Journal of Multivariate Analysis 142 (2015) 86–105

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Half-region depth for stochastic processes

James Kuelbs [a], Joel Zinn [b],*

[a] Department of Mathematics, University of Wisconsin, Madison, WI 53706-1388, United States

[b] Department of Mathematics, Texas A&M University, College Station, TX 77843-3368, United States

ARTICLE INFO

Article history:
Received 3 February 2014
Available online 21 August 2015

AMS 2010 subject classifications:
primary 60F05
secondary 60F17
62E20

Keywords:
Depth
Consistency
Central limit theorems
Empirical processes

ABSTRACT

We study the concept of half-region depth, introduced in López-Pintado and Romo (2011). We show that for a wide variety of standard stochastic processes, such as Brownian motion and other symmetric stable processes with stationary independent increments tied down at 0, half-region depth assigns depth zero to all sample functions. To alleviate this difficulty we introduce a method of smoothing, which often not only eliminates the problem of zero depth, but allows us to extend the theoretical results on consistency in that paper up to the $\sqrt{n}$ level for many smoothed processes.

1. Introduction and some notation

A number of depth functions are available to provide an ordering of finite dimensional data, and more recently in [14] the interesting notion of half-region depth for stochastic processes was introduced. This depth applies to data given in terms of infinite sequences, as functions defined on some interval, and even in more general settings.

In this paper we focus on three items. The first is to show (see Section 2) that for many standard data sources this depth is identically zero, and hence one needs to be cautious when employing it. In particular, we will see sample continuous Brownian motion, tied down to be zero at $t = 0$ with probability one, assigns zero half-region depth to all functions $h \in C[0, 1]$, but we show this sort of behavior also holds for many other random processes widely used to model data in a variety of settings. A second item we examine is how the difficulty of zero half-region depth can be avoided, and fortunately in many situations smoothing the process by adding an independent real valued random variable Z with a density as in (28) (also see Proposition 4) changes things dramatically for half-region depth. In particular, it allows us to establish positivity for this depth and, as can be seen from Remark 4, the smoothed data remains a good approximation of the original input by taking $E(|Z|)$ small. Using Proposition 4 as in Remark 5, we also provide some sufficient conditions where smoothing is unnecessary for positive half-region depth.

The third item we consider involves limit theorems for the empirical half-region depth of these smoothed processes, and Theorem 1 is a basic consistency result with Theorem 2 and Corollary 6 providing some rates of convergence for this consistency. Moreover, a sub-Gaussian tail bound is obtained in Corollary 6. Theorem 3 implies a consistency result and $\sqrt{n}$-rates for half-region depth over all finite subsets of T of cardinality less than or equal to a fixed r. Hence, Theorem 3 extends the consistency result for random Tukey depth in [2,3] to half-region depth under weaker conditions, i.e. less

* Corresponding author.
E-mail addresses: kuelbs@math.wisc.edu (J. Kuelbs), jzinn@math.tamu.edu (J. Zinn).

http://dx.doi.org/10.1016/j.jmva.2015.07.012
0047-259X/

independence is used and the depth is computed using multi-dimensional marginals rather than those of one dimension. Now we turn to the notation used throughout the paper, and following that we indicate some additional details on our results and how they relate to other recent papers.

To fix some notation let $X := \{X(t) = X_t : t \in T\}$ be a stochastic process on the probability space $(\Omega, \mathcal{F}, P)$, all of whose sample paths are in $M(T)$, a linear space of real valued functions on T which we assume to contain the constant functions. To handle measurability issues, we also **always** assume that $h \in M(T)$ implies

$$\sup_{t\in T} h(t) = \sup_{t\in T_0} h(t) < \infty, \tag{1}$$

where T_0 is a fixed countable subset of T. Typical examples of $M(T)$ are the uniformly bounded continuous functions on T when T is a separable metric space, or the space of cadlag functions on T for T a compact interval of the real line. In either of these situations T_0 could be any countable dense subset of T. It should also be observed that since (1) holds on the linear space $M(T)$, then $h \in M(T)$ implies

$$\inf_{t\in T} h(t) = \inf_{t\in T_0} h(t) > -\infty \quad \text{and} \quad \|h\|_\infty \equiv \sup_{t\in T} |h(t)| = \sup_{t\in T_0} |h(t)| < \infty. \tag{2}$$

If $g, h : T \to \mathbb{R}$ and $S \subseteq T$, let $g \preceq_S h$ (resp., $g \succeq_S h$), denote that $g(t) \le h(t)$ (resp., $g(t) \ge h(t)$) for all $t \in S$. When $S = T$ we will simply write $g \preceq h$ (resp., $g \succeq h$). Then, for a function $h \in M(T)$, the half-region depth with respect to P is defined as

$$D(h, P) := D_{HR}(h, P) := \min(P(X \succeq h), P(X \preceq h)). \tag{3}$$

To simplify, we also will write $D(h)$ for $D(h, P)$ when the probability measure P is understood. Since $M(T)$ is a linear space with (1) and (2) holding, and the sample paths of the stochastic process X are in $M(T)$, we see for each $h \in M(T)$ that

$$\{X \preceq h\} = \{X \preceq_{T_0} h\} \quad \text{and} \quad \{X \succeq h\} = \{X \succeq_{T_0} h\}. \tag{4}$$

Thus the events in (3) are in $\mathcal{F}$ and the probabilities are defined.

Assume that $X, X_1, X_2, \ldots$ are i.i.d. copies of the process X defined on the probability space $(\Omega, \mathcal{F}, P)$ suitably enlarged, if necessary, such that all sample paths of each X_j are in $M(T)$. Then, the empirical half-region depth of $h \in M(T)$ based on the i.i.d. copies $X_1, \ldots, X_n$ is

$$D_n(h) = \min\left\{\frac{1}{n}\sum_{j=1}^{n} I(X_j \succeq h),\ \frac{1}{n}\sum_{j=1}^{n} I(X_j \preceq h)\right\}. \tag{5}$$

Throughout this paper to be certain the half-region depth is not degenerate at zero the smoothing we use is as in Proposition 4. However, the reader may care to notice that in Theorems 1 and 2 we actually assume more on the density $f_Z(\cdot)$, but those assumptions are only required to facilitate their proofs. The positivity of the half-region depth already holds under the weaker assumptions on $f_Z(\cdot)$ of Proposition 4. Other forms of smoothing may also be beneficial when seeking to avoid the problem of the depth being degenerate at zero, and some work is currently being done in this direction. To deal with the 0-depth problem López-Pintado and Romo [14] consider another depth, which they call modified half-region depth (see the definition below). There the depth itself is changed so as to be less restrictive and non-degenerate at zero, whereas here we retain the depth, but apply it to data which has been smoothed. One reason which motivates our choice, at least for us, is that there are examples where the ordering produced by modified half-region depth produces multiple medians, contrary to what one would intuitively expect. Furthermore, half-region depth typically orders the original paths or suitably smoothed paths in these examples so as to identify the intuitive median as being the unique median. To make this more precise, we consider the following simple examples.

In the first two examples $T = [0, 1]$, $\rho(\cdot)$ denotes Lebesgue measure on T, and we assume the sample functions of the stochastic process $\{Y(t) : t \in T\}$ are jointly measurable in (t, ω) with respect to Lebesgue measure on T and the probability $P = \mathcal{L}(Y)$. Then, if $h(\cdot)$ is a Lebesgue measurable function on T, the ρ-modified half-region depth of $h(\cdot)$ is

$$MD(h, P, \rho) = \min\left[\int_T P(h(t) \le Y(t))d\rho(t),\ \int_T P(h(t) \ge Y(t))d\rho(t)\right].$$

Example 1. If $\sup_{t\in[0,1]} |Y(t)| \le \lambda < \infty$ and $Y(t)$ has continuous distribution function for all $t \in [0, 1]$, then modified half-region depth based on Lebesgue measure on $[0, 1]$, never has a unique median. For any subset $A \subset [0, 1]$ with measure $1/2$, one considers $h_A := \lambda(2I_A - 1)$. Then, for each $t \in [0, 1]$, we have $P(Y(t) = -\lambda) = 0$ and that

$$P(Y(t) \le h_A(t)) = I_A(t) + I_{A^c}(t)P(Y(t) = -\lambda) = I_A(t).$$

Similarly, $P(Y(t) \ge h_A(t)) = I_{A^c}(t)$, and therefore the modified half-region depth of h_A is $1/2$. Since, $1/2$ is the maximal value of this depth when the distribution function of $Y(t)$ is continuous for all t, h_A is a median. In particular, among these medians we have $h_{(0,1/2)} \cdot h_{(1/2,1)} = 0$, and if the distribution of Y is symmetric enough around the zero function, neither of these functions seems an intuitive median. Furthermore, if we smooth the process Y as in Proposition 4, then the half-region depth of the smoothed process is positive, but unless we know more about the Y process it is still hard to determine the median for this half-region depth. Our next two examples are more specific, and allow us to make such determinations.

The next example is a special case of those above, but the extra details allow to show that not only does it have multiple modified half-region medians as in Example 1, it also has the same unique half-region median for both the original data, and suitably smoothed data. Moreover, this unique median is what one would intuitively expect.

Example 2. For $t \in [0, 1]$ and U a uniform random variable with values in $(-\frac{1}{2}, \frac{1}{2})$, let

$$Y(t) = U.$$

Also, assume Lebesgue measure on $[0, 1]$ is used to determine the modified form of the depth. Then, one can easily check that for $h(t) = 0, 0 \leq t \leq 1$, the half-region and modified half-region depth of the function h with respect to the probability law of Y are both $\frac{1}{2}$, and since the distribution function of $Y(t)$ is continuous for each $t \in [0, 1]$, h is also a median. Moreover, h is the unique median for the half-region depth based on Y.

To define the smoothed data we take Z to be Gaussian with mean zero, variance $\sigma^2 > 0$, and independent of $\{Y(t) : t \in [0, 1]\}$. Then,

$$X(t) = U + Z, \quad t \in [0, 1],$$

and the half-region depth of the function h with respect to the law of X is also $\frac{1}{2}$. Of course, Z need only be symmetric about zero for this to hold, but for Z centered Gaussian and $\sigma^2 > 0$ small, on average

$$|Z| = \sup_{t \in [0,1]} |X(t) - Y(t)|$$

is quantifiably small. Moreover, since the distribution function for $X(t)$ is continuous for each $t \in [0, 1]$, this depth is at most $\frac{1}{2}$, and h is the unique median with respect to the half-region depth given by X.

In addition, the process Y has many (actually infinitely many) modified half-region depth medians beyond those already obtained as in Example 1. These are of interest as they are also continuous functions on $[0, 1]$. For example, let $k(t) : [0, 1/2] \to [0, 1/2]$ be continuous with $k(1/2) = 0$, and define

$$H(t) = \begin{cases} -k(t), & 0 \leq t \leq 1/2 \\ k(1-t), & 1/2 \leq t \leq 1. \end{cases}$$

Then, H is continuous, and with $V = U + \frac{1}{2}$ we have

$$\int_0^{1/2} P(U \leq H(t))\, dt = \int_0^{1/2} P(V \leq 1/2 - k(t))\, dt = \int_0^{1/2} (1/2 - k(t))\, dt$$

and

$$\int_{1/2}^{1} P(U \leq H(t))\, dt = \int_{1/2}^{1} P(V \leq 1/2 + k(1-t))\, dt = \int_0^{1/2} (1/2 + k(s))\, ds.$$

The sum of these is $\int_0^{1/2} 1\, dt = 1/2$, and as before, each such H is a continuous median.

It is easy to find other such medians for modified half-region depth in this situation (simply take k to be something other than continuous), and perhaps one suspects that this is because the process Y is very special. This may be part of the story, but not all of it, as there are other processes Y that present similar problems. Here is one for which the modified half-region depth has some unusual properties.

