathbf{s}}) \leq 0$. By (2) and definitions of $M(\mathbf{s}, \mathbf{v})$ and $\widehat{\mathbf{s}}$, $M(\widehat{\mathbf{s}}) \geq 0$, we conclude that $M(\widehat{\mathbf{s}}) = 0$. This further implies that $\forall\ \mathbf{u} \in V(\mathbf{0})$ and $\forall\ \mathbf{s} \in K$,

$$E_P(f(y, \mathbf{w}, \mathbf{0}, \mathbf{u}) \geq 0) + \mathbf{g}(\mathbf{u}) \cdot \mathbf{s} = 0 \text{ and } V(\mathbf{0}) = \mathscr{V}(\widehat{\mathbf{s}}). \tag{5}$$

Now assume that there is another vector $\mathbf{v}_1 (\neq \pm\mathbf{v}) \in V(\mathbf{0})$, then $\mathbf{g}(\mathbf{v}_1) = k\mathbf{g}(\mathbf{v})$ for some constant $k$; otherwise $\mathbf{g}(\mathbf{v}_1)$ and $\mathbf{g}(\mathbf{v})$ are linearly independent. (5) implies that

$$E_P(f(y, \mathbf{w}, \mathbf{0}, \mathbf{v}_1) \geq 0) = kE_P(f(y, \mathbf{w}, \mathbf{0}, \mathbf{v}) \geq 0).$$

Write $X$ and $Y$ for $E_P(f(y, \mathbf{w}, \mathbf{0}, \mathbf{v}_1) \geq 0)$ and $E_P(f(y, \mathbf{w}, \mathbf{0}, \mathbf{v}) \geq 0)$, respectively. Then by P84 (page 149), $X$ and $Y$ have a joint bivariate normal distribution. This, however, is impossible (see (7) in the main body) since the covariance matrix between $X$ and $Y$ has no inverse. This implies that $\mathbb{S}^{p-1} = V(\mathbf{0}) = \{\mathbf{v}, -\mathbf{v}\}$, which can happen only if $p = 1$. Namely, both $\mathbf{g}(\mathbf{v})$ and $\widehat{\mathbf{s}}$ are one-dimensional. The uniqueness of $\widehat{\mathbf{s}}$ follows in a straightforward fashion from (5).

We now assume that $2 \leq r \leq p$. Assume that $\mathbf{g}(\mathbf{v_1}), \cdots, \mathbf{g}(\mathbf{v_r})$ are linearly independent and belong to $\mathscr{G}$ and $\mathbf{v}_i \in \mathscr{V}(\alpha\widehat{\mathbf{s}} + (1-\alpha)\widehat{\mathbf{t}})$ for an $\alpha \in (0,1)$. Let $S$ be any space that contains both $\widehat{\mathbf{s}}$ and $\widehat{\mathbf{t}}$, then both $\widehat{\mathbf{s}}$ and $\widehat{\mathbf{t}}$ satisfy the following linear system of equations:

$$-\mathbf{g}(\mathbf{v_i}) \cdot \mathbf{s} = E_P(f(y, \mathbf{w}, \mathbf{0}, \mathbf{v_i}) \geq 0) - M(\widehat{\mathbf{s}}), \quad i = 1, \cdots, r, \quad \mathbf{s} \in S$$

which immediately implies that $\widehat{\mathbf{s}} - \widehat{\mathbf{t}} = 0$ is the only solution of the linear system $-\mathbf{g}(\mathbf{v_i}) \cdot (\widehat{\mathbf{s}} - \widehat{\mathbf{t}}) = 0$, $i = 1, \cdots, r$. That is, $\widehat{\mathbf{s}}$ is unique. ■

We have verified (B) completely. As we noticed above $\widehat{\mathbf{s}}_n := n^{1/2}\boldsymbol{\beta}_n^*$ maximizes $M_n(\mathbf{s})$. To verify (A) and thus complete the proof of Theorem 5.1, we need only show that $M_n(\mathbf{s}) \xrightarrow{d} M(\mathbf{s})$ uniformly in $\mathbf{s} \in K$, where $K \subset \mathbb{R}^p$ is a compact set. Note that by (11) in the main body

