Electronic Journal of Statistics
Vol. 16 (2022) 6584–6630
ISSN: 1935-7524
https://doi.org/10.1214/22-EJS2095

Depth level set estimation and associated risk measures*

Sara Armaut and Roland Diel and Thomas Laloë

Université Côte d'Azur,
Laboratoire Jean Alexandre Dieudonné,
Parc Valrose, 06108 NICE Cedex 02, France
e-mail: armaut@unice.fr; diel@unice.fr; laloe@unice.fr

Abstract: Depth functions have become increasingly powerful tools in non-parametric inference for multivariate data as they measure a degree of centrality of a point with respect to a distribution. A multivariate risk scenario is then represented by a depth-based lower level set of the risk factors, meaning that we consider a non-compact setting. The aim of this paper is to study the asymptotic behavior of level sets of a general multivariate depth function and a particular multivariate risk measure, the Covariate-Conditional-Tail-Expectation (CCTE) based on a depth function. More precisely, given a probability measure P on $\mathbb{R}^d$ and a depth function $D(\cdot, P)$, we are interested in the α-lower level set $\mathcal{L}_D(\alpha) := \left\{z \in \mathbb{R}^d : D(z, P) \leq \alpha\right\}$. First, we present a plug-in approach in order to estimate $\mathcal{L}_D(\alpha)$, then we derive consistency of its estimator under some regularity conditions. In a second part, we provide a consistent estimator of the CCTE for a general depth function with a rate of convergence and we consider the particular case of Mahalanobis depth. Finally, a simulation study complements the performances of our estimator and an application on real data is presented.

MSC2020 subject classifications: 62H12, 62G05, 60G32.
Keywords and phrases: Multivariate depth function, level set estimation, plug-in, risk measure, Mahalanobis depth.

Received March 2022.

Contents

*This work has been supported by the project ANR McLaren (ANR-20-CE23-0011).

1. Introduction

The estimation of level sets $\mathcal{L}_G(c) := \left\{x \in \mathbb{R}^d : G(x) \geq c\right\}$, where G is an unknown function on $\mathbb{R}^d$, has received attention recently (Cuevas et al. [16], Goldstein and Messer [36], Molchanov [54]). The motivation of studying such level sets lays on the various possible applications. For example, for the density function we can cite the work of Müller and Sawitzki [56], Polonik [61] in mode estimation, Hartigan [37], Cuevas et al. [14, 15] for clustering, and Devroye and Wise [22], Baillo et al. [2], Desforges et al. [21], Scott and Nowak [64], Park et al. [59] for pattern recognition or detection of abnormal behavior in a system. For regression level set estimation, it is known that nonparametric regression (see e.g. [46]) provides a natural model for image analysis; for example, the estimation of regression level sets has been studied by Cavalier [7], Laloë and Servien [48]. For cumulative distribution function (c.d.f) level set estimation, in Coblenz et al. [12] a nonparametric estimation procedure for particular level sets of copulas is presented. In Coblenz et al. [11] confidence regions for these level sets are provided. Additionally, Di Bernardino and Rulliere [25], Di Bernardino et al. [23, 24] propose to define risk measures in multivariate hydrological or financial models using c.d.f level sets as risk regions. One classical estimator of such level sets is the *plug-in* one: if $\hat{G}_n(x)$ is an available estimator of $G(x)$, then one can estimate $\mathcal{L}_G(c)$ by $\mathcal{L}_{\hat{G}_n}(c) := \left\{x \in \mathbb{R}^d : \hat{G}_n(x) \geq c\right\}$.

In this paper, we first consider the problem of depth level sets estimation. Indeed, depth based level sets have gained attention in the last three decades as a key tool for the visualization and exploration of multivariate data (see Donoho and Gasko [26]). Depth level sets are also a natural extension of the notion of quantiles to multivariate data, and can be used for outlier detection (see e.g. Febrero et al. [33], Dai and Genton [17]), as well as for supervised classification (see Ruts and Rousseeuw [63], Hubert et al. [42], Jörnsten [44]) or rank and sign testing (see Liu and Singh [50], Hettmansperger and Oja [40]). Further, Müller [55] derived simple distribution-free tests for regression based on the so-called likelihood depth which is the density function, while Denecke and Müller [19] studied convergence of the latter depth for the correlation coefficient. Roughly speaking, a depth is a function D which provides, for any probability measure, a *center*-outward ordering of points in $\mathbb{R}^d$ (see for instance [29, 70, 49]). He and Einmahl [39] have recently studied extreme quantile regions induced by the halfspace depth and showed their interest in a real world financial application using the daily international market price indices of three countries (USA, UK and Japan). In this paper, we both address the problem of the estimation of depth based level set and the definition and estimation of one depth based risk measure.

Broadly speaking, a risk measure is a mapping from a set of d-dimensional random variables ($d \geq 1$) to $\mathbb{R}$, and is used to determine the amount of an asset (or assets/goods) to be kept in reserve in order to cover for unexpected losses. The Conditional-Tail-Expectation (CTE) [20], also known as "expected shortfall", has been widely used in the literature on risk measures. It characterizes the conditional expected loss given that the loss exceeds a critical loss threshold. Formally, given a real random variable X with distribution function F_{X}, the CTE at level $\alpha \in (0, 1)$ is defined as:

$$\mathrm{CTE}_{\alpha}(\mathrm{X}) := \mathbb{E}[\mathrm{X} \,|\, \mathrm{X} > \mathrm{VaR}(\alpha)], \tag{1.1}$$

where

$$\mathrm{VaR}(\alpha) := \inf\{t \in \mathbb{R} : F_{\mathrm{X}}(t) \geq 1 - \alpha\}$$

is the well-known *Value at Risk* which corresponds to the univariate quantile of order $1 - \alpha$ of X (α is small). Thus, the CTE is nothing but the mathematical description of an average loss in the worst $100 \cdot \alpha\%$ risk scenario. Further, the CTE has a close relation to the zonoid depth, introduced by Koshevoy and Mosler [47]: in the univariate case the α-trimmed region of the zonoid depth is given by

$$\Big[\mathbb{E}[\mathrm{X} \,|\, \mathrm{X} \leq Q_{\mathrm{X}}(\alpha)], \mathbb{E}[\mathrm{X} \,|\, \mathrm{X} \geq Q_{\mathrm{X}}(1 - \alpha)]\Big],$$

where $Q_{\mathrm{X}}(\alpha)$ denotes the quantile function of X. This link is established in Cascos and Molchanov [6].

However, considering a single risk factor is restrictive, as we can easily imagine correlated risk factors that could be studied together. One possibility is to consider quantile regions of the risk factors distribution. In the multivariate case, a wide panel of multivariare quantiles has been reviewed in the literature [8, 9, 39]. The study of multivariate quantile regions has increasingly been pursued in the last decades as a tool to model multivariate risk regions, especially those based on a multivariate distribution function ([3, 18, 13]), or on a depth function.

An interesting classical problem in the theory of risk is to study the behavior of an expected cost $\mathrm{Y} \in \mathbb{R}$ associated to $d \geq 1$ risk factors which are heterogeneous in nature. In econometrics, for instance, one can be interested in an average return (which measures the performance of a portfolio for a certain period of time) with respect to $d \geq 1$ risk factors $\mathbf{X} \in \mathbb{R}^d$. On another note, one can also be interested in the impact of climate change (via d risk factors) on high temperatures. Regarding the risk literature, conditional quantiles have been widely studied, especially to provide predictive environmental models. We can mention the work of Wang and Li [67] who studied extreme conditional quantiles of a response variable Y given covariates $\mathbf{X} = \boldsymbol{x}$ and demonstrates their usefulness through a power transformation model. More recently, Girard et al. [35] studied conditional extremes Y given a covariate $\mathbf{X}$ by means of conditional expectiles in heteroscedastic regression models with heavy-tailed noise and showcased their fruitfulness on an actuarial and financial real dataset. Early attempts at tackling conditional quantiles can be found in Gardes and Girard [34], Wang et al. [68] and references therein.