Example 3. Here we assume $T = \{1, 2, 3, \ldots\}$, and $M(T) = \ell^\infty(T)$, the linear space of bounded real sequences, with norm $\|\mathbf{a}\|_\infty = \sup_{t \in T} |a(t)|$ for $\mathbf{a} = \{a(t) : t \geq 1\} \in M(T)$. For $G(t), t \geq 1$, i.i.d. centered Gaussian random variables with variance one, we define $Y(t) = G(t) \wedge 1$ if $G(t) \geq 0$, and $Y(t) = G(t) \vee (-1)$ if $Y(t) \leq 0$. We let $P = \mathcal{L}(Y)$ and for Z also $N(0, 1)$, and independent of $\{G(t) : t \in T\}$, we define $X(t) = Y(t) + Z$ for $t \in T$, and let Q denote the law of $\{X(t) : t \in T\}$ on $M(T)$. Then, $D(\mathbf{a}, P)$ and $D(\mathbf{a}, Q)$ defined as in (3) denote the half-region depths of $\mathbf{a} \in M(T)$ with respect to P and Q, and the process $\{X(t) : t \in T\}$ is the smoothed version of the process $\{Y(t) : t \in T\}$ given as in Proposition 4.

If $A \subseteq T$, we consider the probability $\rho(A) = \sum_{t \in A} q_t$, where each $q_t > 0$ and $\rho(T) = 1$. Then, for $\mathbf{a} = \{a(t) : t \in T\}$ we define the ρ-modified half-region depth of $\mathbf{a}$ with respect to P to be

$$MD(\mathbf{a}, P, \rho) := \min\left[\int_T P(Y(t) \geq a(t))d\rho(t), \int_T P(Y(t) \leq a(t))d\rho(t)\right]. \tag{6}$$

The depths of these processes have a number of interesting properties, and taking T countable simplifies some of the details in their verification. They appear as (I)–(VI) below along with a few details indicating how the conclusions are obtained, but the full proofs are in [12]. In particular, they show that even for data that is symmetric about zero, the zero vector is not the median for the modified half-region depth with respect to P, but it is the median for half-region depth for Q, the law of the smoothed process X. Moreover, we indicate in (VI) how one can embed such examples into step functions with countably many jumps on $[0, 1)$ and use Lebesgue measure for the modified half-region depth as in Examples 1 and 2.

(I) $D(\mathbf{a}, P) = 0$ and $D(\mathbf{a}, Q) > 0$ for all $\mathbf{a} \in M(T)$.

(II) If $\mathbf{a} = \{a(t) : t \in T\}$ and $a(t) = c$ for all $t \in T$, then

$$\begin{aligned} D(\mathbf{a}, Q) &= \min[Q(Y(t) \geq c - Z,\ \forall t \in T), Q(Y(t) \leq c - Z,\ \forall t \in T)] \\ &= \min[\Phi(-1-c), \Phi(-1+c)]. \end{aligned} \tag{7}$$

In particular, if $\mathbf{a} = \mathbf{0}$ is the zero vector in $M(T)$, then $D(\mathbf{0}, Q) = \Phi(-1)$, and this maximizes the half-region depth over the constant vectors.

(III) If $\mathbf{a} = \{a(t) : t \in T\}$ and $\{a(t) : t \in T\}$ has a subsequence $\{a(t_k) : k \geq 1\}$ converging to some constant c, then

$$D(\mathbf{a}, Q) \leq \min[\Phi(-1-c), \Phi(-1+c)]. \tag{8}$$

Furthermore, since $M(T)$ consists of uniformly bounded sequences, every $\mathbf{a} \in M(T)$ satisfies these assumptions, we therefore have

$$D(\mathbf{a}, Q) \leq \min[\Phi(-1-c), \Phi(-1+c)], \quad \mathbf{a} \in M(T), \tag{9}$$

which implies

$$\sup_{\mathbf{a} \in T} D(\mathbf{a}, Q) \leq \Phi(-1) = D(\mathbf{0}, Q). \tag{10}$$

Hence the zero vector is the unique median for the smoothed half-region depth when Z is $N(0, 1)$.

(IV) If $\mathbf{a} \in M(T)$, then the ρ-modified half-region depth of $\mathbf{a}$ with respect to P is such that

$$MD(\mathbf{a}, P, \rho) \leq \frac{1}{2}\left[1 + \sum_{\{t : a(t) = \pm 1\}} q_t \Phi(-1)\right]. \tag{11}$$

Furthermore, if $A \subseteq T$ with $\mathbf{a}_A = \{a_A(t) = (2I_A(t) - 1) : t \in T\}$, then $a_A(t) = \pm 1$ for all $t \in T$ and

$$\begin{aligned} MD(\mathbf{a}_A, P, \rho) &= \min\left[\sum_{t \in A} \Phi(-1) q_t + \sum_{t \in A^c} q_t, \sum_{t \in A} q_t + \sum_{t \in A^c} \Phi(-1) q_t\right] \\ &= \min[\Phi(-1)\rho(A) + \rho(A^c), \rho(A) + \Phi(-1)\rho(A^c)]. \end{aligned} \tag{12}$$

Hence, if ρ is such that there exists $A \subseteq T$ with $\rho(A) = \frac{1}{2}$, then

$$MD(\mathbf{a}_A, P, \rho) = MD(\mathbf{a}_{A^c}, P, \rho) = \frac{1}{2}[1 + \Phi(-1)]. \tag{13}$$

(V) If ρ is such that for $A \subseteq T$ we have $\rho(A) = \frac{1}{2}$, then both $\mathbf{a}_A$ and $\mathbf{a}_{A^c}$ are medians for the ρ-modified half-region depth with ρ-modified half-region depth $\frac{1}{2}[1 + \Phi(-1)]$. Moreover, the zero vector is such that $MD(\mathbf{0}, P, \rho) = \frac{1}{2}$, and hence is not a median. Given the symmetry about zero in this particular example, that zero is not a median indicates modified half-region depths lacks a property one might readily expect. Furthermore, if ρ is such that the q_t are small dyadic rationals, possibly some of them repeated and summing to one, then there can be several choices for the set A with $\rho(A) = \frac{1}{2}$.

(VI) The example discussed in (I)–(V) can be realized in the space of step functions on $[0, 1)$ with the modified half-region depth being with respect to Lebesgue measure on $[0, 1)$. That is, let $q_t = \rho(\{t\}) > 0$ for $t \in T = \{1, 2, \ldots\}$, and assume $\rho(T) = 1$. Set $I_1 = [0, q_1)$, $I_2 = [q_1, q_1 + q_2), \ldots,$ where $I_n = [q_1 + \cdots + q_{n-1}, q_1 + \cdots + q_n)$ for $n \geq 3$, and define

$$M([0, 1)) = \{a(t) : a(t) = a_j, t \in I_j, j \geq 1, \sup_{j \geq 1} |a_j| < \infty\}.$$

Now let $g_1, g_2, \ldots$ be i.i.d. $N(0, 1)$ random variables, and define $Y(t) = g_j \wedge 1, t \in I_j, j \geq 1$, when $g_j \geq 0$, and $Y(t) = g_j \vee 1, t \in I_j, j \geq 1$, when $g_j < 0$. Then, the properties established in (I)–(V) also hold for this model.

We now conclude the introduction with a few comments and some connections to other papers. The zero depth results we obtain in Section 2.1 are such that every function in the natural support of the process has half-region depth zero. The results in [5,1], which we found as we were in the final writing of this paper, differ in that they show almost every function (with respect to the law of the process) has zero half-region depth. We also observe in Section 2.1 that the size of the collection of evaluation maps used in formulating a depth in the infinite dimensional setting, can make an enormous difference. If the collection is too large it is likely the depth will be degenerate, and if it is too small the depth may not reveal details of importance in the data. This phenomenon also appears in connection with the central limit theorems we obtained for empirical processes and empirical quantile processes in [9,11], where these CLTs may fail if the class of sets is too large, or there are degeneracies in the sample paths, as with Brownian motion tied down at zero. Of course, smoothing helps, but the exact form of the depth and the evaluation maps used to define it can still produce unusual behavior. For example, in the setting of half-region depth the symmetric stable processes with stationary independent increments, cadlag paths on $[0, 1]$, and tied down at $t = 0$, are such that all cadlag paths on $[0, 1]$ have half-region depth zero (Corollary 4), whereas by Proposition 4 these processes smoothed as in (28) have positive depth. Moreover, they satisfy the consistency results and $\sqrt{n}$-asymptotics provided in Theorems 1, 2, and 3. However, if we look at the increment half-region depth formed by differences of evaluations over only countably many disjoint subintervals of $[0, 1]$ as in Corollary 3, we see that both the smoothed and the unsmoothed version of these processes yield zero increment half-region depth for every function on $[0, 1]$.

2. Zero half-region depth and how it can be eliminated

For many stochastic processes used in modeling data, half-region depth may be identically zero, but if we smooth the processes as in Proposition 4, this problem is eliminated. Section 2.1 deals with explicit classes of examples, and although these results demonstrate that zero half-region depth is a common phenomenon for many standard processes, the tools developed there should be useful when examining other processes for this problem. Furthermore, it should also be observed that the smoothing result in Section 2.2, and the consistency and $\sqrt{n}$-asymptotics of Sections 3 and 4, are independent of the proofs in Section 2.1.

2.1. Some examples

The half-region depths we examine first are for product probabilities P on the space of all real sequences $R(T)$, where $T = \{t : t = 1, 2 \ldots\}$, and for each $h \in R(T)$ the half-region depth remains to be defined as in (3). As before we will write $D(h)$ for $D(h, P)$ when the probability measure P is understood.

For many such P the uniformly bounded sequences $M(T)$ have probability zero, yet we still want to examine such situations as they are natural models of data sources, and they also can be used (as in Corollaries 2 and 3) to determine when a half-region depth may be zero. For example, if P is the product probability whose coordinates are i.i.d. centered Gaussian with variance one, then every coordinate-wise bounded sequence in $R(T)$ has half-region depth equal to zero with respect to this P. Although the set of all such sequences has P-probability zero in this example, a little thought suggests much more may be true, and our next proposition shows that under rather broad circumstances the half-region depth may be zero for all sequences in $R(T)$. In particular, it applies to the Gaussian example we mentioned, and in Corollary 1 it also allows us to examine the situation for sequences converging to zero, which are relevant when P assigns mass one to a Banach sequence space such as c_0 or ℓ_p, $1 \leq p < \infty$.

Furthermore, if P assigns probability one to $M(T)$, then using Proposition 4 at the end of this section we can show the half-region depth of every $h \in M(T)$ can be strictly positive for a smoothed version of the input data. This latter result applies to data indexed by countable or uncountable T, and $M(T)$ is as defined earlier. Of course, if T is countably infinite, then $M(T)$ is a subset of the sequence space ℓ_∞, but our results also apply to many standard stochastic processes indexed by uncountable T.

Our first result provides necessary and sufficient conditions for half-region depth to be identically zero for P a product measure on the sequence space $R(T)$. In contrast, a sufficient condition that implies a half-space depth is zero with P-probability one in $R(T)$ for various probabilities P, can be found in [10]. However, these half-space depths are not zero everywhere, so determining when they are zero, when they are positive, and consistency issues for the related empirical depth are the main concerns there.