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

$$+ E_n(f(y, \mathbf{w}, \mathbf{0}, \mathbf{v}) \geq 0) + o_p(1)\Big), \tag{6}$$

Write

$$\lambda_n(\mathbf{v},\mathbf{s}) := n^{1/2}\big(P(f(y,\mathbf{w},\mathbf{0},\mathbf{v})\geq 0)-\alpha^*\big)+\mathbf{g}(\mathbf{v})\cdot\mathbf{s}+E_n(f(y,\mathbf{w},\mathbf{0},\mathbf{v})\geq 0), \tag{7}$$
$$M_n^1(\mathbf{s}) := \inf_{\mathbf{v}\in\mathbb{S}^{p-1}}\lambda_n(\mathbf{v},\mathbf{s}). \tag{8}$$

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

**Proof**: We employ two steps to prove the Lemma.

(i) First, we show $\sup_{\mathbf{s}\in K}|M_n(\mathbf{s})-M_n^1(\mathbf{s})|=o_p(1)$. In light of (6) and (7), we have

$$\begin{aligned}\sup_{\mathbf{s}\in K}\big|M_n(\mathbf{s})-M_n^1(\mathbf{s})\big| &= \sup_{\mathbf{s}\in K}\Big|\inf_{\mathbf{v}\in\mathbb{S}^{p-1}}(\lambda_n(\mathbf{v},\mathbf{s})+o(\|\mathbf{s}\|)+o_p(1))-\inf_{\mathbf{v}\in\mathbb{S}^{p-1}}\lambda_n(\mathbf{v},\mathbf{s})\Big| \\ &\leq \sup_{\mathbf{s}\in K}\sup_{\mathbf{v}\in\mathbb{S}^{p-1}}|o(\|\mathbf{s}\|)+o_p(1)|=o_p(1),\end{aligned}$$

where the last equality follows from two facts: (1) the term $o(\|\mathbf{s}\|)$ in (6) is $o(1)$ uniformly in $\mathbf{s}$ over $K$, and (2) the term $o_p(1)$ in (6) holds uniformly in $\mathbf{s}$ over $K$ for large enough $n$, because it is obtained from application of stochastic equicontinuity over a class of functions whose members are close enough in the sense that each other is within a distance $\delta>0$ w.r.t. seminorm $\rho_P$ (see Lemma VII. 15 of P84). Thus (i) follows.

(ii) Second, we show that $M_n(\mathbf{s})\xrightarrow{d}M(\mathbf{s})$ uniformly over $\mathbf{s}\in K$. By virtue of (i), it suffices to show that $M_n^1(\mathbf{s})\xrightarrow{d}M(\mathbf{s})$ uniformly over $\mathbf{s}\in K$. Notice that by **A2**, $V(\mathbf{0})=\mathbb{S}^{p-1}$ and $P(f(y,\mathbf{w},\mathbf{0},\mathbf{v}))-\alpha^*=0$ for any $\mathbf{v}\in V(0)$. Therefore,

$$\begin{aligned}M_n^1(\mathbf{s}) &= \inf_{v\in V(\mathbf{0})}\big(E_n(f(y,\mathbf{w},\mathbf{0},\mathbf{v})\geq 0)+\mathbf{g}(\mathbf{v})\cdot\mathbf{s}+n^{1/2}(P(f(y,\mathbf{w},\mathbf{0},\mathbf{v})\geq 0)-\alpha^*)\big) \\ &= \inf_{v\in V(\mathbf{0})}\big(E_n(f(y,\mathbf{w},\mathbf{0},\mathbf{v})\geq 0)+\mathbf{g}(\mathbf{v})\cdot\mathbf{s}\big)\xrightarrow{d}\inf_{v\in V(\mathbf{0})}\big(E_P(f(y,\mathbf{w},\mathbf{0},\mathbf{v})\geq 0)+\mathbf{g}(\mathbf{v})\cdot\mathbf{s}\big),\end{aligned}$$

where the last step follows from the central limit theorem for empirical process (Theorem VII. 21 of P84) and the continuous mapping theorem. The steps above hold uniformly for $\mathbf{s}\in K$. (A) has been verified completely. ■

# 2 Stochastic Equicontinuity Lemma and VC-classes of sets

The main reference of this part is P84. Similar materials could also be found in VDV98 and Van Der Vaart and Wellner (1996) (VW96).

