

---

## Supplementary Material to the Article

### The Area of the Convex Hull of Sampled Curves: a Robust Functional Statistical Depth Measure

---

First, Section **A** collects technical proofs omitted in the body of the article. Then, Section **B** provides the exact algorithm for approximate computation of the proposed depth notion. Finally, Section **C** collects further experimental results mentioned in the article.

#### A Technical proofs

This part presents the proofs of Proposition 3.1, Theorems 3.1 and 3.2 as well as the counter examples for the non-satisfied properties. Most of the proofs are done for both  $D_J$  and  $\overline{D}_J$ .

##### A.1 Proof of Proposition 3.1

###### A.1.1 Affine-invariance

Let  $a, b \in \mathbb{R}$ , it is clear that

$$\begin{aligned} & \text{conv}(\text{graph}(\{aX_1 + b, \dots, aX_n + b\})) \\ &= a \times \text{conv}(\text{graph}(\{X_1, \dots, X_n\})) + b \end{aligned}$$

where  $a \times \text{conv}(\text{graph}(\{X_1, \dots, X_n\})) + b = \{(t, ax + b) : (t, x) \in \text{conv}(\text{graph}(\{X_1, \dots, X_n\}))\}$ . Following this, and by properties of Lebesgue measure, we have

$$\begin{aligned} \frac{\lambda_2(a \times \text{conv}(\text{graph}(\{X_1, \dots, X_n\})) + b)}{\lambda_2(a \times \text{conv}(\text{graph}(\{X_1, \dots, X_n, z\})) + b)} &= \frac{\lambda_2(a \times \text{conv}(\text{graph}(\{X_1, \dots, X_n\})))}{\lambda_2(a \times \text{conv}(\text{graph}(\{X_1, \dots, X_n, z\})))} \\ &= \frac{a \times \lambda_2(\text{conv}(\text{graph}(\{X_1, \dots, X_n\})))}{a \times \lambda_2(\text{conv}(\text{graph}(\{X_1, \dots, X_n, z\})))} \\ &= \frac{\lambda_2(\text{conv}(\text{graph}(\{X_1, \dots, X_n\})))}{\lambda_2(\text{conv}(\text{graph}(\{X_1, \dots, X_n, z\})))} \end{aligned}$$

###### The case of $a, b \in \mathcal{X}$ :

Now, we just take a counter example to prove that it is not true if  $b$  belongs to  $\mathcal{X}$ , the case where  $a \in \mathcal{X}$  is trivial. For the sake of simplicity, let  $I = [0, 1]$  and  $J = 2$ . If we take  $x \equiv 0, x_1 \equiv 1, x_2 \equiv 2$  and  $X$  a random variable with distribution  $P$  such that  $\mathbb{P}(X \equiv x_1) = \frac{1}{2}$  and  $\mathbb{P}(X \equiv x_2) = \frac{1}{2}$ . Let  $X_1, X_2$  be samples from  $P$  and  $b$  be continuous function  $t \mapsto (10t - 4)\mathbb{1}_{(0.4, 0.5)} + (-10t + 6)\mathbb{1}_{(0.5, 0.6)}$ . It is easy to see that  $D_J(x|P) = \frac{1}{8} \neq D_J(x + b|P_{X+b})$  since

$$\begin{aligned} D_J(x + b|P_{X+b}) &= \frac{1}{2} \times \left( \frac{1}{2} \times \frac{0.5}{1.5} + \frac{1}{2} \times \frac{0.5}{2.5} \right)^{j=1} + \frac{1}{2} \times \left( \frac{1}{4} \times \frac{0.5}{1.5} + \frac{1}{4} \times \frac{0.5}{2.5} + \frac{1}{2} \times \frac{1.5}{2.5} \right)^{j=2} \\ &= \frac{8}{60} + \frac{9}{60} \\ &= \frac{17}{60}. \end{aligned}$$

Note that even if we set  $j > 1$  to avoid the fact that the convex hull of constant function have null Lebesgue measure,  $D_J(x|P)$  and  $D_J(x + b|P_{X+b})$  remain different, see Fig. **1**.