Proposition 1. *Let $\{Z_t : t \geq 1\}$ be independent rv's on the probability space $(\Omega, \mathcal{F}, P)$ with distribution functions F_t, and assume $\mathbf{a} = \{a_t\}_{t=1}^{\infty}$ is any sequence in $R(T)$. Then, $D(\mathbf{a}, P) = 0$ if and only if*

(i) *for at least one $t \in T$, $P(Z_t \geq a_t) = 0$ or $P(Z_t \leq a_t) = 0$, or*
(ii) *for all $t \in T$, $P(Z_t \geq a_t) > 0$ and $P(Z_t \leq a_t) > 0$, and*

$$\sum_{t \in T} P(Z_t \neq a_t) = \infty. \tag{14}$$

Remark 1. Under the conditions of Proposition 1, it is immediate that the conclusion of Proposition 1 is equivalent to the claim that $D(\mathbf{a}, P) > 0$ if and only if for all $t \in T$, $P(Z_t \geq a_t) > 0$ and $P(Z_t \leq a_t) > 0$, and

$$\sum_{t \in T} P(Z_t \neq a_t) < \infty. \tag{15}$$

Proof. Under the assumptions of Proposition 1, it suffices to prove Remark 1. To do this we first we note that

$$\begin{aligned} D(\mathbf{a}, P) &= \min(P(Z_t \leq a_t \text{ for all } t \geq 1), P(Z_t \geq a_t \text{ for all } t \geq 1)) \\ &= \min\left(\prod_{t \geq 1} F_t(a_t), \prod_{t \geq 1} (1 - F_t^-(a_t))\right), \end{aligned}$$

where $F_t^-(x)$ is the left limit at $x \in \mathbb{R}$.

Hence, $D(\mathbf{a}, P) > 0$ if and only if for all $t \in T$ we have $P(Z_t \geq a_t) > 0$ and $P(Z_t \leq a_t) > 0$, and both the products

$$\prod_{t\geq 1} F_t(a_t) = \prod_{t\geq 1}(1 - P(Z_t > a_t)), \tag{16}$$

and

$$\prod_{t\geq 1}(1 - F_t^-(a_t)) = \prod_{t\geq 1}(1 - P(Z_t < a_t)) \tag{17}$$

are strictly positive. Since $P(Z_t \geq a_t) > 0$ and $P(Z_t \leq a_t) > 0$ for all $t \in T$, the products in (16) and (17) are strictly positive if and only if

$$\sum_{t\in T} P(Z_t > a_t) < \infty \quad \text{and} \quad \sum_{t\in T} P(Z_t < a_t) < \infty, \tag{18}$$

respectively. Now both series converging in (18) is equivalent to (15), and hence the proof is complete. □

Corollary 1. *Let $\{Z_t : t \geq 1\}$ be independent rv's on the probability space $(\Omega, \mathcal{F}, P)$ with continuous distribution functions F_t for $t \in T_1$, where T_1 is an infinite subset of T. Then, $D(\mathbf{a}, P) = 0$ for* **all** *sequences $\mathbf{a} = \{a_t\}_{t=1}^{\infty}$ in $R(T)$. Furthermore, if for $t \in T_1$ and some $\delta > 0$ we weaken the continuity assumption to F_t being continuous on $(-\delta, \delta)$, then $D(\mathbf{a}, P) = 0$ for* **all** *sequences $\mathbf{a} = \{a_t\}_{t=1}^{\infty}$ such that $\lim_{t\to\infty} |a_t| = 0$.*

Proof. If the distribution functions F_t are continuous on $\mathbb{R}$ for all $t \in T_1$, where T_1 is an infinite subset of T, then $P(Z_t \neq a_t) = 1$ for all such t's and (14) holds. Thus $P(Z_t \geq a_t) > 0$ and $P(Z_t \leq a_t) > 0$ for all $t \in T$, and part (ii) of Proposition 1, implies $D(\mathbf{a}, P) = 0$. Of course, if it is not the case that $P(Z_t \geq a_t) > 0$ and $P(Z_t \leq a_t) > 0$ for all $t \in T$, then we also have $D(\mathbf{a}, P) = 0$.

If the assumption of continuity is weakened as indicated, then an entirely similar argument applies for all sequences converging to zero. □

Remark 2. In the previous corollary continuity of the distributions $F_t, t \in T$, played an important role in showing zero half-region depth, but it clearly is not a necessary condition. For example, if $\{Z_t : t \in T\}$ are independent random variables with $P(Z_t = \pm c_t) = d_t, t \in T$, where $\{c_t : t \in T\}$ are strictly positive constants, $\sum_{t\in T} d_t = \infty$, and $F_t, t \in T$, is arbitrary otherwise, then Proposition 1 immediately implies for any sequence $\mathbf{a} = \{a_t : t \in T\}$

$$D(\mathbf{a}, P) = 0.$$

It is also easy to formulate two immediate consequences of Corollary 1, where natural sequential half-region depths will always be zero for probabilities which behave well in many instances, and are important in many modeling situations. Since more restrictions in the definition of a half-region depth make it easier for the depth to be zero, it is interesting to observe that in both examples the class of evaluation maps used to define the depths is again countably infinite. In the first we assume P is a centered Gaussian probability measure on a separable Banach space with infinite dimensional support. Then, it is well known that there are many sequences of continuous linear functionals $\mathcal{A} = \{\alpha_t : t \in T\} \subseteq B^*$ that are i.i.d. centered Gaussian random variables with $\int_B \alpha_t^2(x)dP(x) = 1$, and for P-almost all $x \in B$

$$\lim_{n\to\infty} \left\| x - \sum_{t=1}^{n} \alpha_t(x) S\alpha_t \right\| = 0,$$

where $\|\cdot\|$ is the norm on B, and for each $\alpha \in B^*$, $S\alpha$ is the Bochner integral $\int_B x\alpha(x)dP(x)$. Hence, with P-probability one the sequence $\mathcal{A} = \{\alpha_t : t \in T\}$ determines $x \in B$ in the sense that above series converges to x, and we define the $\mathcal{A}$-half-region depth of a vector $\mathbf{a} \in B$ to be

$$D_{\mathcal{A}}(\mathbf{a}, P) = \min\{P(\alpha_t(x) \geq \alpha_t(\mathbf{a}),\ \forall t \in T), P(\alpha_t(x) \leq \alpha_t(\mathbf{a}),\ \forall t \in T)\}. \tag{19}$$

Corollary 2. *If P is a centered Gaussian measure on a separable Banach space with infinite dimensional support, and $\mathcal{A} = \{\alpha_t : t \in T\} \subseteq B^*$ is as above, then for all $\mathbf{a} \in B$*

$$D_{\mathcal{A}}(\mathbf{a}, P) = 0. \tag{20}$$

In the second application of Proposition 1 we let $X = \{X(t) : t \in [0, 1]\}$ be a symmetric non-degenerate stable process with stationary independent increments and cadlag sample paths on $[0, 1]$. If X is tied down at $t = 0$, then Proposition 3 shows that the half-region depth of every cadlag path on $[0, 1]$ is zero with respect to P, and here we examine what might be considered a natural depth for the increments of these processes. Unfortunately, this depth is also zero for every function on $[0, 1]$.

Corollary 3. *Let $\mathcal{I} = \{I_j = [u_j, v_j], j \geq 1\}$ consist of disjoint intervals of* $[0, 1]$*, and define the increment half-region depth for every function h on* $[0, 1]$ *with respect to $P = \mathcal{L}(X)$ and $\mathcal{I}$ by*

$$D_{\mathcal{I}}(h, P) = \min\{P(X(I_j) \geq h(I_j),\ \forall j \geq 1), P(X(I_j) \leq h(I_j),\ \forall j \geq 1)\},$$

where $f(I_j) = f(v_j) - f(u_j)$ for every function f on $[0, 1]$*. Then,*

$$D_{\mathcal{I}}(\mathbf{a}, P) = 0. \tag{21}$$

As mentioned above, both Corollaries 2 and 3 are immediate from Corollary 1, and the continuity of the relevant distribution functions.

The next proposition will allow us to obtain several more typical examples of "zero half-region depth".

Proposition 2. *Let $\{X(t) : t \in T\}$ and $\{Y(t) : t \in T\}$ be i.i.d. stochastic processes on $(\Omega, \mathcal{F}, P)$, all of whose sample paths are in the linear space of functions $M(T)$. If $h \in M(T)$ and*

$$P(X - Y \preceq_S 0) = 0 \tag{22}$$

for some subset S of T_0, then $D(h, P) = 0$.

Proof. If the depth of $h \in M(T)$ is positive, then the product, $P(h \preceq X) \cdot P(X \preceq h)$, is positive. So, since we always are assuming (1), (4) and (22), we then have

$$\begin{aligned} 0 &< P(h \preceq_{T_0} X) \cdot P(X \preceq_{T_0} h) = P(h \preceq_{T_0} X, Y \preceq_{T_0} h) \\ &\leq P(Y \preceq_{T_0} X) \leq P(Y - X \preceq_S 0) = 0. \quad \square \end{aligned} \tag{23}$$

Corollary 4. *Let X be an independent increment process with paths in the Skorohod space $D[0, 1]$ such that*

1. *the increments have a continuous distribution, and*
2. *$P(X(0) = 0) = 1$.*

If $h \in D[0, 1]$, then $D(h, P) = 0$.

Proof. Let $Z = X - Y$, where X and Y are defined on the probability space $(\Omega, \mathcal{F}, P)$, Y is an independent copy of X, and X and Y have sample paths in $D[0, 1]$. Using Proposition 2, with T_0 the rational numbers in $[0, 1]$ and $S = \{\frac{1}{k} : k = 1, 2, \ldots\}$, we only have to check that $P(Z \preceq_S 0) = 0$. We will assume not. But, by the (right) continuity at $t = 0$ and telescoping terms we have

$$Z\left(\frac{1}{k}\right) = \lim_{r\to\infty}\left[Z\left(\frac{1}{k}\right) - Z\left(\frac{1}{r+1}\right)\right] = \lim_{r\to\infty}\sum_{j=k}^{r}\Delta_j(Z) = \sum_{j=k}^{\infty}\Delta_j(Z), \tag{24}$$

where $\Delta_j(Z) = [Z(\frac{1}{j}) - Z(\frac{1}{j+1})]$. Therefore, by our choice of S and (24)

$$\begin{aligned} 0 &< P(Z \preceq_S 0) = P\left(\sum_{j=k}^{\infty}\Delta_j(Z) \leq 0, \forall k \geq 1\right) \\ &\leq P\left(\sum_{j=k}^{\infty}\Delta_j(Z) \leq 0,\ \text{eventually in } k\right). \end{aligned}$$

This last event is in the tail σ-field of $\{Z(\frac{1}{j}) - Z(\frac{1}{j+1}) : j \geq 1\}$, so by Kolmogorov's zero–one law and the symmetry of Z, we have

$$P\left(\sum_{j=k}^{\infty}\Delta_j(Z) \leq 0,\ \text{eventually in } k\right) = P\left(\sum_{j=k}^{\infty}\Delta_j(Z) \geq 0,\ \text{eventually in } k\right),$$

and both probabilities are one. Hence, by (24) we have

$$P\left(Z\left(\frac{1}{k}\right) = 0 \text{ eventually in } k\right) = 1,$$

and therefore $P(Z(1/k) - Z(1/k+1) = 0 \text{ eventually in } k) = 1$. By the independence of the increments this last statement is equivalent to

$$\sum_{k=1}^{\infty} P(Z(1/k) - Z(1/k+1) \neq 0) < \infty.$$

Since each term is 1, we have a contradiction. $\square$

Remark 3. Let $X = \{X(t) : t \in [0, 1]\}$ be a symmetric stable process with parameter $r \in (0, 2]$, and stationary independent increments with paths in $D[0, 1]$. If we also have $P(X(0) = 0) = 1$, then the conclusion of Corollary 4 immediately holds. If $r = 2$ and X is Brownian motion with continuous sample paths, then the result also holds in that setting. However, if X is a Poisson process with parameter $\lambda > 0$, then the first condition of Corollary 4 does not hold. And, if ξ has an exponential distribution with mean λ, then

$$P(X(t) \le 0 \text{ for all } t \in [0, 1]) = P(\xi > 1) > 0.$$

Therefore, the half-space depth of the 0 function is positive. Of course, the same conclusion is valid for compound Poisson processes starting at zero with probability one.