To address this kind of risk assessment problems, Di Bernardino et al. [24] proposed studying the behavior of a covariate variable Y on the level sets of the distribution of a d-dimensional vector of risk factors $\mathbf{X}$. More precisely, they define and estimate the multivariate Covariate-Conditional-Tail-Expectation (CCTE) given by:

$$\mathrm{CCTE}_\alpha(\mathrm{Y}, \mathbf{X}) := \mathbb{E}[\mathrm{Y} \,|\, \mathbf{X} \in \mathcal{L}_{F_{\mathbf{X}}}(\alpha)],\ \ \alpha \in (0,1), \tag{1.2}$$

where

$$\mathcal{L}_{F_{\mathbf{X}}}(\alpha) := \left\{ x \in \mathbb{R}^d : F_{\mathbf{X}}(x) \geq 1 - \alpha \right\},$$

is the α-upper level set of $F_{\mathbf{X}}$. However, this CCTE based on the distribution function only considers canonical directions. For instance, it could consider an average cost associated to high or low temperatures, but not to high and low temperatures at the same time. Therefore, instead of studying the level sets $\mathcal{L}_{F_{\mathbf{X}}}(\alpha)$, Torres et al. [65] studied the level sets $\mathcal{L}_{F_{R\,\mathbf{X}}}(\alpha)$ of a rotation R of the distribution. In other words, oriented orthant are considered in order to investigate other risk regions. We propose here a more general approach, replacing the distribution function by a depth function. This can be useful in financial or environmental applications and here we present in Section 4.3 a real world environmental application by collecting data of power consumption and temperatures for the region of Nice, France. This can be interesting in forecasting problems of electricity loads in order to help mitigate energy supply interruption risks and allow for long-term forecasts to plan future capacity investments. Moreover, electric companies could use forecasted heat/cold waves (such like the intense heat events that France has experienced in 2022) to anticipate future power consumption. This may be useful in order to calibrate the electricity production over the period of interest. On the mathematical side for instance, Dudek [27] and Oreshkin et al. [57] proposed models for short and mid-term electricity load forecasting respectively, based on patterns of daily cycles (or lags) of load time series.

In order to deal with risk regions, we consider here the lower-level sets of a depth function and propose a depth-based CCTE defined by:

$$\mathrm{CCTE}_{D,\alpha}(\mathrm{Y}, \mathbf{X}) := \mathbb{E}[\mathrm{Y} \,|\, \mathbf{X} \in \mathcal{L}_D(\alpha)],\ \ \alpha > 0, \tag{1.3}$$

where $\mathcal{L}_D(\alpha) = \left\{ x \in \mathbb{R}^d : D(x, P_{\mathbf{X}}) \leq \alpha \right\}$ is the α-depth-based lower level set. We derive rates of convergence for an estimator of CCTE_D which is closely linked to the rate of convergence of the depth level set in terms of symmetric difference, or the Hausdorff distance in smooth depth cases.

The paper is organized as follows. In Section 2, we introduce some notations, tools and the mathematical definition of a depth function. Section 3 is devoted to our main results: in Section 3.1 we study the general asymptotic behavior of our depth level set estimator, in Section 3.2 a construction and consistency and convergence rates of our CCTE_D estimator are given. In subsection 3.2.1 results are derived in a general setting in which we discuss the existing classical depths, while in subsection 3.2.2 we provide consistency and convergence rates of the CCTE_D in the particular case of *Mahalanobis* depth. Illustrations and simulations are presented in Section 4 along with a real world application. Section 5 wraps up our work as a conclusion part. Finally, proofs are postponed to Section 6 where Section 6.1 complements level set estimation Section 3.1 with more details.

2. Notations and definitions

This section is dedicated to introducing some useful notations and tools. We begin by general notations before focusing on depth functions.

2.1. General notations

In the sequel, we denote by $\mathbb{N}$ the set of positive integers and $a \vee b = \max(a, b)$. Let $A \Delta B = (A \backslash B) \cup (B \backslash A)$ be the symmetric difference between the sets A and B, and λ_d designate the Lebesgue measure on $\mathbb{R}^d$, $d \geq 1$. We consider $\mathbb{R}^d$ endowed with the Euclidean norm $\|\cdot\|$ and its unit sphere $S^{d-1} := \{u \in \mathbb{R}^d : \|u\| = 1\}$. Also, ${}^t x$ is the transpose of a d-dimensional real vector x. Let

$$\|f\|_\infty := \text{ess sup}_{x \in \mathbb{R}^d} |f(x)|$$

be the essential supremum of a given function f.

When dealing with random variables, we assume that they are defined on a common underlying probability space $(\Omega, \mathscr{A}, \mathbb{P})$, and we assume that $\mathbb{R}^d$, $d \geq 1$, is equipped with the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$. Furthermore, for any real number $q > 0$, let $\mathbb{L}^q(\Omega) := \mathbb{L}^q(\Omega, \mathscr{A}, \mathbb{P})$ denote the vector space of real-valued random variables U for which $\mathbb{E}[|U|^q] < +\infty$. Let $\mathscr{P} := \mathscr{P}(\mathbb{R}^d)$ be the set of all probability measures on $\mathbb{R}^d$. Given an i.i.d sample $\tilde{S}_n := (\tilde{\mathbf{X}}_i)_{1 \leq i \leq n}$ of size $n \in \mathbb{N}$ with $\mathbb{R}^d$-valued observations from $P \in \mathscr{P}$, we denote by $\tilde{P}_n := \frac{1}{n} \sum_{i=1}^n \delta_{\tilde{\mathbf{X}}_i}$ the empirical measure based on this finite sample. Sometimes, when there is no ambiguity, we simply denote by P the law $P_{\mathbf{X}}$ of the random vector $\mathbf{X}$ meaning $P(A) := P_{\mathbf{X}}(A) = \mathbb{P}(\mathbf{X} \in A)$. Akin to the latter, $P_{(\mathrm{Y}, \mathbf{X})}$ denominates the joint law of the vector $(\mathrm{Y}, \mathbf{X}) \in \mathbb{R} \times \mathbb{R}^d$. Subsequently, for notational convenience, we introduce $\mathbb{E}_P$ the mathematical expectation under P, and $\mathbb{E}_{\tilde{S}_n}[Z] := \mathbb{E}[Z | \tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_n]$ the conditional expectation of Z knowing $\tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_n$. Moreover, we denote by $P_{\tilde{S}_n}(A) := \mathbb{E}_{\tilde{S}_n}[\mathbb{1}_A]$, where $A := A(\tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_n)$ is a subset of $\mathbb{R}^d$ which depends on the data $\mathbf{X}_1, \cdots, \mathbf{X}_n$. Thus, $A(\tilde{\mathbf{X}}_1, \cdots, \tilde{\mathbf{X}}_n)$ is a random subset, so that $P_{\tilde{S}_n}(A)$ is a r.v.