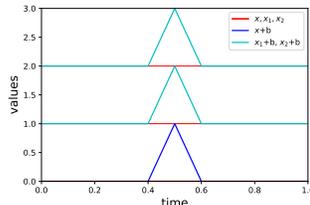


Figure 1: Plots of the functions used in the case of  $\mathbf{a}, \mathbf{b} \in \mathcal{X}$ . The three red lines come from  $x_1, x_2, x$ . The cyan curves correspond to  $x_1 + b$  and  $x_2 + b$  and blue curve to  $x + b$ .

### A.1.2 Vanishing at infinity

Let  $J$  be fixed and  $x_n$  a sequence of function such that  $\|x_n\|$  tends to infinity when  $n$  grows, for every  $j \in \{1, \dots, J\}$  we define :

$$\begin{aligned} \Phi_{x_n} : \mathcal{X}^{\otimes j} &\longrightarrow [0, 1] \\ (x_1, \dots, x_j) &\longmapsto \frac{\lambda_2(\text{conv}(\text{graph}(\{x_1, \dots, x_j\})))}{\lambda_2(\text{conv}(\text{graph}(\{x_1, \dots, x_j, x_n\})))} \end{aligned}$$

As a continuous function on compact set,  $x_1, \dots, x_j$  are bounded, then  $\Phi_{x_n} \xrightarrow{\|x_n\|_\infty \rightarrow \infty} 0$ . The result follows from dominated convergence theorem since  $\Phi_{x_n}$  is bounded by 1.

### A.1.3 Continuity in $x$

Let  $x_1, \dots, x_j, j \in \{1, \dots, J\}$  be fixed curves in  $\mathcal{C}(I)$  with at least two different curves, i.e., there exists a  $t \in I$  and  $l, k \in 1, \dots, j$  such that  $x_k(t) \neq x_l(t)$ . If  $j = 1$ , we need that  $x_1$  is not a constant function. From Lemma A.6.1, we know that the function

$$f : x \longmapsto \text{Band}(x_1, \dots, x_j, x)$$

is continuous.

Let  $\mathcal{K}_2$  be the set of all compact set in  $\mathbb{R}^2$  and  $\mathcal{K}_2^C$  the set of all convex bodies (compact, convex set with non-empty interior). We equip both spaces with the Hausdorff distance. We know that :

$$\begin{aligned} \text{conv} : \mathcal{K}_2 &\longrightarrow \mathcal{K}_2 \\ K_2 &\longmapsto \text{conv}(K_2) \end{aligned}$$

and

$$\begin{aligned} \psi : \mathcal{K}_2^C &\longrightarrow \mathcal{K}_2^C \\ K_2^c &\longmapsto \lambda_2(K_2^c) \end{aligned}$$

are continuous with respect to the Hausdorff distance. See for example, Theorems 12.3.5 and 12.3.6 in [Schneider and Weil \(2008\)](#) for  $g$  and Theorem 1.8.16 in [Schneider \(2013\)](#) for  $\psi$ . Then  $\Phi := \psi \circ \text{conv} \circ f : x \mapsto \lambda_2(\text{conv}(\text{graph}\{x_1, \dots, x_j, x\}))$  is continuous.

It is straightforward to show that

$$\phi : x \longmapsto \frac{\lambda_2(\text{conv}(\text{graph}(\{x_1, \dots, x_j\})))}{\lambda_2(\text{conv}(\text{graph}(\{x_1, \dots, x_j, x\}))})}$$

is continuous. Now, we just have to prove that

$$x \longmapsto \mathbb{E} \left[ \frac{\lambda_2(\text{conv}(\text{graph}(\{x_1, \dots, x_j\})))}{\lambda_2(\text{conv}(\text{graph}(\{x_1, \dots, x_j, x\}))})} \right]$$

is continuous which is true by dominated convergence theorem. We conclude the proof with the continuity of the sum of continuous functions.