### Stochastic equicontinuity

*Stochastic equicontinuity* refers to a sequence of stochastic processes $\{Z_n(t):t\in T\}$ whose shared index set $T$ comes equipped with a semi metric $d(\cdot,\cdot)$. (a semi metric has all the properties of a metric except that $d(s,t)=0$ need not imply that $s$ equals $t$.)

**Definition 1** [IIV. 1, Def. 2, P84 ]. Call $Z_n$ stochastically equicontinuous at $t_0$ if for each $\eta > 0$ and $\epsilon > 0$ there exists a neighborhood $U$ of $t_0$ for which

$$\limsup P\left(\sup_U |Z_n(t) - Z_n(t_0)| > \eta\right) < \epsilon. \tag{9}$$

■

Because stochastic equicontinuity bounds $Z_n$ uniformly over the neighborhood $U$, it also applies to any randomly chosen point in the neighborhood. If $\tau_n$ is a sequence of random elements of $T$ that converges in probability to $t_0$, then

$$Z_n(\tau_n) - Z(t_0) \to 0 \text{ in probability,} \tag{10}$$

because, with probability tending to one, $\tau_n$ will belong to each $U$. The form above will be easier to apply, especially when behavior of a particular $\tau_n$ sequence is under investigation. This also is the form used in the Theorem 5.1.

To establish (9), we need the *chaining technique* to prove maximum inequalities, which involves the *covering number* (IIV. 2, P84). Chaining is a technique for proving maximal inequalities for stochastic processes, the sorts of things required if we want to check the stochastic equicontinuity condition defined in Definition 1. It applies to any process $\{Z(t) : t \in T\}$ whose index set is equipped with a semimetric $d(\cdot, \cdot)$ that controls the increments:

$$P\left(|Z(s) - Z(t)| > \eta\right) \leq \Delta(\eta, d(s,t)) \text{ for } \eta > 0.$$

It works best when $\Delta(\cdot, \cdot)$ takes the form

$$\Delta(\eta, \delta) = 2\exp\left(-\frac{1}{2}\eta^2/D^2\delta^2\right),$$

with $D$ a positive constant. Under some assumptions about covering numbers for $T$, the chaining technique will lead to an economical bound on the tail probabilities for a supremum of $\|Z(s) - Z(t)\|$ over pairs $(s,t)$.

**Covering number**

**Definition 2** [IIV. 2, Def. 8, P84]. The covering number $N(\delta, d, T)$ is the size of the smallest $\delta$-net for $T$. That is, $N(\delta, d, T)$ equals the smallest $m$ for which there exist points $t_1, \cdots, t_m$ with $\min_i d(t, t_i) \leq \delta$ for every $t \in T$. The associated covering integral is

$$J(\delta, d, T) = \int_0^\delta [2\log\big(N(\delta, d, T)^2/u\big)]^{1/2} du \quad \text{for} \quad 0 < \delta < 1. \tag{11}$$

**Chaining Lemma** [VII. 2. Lemma 9, P84]. Let $\{Z(t) : t \in T\}$ be a stochastic process whose index set has a finite covering integral $J(\delta, d, T)$. Suppose there exists a constant $D$ such that, for all $s$ and $t$

$$P\big(|Z(s) - Z(t)| > \eta d(s,t)\big) \leq 2\exp\big(-\eta^2/D^2\big) \quad \text{for} \quad \eta > O.$$

Then there exists a countable dense subset $T^*$ of $T$ such that, for $0 < \epsilon < 1$,

$$P\Big(|Z(s) - Z(t)| > 26DJ(d(s,t)) \text{ for some } s,\, t \text{ in } T* \text{ with } d(s,t) \le \epsilon\Big) \le 2\epsilon$$