As long as the asymptotics are concerned, we recall that a sequence $(\mathrm{X}_n)_{n \in \mathbb{N}}$ of random variables (r.v.) converges in probability towards the r.v. X if for any

$\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| \geq \varepsilon\right) = 0.$$

Thereafter, convergence in probability is denoted by $\mathrm{X}_n \xrightarrow[n\to\infty]{\mathbb{P}} \mathrm{X}$. Recall also that given two sequences of real numbers $(u_n)_{n\geq 1}$ and $(v_n)_{n\geq 1}$, $u_n = O_n(v_n)$ means that there exist a constant $C > 0$ and $N \in \mathbb{N}$ s.t. for all $n \geq N$, $|u_n| \leq C|v_n|$. Finally, let $(\mathbf{X}_{\boldsymbol{n}})_{\boldsymbol{n}\in\mathbb{N}^r}$, $r \geq 1$, be a set of random variables and $(u_n)_{n\in\mathbb{N}^r}$ be a deterministic set of positive real numbers, we recall the following classical notation of stochastic boundedness

$$\mathbf{X}_{\boldsymbol{n}} = \boldsymbol{\mathcal{O}}_{\mathbb{P},\boldsymbol{n}}(\boldsymbol{u_n}) \overset{\text{def}}{\Longleftrightarrow}$$

$$\forall\ \varepsilon > 0,\ \exists\ M_\varepsilon > 0,\ \exists\ N_\varepsilon \geq 1,\ \forall\ n := (n_1, \cdots, n_r) \in \mathbb{N}^r,$$
$$\min_{1\leq i\leq r} n_i \geq N_\varepsilon \Rightarrow \mathbb{P}\left(\mathbf{X}_{\boldsymbol{n}} \geq M_\varepsilon \cdot u_n\right) \leq \varepsilon.$$

2.2. Depth functions

A depth function is a mapping which measures a degree of centrality of a point w.r.t. an arbitrary distribution. Many depths have been proposed in the literature, e.g., the halfspace depth (Tukey, 1975), the simplicial depth (Liu, 1990), Mahalanobis depth (Mahalanobis, 1936), the zonoid depth (Koshevoy and Mosler, 1997) and others. These depths differ in many aspects, mainly in the shape of trimmed regions or the *deepest point* (point of maximum depth). However, they share some "desirable" properties which every depth should satisfy. In our study, we follow the axiomatization of *multivariate depth function* as stated in Dyckerhoff (2004) (Definition 1 in [29]), which is an alternative to the original definition of depth introduced by Zuo and Serfling (2000) (Definition 2.1 in [70]).

Definition 2.1 (Dyckerhoff [29])**.** A statistical depth function is a mapping $D : \mathbb{R}^d \times \mathscr{P} \to \mathbb{R}$ which is bounded, non negative, measurable in its first argument and satisfying:
(D1) Affine invariance: for any $P_{\mathbf{X}} \in \mathscr{P}$, $b \in \mathbb{R}^d$, and any invertible size d matrix A, $D(Ax + b, P_{A\,\mathbf{X}+b}) = D(x, P_{\mathbf{X}})$.
(D2) Upper semicontinuity: for any $P_{\mathbf{X}} \in \mathscr{P}$ and any $\alpha > 0$ the upper level set $\left\{x \in \mathbb{R}^d : D(x, P_{\mathbf{X}}) \geq \alpha\right\}$ is closed.
(D3) Monotone on rays: for each $P_{\mathbf{X}} \in \mathscr{P}$, each x_0 of maximal depth i.e. $D(x_0, P_{\mathbf{X}}) = \sup_{x\in\mathbb{R}^d} D(x, P_{\mathbf{X}})$ and each $r \in S^{d-1}$, the function $\lambda \in \mathbb{R}_+ \mapsto D(x_0 + \lambda r, P_{\mathbf{X}})$ is monotone decreasing.
(D4) Vanishing at infinity: $D(x, P_{\mathbf{X}}) \to 0$ as $\|x\| \to \infty$, for each $P_{\mathbf{X}} \in \mathscr{P}$.

Let us note that the axiomatization of depth of Dyckerhoff [29] differs from the axiomatization of Zuo and Serfling [70] only by property **(D2)**, which is replaced by a property of maximality at center. According to Dyckerhoff [29] (Proposition 4), when $\mathbf{X}$ is centrally symmetric about some point $x_0 \in \mathbb{R}^d$ (a random vector $\mathbf{X}$ is said to be centrally symmetric distributed with center x_0, if

$\mathbf{X} - x_0$ and $x_0 - \mathbf{X}$ have the same distribution), then a depth function (in sense of definition 2.1) assume maximum value at this center. So, when the definition of depth is restricted to the set of centrally symmetric distributions, axiomatization of Dyckerhoff implies axiomatization of Zuo and Serfling. For further details about depth functions, the interested reader can refer to [70, 71, 72, 49].

Informally, the first property of a depth **(D1)** suggests that the depth of a point $x \in \mathbb{R}^d$ does not depend on the underlying coordinate system. Property **(D2)** is a regularity assumption which is central in our results (see Section 3.1), making the axiomatization of Dyckerhoff [29] better suited to our setting. Property **(D3)** illustrates the fact that as a point $x \in \mathbb{R}^d$ moves away from the point of maximal depth (for instance the "center" of a distribution) along any fixed ray through the center, the depth at x should decrease monotonically. Note that **(D3)** and **(D4)** mean that the upper level sets

$$\left\{x \in \mathbb{R}^d : D(x, P_{\mathbf{X}}) \geq \alpha\right\}, \ \alpha > 0,$$

are bounded and starshaped about the point of maximum depth.

Let us note that properties **(D2)** and **(D4)** play a crucial role in our main results and **(D4)** enables an interpretation of CCTE_D as a tail-conditioned expectation.

3. Main results

In this section, we study the asymptotic behavior of general depth level sets in terms of the *pseudo-metric* of the symmetric difference (i.e. the probability under $P \in \mathscr{P}$ of the symmetric difference) in the case of smooth depths. Then, we define a risk measure based on a general depth function, the Covariate-Conditional-Tail-Expectation (CCTE_D) and we propose an estimator of the CCTE_D using a plug-in estimator of the level set. We study the asymptotic behavior of the CCTE_D when consistency of the level sets in terms of the volume and the pseudo-metric of the symmetric difference is provided. Finally, we derive results in the smooth case of Mahalanobis depth, and discuss when general non-smooth depths are at hand.

3.1. General depth level set estimation

In this section, the problem of interest is to study depth level sets, their estimation and consistency in some sense that will be specified below. Fix a depth function $D : \mathbb{R}^d \times \mathscr{P} \to \mathbb{R}$ and a distribution $P \in \mathscr{P}$. We denote

$$\alpha_{\max}(P) := \sup_{x \in \mathbb{R}^d} D(x, P).$$

For a fixed $\alpha \in (0, \alpha_{\max}(P))$, we define the α-lower level set of D and its plug-in estimator based on an *i.i.d* sample $\tilde{S}_n := (\tilde{\mathbf{X}}_i)_{i=1,..,n}$ from P of size $n \geq 1$:

$$\mathcal{L}_D(\alpha) := \mathcal{L}_D(\alpha, P) := \left\{x \in \mathbb{R}^d : D(x, P) \leq \alpha\right\}, \text{and}$$

$$\mathcal{L}_n(\alpha) := \mathcal{L}_D(\alpha, \tilde{P}_n) := \left\{x \in \mathbb{R}^d : D_n(x) := D(x, \tilde{P}_n) \leq \alpha\right\},$$

where $\tilde{P}_n$ is the empirical measure based on the sample $\tilde{S}_n := (\tilde{\boldsymbol{X}}_{\boldsymbol{i}})_{1\leq i\leq n}$. Regarding convergence of depth level sets, Zuo and Serfling [72] and He and Wang [38] showed uniform depth contours convergence under some conditions on the depth measure while Jeankyung [43] proved their $n^{-\frac{1}{2}}$ rate of convergence. More recently, Brunel [4] derived concentration inequalities for the Hausdorff distance in the case of the halfspace depth, under mild conditions on the distribution.