Corollary 5. *Let $X = \{X(t) : t = (t_1, t_2) \in [0, 1] \times [0, 1]\}$ be a centered Brownian sheet with covariance*

$$E(X(t_1, t_2)X(s_1, s_2)) = \min\{s_1, t_1\} \min\{s_2, t_2\}, \tag{25}$$

and continuous paths on $T = [0, 1] \times [0, 1]$. If h is a continuous function on T and P is the law of X, then the half-region depth $D(h, P) = 0$.

Proof. Let $Z = X - Y$, where X and Y are defined on the probability space $(\Omega, \mathcal{F}, P)$, Y is an independent copy of X, and X and Y have sample paths in $C(T)$. Let T_0 be the subset of T consisting of points with both coordinates rational numbers in $[0, 1]$ and let $S = \{t \in T_0 : t = (t_1, t_1)\}$. Using Proposition 2, we only have to check that $P(Z \preceq_S 0) = 0$. We will assume not. Then,

$$0 < P(Z(t) \le 0, \text{ for all } t \in S) = P(B(u) \le 0 \text{ for all, } u \in [0, 1] \cap Q),$$

where Q is the rational numbers and $B(u) = Z(u, u)$, $u \in [0, 1]$. Since $\{B(u) : u \in [0, 1]\}$ is a Brownian motion process with continuous sample paths and $P(B(0) = 0) = 1$, we have

$$P(B(u) \le 0 \text{ for all } u \in [0, 1] \cap Q) = P(B(u) \preceq_{[0,1]} 0) = 0,$$

where the last equality follows from Remark 3. □

The next result applies to many Markov processes with or without independent increments.

Proposition 3. *Assume the stochastic process $X = \{X_t : 0 \le t \le 1\}$ has sample paths in the Skorohod space $D[0, 1]$ and it satisfies*

1. *the Blumenthal zero–one law at $t = 0$, i.e., for every $A \in \mathcal{F}_0^+ := \cap_{t>0} \mathcal{F}_t$ we have $P(A) = 0$ or 1, where $\mathcal{F}_t = \cup_{0\le s\le t} \sigma(X_s)$ and $\sigma(X_s)$ is the minimal sigma-field making X_s measurable, and*
2. *for every $t > 0$, $X(t)$ has a continuous distribution function.*

Then, the half-region depth $D(h, P) = 0$ for every $h \in D[0, 1]$.

Proof. If $D(h, P) > 0$, then

$$P(X(\cdot) \succeq_{[0,1]} h(\cdot)) > 0 \tag{26}$$

and

$$P(X(\cdot) \preceq_{[0,1]} h(\cdot)) > 0. \tag{27}$$

For $n \ge 1$, let

$$E_n = \{X(\cdot) \succeq_{[0,1/n]} h(\cdot)\} \quad \text{and} \quad F_n = \{X(\cdot) \preceq_{[0,1/n]} h(\cdot)\}.$$

Then, for every integer k

$$E = \{X(t) \ge h(t) \text{ eventually as } t \downarrow 0\} = \cup_{n\ge k} E_n,$$

and

$$F = \{X(t) \le h(t) \text{ eventually as } t \downarrow 0\} = \cup_{n\ge k} F_n.$$

This implies $E \in \mathcal{F}_{1/k}$, $F \in \mathcal{F}_{1/k}$ for all $k \ge 1$, and therefore $E, F \in \mathcal{F}_0^+ = \cap_{k=1}^\infty \mathcal{F}_{1/k}$. Now (26) implies $P(E) > 0$ and (27) implies $P(F) > 0$, so the Blumenthal zero–one law implies $P(E) = P(F) = 1$. Since the events E_n and F_n increase in n, we have that there exists a k_0 such that $n \ge k_0$ implies $P(E_n) > 3/4$ and $P(F_n) > 3/4$. Hence

$$P(E_n \cap F_n) > 1/2 \quad \text{for all } n \ge k_0.$$

Since $E_n \cap F_n = \{X(t) = h(t), \ \forall t \in [0, 1/n]\}$, this is a contradiction to the fact that $X(t)$ has a continuous distribution for all $t > 0$. Thus the half-region depth of $h \in D[0, 1]$ must be zero. □

2.2. Eliminating half-region zero depth by smoothing

Although sample continuous Brownian motion, tied down to be zero at $t = 0$ with probability one, assigns zero half-region depth to all functions $h \in C[0, 1]$, by starting the process randomly with a density changes things dramatically. This follows immediately from the next proposition, and hence in order to be assured half-region depth is non-trivial, we use smoothing in the results that follow in subsequent sections. Moreover, the precise assumptions used for smoothing in these later results are also important in other parts of their proofs. The smoothed stochastic process $X = \{X(t) : t \in T\}$ will be such that

$$X(t) = Y(t) + Z, t \in T, \tag{28}$$

where Z is a real valued random variable independent of the process $Y = \{Y(t) : t \in T\}$, Z has density $f_Z(\cdot)$ on $\mathbb{R}$, Y has sample paths in the linear space $M(T)$, and we are assuming $M(T)$ is such that (1) holds. Of course, then (2) also holds, and since we are assuming $M(T)$ contains the constant functions on T, X also has its sample paths in $M(T)$.

Proposition 4. *Let $X(t) = Y(t) + Z, t \in T$, where $Y = \{Y(t) : t \in T\}$ has sample paths in the linear space $M(T)$ satisfying (1) and Z is independent of Y with density f_Z. If $f_Z > 0$ a.s. with respect to Lebesgue measure on $\mathbb{R}$ and $h \in M(T)$, then the half-region depth of h determined by $\{X(t) : t \in T\}$ is strictly positive.*

Proof. Let $h \in M(T)$. Then,

$$P(X \succeq h) = \int_{-\infty}^{\infty} P(Y(t) \geq h(t) - u,\ \forall t \in T | Z = u) f_Z(u)du. \tag{29}$$

Since (1) holds there exists an constant $c > 0$ such that $P(\|Y\|_\infty \leq c) > \frac{1}{2}$ and hence for $u > 2c + \|h\|_\infty$ we have

$$P(Y(t) \geq h(t) - u,\ \forall t \in T | Z = u) \geq P(Y(t) \geq -2c,\ \forall t \in T) > \frac{1}{2}. \tag{30}$$

Since $f_Z > 0$ a.s., by combining (29) and (30) we have

$$P(X \succeq h) \geq \int_{2c+\|h\|_\infty}^{\infty} \frac{1}{2} f_Z(u)du > 0.$$

Similarly, $P(X \preceq h) > 0$ for all $h \in M(T)$, and hence $D(h, P) > 0$ for all $h \in M(T)$. □

Remark 4. If $P(\|Y\|_\infty \leq c) > 0$ for all $c > 0$, then it is easy to see from the proof of the previous proposition that the half-region depth could be strictly positive for some $h \in M(T)$ without the density being strictly positive on all of $\mathbb{R}$. Moreover, X is a good approximation of Y in the sense that $E(\|X - Y\|_\infty) = E(|Z|)$, and we are free to take $E(|Z|) > 0$ arbitrarily small.

Remark 5. Let $T = [a, b], -\infty < a < b < \infty$, and assume $M(T)$ denotes the real-valued cadlag paths on T. If $X = \{X(t) : t \in T\}$ has paths in $M(T)$, and $Z := X(a)$ is independent of $\{Y(t) = X(t) - X(a), t \in T\}$ with density $f_Z > 0$ a.s. with respect to Lebesgue measure on $\mathbb{R}$, then Proposition 4 implies the half-region depth with respect to $P = \mathcal{L}(X)$ is strictly positive on $M(T)$. Hence, under these conditions no smoothing is required to be certain the depth is strictly positive.

3. Consistency for empirical half-region depth

The consistency result we prove depends on two lemmas, which are also important for the $\sqrt{n}$ asymptotics we obtain in Theorem 2. The proof of consistency is an application of empirical process ideas involving the Blum–Dehardt Theorem and bracketing entropy.

Let $X(t) = Y(t) + Z, t \in T$, where $Y = \{Y(t) : t \in T\}$ has sample paths in the linear space $M(T)$ satisfying (1), and Z is independent of Y with density f_Z. Also, assume $X_1, X_2, \ldots$ are i.i.d. copies of the process X with sample paths in $M(T)$ and that $X, X_1, X_2, \ldots$ are defined on the probability space $(\Omega, \mathcal{S}, P)$. Then, with half-region depth and half-region empirical depth defined as in (3) and (5), and since for real numbers a, b, c, d

$$|\min\{a, b\} - \min\{c, d\}| \leq |a - c| + |b - d|, \tag{31}$$

the classical strong law of large numbers implies for each $h \in M(T)$

$$\lim_{n\to\infty} |D_n(h) - D(h)| = 0 \tag{32}$$

with probability one. The theorem below refines (32) to be uniform over $h \in E$, where E is a suitably chosen subset of $M(T)$.

Notation 1. For a function $f : \Omega \to \bar{\mathbb{R}}$ we use the notation f^* to denote a measurable cover function (see Lemma 1.2.1 van der Vaart and Wellner [16]).

Theorem 1. *Let $X(t) = Y(t) + Z, t \in T$, where $Y = \{Y(t) : t \in T\}$ has sample paths in the linear space $M(T)$, and Z is independent of Y with density $f_Z(\cdot)$ on $\mathbb{R}$ that is absolutely continuous and its derivative $f_Z'(\cdot)$ is in $L_1(\mathbb{R})$. Also, assume $X_1, X_2, \ldots$ are i.i.d. copies of the process X with sample paths in $M(T)$ and that $X, X_1, X_2, \ldots$ are defined on the probability space $(\Omega, \mathcal{F}, P)$. If E is subset of $M(T)$ such that for every $r > 0$*

$$E_r = E \cap \{f \in M(T) : \|f\|_\infty \le r\} \tag{33}$$

is a sup-norm compact subset of $M(T)$, then with probability one

$$\lim_{n\to\infty} \sup_{h\in E} |D_n(h) - D(h)|^* = 0. \tag{34}$$

In order to prove this result we first establish some lemmas which will also be useful in our refinements of (34) that follow below. For $h \in M(T)$ we define the stochastic process $\{W_h : h \in M(T)\}$ on $(\Omega, \mathcal{F}, P)$, where

$$W_h \equiv W(h) = \inf_{t\in T}(X(t) - h(t)), \quad h \in M(T). \tag{35}$$

Lemma 1. *Let f be a probability density on $\mathbb{R}$ which is absolutely continuous and such that its derivative f' is in $L^1(\mathbb{R})$. Then,*

$$\int_{\mathbb{R}} |f(x+\delta) - f(x)|dx \le |\delta| \int_{\mathbb{R}} |f'(x)|dx. \tag{36}$$

Proof. If $\delta \ge 0$, then

$$\int_{\mathbb{R}} |f(x+\delta) - f(x)|\, dx \le \int_{\mathbb{R}}\int_{\mathbb{R}} |f'(u)| I_{x\le u\le x+\delta}\, du\, dx = (\text{by Fubini}) \int_{\mathbb{R}} |f'(u)| \int_{\mathbb{R}} I_{x\le u\le x+\delta}\, dx\, du = \delta \int_{\mathbb{R}} |f'(u)|\, du,$$

which gives (36). The case $\delta < 0$ follows similarly. □

Lemma 2. *Let X be as in (28) with Y and Z satisfying the assumptions of Theorem 1, and assume W_h be as in (35). Then, for $h_1, h_2 \in M(T)$ we have*

$$|W_{h_1} - W_{h_2}| \le \|h_1 - h_2\|_\infty. \tag{37}$$

Hence, if $\|h_1 - h_2\|_\infty \le \delta$, then

$$|P(W_{h_1} \ge x) - P(W_{h_2} \ge x)| \le P(x - \delta \le W_{h_1} \le x) + P(x - \delta \le W_{h_2} \le x), \tag{38}$$

and we also have

$$|P(W_{h_1} \ge x) - P(W_{h_2} \ge x)| \le 2\delta \int_{\mathbb{R}} |f_Z'(x)|dx. \tag{39}$$

Proof. First observe that for all $s \in T$

$$\inf_{t\in T} \big(X(t) - h_1(t)\big) \le X(s) - h_1(s) \le X(s) - h_2(s) + \|h_2 - h_1\|_\infty.$$

Hence,

$$\begin{aligned} W_{h_1} &= \inf_{t\in T}\big(X(t) - h_1(t)\big) \le \inf_{s\in T}\big(X(s) - h_2(s)\big) + \|h_2 - h_1\|_\infty \\ &= W_{h_2} + \|h_2 - h_1\|_\infty \end{aligned}$$

and interchanging h_1 and h_2 we have (37).