We can replace $T^*$ by $T$ if $Z$ has continuous sample paths. ■

**Random Covering Numbers**

The symmetrization method (II. 3, P84) relates $P_n - P$ to the random signed measure $P_n^o$ that puts mass $\pm n^{-1}$ at each of $\xi_1 ..., \xi_n$ (random sample from P), the signs being allocated independently plus or minus, each with probability 1/2 (see page 15 of P84). For central limit theorem calculations it is neater to work with the symmetrized empirical process $E_n^o = n^{1/2} P_n^o$. Hoeffding's Inequality gives the clean exponential bound for $E_n^o$ conditional on everything but the random signs. For each fixed function $f \in \mathscr{F}$, a class of functions,

$$\begin{aligned} P\big(|E_n^o f| > \eta \big| \boldsymbol{\xi}\ \big) &= P\Big(\Big|\sum_{i=1}^{n} \pm f(\xi_i)\Big| > \eta n^{1/2} \Big| \boldsymbol{\xi}\Big) \\ &\le 2\exp\Big[-2(\eta n^{l/2})^2 / \sum_{i=1}^{n} 4f(\xi_i)^2\Big] = 2\exp\big[-\frac{1}{2}\eta^2/P_n f^2\big]. \end{aligned} \tag{12}$$

That is, if distances between functions are measured using the $\mathscr{L}^2(P_n)$ seminorm then tail probabilities of $E_n^o$ under $P(\Delta\big|\boldsymbol{\xi})$ satisfy the exponential bound required by the Chaining Lemma, with $D = 1$. For the purposes of the chaining argument, $E_n^o$ will behave very much like the gaussian process $B_P$, except that the bound involves the random covering number calculated using the $\mathscr{L}^2(P_n)$ seminorm ($\rho_{P_n}(f,g) = (\int (f-g)^2 dP_n)^{1/2}$, for $f, g \in \mathscr{F}$). Write

$$J_2(\delta, P_n, \mathscr{F}) = \int_0^\delta \big[2\log\big(N_2(u, P_n, \mathscr{F})^2/u\big)\big]^{1/2} du$$

for the corresponding covering integral, where we interpret P as standing for $\mathscr{L}^2(P)$ semimetrics on $\mathscr{F}$, the notation $N_2(\delta, P_n, \mathscr{F})$ (a random number) agrees with Definition 2 (also see II.6, Def. 32, P84).

Stochastic equicontinuity of the empirical processes $\{E_n\}$ (the signed measure $n^{l/2}(P_n - P)$) at a function $f_0$ in $\mathscr{F}$ means roughly that, with high probability and for all $n$ large enough, $\big|E_n f - E_n f_0\big|$ should be uniformly small for all $f$ close enough to $f_0$. Here closeness should be measured by the $\mathscr{L}^2(P)$ seminorm $\rho_P$. Of course we need $\mathscr{F}$ to be permissible (see Appendix C, Def. 1, P84), i.e. roughly speaking, there is no measurability issue.

**Equicontinuity Lemma** [IIV. 4, Lemma 15, P84]

Let $\mathscr{F}$ be a permissible class of functions with envelope $F$ in $\mathscr{L}^2(P)$ (call each measurable $F$ such that $|f| \le F$, for every $f \in \mathscr{F}$, an envelope for $\mathscr{F}$). Suppose the random covering numbers satisfy the uniformity condition: for each $\eta > 0$ and $\epsilon > 0$ there exists a $\gamma > 0$ such that

$$\limsup P\big(J_2(\gamma, P_n, \mathscr{F}) > \eta\big) < \epsilon. \tag{13}$$

Then there exists a $\delta > 0$ for which

$$\limsup P\Big(\sup_{[\delta]} |E_n(f-g)| > \eta\Big) < \epsilon, \tag{14}$$

where $[\delta] = \{(f,g) : f, g \in \mathscr{F} \text{ and } \rho_P(f-g) \leq \delta\}$. ■

Up to this point, there are two approaches to establish the stochastic equicontinuity: (i) via Definition 1, (ii) via Equicontinuity Lemma. The first approach is usually more challenging, the second one is equivalently to verify the uniformity condition for the random covering numbers.