Here, we are interested in the rate of convergence of the random pseudo-metric of the symmetric difference between $\mathcal{L}_n$ and $\mathcal{L}_D$: $P_{\tilde{S}_n}(\mathcal{L}_n(\alpha)\Delta\mathcal{L}_D(\alpha))$. For absolutely continuous distributions with some density regularity, a sufficient condition to control $P_{\tilde{S}_n}(\mathcal{L}_n(\alpha)\Delta\mathcal{L}_D(\alpha))$ is to find an upper bound for the volume of the symmetric difference as put forward in Theorem 3.1 and Corollary 3.2. Under some smoothness requirements on the depth function, according to Theorem 3.1, the volume $\lambda_d(\mathcal{L}_D(\alpha)\Delta\mathcal{L}_n(\alpha))$ of the symmetric difference between the empirical level set and its population counterpart converges to zero and the quality of our plug-in estimator is obviously related to the quality of our depth estimator D_n. Hereafter we introduce the regularity assumption $(\mathbf{R})$ implying our results following both the approach of Rodríguez-Casal [62] and Cuevas et al. [16] (Proposition 3.1 in the Ph.D. thesis of [62] and Theorem 2 in [16] resp.).

Assumption $(\mathbf{R})$**.** *Let $D : \mathbb{R}^d \times \mathscr{P} \to \mathbb{R}_+$ be a depth as in Definition 2.1. Fix $P \in \mathscr{P}$, $\alpha \in (0, \alpha_{\max}(P))$ and $0 < \varepsilon < \alpha$. Denoting $D(x) := D(x, P)$, we assume the following:*

(i) the function $x \mapsto D(x)$ is continuous on $\mathbb{R}^d$ and of class $\mathscr{C}^2$ on the set $\mathcal{K}_\varepsilon(\alpha) := D^{-1}([\alpha-\varepsilon, \alpha+\varepsilon])$,

(ii) $m_\nabla := m_\nabla(\alpha, \varepsilon, P) := \inf_{x\in\mathcal{K}_\varepsilon(\alpha)} \|(\nabla D)_x\| > 0$, where $(\nabla D)_x$ is the gradient of $D(\cdot)$ at x,

(iii) $\|D_n - D\|_{\infty,\mathbb{R}^d} \xrightarrow[n\to\infty]{\mathbb{P}\text{-}a.s.} 0.$

Under assumption $(\mathbf{R})$, one controls the volume of the symmetric difference between $\mathcal{L}_n$ and $\mathcal{L}_D$ by $\|D_n - D\|_{\infty,\mathbb{R}^d}$. An immediate result on the pseudo-metric is hence obtained in Corollary 3.2. The proofs of both results are postponed to Section 6. As long as validity of Assumption $(\mathbf{R})$ is concerned, it is known that many of the commonly encountered depths in the literature satisfy assumption $(\mathbf{R})$*(iii)* (e.g. halfspace, simplicial, Mahalanobis, projection, depths...). Continuity of the depth function at x is obtained in most cases (mainly for absolutely continuous distributions). Assumptions $(\mathbf{R})(i)$ and $(\mathbf{R})(ii)$, though, are not always satisfied by the above mentioned depths. However, $(\mathbf{R})(i)$ and $(\mathbf{R})(ii)$ can be easily shown to hold for the Mahalanobis and L^2 depths (see Example 2.3 in [70]) for instance. The intuition behind those regularity conditions is essentially the need to avoid flat depths around the level α, since the presence of plateaus generally causes structural issues in the plug-in estimation of the level sets. One can note that a sufficient condition for a depth to have no plateaus is the property of strict monotonicity as studied by [30]. Under some additional assumptions on the distribution P (e.g. absolute continuity, convex support), common depth functions such as Mahalanobis, halfspace, zonoid, weighted mean [31] depths, are strictly monotone [30]. However, in our setting, we need a stronger condition on the depth, namely the differentiability property. We refer the reader to Section 6.1 for more details about those technical results, discussing especially assumption (**R**).

Theorem 3.1. *Under assumption (**R**), it holds*

$$\lambda_d\left(\mathcal{L}_D(\alpha)\Delta\mathcal{L}_n(\alpha)\right) = \underset{n\to\infty}{O}\left(\|D_n - D\|_{\infty,\mathbb{R}^d}\right),\ \mathbb{P}\text{-}a.s.$$

Corollary 3.2. *Let $P \in \mathscr{P}$ be an absolutely continuous distribution with density function $f \in \mathbb{L}^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$ for some $p \in (1, +\infty]$. Under assumption (**R**), it holds*

$$P_{\tilde{S}_n}\left(\mathcal{L}_n(\alpha)\Delta\mathcal{L}_D(\alpha)\right) = \underset{n\to\infty}{O}\left(\|D_n - D\|_{\infty,\mathbb{R}^d}^{1-\frac{1}{p}}\right),\ \mathbb{P}\text{-}a.s.$$

The proof of Theorem 3.1 relies mainly on a technical result we postpone to Section 6.1, namely Theorem 6.1. In the latter we show that if a depth is smooth enough in the sense of Assumption (**R**), its empirical version D_n is only upper semicontinuous and D_n is a consistent estimator of D (uniformly on $\mathbb{R}^d$), then one obtains asymptotic upper bound for the Hausdorff distance $d_H(\partial\mathcal{L}_D(\alpha), \partial\mathcal{L}_n(\alpha))$ between the respective boundaries $\partial\mathcal{L}_D(\alpha)$ and $\partial\mathcal{L}_n(\alpha)$. To be more precise, from the approach of Cuevas et al. [16], $d_H(\partial\mathcal{L}_D(\alpha), \partial\mathcal{L}_n(\alpha))$ is controlled by the supremum norm error $\|D_n - D\|_{\infty,\mathbb{R}^d}$ related to the depth D. Using Weyl's tube formula [69] one obtains that the Hausdorff distance between the boundaries of the level sets is an asymptotic upper bound for the volume of the symmetric difference, and hence so is $\|D_n - D\|_{\infty,\mathbb{R}^d}$.

3.2. Depth based CCTE

3.2.1. General setting

Consider a couple $(\mathrm{Y}, \mathbf{X})$ s.t. Y is a real random variable which is dependent of a random vector $\mathbf{X} \in \mathbb{R}^d$ with distribution $P_{\mathbf{X}}$. In Definition 3.3, we formally define our CCTE_D and propose an estimator of the latter. For $n_1,\ n_2 \geq 1$, let

$$\tilde{S}_{n_1} := (\tilde{\mathbf{X}}_i)_{i=1,..,n_1} \text{ be an i.i.d } n_1\text{-sample from } P_{\mathbf{X}}, \text{ and}$$
$$S_{n_2} := ((\mathrm{Y}_i, \mathbf{X}_i))_{i=1,..,n_2} \text{ be an i.i.d } n_2\text{-sample from } P_{(\mathrm{Y},\mathbf{X})},$$

s.t. $\tilde{S}_{n_1}$ and S_{n_2} are independent.

In what follows, we provide the definition of our CCTE_D and its associated estimator.

Definition 3.3 (Depth-based Covariate-Conditional-Tail-Expectation)**.** Let $\mathbf{X} \in \mathbb{R}^d$ be a random vector with distribution $P \in \mathscr{P}$ and Y be an integrable real random variable (which is dependent on $\mathbf{X}$). Let $\alpha > 0$ and assume $P[\mathcal{L}_D(\alpha)] > 0$.

(i) The depth-based Covariate-Conditional-Tail-Expectation at level α is defined by:

$$\mathrm{CCTE}_{D,\alpha}(\mathrm{Y}, \mathbf{X}) := \mathbb{E}[\mathrm{Y} \,|\, \mathbf{X} \in \mathcal{L}_D(\alpha)].$$

(ii) Its estimator based on the sample $\tilde{S}_{n_1}$ and S_{n_2} is given by:

$$\widehat{\mathrm{CCTE}}_{D,\alpha}^{n_1,n_2}(\mathrm{Y}, \mathbf{X}) := \frac{\sum_{i=1}^{n_2} \mathrm{Y}_i \mathbb{1}_{\mathbf{X}_i \in \mathcal{L}_{n_1}(\alpha)}}{\sum_{i=1}^{n_2} \mathbb{1}_{\mathbf{X}_i \in \mathcal{L}_{n_1}(\alpha)}}, \tag{3.1}$$

with the convention $0/0 = 0$.