Hence, if $\|h_1 - h_2\|_\infty \le \delta$, we then have from (37) that

$$P(W_{h_1} \ge x) \le P(W_{h_2} \ge x) + P(x - \delta \le W_{h_2} \le x) \tag{40}$$

and

$$P(W_{h_2} \ge x) \le P(W_{h_1} \ge x) + P(x - \delta \le W_{h_1} \le x), \tag{41}$$

and (40) and (41) combine to give (38).

To verify (39) we define for $h \in M(T)$

$$F(h, x) = P(W_h \ge x). \tag{42}$$

From (28), $F(h,x) = P(\inf_{t\in T}(Y(t)-h(t)) + Z \geq x)$, and hence the independence of Y and Z implies

$$F(h,x) = \int_{\mathbb{R}} P_Y(\inf_{t\in T}(Y(t)-h(t)) \geq x-y) f_Z(y)dy. \tag{43}$$

Letting $\xi(h) = \inf_{t\in T}(Y(t)-h(t))$, we see (43) implies

$$F(h,x_1) - F(h,x_2) = \int_{\mathbb{R}} P_Y(\xi(h) \geq s)[f_Z(x_1-s) - f_Z(x_2-s)]ds. \tag{44}$$

Therefore,

$$|F(h,x_1) - F(h,x_2)| \leq \int_{\mathbb{R}} |f_Z(x_1-s) - f_Z(x_2-s)|ds, \tag{45}$$

and setting $u = x_1 - s$ we have $x_2 - s = (x_2 - x_1) + u$, so Lemma 1 implies

$$|F(h,x_1) - F(h,x_2)| \leq \int_{\mathbb{R}} |f_Z(u) - f_Z(u + (x_2-x_1))|du \leq |x_1 - x_2| \int_{\mathbb{R}} |f'_Z(x)|dx.$$

Thus the lemma is proven since (38) and the above combine to give (39) when $\|h_1 - h_2\|_\infty \leq \delta$. □

Proof. In order to verify (34) we first will show for every $\epsilon > 0$ there is an $r_0 < \infty$ such that the strong law of large numbers implies with probability one that

$$\limsup_{n\to\infty} \sup_{\{h:\|h\|_\infty \geq r_0\}} D_n(h) \leq P(\|X\|_\infty \geq r_0) \leq \epsilon, \tag{46}$$

and

$$\lim_{r\to\infty} \sup_{\{h:\|h\|_\infty \geq r\}} D(h) = 0. \tag{47}$$

The argument for (46) and (47) is essentially the proof of Proposition 5 in [14], but the details are included below.

To prove (47) we observe

$$\sup_{\|h\|_\infty \geq r} D(h) \leq A_r + B_r,$$

where

$$A_r = \sup_{\|h\|_\infty \geq r, \|h\|_\infty = \sup_{t\in T} h(t)} P(X \succeq h),$$

and

$$B_r = \sup_{\|h\|_\infty \geq r, \|h\|_\infty = \sup_{t\in T}(-h(t))} P(X \preceq h).$$

Thus

$$\begin{aligned} A_r &\leq \sup_{\|h\|_\infty = \sup_{t\in T} h(t) \geq r} P(\sup_{t\in T} X(t) \geq \sup_{t\in T} h(t)) \\ &\leq \sup_{\|h\|_\infty = \sup_{t\in T} h(t) \geq r} P(\|X\|_\infty \geq \|h\|_\infty) \leq P(\|X\|_\infty \geq r), \end{aligned}$$

and

$$\begin{aligned} B_r &\leq \sup_{\|h\|_\infty = \sup_{t\in T}(-h(t)) \geq r} P(\inf_{t\in T} X(t) \leq \inf_{t\in T} h(t)) \\ &\leq \sup_{\|h\|_\infty = \sup_{t\in T}(-h(t)) \geq r} P(\|X\|_\infty \geq \|h\|_\infty) \leq P(\|X\|_\infty \geq r), \end{aligned}$$

and hence we have (47). To prove (46) we note that

$$\sup_{\|h\|_\infty \geq r} D_n(h) \leq A_{r,n} + B_{r,n}$$

where

$$A_{r,n} = \sup_{\|h\|_\infty \ge r, \|h\|_\infty = \sup_{t\in T} h(t)} \frac{1}{n}\sum_{j=1}^n I(X_j \succeq h),$$

and

$$B_{r,n} = \sup_{\|h\|_\infty \ge r, \|h\|_\infty = \sup_{t\in T}(-h(t))} \frac{1}{n}\sum_{j=1}^n I(X_j \preceq h).$$

Thus, in similar fashion it follows that

$$A_{r,n} \le \sup_{\|h\|_\infty = \sup_{t\in T} h(t) \ge r} \frac{1}{n}\sum_{j=1}^n I(\|X_j\|_\infty \ge \|h\|_\infty) \le \frac{1}{n}\sum_{j=1}^n I(\|X_j\|_\infty \ge r),$$

and

$$B_{r,n} \le \sup_{\|h\|_\infty = \sup_{t\in T}(-h(t)) \ge r} \frac{1}{n}\sum_{j=1}^n I(\|X_j\|_\infty \ge \|h\|_\infty) \le \frac{1}{n}\sum_{j=1}^n I(\|X_j\|_\infty \ge r),$$

and therefore we have (46).

Since $\epsilon > 0$ is arbitrary, (47) and (46) combine to imply (34) provided we show for every $r > 0$ that with probability one

$$\lim_{n\to\infty} \sup_{h\in E_r} |D_n(h) - D(h)|^* = 0, \tag{48}$$

where E_r is defined as in (33). The proof of (48) follows from the Blum–Dehardt Theorem using the bracketing entropy for E_r as in [4, p. 235]. That is, since E_r is compact in $M(T)$ with respect to the sup-norm, for every $\delta > 0$ implies there exists finitely many points $\{h_1, \ldots, h_{k(\delta)}\} \subseteq E_r$ such that

$$\sup_{h\in E_r} \inf_{h_j} \|h - h_j\|_\infty \le \delta.$$

In addition, the brackets $F(\delta, h_j) = \{z \in M(T) : h_j(t) - \delta \le z(t) \le h_j(t) + \delta\}$ have union covering E_r with $z \in F(\delta, h_j)$ implying

$$I(X \succeq h_j + \delta) \le I(X \succeq z) \le I(X \succeq h_j - \delta).$$

Hence, for $\epsilon > 0$ fixed, and $\delta = \delta(\epsilon) > 0$ such that $4\delta \int_{\mathbb{R}} |f_Z'(x)|dx \le \epsilon$, we have from (39) that $\mathcal{E}_r \equiv \{I(X \succeq h) : h \in E_r\}$ is a subset of

$$\cup_{j=1}^{k(\delta(\epsilon))} \{I(X \succeq z) : z \in E_r, I(X \succeq h_j + \delta) \le I(X \succeq z) \le I(X \succeq h_j - \delta)\},$$

and $\|I(X \succeq h_j - \delta) - I(X \succeq h_j + \delta)\|_1 \le \epsilon$, where $\|\cdot\|_1$ denotes the L_1 norm with respect to P. Hence, for every $\epsilon > 0$ we have $\mathcal{E}_r$ covered by finitely many L_1-brackets of diameter ϵ. A similar argument can be made for

$$\mathcal{F}_r \equiv \{I(X \preceq h) : h \in E_r\},$$

and hence (48) holds by (31) and the Blum–Dehardt result mentioned above. Combining (47), (46), and (48) we have (34). □

3.1. Some remarks on the C1 condition in [14]

Let $X = \{X_t : t \in [0, 1]\}$ be a sample continuous stochastic process, and assume P is the Borel probability on $C[0, 1]$ induced by X. The main focus of the paper López-Pintado and Romo [14] is the formulation of a consistency result for half-region depth that is uniform over an equicontinuous family of functions on [0, 1], where the depth is with respect to X, or equivalently the probability distribution P. One of the crucial assumptions in this endeavor is that P satisfy their C1 condition, where

C1: Given $\epsilon > 0$, there exists a $\delta > 0$, such that for every pair of functions $h_1, h_2 \in C[0, 1]$ with $\|h_1 - h_2\|_\infty \le \delta$ implies

$$P(h_1 \preceq_{[0,1]} X \preceq_{[0,1]} h_2) < \epsilon. \tag{49}$$

This condition appears on the bottom of page 1687 in [14]. The notation in [14] is slightly different than that above, but (49) is consistent with their use of C1 on page 1688 of López-Pintado and Romo [14]. However, the main problem with (49) as used in [14] is two-fold. First, in their proof of Theorem 3 of López-Pintado and Romo [14] it is applied to functions h_1, h_2 which are not continuous, and secondly it is claimed that for $h_1, h_2 \in C[0, 1]$ with $h_1 \preceq h_2$

$$P(h_1 \preceq X) - P(h_2 \preceq X) = P(h_1 \preceq X \preceq h_2), \tag{50}$$

which is far from being true since

$$P(h_1 \preceq X) - P(h_2 \preceq X) = P(\{h_1 \preceq X\} \cap \{h_2 \preceq X\}^c). \tag{51}$$

Hence there are some major concerns with their proof, and in Theorem 1 we obtained a result that alleviates such concerns. Moreover, we have taken care to discuss when half-region is non-trivial, and how to eliminate the problem of it being trivial by using smoothing. Another question one might ask is whether the quantity

$$|P(h_1 \preceq X) - P(h_2 \preceq X)|, \tag{52}$$

can be made arbitrarily small when $h_1, h_2 \in C[0, 1]$ provided $\|h_1 - h_2\|_\infty$ is sufficiently small. This is an important ingredient in our proof, and we established sufficient conditions for this in Lemma 2, but it is easy to see it may fail in many cases. For example, let

$$X_t = \max\{0, B_t\}, \quad t \in [0, 1],$$

where $\{B(t) : t \in [0, 1]\}$ is a sample continuous Brownian motion such that $P(B(0) = 0) = 1$. Thus for $h_1(t) = 0, t \in [0, 1]$, and $h_2(t) = \delta > 0, t \in [0, 1]$, we have $\|h_1 - h_2\|_\infty = \delta$ and $P(h_1 \preceq X) - P(h_2 \preceq X) = 1$, no matter how small the constant δ.

4. Additional asymptotics for half-region depth

Our next result shows the consistency result of (34) can be refined to include rates of convergence provided we restrict the set E to be a sup-norm compact subset of $M(T)$ satisfying the entropy condition

$$\int_{0+} (\log N(E, \epsilon, \|\cdot\|_\infty))^{\frac{1}{2}} \epsilon^{-\frac{1}{2}} d\epsilon < \infty, \tag{53}$$

where $N(E, \epsilon, \|\cdot\|_\infty)$ is thecovering number of E with ϵ-balls in the $\|\cdot\|_\infty$-norm. In particular, since the processes

$$\{\sqrt{n}(D_n(h) - D(h)) : h \in E\}, \quad n \geq 1, \tag{54}$$

live in $\ell_\infty(E)$, we examine their asymptotic behavior in that setting, and in Corollary 6 produce sub-Gaussian tail bounds that are uniform in n. The basic notation is as in Section 3, and we freely use the empirical process ideas for weak convergence in the space $\ell_\infty(E)$ as presented in [4,16].