## A.2 Continuity in $P$

$(\mathcal{C}([0, 1]), \|\cdot\|_\infty)$  is a polish space and implies that the set of all probability measures on this space with the Lévy-Prohorov metric  $\rho_{LP}$  is still polish. By the portmanteau theorem (*e.g.*, see Theorem 11.3.3 in [Dudley \(2002\)](#)) it follows that  $\rho_{LP}(Q_n, Q) \rightarrow 0$  is equivalent to  $Q_n \xrightarrow{d} Q$  for  $Q, Q_n$  respectively a measure and a sequence of measures on  $\mathcal{C}([0, 1])$ . It implies that

$$\int f dQ_n \rightarrow \int f dQ$$

for every  $f$  bounded continuous real function on  $(\mathcal{C}([0, 1]), \|\cdot\|_\infty)$ .

Let  $j$  be fixed natural number and define the following function

$$\begin{aligned} \Phi_x : \quad \mathcal{X}^{\otimes j} &\rightarrow [0, 1] \\ (x_1, \dots, x_j) &\mapsto \frac{\lambda_2(\text{conv}(\text{graph}(\{x_1, \dots, x_j\})))}{\lambda_2(\text{conv}(\text{graph}(\{x_1, \dots, x_j, x\})))} \end{aligned}$$

If we equip  $\mathcal{C}([0, 1])^{\otimes j}$  with the infinite norm  $\|\cdot\|_{\infty, j}$  defined by

$$\|\|f\|\|_{\infty, j} = \max(\|f_1\|_\infty, \dots, \|f_j\|_\infty),$$

following the same argument from the proof [A.1.3](#),  $\Phi_x$  is bounded and continuous.

Now, let  $J \leq n$  be fixed and  $P_n$  be a sequence of measures on  $\mathcal{C}([0, 1])$  such that  $\rho_{LP}(Q_n, Q) \rightarrow 0$ . we have :

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{j=1}^J D_j(x|Q_n) &= \sum_{j=1}^J \lim_{n \rightarrow \infty} \int_{\mathcal{C}([0, 1])^{\otimes j}} \Phi_x dQ_n^{\otimes j} \\ &= \sum_{j=1}^J \int_{\mathcal{C}([0, 1])^{\otimes j}} \Phi_x dQ^{\otimes j} \\ &= \sum_{j=1}^J D_j(x|Q) \end{aligned}$$

Then the results holds for  $\sum_{j=1}^J D_j$  (and trivially for  $D_J$ ).

## A.3 Proof of the Theorem 3.1

For every  $1 \leq j \leq J$ , the term  $|D_{j,n}(x|P_n) - D_j(x|P)|$  goes to zero almost-surely by U-statistics consistency (see *e.g.* [Hoeffding \(1961\)](#)). Then

$$\mathbb{P}\left(\forall j : \left|D_{j,n}(x|P_n) - D_j(x|P)\right| \rightarrow 0\right) = 1$$

which is equivalent to

$$\mathbb{P}\left(\sum_{j=1}^J \left|D_{j,n}(x|P_n) - D_j(x|P)\right| \rightarrow 0\right) = 1.$$

By triangle inequality, for any  $x \in \mathcal{C}([0, 1])$ ,

$$\begin{aligned} \left|\sum_{j=1}^J D_{j,n}(x|P_n) - D_j(x|P)\right| &\leq \sum_{j=1}^J \left|D_{j,n}(x|P_n) - D_j(x|P)\right| \\ &\leq \sum_{j=1}^J \left|D_{j,n}(x|P_n) - D_j(x|P)\right| \end{aligned}$$

which leads to the result.

#### A.4 Proof of the Theorem 3.2

The result follows from the continuity in  $P$  and Theorem 3 in [Nagy et al. \(2016\)](#).