Our main result, namely Theorem 3.5, links the rate of convergence of the CCTE_D to the one of the symmetric difference between the true and estimated α-level set. We first state the following assumption describing a convergence rate for the level sets:

Assumption (H0). *There exists an increasing sequence of positive real numbers $(v_{n_1})_{n_1 \geq 1}$ s.t.*

$$P_{\tilde{S}_{n_1}}\left(\mathcal{L}_{n_1}(\alpha) \Delta \mathcal{L}_D(\alpha)\right) = \mathcal{O}_{P,n_1}\left(v_{n_1}^{-1}\right).$$

For general (non-smooth) depths, checking Assumption (**H0**) with a rate of convergence seems difficult. So here, in a first place we show that $\widehat{\mathrm{CCTE}}^{n_1,n_2}$ converges to CCTE provided that D_{n_1} converges uniformly to D. This is derived in Theorem 3.4 using mainly the proof of Theorem 3.5. In Section 6.2, we first give the proof of Theorem 3.5. Theorem 3.4 is proved following the guidelines of the proof of Teorem 3.5.

Theorem 3.4. *Let $\alpha > 0$ and $P \in \mathscr{P}$. Assume $P[\mathcal{L}_D(\alpha)] > 0$ and there exists $r \in [2, \infty]$ s.t. $Y \in \mathbb{L}^r(\Omega)$. If the following two conditions are satisfied*

(i) $\|D_{n_1} - D\|_{\infty,\mathbb{R}^d} \xrightarrow[n_1 \to \infty]{\mathbb{P}} 0$,

(ii) $\mathbb{P}(D(\boldsymbol{X}) = \alpha) = 0$, *where $\boldsymbol{X}$ has distribution P,*
then

$$\left|\widehat{\mathrm{CCTE}}_{D,\alpha}^{n_1,n_2}(Y, \boldsymbol{X}) - \mathrm{CCTE}_{D,\alpha}(Y, \boldsymbol{X})\right| \xrightarrow[n_1,n_2 \to \infty]{\mathbb{P}} 0.$$

The key argument of the proof of Theorem 3.4 lays on the fact that for any depth D s.t. $\|D_{n_1} - D\|_{\infty,\mathbb{R}^d} \xrightarrow[n_1 \to \infty]{\mathbb{P}} 0$ and $\mathbb{P}_P(D(\mathbf{X}) = \alpha) = 0$, it holds

$$\mathbb{E}\left[P_{\tilde{S}_{n_1}}\left(\mathcal{L}_D(\alpha) \Delta \mathcal{L}_{n_1}(\alpha)\right)\right] \xrightarrow[n_1 \to \infty]{} 0,$$

so that assumption (**H0**) is satisfied (without an explicit rate of convergence here) and hence the result follows from the guidelines of the proof of Theorem 3.5. Note that assumption $\mathbb{P}_P(D(\mathbf{X}) = \alpha) = 0$ is not too restrictive and can be found in depths literature, for example in Theorem 4.1 in Zuo and Serfling [72] the convergence of sample depth contours is studied. Besides, any common depth D satisfies $\|D_{n_1} - D\|_{\infty,\mathbb{R}^d} \xrightarrow[n_1\to\infty]{\mathbb{P}-\text{a.s.}} 0$, which implies convergence in probability.

In the spirit of Di Bernardino et al. [24], Theorem 3.5 gives the rates of convergence for CCTE_D. Note that in Theorem 3.5, the *r-th* moment of Y is only involved in the rate $(v_{n_1})_{n_1}$.

Theorem 3.5. *Let $\alpha > 0$ and $P \in \mathscr{P}$. Assume $P[\mathcal{L}_D(\alpha)] > 0$, and (**H0**) is satisfied and there exists $r \in [2, \infty]$ s.t. $Y \in \mathbb{L}^r(\Omega)$. Then, it holds that*

$$\left|\widehat{\text{CCTE}}_{D,\alpha}^{n_1,n_2}(Y, \boldsymbol{X}) - \text{CCTE}_{D,\alpha}(Y, \boldsymbol{X})\right| = \mathcal{O}_{P,n_1,n_2}\left(n_2^{-\frac{1}{2}} \vee v_{n_1}^{-(1-\frac{1}{r})}\right).$$

The proofs of the main theorems can be found in Section 6 (sub-section 6.2). Corollary 3.6 below is a straightforward consequence of Corollary 3.2 (Section 3.1) and Theorem 3.5 when studying smooth depth.

Corollary 3.6. *Let $\alpha > 0$ and $P \in \mathscr{P}$ be an absolutely continous distribution with density function in $\mathbb{L}^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$ for some $p \in (1, +\infty]$. Assume $P[\mathcal{L}_D(\alpha)] > 0$ and there exists $r \in [2, \infty]$ s.t. $Y \in \mathbb{L}^r(\Omega)$. Then, under assumptions of Corollary 3.2 it holds*

$$\left|\widehat{\text{CCTE}}_{D,\alpha}^{n_1,n_2}(Y, \boldsymbol{X}) - \text{CCTE}_{D,\alpha}(Y, \boldsymbol{X})\right| = \mathcal{O}_{P,n_1,n_2}\left(n_2^{-\frac{1}{2}} \vee v_{n_1}^{-(1-\frac{1}{r})}\right),$$

with $(v_{n_1})_{n_1}$ s.t. $\|D_{n_1} - D\|_\infty^{1-\frac{1}{p}} = \mathcal{O}_{P,n_1}\left(v_{n_1}^{-1}\right)$.

It is well known that $\sqrt{n}\|F_n - F\|_\infty \xrightarrow[n\to\infty]{\mathbb{P}} 0$ where F is a given c.d.f and F_n its empirical version (see e.g. [66]), a result from which Di Bernardino et al. [24] derived under mild conditions $O(\sqrt{n})$ convergence for c.d.f based CCTE. As far as depths are concerned, the convergence of the U-process based on simplicial depth was shown by Dümbgen [28], while Massé [52] proved that in general the empirical Tukey process does not converge weakly, even though its marginals always do. To the best of our knowledge, the rate at which $\|D_n - D\|_\infty$ goes to zero has not been derived for existing smooth depth functions in the literature (for instance Mahalanobis depth, or L^2-depth). This is why in section 3.2.2 we derive a rate of convergence for Mahalanobis depth.

3.2.2. Mahalanobis depth

The *Mahalanobis* depth function (see Definition 3.7 below) is a depth function in the sense of Definition 2.1 (Example 2.5 in [70]), and is smooth as a function of x (which implies the upper-semicontinuity property in the empirical case as well).