In the proof of these results we have need for the stochastic processes $\{H_{n,1,h} : h \in E\}, n \geq 1$, and $\{H_{n,2,h} : h \in E\}, n \geq 1$, where

$$H_{n,1,h} \equiv \frac{1}{\sqrt{n}} \sum_{j=1}^{n} [I(X_j \succeq h) - P(X_j \succeq h)], \tag{55}$$

and

$$H_{n,2,h} \equiv \frac{1}{\sqrt{n}} \sum_{j=1}^{n} [I(X_j \preceq h) - P(X_j \preceq h)]. \tag{56}$$

The first step of our proof will be to show that each of these processes satisfies the CLT in $\ell_\infty(E)$ with limits that are centered, sample path bounded, Gaussian processes $G_1 = \{G_{1,h} : h \in E\}$ and $G_2 = \{G_{2,h} : h \in E\}$, respectively, that are uniformly continuous on E with respect to their L_2-distances, and have covariance functions

$$E(G_{1,h_1} G_{1,h_2}) = P(X \succeq h_1, X \succeq h_2) - P(X \succeq h_1)P(X \succeq h_2), \quad h_1, h_2 \in E, \tag{57}$$

and

$$E(G_{2,h_1} G_{2,h_2}) = P(X \preceq h_1, X \preceq h_2) - P(X \preceq h_1)P(X \preceq h_2), \quad h_1, h_2 \in E. \tag{58}$$

In the following theorem these Gaussian processes also appear in connection with the limiting finite dimensional distributions of the centered empirical half-region depth processes given in (54), see (60)–(62).

Theorem 2. *Let $X(t) = Y(t) + Z, t \in T$, where $Y = \{Y(t) : t \in T\}$ has sample paths in the linear space $M(T)$ and Z is independent of Y with density $f_Z(\cdot)$ on $\mathbb{R}$ that is absolutely continuous and its derivative $f_Z'(\cdot)$ is in $L_1(\mathbb{R})$. Also, assume $X_1, X_2, \ldots$ are i.i.d. copies of the process X with sample paths in $M(T)$ and that $X, X_1, X_2, \ldots$ are defined on the probability space $(\Omega, \mathcal{F}, P)$. If E is a sup-norm compact subset of $M(T)$ satisfying the entropy condition (53), then*

$$\lim_{r \to \infty} \sup_{n \geq 1} P^*(\sup_{h \in E} \sqrt{n}|D_n(h) - D(h)| \geq r) = 0, \tag{59}$$

where P^* denotes the outer probability for subsets of $(\Omega, \mathcal{F}, P)$. Furthermore, there is a stochastic process $\{\Gamma_h : h \in E\}$ such that the finite dimensional distributions of the processes $\{\sqrt{n}(D_n(h) - D(h)) : h \in E\}, n \geq 1$ converge weakly to $\{\Gamma_h : h \in E\}$, where

$$\mathcal{L}(\Gamma_h) = \mathcal{L}(G_{1,h}) \quad \text{for } h \in E \quad \text{and} \quad P(X \succeq h) < P(X \preceq h), \tag{60}$$

$$\mathcal{L}(\Gamma_h) = \mathcal{L}(G_{2,h}) \quad \text{for } h \in E \quad \text{and} \quad P(X \preceq h) < P(X \succeq h), \tag{61}$$

and

$$\mathcal{L}(\Gamma_h) = \mathcal{L}(\min\{G_{1,h}, G_{2,h}\}) \quad \text{for } h \in E \quad \text{and} \quad P(X \succeq h) = P(X \preceq h). \tag{62}$$

Proof. Since (31) holds and $X, X_1, X_2, \ldots$ are i.i.d. we have

$$\sqrt{n}|D_n(h) - D(h)| \leq |H_{n,1,h}| + |H_{n,2,h}|. \tag{63}$$

Hence (59) will hold provided we show

$$\lim_{r\to\infty} \sup_{n\geq 1} P^*(\sup_{h\in E} |H_{n,1,h}| \geq r) = 0, \tag{64}$$

and

$$\lim_{r\to\infty} \sup_{n\geq 1} P^*(\sup_{h\in E} |H_{n,2,h}| \geq r) = 0. \tag{65}$$

To verify (64) and (65) it suffices to show that the stochastic processes $\{H_{n,1,h} : h \in E\}$ and $\{H_{n,2,h} : h \in E\}$ converge weakly in $\ell_\infty(E)$ to the centered Gaussian processes G_1 and G_2, respectively. That is, once these CLT's hold, then item (iii) of Theorem 1.3.4 of [16] provides the conclusion we need.

In order to formulate these CLTs in $\ell_\infty(E)$ we let $\mathcal{C}$ be a family of subsets of $M(T)$ indexed by E, where

$$\mathcal{C} = \mathcal{C}_{\inf} \cup \mathcal{C}_{\sup}, \tag{66}$$

$$\mathcal{C}_{\inf} = \{C_h : h \in E\} \quad \text{and} \quad \mathcal{C}_{\sup} = \{\hat{C}_h : h \in E\}, \tag{67}$$

$$C_h = \{z \in M(T) : \inf_{t\in T}(z(t) - h(t)) \geq 0\}, \tag{68}$$

and

$$\hat{C}_h = \{z \in M(T) : \sup_{t\in T}(z(t) - h(t)) \leq 0\}. \tag{69}$$

Of course, since we are assuming $M(T)$ is a linear space such that (1) holds we have the inf and sup defining the sets C_h and D_h, respectively, are the same when $t \in T$ is replaced by $t \in T_0$.

Since $\ell_\infty(E)$ is a separable Banach space only when E is finite, we need to use weak convergence in the non-separable setting, and proceed to verify that $\mathcal{C}_{\inf}$ and $\mathcal{C}_{\sup}$ are both P-Donsker classes of sets. Then, since a finite union of P-Donsker classes is P-Donsker, we will have $\mathcal{C}$ also P-Donsker.

To show $\mathcal{C}_{\inf}$ is P-Donsker we recall the stochastic process $\{W_h : h \in M(T)\}$ on $(\Omega, \mathcal{F}, P)$ given in (35). Then, the path $X(t, \cdot)$ is in C_h if and only if $W_h(\cdot) \geq 0$, and we also have $X \succeq h$ on T if and only if $W_h \geq 0$. Therefore, $\mathcal{C}_{\inf}$ P-Donsker will imply that the empirical processes $\{H_{n,1,h} : h \in E\}$ as given in (55) converge in distribution on $\ell_\infty(E)$ to a centered Gaussian measure $\gamma_{\inf}$ with separable support in $\ell_\infty(E)$. Furthermore, $\gamma_{\inf}$ is induced by the Gaussian process G_1 as indicated above.

Since (53) holds, for every $\delta > 0$ there are $N_\delta \equiv N(E, \delta, \|\cdot\|_\infty)$ functions $h_1, \ldots, h_{N_\delta}$ in E such that the brackets

$$F(\delta, h_j) = \{z \in M(T) : h_j(t) - \delta \leq z(t) \leq h_j(t) + \delta\}, \quad j = 1, \ldots, N_\delta, \tag{70}$$

have union covering E. Furthermore, $z \in F(\delta, h_j)$ implies

$$I(X \succeq h_j + \delta) \leq I(X \succeq z) \leq I(X \succeq h_j - \delta).$$

Hence, for $\delta > 0$ fixed we have from (39) that

$$\mathcal{E} \equiv \{I(X \succeq h) : h \in E\}$$

is a subset of

$$\cup_{j=1}^{N_\delta}\{I(X \succeq z) : z \in E, I(X \succeq h_j + \delta) \leq I(X \succeq z) \leq I(X \succeq h_j - \delta)\}.$$

Furthermore,

$$\begin{aligned}\|I(X \succeq h_j - \delta) - I(X \succeq h_j + \delta)\|_2^2 &= \|I(X \succeq h_j - \delta) - I(X \succeq h_j + \delta)\|_1 \\ &= P(X \succeq h_j - \delta) - P(X_j \succeq h_j + \delta) \leq 4\delta \int_{\mathbb{R}} |f_Z'(x)|dx,\end{aligned}$$

where $\|\cdot\|_p$ denotes the L_p norm with respect to P, and the inequality follows from (39) and (70). Now

$$\int_{0^+} (\log N(\mathcal{E}, x, \|\cdot\|_2))^{\frac{1}{2}} dx = \int_{0^+} (\log N(\mathcal{E}, x^2, \|\cdot\|_1))^{\frac{1}{2}} dx \le \int_{0^+} (\log N(\mathcal{E}, x^2, \|\cdot\|_\infty))^{\frac{1}{2}} dx$$

where the inequality follows since $\|\cdot\|_1 \le \|\cdot\|_\infty$ on $M(T)$. Letting $s = x^2$ in the right most integral above and applying (53) we have

$$\int_{0^+} (\log N(\mathcal{E}, x, \|\cdot\|_2))^{\frac{1}{2}} dx \le \int_{0^+} (\log N(\mathcal{E}, s, \|\cdot\|_\infty))^{\frac{1}{2}} s^{-\frac{1}{2}} ds < \infty. \tag{71}$$

Hence by Ossiander's CLT with bracketing Ossiander [15], or as in [4], p. 239, we have $\mathcal{E}$ a P-Donsker class of functions, which implies $\mathcal{C}_{\inf}$ is a P-Donsker class of sets. Hence the empirical processes $\{H_{n,1,h} : h \in E\}$ given in (55) converge weakly in $\ell_\infty(E)$ to the centered Gaussian process G_1 induced by the Radon Gaussian measure $\gamma_{\inf}$ and has covariance as indicated in (4.5). A similar result holds for the empirical processes $\{H_{n,2,h} : h \in E\}$ given in (56), which therefore satisfy the CLT in $\ell_\infty(E)$ with centered Gaussian limit G_2. Hence (59) is proven.

The next step of our proof is to show the finite dimensional distributions of the stochastic processes in (54) converge. To check this we set

$$F_n(h) = \frac{1}{n}\sum_{j=1}^n I(X_j \preceq h), \quad F(h) = P(X \preceq h),$$

and

$$G_n(h) = \frac{1}{n}\sum_{j=1}^n I(X_j \succeq h), \quad G(h) = P(X \succeq h).$$

Hence, let $I = I_1 \cup I_2 \cup I_3$, where I_1, I_2, I_3 are disjoint,

$$I_1 = \{h_1, \dots, h_{r_1}\}, \qquad I_2 = \{h_{r_1+1}, \dots, h_{r_2}\}, \qquad I_3 = \{h_{r_2+1}, \dots, h_r\},$$

and

$$I_1 = \{h \in I : F(h) < G(h)\}, \qquad I_2 = \{h \in I : F(h) > G(h)\},$$

and

$$I_3 = \{h \in I : F(h) = G(h)\}.$$

Setting

$$V_n(h) = \sqrt{n}(D_n(h) - D(h)),$$

we have

$$V_n(h) = \sqrt{n}[\min(F_n(h), G_n(h)) - \min(F(h), G(h))],$$

and since I is an arbitrary subset of E to prove the finite dimensional distributions of the processes in (54) we need to show

$$(V_n(h_1), \dots, V_n(h_r)),$$

converges in distribution on $\mathbb{R}^r$.