A sufficient condition for the latter is the graphs of the functions in $\mathscr{F}$ have polynomial discrimination. The graph of a real-valued function $f$ on a set $S$ is defined as the subset

$$G_f = \{(s,t) : 0 \leq t \leq f(s) \text{ or } f(s) \leq t \leq 0, s \in S\}.$$

If the graphs of the functions in $\mathscr{F}$ have polynomial discrimination, then $N_2(u, P_n, \mathscr{F})$ is bounded by a polynomial $A(u^{-1})^W$ in $u^{-1}$ with $A$ and $W$ not depending on $P_n$ (Lemma II. 36, P84), which amply suffices for the Equicontinuity Lemma: For each $\eta > 0$, there is a $\gamma > 0$ so that $J_2(\gamma, P_n, \mathscr{F}) \leq \eta$ for every $P_n$. Therefore, the graphs of the functions in $\mathscr{F}$ having polynomial discrimination becomes a key point for Equicontinuity Lemma.

### Polynomial discrimination

**Definition 3** [II.4, Def.13, P84]. Let $\mathscr{D}$ be a class of subsets of some space $S$. It is said to have polynomial discrimination (of degree $v$) if there exists a polynomial $\rho(\cdot)$ (of degree $v$) such that, from every set of $N$ points in $S$, the class picks out at most $\rho(N)$ distinct subsets. Formally, if $S_0$ consists of $N$ points, then there are at most $\rho(N)$ distinct sets of the form $S_0 \cap D$ with $D \in \mathscr{D}$. Call $\rho(\cdot)$ the discriminating polynomial for $\mathscr{D}$. $\mathscr{D}$ is also called a VC-class of sets (see Vapnik and Chervonenkis, 1971).

### Generalized Glivenko-Cantelli theorem

**Theorem 1** [II.4, Th.14, P84]. Let $P$ be a probability measure on a space $S$. For every permissible class $\mathscr{D}$ of subsets of $S$ with polynomial discrimination,

$$\sup_{\mathscr{D}} |P_n D - PD| \to 0 \ \text{ almost surely.}$$

### Examples

**1**. Let $\mathscr{D} = \{(-\infty, t], t \in \mathbb{R}\}$. The collection of sets is the one in the traditional Glivenko-Cantelli theorem in one-dimension. The $\mathscr{D}$ can pick at most $(n+1)$ subsets for any set of $n$ points on the line. $\mathscr{D}$ has polynomial discrimination. Theorem 1 holds true.

**2**. Let $\mathscr{D} = \{(-\infty, \boldsymbol{t}], \boldsymbol{t} \in \mathbb{R}^2\}$. The collection of all quadrants of the form $(-\infty, \boldsymbol{t}]$ in $\mathbb{R}^2$, which can pick fewer than $(n+1)^2$ subsets from a set of $n$ points on the plane. $\mathscr{D}$ has polynomial discrimination. Theorem 1 holds true.

**3**. Let $\mathscr{D}$ be the class of all closed halfspaces in $\mathbb{R}^d$, then it can pick at most $O(n^2)$ subsets from a set of $n$ points in $\mathbb{R}^d$. $\mathscr{D}$ has polynomial discrimination. Theorem 1 applies.

**4**. Let $\mathscr{D}$ be the class of closed, convex sets. From every collection of n points lying on the circumference of a circle in $\mathbb{R}^2$, it can pick out all $2^n$ subsets. $\mathscr{D}$ no longer has polynomial discrimination. ■

Back to the Equicontinuity Lemma, a sufficient condition for the uniformity condition in the lemma is that the graphs of the functions in $\mathscr{F}$ have polynomial discrimination. How to verify the latter becomes the key point. It turns out that this can be done straightforwardly by the following lemma.

**Lemma 1** [II.5, Lemma 28, P84]. Let $\mathscr{F}$ be a finite-dimensional vector space of real functions on S. The class of graphs of functions in $\mathscr{F}$ has polynomial discrimination.

The following lemmas are equally useful as well.

**Lemma 2** [II.4, Lemma 18, P84]. Let $\mathscr{G}$ be a finite-dimensional vector space of real functions on $S$. The class of sets of the form $\{g \geq 0\}$, for g in $\mathscr{G}$, has polynomial discrimination of degree no greater than the dimension of $\mathscr{G}$.

**Lemma 3** [II.4, Lemma 15, P84]. If $\mathscr{C}$ and $\mathscr{D}$ have polynomial discrimination, then so do each of: (i) $\{D^c : D \in \mathscr{D}\}$; (ii) $\{C \cup D : C \in \mathscr{C}$ and $D \in \mathscr{D}\}$; (iii) $\{C \cap D : C \in \mathscr{C}$ and $D \in \mathscr{D}\}$.