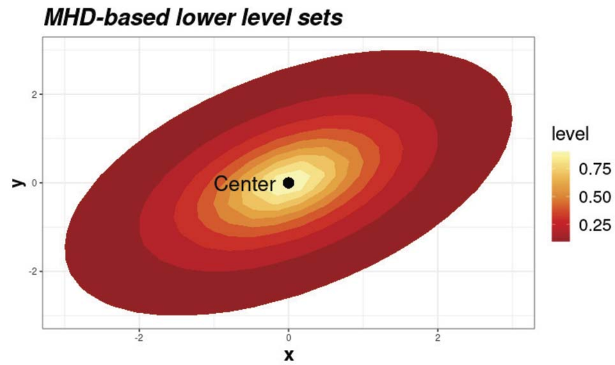

Fig 1. *Theoretical lower-level sets based on $MHD(\cdot, P)$, with P the law of a Gaussian vector in $\mathbb{R}^2$.*

Definition 3.7 (*Mahalanobis depth*, Zuo and Serfling [70])**.** Let $\mathbf{X} \in \mathbb{R}^d$ be a random vector with distribution $P \in \mathscr{P}$. The *Mahalanobis* depth is defined by

$$MHD(x, P) = \begin{cases} \left(1 + d^2_{\Sigma_{\mathbf{X}}}(x, \mu_{\mathbf{X}})\right)^{-1} & \text{if } \mathbb{E}_P[\| \mathbf{X} \|^2] < +\infty \text{ and } \Sigma_{\mathbf{X}} \textit{ invertible} \\ 0 & \text{if } \mathbb{E}_P[\| \mathbf{X} \|^2] = +\infty \text{ or } \Sigma_{\mathbf{X}} \text{ is singular} \end{cases}$$

where $\mu_{\mathbf{X}} = \mathbb{E}_P[\mathbf{X}]$ is the mean vector of $\mathbf{X}$ and $\Sigma_{\mathbf{X}}$ is its covariance matrix and

$$d^2_{\Sigma_{\mathbf{X}}}(x, \mu_{\mathbf{X}}) := \|x - \mu_{\mathbf{X}}\|^2_{\Sigma_{\mathbf{X}}} := {}^t(x - \mu_{\mathbf{X}})\Sigma_{\mathbf{X}}^{-1}(x - \mu_{\mathbf{X}})$$

is the *Mahalanobis* distance.

Remark 3.8. Note that the above definition of MHD is introduced as such in order to highlight the fact that it is restricted to distributions with invertible covariance matrix while still remaining a depth function in the sense of Definition 2.1. Furthermore, for a fixed distribution P with invertible covariance matrix, the function $x \in \mathbb{R}^d \mapsto MHD(x, P)$ is infinitely differentiable, concave, and has $x = \mu_{\mathbf{X}}$ as unique critical point, thus $\mu_{\mathbf{X}} = \arg\max_{x \in \mathbb{R}^d} MHD(x, P)$. And $\alpha_{\max}(P) := \max_{x \in \mathbb{R}^d} MHD(x, P) = 1$.

It is well known that Mahalanobis depth is a depth in sense of the axiomatization of Dyckerhoff [29] of which the level sets have elliptical shape around the mean (see Figure 1). A natural estimator of MHD is given by

$$MHD_n(x) := MHD(x, \tilde{P}_n) = \left(1 + {}^t(x - \hat{\mu}_n)\hat{\Sigma}_n^{-1}(x - \hat{\mu}_n))\right)^{-1}, \qquad (3.2)$$

where $\hat{\mu}_n$ and $\hat{\Sigma}_n$ are respectively the empirical mean vector and empirical covariance matrix based on $\tilde{P}_n$. In order to study the rate of convergence of the CCTE$_D$ estimator based on $D = MHD$, we check here Assumption (**H0**). According to Section 3.1, in the case of absolutely continuous distributions and $\mathscr{C}^2$ depths, the problem reduces to studying the rate of convergence of $\|D_n - D\|_\infty$ to zero, in probability (c.f. Section 3.1, Corollary 3.2). In Theorem 3.9

we provide the rate of convergence (in probability) of MHD_n to its population version MHD uniformly on $\mathbb{R}^d$.

Theorem 3.9. *Let $\boldsymbol{X}$ be a random vector with distribution $P \in \mathscr{P}$ and invertible covariance matrix, s.t. $\boldsymbol{X}$ satisfies $\mathbb{E}_P[\| \boldsymbol{X} \|^4] < \infty$. Then, it holds that*

$$\|MHD_n - MHD\|_{\infty,\mathbb{R}^d} = \mathcal{O}_{P,n}\left(n^{-\frac{1}{2}}\right).$$

Finally, in Corollary 3.10, we derive the specific rate of convergence for the CCTE based on MHD-depth. Let us note that regularity assumption (**R**) as stated in Section 3.1 for Corollary 3.2 is satisfied by Mahalanobis depth (see Proposition 6.6, Section 6.1).

Corollary 3.10. *Let $P \in \mathscr{P}$, $D(\cdot, P) = MHD(\cdot, P)$ and $\alpha \in (0, 1)$. Assume $P[\mathcal{L}_D(\alpha)] > 0$. Under assumptions of Corollary 3.6 and Theorem 3.9, it holds that*

$$\left|\widehat{\mathrm{CCTE}}_{D,\alpha}^{n_1,n_2}(Y, \boldsymbol{X}) - \mathrm{CCTE}_\alpha(Y, \boldsymbol{X})\right| = \mathcal{O}_{P,n_1,n_2}\left(n_2^{-\frac{1}{2}} \vee n_1^{-\frac{1}{2}\left(1-\frac{1}{r}\right)\left(1-\frac{1}{p}\right)}\right).$$

4. Simulations and illustrations

The present section is divided into three parts. First, Section 4.1 is dedicated to illustrating consistency of depth level sets in terms of the pseudo-metric (see Section 3.1) and Section 4.2 to consistency of depth based CCTE. In both sections we work with Mahalanobis, projection and halfspace depths as a non-exhaustive list of depth examples. About the estimation of each of the previous depth, we use the R package `ddalpha` (Pokotylo et al. [60]) which has good algorithms for computing these functions. Finally, in Section 4.3 we present a real world environmental application. Subsequently, all the simulations are done in dimension $d = 2$.

4.1. Level set estimation

The aim of the current section is to illustrate the consistency of depth level sets in terms of the pseudo-metric as stated in Corollary 3.2, for Mahalanobis (MHD), projection (D_{PJ}) and halfspace (D_{HS}) depths, and to exhibit the corresponding rates of convergence. This convergence rate is provided for MHD in Theorem 3.9 along with Corollary 3.2 since it fulfills the smoothness requirements stated in assumption (**R**) of Corollary 3.2, but this is clearly not the case for both of the other depths. Recall nevertheless that we showed $\mathbb{L}^1$-consistency of the pseudo-metric without an explicit rate of convergence for any general depth (see the proof of Theorem 3.4, in Section 6.2).

The projection depth [70] is defined by

$$D_{PJ}(x, P) = \left(1 + \sup_{u \in S^{d-1}} \frac{|\langle u, x\rangle - \mathrm{Med}(\langle u, \mathbf{X}\rangle)|}{\mathrm{MAD}(\langle u, \mathbf{X}\rangle)}\right)^{-1},$$

where $\mathbf{X} \sim P$, Med is the median of a real r.v. and $\mathrm{MAD}(Z) = \mathrm{Med}\,(|Z - \mathrm{Med}\,(Z)|)$ is the median absolute deviation of a real r.v. Z. The halfspace depth is defined by

$$D_{HS}(x, P) := \inf_{u \in S^{d-1}} \mathbb{P}_P\left(\langle u, x\rangle \geq \langle u, \mathbf{X}\rangle\right).$$

The empirical counterparts of $D_{PJ}(x, P)$ and $D_{HS}(x, P)$ are then given by taking $P = P_n$ the empirical measure based on a sample of size n from P. In practice, one cannot always compute the population versions of the projection or halfspace depth (at least when $d \geq 2$). So, in our simulations hereafter, we considered that the population projection/halfspace depth is computed with a "large" sample of size 50000 in order to get an estimation of its associated level sets (note that the halfspace calculation has exponential complexity in the sample size n and the dimension d, see e.g. [63], thus it was not possible to use a larger sample in a reasonable time). Now, when $D = MHD$, we perform a deterministic approximation (of precision 10^{-8}) of the theoretical mean vector and covariance matrix, m and Σ respectively. This will provide a "population" version of MHD depth. To be precise, we estimate the mean pseudo-metric $p^n(D, \alpha) := \mathbb{E}[P_{\tilde{S}_n}(\mathcal{L}_D(\alpha)\Delta\mathcal{L}_n(\alpha))]$ by

$$\hat{p}^n_{N,M}(D, \alpha) := \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\mathbb{1}_{\mathbf{X}_i \in \mathcal{L}_D(\alpha)\Delta\mathcal{L}^j_n(\alpha)},$$

where $\mathcal{L}^j_n(\alpha)$ is a j-th estimation of the empirical level set $\mathcal{L}_n(\alpha)$. For all three depths above, we took $N = M = 400$ and different values of n and α (see Tables 1 to 3 in the appendix).