For $n \ge 1$ let

$$\begin{aligned} U_n(h) &= \sqrt{n}(F_n(h) - F(h)), \quad h \in I_1, \\ U_n(h) &= \sqrt{n}(G_n(h) - G(h)), \quad h \in I_2 \\ U_n(h) &= \sqrt{n}\min(F_n(h) - F(h), G_n(h) - G(h)), \quad h \in I_3, \end{aligned}$$

and take

$$N(\omega) = \min\{m \ge 1 : U_n(h_i) = V_n(h_i), 1 \le i \le r_2,\ n \ge m\}.$$

Then, the strong law of large numbers implies $P(N < \infty) = 1$, and $U_n(h) = V_n(h)$ for all $h \in I$ and all $n \ge N$. Therefore,

$$\lim_{n\to\infty} P(\sup_{h\in I} |U_n(h) - V_n(h)| \ge \epsilon) \le \lim_{n\to\infty} P(N > n) = 0,$$

and the convergence of the finite dimensional distributions will hold if we show

$$T_n = u_1 U_n(h_1) + \cdots + u_r U_n(h_r)$$

converges in distribution for all $(u_1, \ldots, u_r) \in \mathbb{R}^r$. Setting

$$S_n = \sum_{j=1}^{r_1} u_j (F_n(h_j) - F(h_j)) + \sum_{j=r_1+1}^{r_2} u_j (G_n(h_j) - G(h_j)),$$

we have

$$\begin{aligned} T_n &= \sqrt{n}\big[\min[S_n + u_{r_2+1}[F_n(h_{r_2+1}) - F(h_{r_2+1})], S_n + u_{r_2+1}[G_n(h_{r_2+1}) - G(h_{r_2+1})]]\big] \\ &\quad + \sqrt{n} \sum_{j=r_2+2}^{r} u_j \min[F_n(h_j) - F(h_j), G_n(h_j) - G(h_j)]. \end{aligned}$$

If

$$\Lambda(a_1, b_1, a_2, b_2, \ldots, a_k, b_k) = \sum_{i=1}^{k} \min[a_i, b_i],$$

then Λ is continuous from $\mathbb{R}^{2k}$ to $\mathbb{R}^k$. Therefore, if $k = r - r_2$ with

$$R_n = (a_{n,1}, b_{n,1}, \ldots, a_{n,r-r_2}, b_{n,r-r_2})$$

and

$$\begin{aligned} a_{n,1} &= \sqrt{n}\big(S_n + u_{r_2+1}[F_n(h_{r_2+1}) - F(h_{r_2+1})]\big), \\ b_{n,1} &= \sqrt{n}\big(S_n + u_{r_2+1}[G_n(h_{r_2+1}) - G(h_{r_2+1})]\big), \\ a_{n,i} &= \sqrt{n} u_{r_2+i}[F_n(h_{r_2+i}) - F(h_{r_2+i})], \quad i = 2, \ldots, r - r_2, \\ b_{n,i} &= \sqrt{n} u_{r_2+i}[G_n(h_{r_2+i}) - G(h_{r_2+i})], \quad i = 2, \ldots, r - r_2, \end{aligned}$$

we have R_n converging weakly to a centered Gaussian random variable, i.e. it is a sum of independent vectors in $\mathbb{R}^{2(r-r_2)}$ whose summands are indicator functions multiplied by u_j's. Now $\Lambda(R_n) = T_n$, and thus the continuous mapping theorem implies T_n converges in distribution. Since the vector $(u_1, \ldots, u_r) \in \mathbb{R}^r$ is arbitrary, the finite dimensional distributions converge. Of course, the claims in (60), (61), and (62) involving the one dimensional distributions are also now proven. □

Next we turn to a corollary of Theorem 2, which provides sub-Gaussian tail bounds for the convergence to zero in (59). To avoid measurability issues arising in its proof, we assume the set E is countable. Of course, under the assumption (1) and that $M(T)$ is a linear space, we have that the random vectors (stochastic processes)

$$\{(D_n(h)) - D(h) : h \in E\} \quad \text{and} \quad H_{n,i} := \{H_{n,i,h} : h \in E\}, \quad i = 1, 2,$$

given in (55) and (56), take values in the Banach space $\ell_\infty(E)$ with norm $\|x\|_\infty = \sup_{h \in E} |x_h|$ for $x = \{x_h\} \in \ell_\infty(E)$. Hence, the assumption E is countable implies these random vectors on $(\Omega, \mathcal{F}, P)$ are $\ell_\infty(E)$ valued in the sense used in [13], so for the convenience of the reader we freely quote from this single source a number of results used in the proof. However, from a historical point of view it should be observed that an important first step in these results involves the Hoffmann-Jørgenesen inequalities obtained in [6], and for series and a.s. normalized partial sums of sequences of independent random vectors, some results of a similar nature appeared in [7,8].

Notation 2. Let X take values in a Banach space B with norm $\|x\| = \sup_{f \in D} |f(x)|, x \in B$, where D is a countable subset of the unit ball of the dual space of B and $f(X)$ is measurable for each $f \in D$. Then, we are in the setting used in Chapter 6 of Ledoux and Talagrand [13], and the ψ_2-Orlicz norm of $\|X\|$ is given by

$$\|X\|_{\psi_2} = \inf\left\{c > 0 : E\left(\exp\left\{\left(\frac{\|X\|}{c}\right)^2\right\}\right) \leq 2\right\}.$$

Corollary 6. *Let $H_{n,i} := \{H_{n,i,h} : h \in E\}, i = 1, 2$, be the stochastic processes in (55) and (56), and for $i = 1, 2$*

$$\|H_{n,i}\|_\infty = \sup_{h \in E} |H_{n,i,h}|. \tag{72}$$

Then, under the assumptions of Theorem 2 and that E is countable, we have

$$\hat{k} = \sup_{n \geq 1, i=1,2} E(\|H_{n,i}\|_\infty) < \infty, \tag{73}$$

and there exists an absolute constant $k_2 < \infty$ *such that for any* $r > 0$

$$\sup_{n\geq 1} P(\sup_{h\in E} \sqrt{n}|D_n(h) - D(h)| \geq r) \leq 4\exp\{-\alpha r^2\} \tag{74}$$

provided $\alpha > 0$ *is sufficiently small that*

$$\sqrt{4\alpha}k_2(\hat{k} + 2) < 1. \tag{75}$$

Proof. From (63)

$$P(\sup_{h\in E} \sqrt{n}|D_n(h) - D(h)| \geq r) \leq P\left(\|H_{n,1}\|_\infty \geq \frac{r}{2}\right) + P\left(\|H_{n,2}\|_\infty \geq \frac{r}{2}\right).$$

Hence, Markov's inequality implies

$$P(\sup_{h\in E} \sqrt{n}|D_n(h) - D(h)| \geq r) \leq \exp\{-\alpha r^2\} \sum_{i=1}^{2} E(\exp\{4\alpha\|H_{n,i}\|_\infty^2\}),$$

and (74) holds provided $\alpha > 0$ is sufficiently small that

$$\sqrt{4\alpha} \sup_{n\geq 1, i=1,2} \|H_{n,i}\|_{\infty,\psi_2} < 1, \tag{76}$$

where we write $\|H_{n,i}\|_{\infty,\psi_2}$ to denote the ψ_2-norm of $\|H_{n,i}\|_\infty$. Now Theorem 6.21 of Ledoux and Talagrand [13] implies there exists an absolute constant $k_2 < \infty$ such that

$$\|H_{n,i}\|_{\infty,\psi_2} \leq k_2 \left[E(\|H_{n,i}\|_\infty) + \left(\sum_{j=1}^{n} \left\| \frac{Y_j}{\sqrt{n}} \right\|_{\infty,\psi_2}^2 \right)^{\frac{1}{2}} \right], \tag{77}$$

where $\{Y_j : j \geq 1\}$ are independent, mean zero, $\ell_\infty(E)$ valued random vectors with $Y_j = \{I(X_j \succeq h) - P(X_j \succeq h) : h \in E\}$ for $j \geq 1$ when $i = 1$, and $Y_j = \{I(X_j \preceq h) - P(X_j \preceq h) : h \in E\}$ for $j \geq 1$ when $i = 2$. Since $\|Y_j\|_\infty \leq 1$, we have $\|\frac{Y_j}{\sqrt{n}}\|_{\infty,\psi_2} \leq \frac{1}{\sqrt{n\log(2)}}$, and (77) implies

$$\|H_{n,i}\|_{\infty,\psi_2} \leq k_2 \left[E(\|H_{n,i}\|_\infty) + \left(\frac{1}{\log(2)} \right)^{\frac{1}{2}} \right]. \tag{78}$$

Now from (76) and (78) we have (74) for $\alpha > 0$ sufficiently small that

$$\sqrt{4\alpha}k_2 \left[\hat{k} + \left(\frac{1}{\log(2)} \right)^{\frac{1}{2}} \right] < 1, \tag{79}$$

provided (73) holds.

Hence, to complete the proof we must prove $\hat{k} < \infty$. To accomplish this we first show

$$\sup_{n\geq 1} E(\|H_{n,1}\|_\infty) < \infty. \tag{80}$$

This follows from Proposition 6.8 of Ledoux and Talagrand [13] applied to the partial sums S_k of the $\{Y_j : j \geq 1\}$ with $p = 1$ provided we show $\sup_{n\geq 1} t_{0,n} < \infty$, where

$$t_{0,n} = \inf\left\{ t > 0 : P\left(\max_{1\leq k\leq n} \left\| \frac{S_k}{\sqrt{n}} \right\|_\infty > t \right) \leq \frac{1}{8} \right\}, \quad n \geq 1. \tag{81}$$

Moreover, E countable implies Ottaviani's inequality is available as in Lemma 6.2 of Ledoux and Talagrand [13], and hence for every $u, v > 0$

$$P\left(\max_{1\leq k\leq n} \left\| \frac{S_k}{\sqrt{n}} \right\|_\infty > u + v \right) \leq \frac{P\left(\left\| \frac{S_n}{\sqrt{n}} \right\|_\infty > v \right)}{1 - \max_{1\leq k\leq n} P\left(\left\| \frac{S_n - S_k}{\sqrt{n}} \right\|_\infty > u \right)}. \tag{82}$$

Furthermore, since $\frac{S_n}{\sqrt{n}} = H_{n,1}$ for $n \geq 1$, and the proof of Theorem 2 implies $\{H_{n,1} : n \geq 1\}$ satisfies the central limit theorem in $\ell_\infty(E)$, the Portmanteau Theorem (applied to closed sets) implies there exists $u_0 < \infty$ such that for $u \geq u_0$

$$P\left(\left\|\frac{S_m}{\sqrt{m}}\right\|_\infty \geq u\right) \leq \frac{1}{2}$$

for all $m \in [m_0, \infty)$. Therefore, there exists $u_1 \in [u_0, \infty)$ such that

$$\sup_{m\geq 1} P\left(\left\|\frac{S_m}{\sqrt{m}}\right\|_\infty \geq u_1\right) \leq \frac{1}{2},$$

and hence

$$\sup_{n\geq 1}\max_{1\leq k\leq n} P\left(\left\|\frac{S_n - S_k}{\sqrt{n}}\right\|_\infty \geq u_1\right) \leq \sup_{n\geq 1}\max_{1\leq m\leq n} P\left(\left\|\frac{S_m}{\sqrt{m}}\right\|_\infty \geq u_1\right) \leq \frac{1}{2}.$$

Thus (82) implies for all $v > 0$ and $n \geq 1$ that

$$P\left(\max_{1\leq k\leq n}\left\|\frac{S_k}{\sqrt{n}}\right\|_\infty > u_1 + v\right) \leq 2P\left(\left\|\frac{S_n}{\sqrt{n}}\right\|_\infty > v\right). \tag{83}$$

Again, by the central limit theorem there exists $v_1 < \infty$ such that $v \geq v_1$ implies

$$2\sup_{n\geq 1} P\left(\left\|\frac{S_n}{\sqrt{n}}\right\|_\infty \geq v\right) \leq \frac{1}{8},$$

and hence we see from (83) that $\sup_{n\geq 1} t_{0,n} \leq u_1 + v_1 < \infty$ when $i = 1$ (and the partial sums come from the $\{Y_j : j \geq 1\}$). However, the same proof applies when $i = 2$ and the partial sums are formed from $\{Z_j : j \geq 1\}$, where $Z_j = \{I(X_j \preceq h) - P(X_j \preceq h) : h \in E\}$ for $j \geq 1$. Hence the proof is complete. □

5. Half-region depth over finite subsets

In order to make half-region depth more amenable to discrete computations we now define half-region depth over finite sets, and prove a uniform consistency result in this setting.