#### A.5 Counter-examples for the not-satisfied properties

##### A.5.1 Maximality at the center

We restrict ourselves for simplicity to  $J = 2$  and  $I = [0, 1]$ . Let  $X \sim P$  be a distribution such that  $P(X \equiv y_1) = P(X \equiv y_2) = \frac{1}{2}$  with

$$\begin{aligned} y_1 &= (-2t + 1)\mathbb{1}_{[0,0.25]} + (2t)\mathbb{1}_{[0.25,0.5]} + (-2t + 2)\mathbb{1}_{[0.5,0.75]} + (2t - 1)\mathbb{1}_{[0.75,1]}, \\ y_2 &= -y_1. \end{aligned}$$

The distribution is clearly centrally and halfspace symmetric around  $\theta \equiv 0$  but we have

$$D_J(\theta, P) < D_J(y_1, P) = D_J(y_2, P).$$

Since

$$\begin{aligned} D_J(0, P) &= \frac{1}{2} \times \left( \frac{1}{2} \times \frac{3}{8} + \frac{1}{2} \times \frac{3}{8} \right)^{j=1} + \frac{1}{2} \times \left( \frac{1}{2} + \frac{1}{4} \times \frac{3}{8} + \frac{1}{4} \times \frac{3}{8} \right)^{j=2} \\ &= \frac{17}{32} \approx 0.53 \end{aligned}$$

and

$$\begin{aligned} D_J(y_1, P) &= \frac{1}{2} \times \left( \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{3}{16} \right)^{j=1} + \frac{1}{2} \times \left( \frac{1}{2} + \frac{1}{4} + \frac{1}{4} \times \frac{3}{16} \right)^{j=2} \\ &= \frac{70}{128} \approx 0.546 \end{aligned}$$

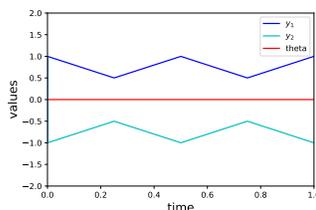


Figure 2: Plot of the functions used in the counter example of the maximality at the center property.  $y_1$  (blue curve) and  $y_2$  (cyan curve) correspond to the distribution and  $\theta \equiv 0$  corresponds to the red curve.

##### A.5.2 Decreasing w.r.t. the deepest point

We restrict ourselves  $I$  to  $[0, 1]$  and  $J = 2$  for the sake of comprehension but the example still works on every  $I$ . Let  $P$  be the distribution such that

$$P(X \equiv 0) = P(X \equiv 1) = P(X \equiv -1) = \frac{1}{3}.$$

It is clear from this distribution that  $0 \in \sup_{x \in \mathcal{X}} D(x, P)$  and, if we write  $z \equiv 0$ ,  $D_J(z, P) = \frac{1}{4}$ . We define  $y \equiv 1.5$  and  $x(t) = 4t\mathbb{1}_{[0,0.5]}(t) + (-4t + 4)\mathbb{1}_{[0.5,1]}(t)$ . We have  $d(x, z) = 2$ ,  $d(x, y) = 0.5$  and  $d(y, z) = 1.5$ . If we compute the depth of  $x$  and  $y$  we have :

$$D_J(y, P) = \frac{1}{2} \times \frac{2}{9} \times \left( \frac{4}{5} + \frac{2}{5} + \frac{2}{3} \right) = \frac{23}{135}$$

and

$$D_J(x, P) = \frac{1}{2} \times \frac{2}{9} \times \left( \frac{4}{5} + \frac{1}{2} + \frac{8}{9} \right) = \frac{197}{810}.$$

The result follows. Notice that the result remains true if *conv* is replaced by the *band* function.

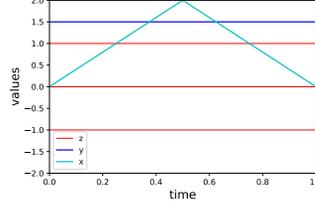


Figure 3: Plots of the functions used in the counter example of the decreasing property. The three red lines come from the distribution, the thicker red curve corresponds to the maximal depth. The cyan curve corresponds to  $x$  and the blue curve to  $y$ .