Thereafter, in all the simulations we consider dependent risk factors $\mathbf{X}_1$ and $\mathbf{X}_2$ via a bivariate Frank copula (with parameter -5) with Gumbel marginals with parameter $(\mu_1, \beta_1) = (0, 0.25)$ and $(\mu_2, \beta_2) = (-0.5, 0.25)$ respectively (Figure 2). The bivariate Frank copula with parameter $\theta \neq 0$ is defined by:

$$C_\theta(u, v) := -\frac{1}{\theta}\log\left(1 + \frac{\left(1 - e^{-\theta u}\right)\left(1 - e^{-\theta v}\right)}{e^{-\theta} - 1}\right),\ (u, v) \in \mathbb{R}^2$$

so that the bivariate sample introduced above has a c.d.f given by:

$$F(x_1, x_2) := C_{\theta=-5}(F_{\mu_1,\beta_1}(x_1), F_{\mu_2,\beta_2}(x_2)),$$

where $F_{\mu,\beta}(x) := \exp\left(-e^{\frac{\mu - x}{\beta}}\right)$ is the univariate c.d.f. of a Gumbel distribution with parameter (μ, β).

Let us precise that we choose MHD because of its simplicity and the other depths because they capture more or less the shape of the above sample as illustrated in Figure 2. Recall that the convergence of $p^n(D, \alpha)$ to zero as $n \to \infty$ is a result we proved in Theorem 3.4. As long as the speed rate of $p^n(D, \alpha)$ is concerned, Figure 3 seems to confirm that the case $\sqrt{n}$ is a limit regime ($\varepsilon = 0$ corresponding to dotted lines), at least for both projection and

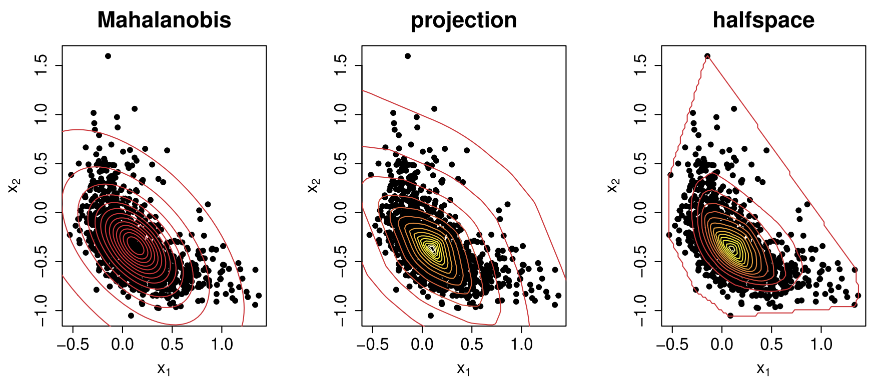

FIG 2. *Depth contours based on a sample of dependent Gumbel marginals via a Frank Copula, for Mahalanobis, projection and halfspace depth respectively.*

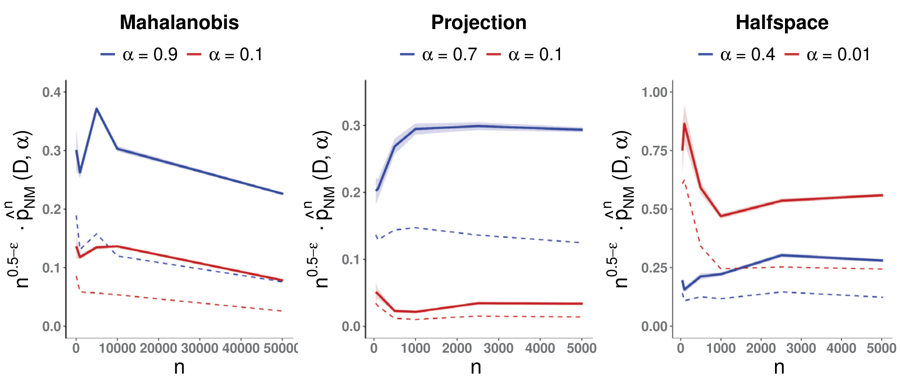

FIG 3. *Estimated $n^{1/2-\varepsilon} \cdot \hat{p}^n_{N,M}(D,\alpha)$ based on a sample of dependent Gumbel marginals via a Frank Copula, for $D = MHD, D_{PJ}$ and D_{HS} respectively. The colored ribbons display a 95% confidence region for each considered level. For each level α, the filled lines correspond to $\varepsilon = 0$ while the dashed ones to $\varepsilon = 0.1$.*

halfspace depths (this illustrates condition (**H0**) in Section 3.2). This critical rate is obtained from our results when $D = MHD$ (see Theorem 3.9 along with Corollary 3.2) but it is just an observation from the simulations for D_{HS} and D_{PJ}. One could still hope for similar regimes for larger values of n or/and other sample types, and thus for theoretical speed rates to hold for both of projection and halfspace depths. Notice however, for low levels α, the halfspace depth exhibit greater values of convergence error (solid red line in Figure 3, halfspace graph with $\alpha = 0.01$), contrary to what is observed for both MHD and D_{PJ} depths (first two graphs on Figure 3, $\alpha = 0.1$). One explanation could be that the level $\alpha = 0.01$ is too extreme compared to the sample size $n = 5000$ chosen for the halfspace depth so that the estimation of the level sets and hence the pseudo-distance leads to greater error values. Not to forget that population level

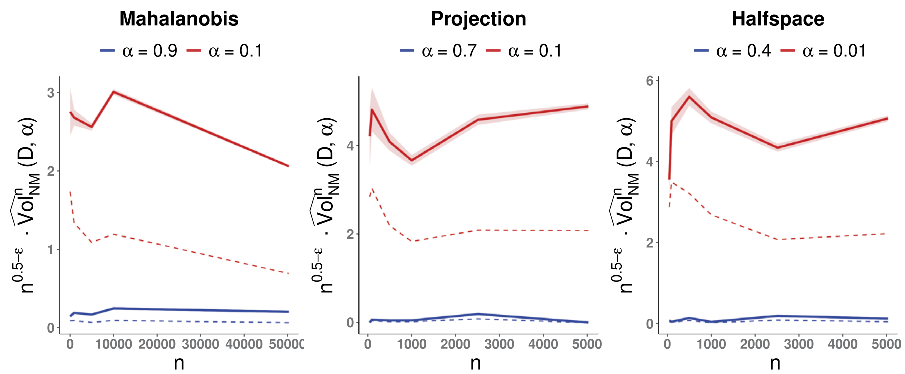

Fig 4. *Estimated $n^{1/2-\varepsilon} \cdot \widehat{\text{Vol}}^n_{N,M}(D,\alpha)$ based on a sample of dependent Gumbel marginals via a Frank Copula, for $D = MHD, D_{PJ}$ and D_{HS} respectively. The colored ribbons display a 95% confidence region for each considered level. For each level α, the filled lines correspond to $\varepsilon = 0$ while the dashed ones to $\varepsilon = 0.1$.*

sets aren't at hand which makes it difficult to completely conclude our study.