## 3 More details for the proofs of Theorems 4.2 and 5.1

Consider three classes of functions that appear in **Theorems 5.1** and **4.2**,

$$\mathscr{F}_{\mathbf{v}} = \{f(\cdot, \cdot, \mathbf{0}, \mathbf{v}), \mathbf{v} \in \mathbb{S}^{p-1}\},$$

$$\mathscr{F}_{\boldsymbol{\beta}} = \{f(\cdot, \cdot, \boldsymbol{\beta}, \mathbf{v}_0), \boldsymbol{\beta} \in \mathbb{R}^p, \mathbf{v}_0 \in \mathbb{S}^{p-1} \text{ is fixed}\},$$

$$\mathscr{F}_{\boldsymbol{\beta},\mathbf{v}} = \{f(\cdot, \cdot, \boldsymbol{\beta}, \mathbf{v}), \boldsymbol{\beta} \in \mathbb{R}^p, \mathbf{v} \in \mathbb{S}^{p-1}\}.$$

Under **A0** given in Theorem 4.2, the three classes have a square integrable envelope F. We want to show that the graphs of the functions in these classes have polynomial discrimination.

It suffices to show this for the $\mathscr{F}_{\boldsymbol{\beta},\mathbf{v}}$ since other are just special cases.

The graph of a function in $\mathscr{F}_{\boldsymbol{\beta},\mathbf{v}}$ contains a point $((y, \mathbf{w}), t)$ if and only if $0 \leq t \leq f(y, \mathbf{w}, \boldsymbol{\beta}, \mathbf{v})$ or $f(y, \mathbf{w}, \boldsymbol{\beta}, \mathbf{v}) \leq t \leq 0$. Therefore, the total number of subsets of given $n$ points $((y_i, \mathbf{w}_i), t_i)$ that can be picked out by the graphs of functions in $\mathscr{F}_{\boldsymbol{\beta},\mathbf{v}}$ is less than the total number of those picked out by the union of classes of sets $\{f(y_i, \mathbf{w}_i, \boldsymbol{\beta}, \mathbf{v}) \geq 0\} \cup \{f(y_i, \mathbf{w}_i, \boldsymbol{\beta}, \mathbf{v}) \leq 0\}$. We now show that each class of sets has polynomial discrimination, so does the union (by Lemma 3) and consequently so do the graphs of functions in $\mathscr{F}_{\boldsymbol{\beta},\mathbf{v}}$.

Note that $f(y, \mathbf{w}, \boldsymbol{\beta}, \mathbf{v}) = (y - \boldsymbol{\beta}'\mathbf{w})\mathbf{v}'\mathbf{w}$ with $\mathbf{w}' = (1, \mathbf{x}')$. It suffices to just treat the class of sets $\{f(y_i, \mathbf{w}_i, \boldsymbol{\beta}, \mathbf{v}) \geq 0\}$. We have that

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

The RHS is built up from sets of the form $\{g \geq 0\}$ with $g$ in the finite-dimensional vector space of functions. There are four classes of functions on the RHS each forming a finite dimensional vector space of real functions. By Lemmas 1 and 3, we conclude that the graphs of the functions in $\mathscr{F}_{\boldsymbol{\beta},\mathbf{v}}$ have polynomial discrimination. So do the other classes of functions.

That is, they are VC classes, or polynomial classes of functions in the terminology of P84. We have all needed for invoking Equicontinuity Lemma.

However, to invoke the Corollary 3.2 in Kim and Pollard (1990), as did in the proof of Theorem 4.2, we actually need to show that $\mathscr{F}_{\boldsymbol{\beta},\mathbf{v}}$ is a *manageable class* of functions, a notion defined in Pollard, 1989 (P89). On the other hand, every VC class is a Euclidean class (a term introduced in Nolan and Pollard, 1987), luckily enough, every Euclidean class is manageable (P89). ■

# References


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

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

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

[4] Nolan, D. and Pollard, D. (1987), "U-Processes: Rates of Convergence", *Ann. Statist.*, 15(2), 780-799.

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

[6] Pollard, D. (1989), "Asymptotics via Empirical Processes", *Statist. Sci.* 4(4), 341-354.

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

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