## A.6 Technical requirements

**Lemma A.6.1** *Let  $x_1, \dots, x_j$ ,  $j \in \{1, \dots, J\}$  be fixed curves in  $\mathcal{C}([0, T])$ . The function*

$$\begin{aligned} \mathcal{C}([0, T]) &\longrightarrow \mathcal{K}_2 \\ x &\mapsto \text{Band}(x_1, \dots, x_j, x) \end{aligned}$$

*is continuous if we equip  $\mathcal{K}_2$  with the Hausdorff distance  $d_H$ .*

**Proof.** Let  $x_0 \in \mathcal{C}([0, T])$  and  $j$  be fixed in  $\{1, \dots, J\}$ . We want to prove

$$\forall \epsilon > 0, \exists \delta : \forall x \in \mathcal{C}([0, T]) : \|x - x_0\|_\infty < \delta \Rightarrow d_H(\text{Band}(x_1, \dots, x_j, x), \text{Band}(x_1, \dots, x_j, x_0)) \leq \epsilon$$

Let  $\epsilon > 0$ , and write  $B_x := \text{Band}(x_1, \dots, x_j, x)$  and  $B_{x_0} := \text{Band}(x_1, \dots, x_j, x_0)$  for the simplicity of notation. We have :

$$d_H(B_x, B(x_0)) = \max \left( \sup_{z \in B_x} d(z, B_{x_0}), \sup_{z \in B_{x_0}} d(z, B_x) \right)$$

with  $d$  being the distance induced by  $\|\cdot\|_\infty$ .

It is easy to see that for any  $z \in B_x$ ,  $\inf_{y \in B_{x_0}} \|z - y\|_\infty$  is minimized by the function  $y^*(t) = z(t)\mathbb{1}_{(z(t) \in B_{x_0})} + \max(x_1(t), \dots, x_j(t), x(t))\mathbb{1}_{(z(t) \notin B_{x_0})}$ .

Following this, we have :

$$\begin{aligned} d(z, B_{x_0}) &= \|z - y^*\|_\infty \\ &= \max \left( \sup_{t: z(t) \notin B_{x_0}, z(t) > B_{x_0}} |z(t) - \max(x_1(t), \dots, x_j(t), x(t))|, \right. \\ &\quad \left. \sup_{t: z(t) \notin B_{x_0}, z(t) < B_{x_0}} |z(t) - \min(x_1(t), \dots, x_j(t), x(t))| \right) \end{aligned}$$

$z \in B_x$  implies that  $\forall t$ ,  $\min \left( x(t), \min_{i=1, \dots, j} X_i(t) \right) \leq z(t) \leq \max \left( x(t), \max_{i=1, \dots, j} X_i(t) \right)$ . If  $z(t) > B_{x_0}$  too,  $\max(\max_{i=1, \dots, j} X_i(t), x_0(t)) < z(t) \leq x(t)$  and

$$\sup_{t: z(t) \notin B_{x_0}, z(t) > B_{x_0}} |z(t) - \max(x_1(t), \dots, x_j(t), x(t))| = \sup_{t: z(t) \notin B_{x_0}, z(t) > B_{x_0}} |z(t) - x(t)|.$$

With the same argument we have :

$$\sup_{t: z(t) \notin B_{x_0}, z(t) < B_{x_0}} |z(t) - \max(x_1(t), \dots, x_j(t), x(t))| = \sup_{t: z(t) \notin B_{x_0}, z(t) < B_{x_0}} |z(t) - x(t)|.$$

It follows that for every  $z \in B_x$ ,  $d(z, B_{x_0}) \leq \|x - x_0\|_\infty \leq \epsilon$  (simply by taking  $\delta = \epsilon$ ). We then have

$$\sup_{z \in B_x} d(z, B_{x_0}) \leq \epsilon.$$

We can prove that

$$\sup_{z \in B_{x_0}} d(z, B_x) \leq \epsilon$$

with the same argument which lead us to the final result.