As before we assume $X := \{X(t) = X_t : t \in T\}$ is a stochastic process on the probability space $(\Omega, \mathcal{F}, P)$, all of whose sample paths are in $M(T)$.

If $h \in M(T)$ we define the half-region P-depth of h with respect to $J \subseteq T$ to be

$$D_J(h) = \min\{P(X \succeq_J h), P(X \preceq_J h)\}, \tag{84}$$

where $h_1 \succeq_J h_2$ ($h_1 \preceq_J h_2$) holds for functions h_1, h_2 defined on T if $h_1(t) \geq h_2(t)$ ($h_1(t) \leq h_2(t)$) for all $t \in J$.

Let $X_1, X_2, \ldots$ be i.i.d. copies of the process X, and assume $X, X_1, X_2, \ldots$ are defined on $(\Omega, \mathcal{F}, P)$ suitably enlarged, if necessary, and that all sample paths of each X_j are in $M(T)$. Then, the empirical half-region depth of $h \in M(T)$ over a set $J \subseteq T$ is given by

$$D_{n,J}(h) = \min\left\{\frac{1}{n}\sum_{j=1}^n I(X_j \succeq_J h), \frac{1}{n}\sum_{j=1}^n I(X_j \preceq_J h)\right\}. \tag{85}$$

For $h \in M(T)$ and J any finite subset of T, the probabilities in (84) are defined, and the events in (85) are in $\mathcal{F}$. Therefore, the classical law of large numbers implies with probability one

$$\lim_{n\to\infty} |D_{n,J}(h) - D_J(h)| = 0. \tag{86}$$

The next theorem refines (86) to be uniform over h and J, as long as $J \in \mathcal{J}_r$, where for each integer $r \geq 1$,

$$\mathcal{J}_r = \{J \subseteq T : \#J \leq r\},$$

and $\#J$ denotes the cardinality of the set J.

Theorem 3. *Let $X, X_1, X_2, \ldots$ be i.i.d. copies of the stochastic process $X = \{X(t) : t \in T\}$ defined on the probability space $(\Omega, \mathcal{F}, P)$, and all of whose sample paths are in the linear space $M(T)$. Let*

$$\mathcal{C} = \{C_{t,y} : t \in T, y \in \mathbb{R}\}, \tag{87}$$

where $C_{t,y} = \{z \in M(T) : z(t) \leq y\}$, and assume the empirical CLT holds with respect to the probability $\mathcal{L}(X)$ over $\mathcal{C}$. Then, for every integer $r \geq 1$ fixed we have with probability one that

$$\lim_{n\to\infty} [\sup_{h\in M(T)}\sup_{J\in\mathcal{J}_r} |D_{n,J}(h) - D_J(h)|]^* = 0. \tag{88}$$

Moreover,

$$\lim_{u\to\infty}\sup_{n\ge 1} P([\sup_{h\in M(T)}\sup_{J\in\mathcal{J}_r}\sqrt{n}|D_{n,J}(h)-D_J(h)|]^*\ge u)=0. \tag{89}$$

Remark 6. The implication in (88) does not follow from the corresponding finite dimensional result for half-region depth since the finite set J is not fixed, but it is also the case that the assumption of an empirical CLT over $\mathcal{C}$ is non-trivial. Fortunately Kuelbs et al. [9] and Kuelbs and Zinn [11] provide many examples of processes that satisfy this empirical CLT, and to which Theorem 3 applies. These include a broad collection of Gaussian processes, compound Poisson processes, stationary independent increment stable processes, and martingales. Moreover, if $J:\Theta\to\mathcal{J}_r$, then

$$[\sup_{h\in M(T)}\sup_{\theta\in\Theta}|D_{n,J(\theta)}(h)-D_{J(\theta)}(h)|]^*\le[\sup_{h\in M(T)}\sup_{J\in\mathcal{J}_r}|D_{n,J}(h)-D_J(h)|]^*,$$

and hence it is immediate that (88) holds when the choice of J is arbitrarily parameterized by Θ as long as $\#J(\theta)\le r,\theta\in\Theta$. In fact, the choice of subsets $J(\theta),\theta\in\Theta$, can be completely random. For example, let $\Theta=T^r$ be the r-fold product of T, assume Q is a probability on $(\Theta,\mathcal{F}_\Theta)$ where $\mathcal{F}_\Theta$ makes the r coordinate maps on T^r to T measurable, and for $\theta=(t_1,\dots,t_r)\in\Theta$ let $J(\theta)$ denote the subset of T defined by θ to be

$$J(\theta)=\{t_1,\dots,t_r\}.$$

Since some of the points in θ may be repeated, we have $\#J(\theta)\le r$ and $J(\theta)\in\mathcal{J}_r$ for all $\theta\in\Theta$. Furthermore, if $\hat{P}$ is any probability on $(\Omega\times\Theta,\mathcal{F}\times\mathcal{F}_\Theta)$ with marginals P and Q, then the measurable cover function

$$[\sup_{h\in M(T)}\sup_{\theta\in\Theta}|D_{n,J(\theta)}(h)-D_{J(\theta)}(h)|]^*$$

computed on $(\Omega\times\Theta,\mathcal{F}\times\mathcal{F}_\Theta,\hat{P})$ is less than or equal to

$$[\sup_{h\in M(T)}\sup_{J\in\mathcal{J}_r}|D_{n,J}(h)-D_J(h)|]^*$$

computed on $(\Omega,\mathcal{F},P)$. Hence (88) of Theorem 3 implies consistency in this setting as well.

Remark 7. Theorem 3 differs from [2,3] in two significant ways. In those papers the authors consider one dimensional projections and those projections are i.i.d., whereas here the depth is computed using multi-dimensional marginals without a need for independent observations since our results are uniform over all finite dimensional projections with dimension bounded by r.

Proof. Since (31) holds, we have

$$|D_{n,J}(h)-D_J(h)|\le A_n(J,h)+B_n(J,h),$$

where

$$A_n(J,h)=\left|\frac{1}{n}\sum_{j=1}^n I(X_j\succeq_J h)-P(X\succeq_J h)\right|$$

and

$$B_n(J,h)=\left|\frac{1}{n}\sum_{j=1}^n I(X_j\preceq_J h)-P(X\preceq_J h)\right|.$$

Therefore, sub-additivity of measurable cover functions implies (88) once we verify that with probability one

$$\lim_{n\to\infty}[\sup_{h\in M(T)}\sup_{J\in\mathcal{J}_r}A_n(J,h)]^*=\lim_{n\to\infty}[\sup_{h\in C[0,1]}\sup_{J\in\mathcal{J}_r}B_n(J,h)]^*=0.$$

Fix an integer $r\ge 1$, and set

$$\phi(u_1,\dots,u_r)=\min\{u_1,\dots,u_r\}.$$

Then, for $J\in\mathcal{J}_r$, $h\in M(T)$, and $j=1,\dots,r$, define

$$f_j=I_{C_{t_j,h(t_j)}},$$

which implies

$$D_{J,h}=\{z\in M(T):z(t)\le h(t),t\in J\}=\cap_{j=1}^r C_{t_j,h(t_j)},$$

and

$$I_{D_{J,h}} = \phi(f_1, \ldots, f_r).$$

Since $\mathcal{C}$ is a Donsker class with respect to P and $r \geq 1$ is fixed, Theorem 2.10.6 of van der Vaart and Wellner [16] implies

$$\mathcal{D} = \{D_{J,h} : J \in \mathcal{J}_r, h \in M(T)\}$$

is also P-Donsker with respect to P. Thus by Lemma 2.10.14 of van der Vaart and Wellner [16] we have almost surely that

$$\lim_{n\to\infty} [\sup_{h\in M(T)} \sup_{J\in\mathcal{J}_r} B_n(J, h)]^* = 0. \tag{90}$$

Since $\tilde{\mathcal{D}} = \{D^c_{J,h} : D_{J,h} \in \mathcal{J}_r\}$ is then also a Donsker class, the above argument implies that

$$\lim_{n\to\infty} [\sup_{h\in M(T)} \sup_{J\in\mathcal{J}_r} A_n(J, h)]^* = 0. \tag{91}$$

Combining (90) and (91) implies (88).

Since both $\mathcal{D}$ and $\tilde{\mathcal{D}}$ are P-Donsker, the proof of (89) follows as in the proof of (59) using appropriate modifications of (63), (64), and (65) to apply to J half-region depth. □

Acknowledgments

It is a pleasure to thank the referees. Their comments and suggestions led to a number of improvements in the exposition. Second author was partially supported by NSF grants DMS-1208962 and DMS-1205781.

References

[1] A. Chakraborty, P. Chaudhuri, On data depth in infinite dimensional spaces, Ann. Inst. Statist. Math. 66 (2) (2014) 303–324.
[2] J. Cuesta-Albertos, A. Nieto-Reyes, A random functional depth, in: Functional and Operatorial Statistics, in: Contrib. Statist, Physica-Verlag/Springer, Heidelberg, 2008, pp. 121–126.
[3] J.A. Cuesta-Albertos, A. Nieto-Reyes, The random Tukey depth, Comput. Statist. Data Anal. 52 (11) (2008) 4979–4988.
[4] R.M. Dudley, Uniform Central Limit Theorems, in: Cambridge Studies in Advanced Mathematics, vol. 63, Cambridge University Press, Cambridge, 1999.
[5] S. Dutta, A.K. Ghosh, P. Chaudhuri, Some intriguing properties of Tukey's half-space depth, Bernoulli 17 (4) (2011) 1420–1434. URL: http://dx.doi.org.lib-ezproxy.tamu.edu:2048/10.3150/10-BEJ322.
[6] J. Hoffmann-Jørgensen, Sums of independent Banach space valued random variables, Studia Math. 52 (1974) 159–186.
[7] N.C. Jain, M.B. Marcus, Integrability of infinite sums of independent vector-valued random variables, Trans. Amer. Math. Soc. 212 (1975) 1–36.
[8] J. Kuelbs, Some exponential moments of sums of independent random variables, Trans. Amer. Math. Soc. 240 (1978) 145–162.
[9] J. Kuelbs, T. Kurtz, J. Zinn, A CLT for empirical processes involving time-dependent data, Ann. Probab. 41 (2) (2013) 785–816. URL: http://dx.doi.org.lib-ezproxy.tamu.edu:2048/10.1214/11-AOP711.
[10] J. Kuelbs, J. Zinn, Concerns with functional depth, ALEA Lat. Am. J. Probab. Math. Stat. 10 (2) (2013) 831–855.
[11] J. Kuelbs, J. Zinn, Empirical quantile clts for time-dependent data, in: Christian Houdre, J.R. David Mason, J. Wellner (Eds.), High Dimensional Probabiity VI, The Banff Volume, in: Progress in Probability, vol. 66, Birkhäuser, Springer Basel, 2013, pp. 167–194.
[12] J. Kuelbs, J. Zinn, Addendum to "Half-region depth for stochastic processes", 2015. URL: http://www.math.tamu.edu/~joel.zinn/compl-pubs.html.
[13] M. Ledoux, M. Talagrand, Probability in Banach Spaces. Isoperimetry and Processes, in: Ergebnisse der Mathematik und ihrer Grenzgebiete (3) ([Results in Mathematics and Related Areas (3)]), vol. 23, Springer-Verlag, Berlin, 1991.
[14] S. López-Pintado, J. Romo, A half-region depth for functional data, Comput. Statist. Data Anal. 55 (4) (2011) 1679–1695.
[15] M. Ossiander, A central limit theorem under metric entropy with L_2 bracketing, Ann. Probab. 15 (3) (1987) 897–919.
[16] A.W. van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes. With Applications to Statistics, in: Springer Series in Statistics, Springer-Verlag, New York, 1996.