## B Monte-Carlo approximation of the average version of the empirical ACH depth.

The procedure given in Section 3.2 can be summarize by the Algorithm 1.

**Algorithm 1 Input:**  $\tilde{S}_n = \{X'_1, \dots, X'_n\}$ -dataset, the observed curve  $x'$ ,  $K$ ,  $1 \leq J \leq n$  and the vector of weights  $\mathbf{w} = (w_j)_{1 \leq j \leq J}$  such that  $\mathbf{w}_j = \frac{\binom{j}{n}}{\sum_{m=1}^J \binom{m}{n}}$ .

1. For  $k = 1, \dots, K$  do:

(i) Select  $l \in \{1, \dots, J\}$  according to  $\mathbf{w}$ .

(ii) Select randomly and uniformly  $(i_1, \dots, i_l) \in \{1, \dots, n\}$ .

(iii)  $s(x') \leftarrow s(x') + \frac{\lambda_2(\text{conv}(\text{graph}(\{X'_{i_1}, \dots, X'_{i_l}\})))}{\lambda_2(\text{conv}(\text{graph}(\{X'_{i_1}, \dots, X'_{i_l}, x'\}))})}$ .

2. **Return:**  $\bar{D}_{J,n}(x' | \mathcal{S}'_n) = \frac{1}{K} s(x')$ .

## C Additional Experiences

### C.1 Asymptotic variance of the exact and approximate versions

To obtain further insights about the stability of the proposed depth notion, we explore its asymptotic variance. For this, we compute (exact and approximate) ACHD of  $x_i$ ,  $i = 1, 2, 3, 4$  for different sample sizes. The boxplots over 100 repeated simulation for data sets (a) and (b) are indicated in Fig. 4 and 5. One observes not only stable decrease of the variance of ACHD with the sample size, but also the similarity between exact and approximate versions, which hints on stability and precision of the exact algorithm even when exploring a small portion of combinations (e.g., when  $n = 500$  only 2% of all pairs are explored for  $K = 5n$ ).

The dataset Octane, Wine, and EOG used in the Section 4.4 are represented in Figure 7.

### C.2 Robustness

Additional results on the robustness experiment are given in Table 1 for the location anomalies with the dataset (b) and the isolated anomalies for the data set (a).

### C.3 Anomaly detection

Additional experiments with location anomalies from the data set (b) and isolated anomalies from data set (a) are given in Figure 6.

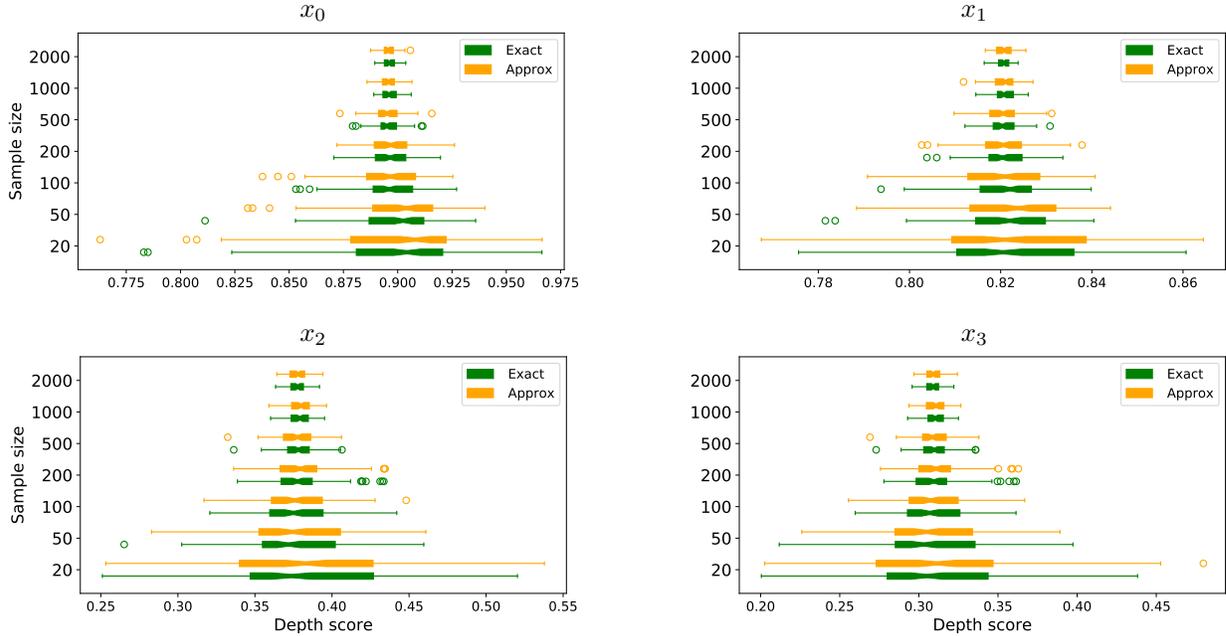


Figure 4: Boxplot (over 100 repetitions) of the depth score for the observations  $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$  for the two following settings on the data set (a): the green boxplots represent the exact computation while the orange boxplots represent the approximation both with  $J = 2$ .

		$d_\tau(\sigma_0, \sigma_\alpha)(\times 10^{-2})$						
		$\alpha$	0	5	10	15	25	30
ACHD	Location		0	0.7	1.5	2.3	4.2	5.1
	Isolated		0	1.3	1.8	1.6	2.4	3.2
FSDO	Location		0	1.5	3.1	5.1	8.8	11
	Isolated		0	0.9	1	1.1	1.5	1.6
FT	Location		0	0.6	1.5	3	6.4	8
	Isolated		0	1.3	0.8	0.9	1.1	1.5
FIF	Location		0	14	15	15	16	15
	Isolated		0	6.9	7.3	7	8.2	8.1

Table 1: Kendall’s tau distances between the rank returned with normal data and contaminated data (over different size of contamination with location and isolated anomalies) for the area of the convex hull depth measure and three others state-of-art methods.

## References

Dudley, R. M. (2002). *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition.

Hoeffding, W. (1961). The strong law of large numbers for u-statistics. Technical report, North Carolina State University. Dept. of Statistics.

Nagy, S., Gijbels, I., and Hlubinka, D. (2016). Weak convergence of discretely observed functional data with applications. *Journal of Multivariate Analysis*, 146:46 – 62.

Schneider, R. (2013). *Convex Bodies: The Brunn-Minkowski Theory*. Cambridge University Press, Cambridge.

Schneider, R. and Weil, W. (2008). *Stochastic and Integral Geometry*. Springer-Verlag, Berlin Heidelberg.

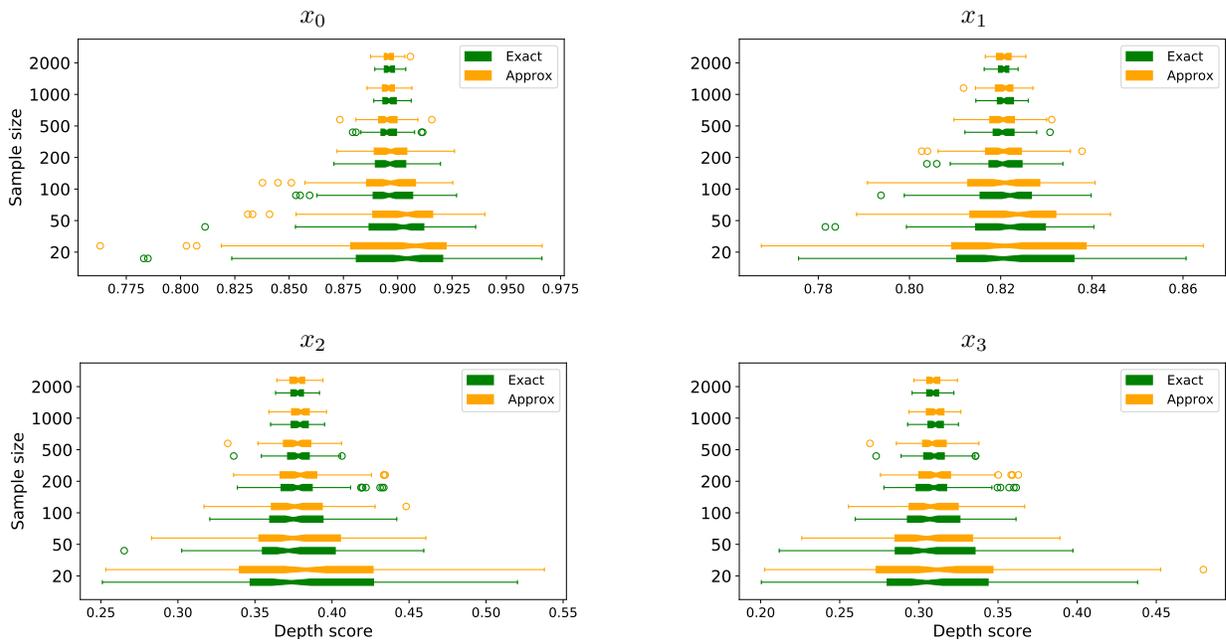


Figure 5: Boxplot (over 100 repetitions) of the depth score for the observations  $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$  for the two following settings on the data set (b): the green boxplots represent the exact computation while the orange boxplots represent the approximation both with  $J = 2$ .

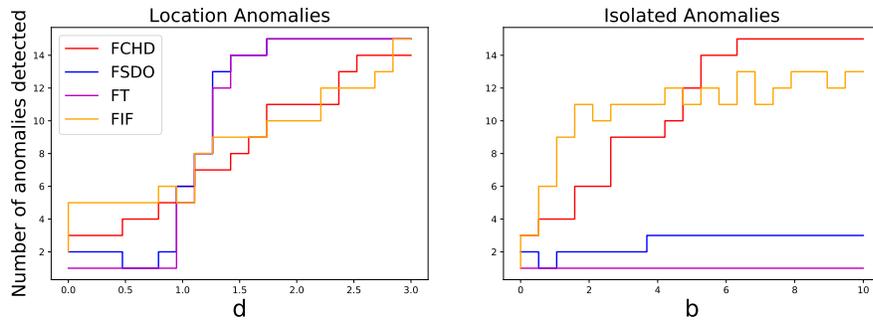


Figure 6: Number of anomalies detected over a grid of parameters for two types of anomalies, location and isolated anomalies for ACHD and three others state-of-art methods.

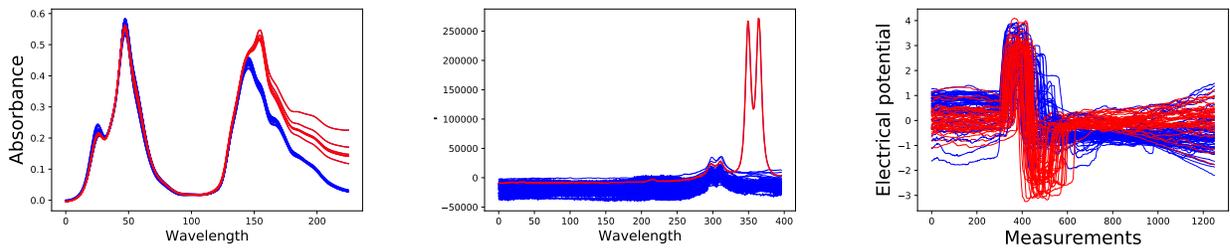


Figure 7: The data sets Octane, Wine, and EOG used in Section 4.5.