According to Theorem 3.1 (see Section 3.1), another natural way to study convergence between the level sets is in the pseudo-metric defined by the d-dimensional volume $\lambda_d(\mathcal{L}_D(\alpha)\Delta\mathcal{L}_n(\alpha))$. Similarly, using a Monte-Carlo procedure, we estimate the mean pseudo-metric $\mathbb{E}[\lambda_2(\mathcal{L}_D(\alpha)\Delta\mathcal{L}_n(\alpha))]$ by

$$\widehat{\text{Vol}}^n_{N,M}(D,\alpha) := \frac{v_2}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\mathbb{1}_{\mathbf{U}_i\in\mathcal{L}_D(\alpha)\Delta\mathcal{L}^j_n(\alpha)},$$

where $(\mathbf{U}_i)_{1\le i\le N}$ are i.i.d. bivariate uniform r.v. over the square $[-1,2]^2$ which contains the simulated gumbel samples (see Figure 2) as well as the observed symmetric difference. Here, v_2 designates the surface area of the square $[-1,2]^2$, that is $v_2 = 9$. Also, $\mathcal{L}^j_n(\alpha)$ is a j-th estimation of the empirical level set $\mathcal{L}_n(\alpha)$. As observed earlier, Figure 4 seems to depict the $\sqrt{n}$-speed rate of the estimator $\widehat{\text{Vol}}^n_{N,M}(D,\alpha)$ as a critical regime for all three depths previously considered (see also Tables 4 to 6 in the appendix). As mentioned previously, we recall that this $\sqrt{n}$-speed rate has been theoretically derived when $D = MHD$. However, it is simply an observation from the simulations when $D = D_{HS}, D_{PJ}$ which is not inconsistent with $\sqrt{n}$ being a critical regime for both depths. Notice further that the pseudo-metric $\lambda_2(\mathcal{L}_D(\alpha)\Delta\mathcal{L}_n(\alpha))$ makes an increase in the estimated error compared to the probability-based pseudo-metric $P_{\mathbf{X}}(\mathcal{L}_D(\alpha)\Delta\mathcal{L}_n(\alpha))$ (as seen in Figure 3). This phenomenon is consistent with the fact that λ_2 can account for small probability mass regions under the law of $\mathbf{X}$.

4.2. CCTE estimation

In this section, we provide an illustration of Corollary 3.10 i.e. consistency of MHD-based CCTE. The latter provides CCTE consistency in probability whereas we perform $\mathbb{L}^1$-estimation in practice, hence we are actually illustrating a stronger result than Corollary 3.10. For purposes of comparison, we perform $\mathbb{L}^1$-estimation of the CCTE based on D_{PJ} and D_{HS} depths as well. According to Theorem 3.5, the rate of convergence of CCTE is linked to the one of the pseudo-metric. So, since convergence rates of the pseudo-metric haven't been derived theoretically for D_{PJ} and D_{HS} (see the beginning of Section 4.1), we could not derive those associated to CCTE. Yet, we made similar observations as for MHD for both of these depths (see Figure 5). CCTE calculations based on the three depths above are given in Tables 7 to 12 in the appendix. Hereafter the detailed setting is presented.

We study the estimated CCTE_D for cost variables Y which are dependent on the law of $\mathbf{X} := (X_1, X_2) \in \mathbb{R}^2$ and having the form:

$$\mathrm{Y} = \| \mathbf{X} \|^2 + \varepsilon,$$

where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, $\sigma^2 > 0$, is a gaussian noise which is independent of $\mathbf{X}$. In our simulations we will take $\sigma^2 = 0.005$. Here, we choose the squared euclidian norm defined by $\|\boldsymbol{x}\|^2 = |x_1|^2 + |x_2|^2$. The bivariate risk factors $\mathbf{X}$ are sampled from Gumbel data with a dependency structure just as in Section 4.1. Note that the above example satisfies the assumptions of Corollary 3.10. First, we compare $\widehat{\text{CCTE}}^{n_1,n_2}_{MHD,\alpha}$ with the theoretical $\text{CCTE}_{MHD,\alpha}$ for *Mahalanobis* depth. For the sake of simplicity, we take $n_1 = n_2 = n$. With the same sample used in Section 4.1, p, r are arbitrarily large so that $|\widehat{\text{CCTE}}^{n}_{MHD,\alpha} - \text{CCTE}_{MHD,\alpha}|$ decays to zero at most with a convergence rate $O(\sqrt{n})$. On another note, due to the complexity of the level-sets as domains of integration in the computation of the CCTE, we perform a Monte Carlo procedure to fix the "true" value of the CCTE based on a sample of size 10^8 (without noise), that is:

$$\frac{\sum_{i=1}^{10^8} \mathrm{Y}_i \mathbb{1}_{\mathbf{X}_i \in \mathcal{L}_{MHD}(\alpha)}}{\sum_{i=1}^{10^8} \mathbb{1}_{\mathbf{X}_i \in \mathcal{L}_{MHD}(\alpha)}}.$$

Recall that for the sake of computational simplicity, we provide $\mathbb{L}^1$-estimation for the CCTE_D (which implies convergence results in probability). More precisely, we denote $\overline{\widehat{\text{CCTE}}^{n}_{\alpha}} := \overline{\widehat{\text{CCTE}}^{n}_{\alpha,MHD}}$ the mean of the $\widehat{\text{CCTE}}^{n}_{\alpha,MHD}$ based on 400 simulations. The empirical standard deviation is

$$\hat{\sigma} = \sqrt{\frac{1}{N_{mc} - 1} \sum_{j=1}^{N_{mc}} \left(\widehat{\text{CCTE}}^{n}_{\alpha,j} - \overline{\widehat{\text{CCTE}}^{n}_{\alpha}} \right)^2},$$

while the relative mean absolute error associated to $\widehat{\text{CCTE}}^{n}_{\alpha}$, denoted by RMAE,

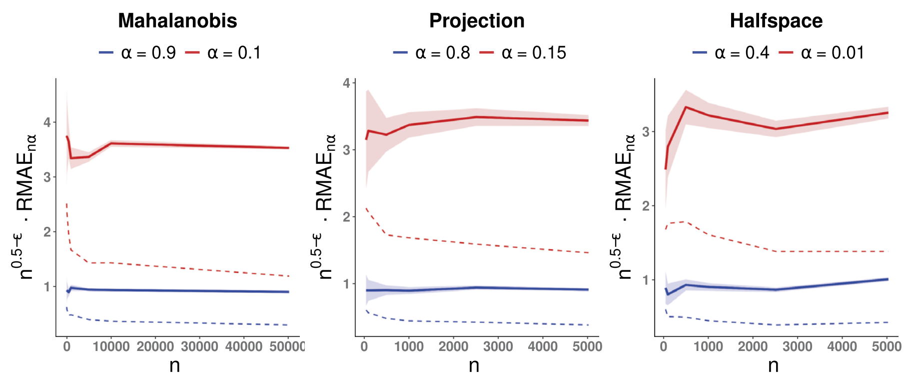

FIG 5. *Convergence rates for Mahalanobis and halfspace depth based* CCTE *for bivariate Frank Copulas with Gumbel marginals. The colored ribbons display a* 95% *confidence region for each considered level. For each level* α*, the filled lines correspond to* $\varepsilon = 0$ *whereas the dashed ones to* $\varepsilon = 0.1$.

is defined as follows:

$$\mathrm{RMAE}_{MHD} := \mathrm{RMAE}_{n,\alpha} = \frac{1}{N_{mc}} \sum_{j=1}^{N_{mc}} \frac{\left|\widehat{\mathrm{CCTE}}^{n}_{\alpha,j} - \mathrm{CCTE}_{MHD,\alpha}(\mathrm{Y}, \mathbf{X}))\right|}{|\,\mathrm{CCTE}_{MHD,\alpha}(\mathrm{Y}, \mathbf{X})|},$$

with $N_{mc} = 400$. Note that, most of the times, one uses the Relative Mean Squared Error (RMSE) rather than the RMAE. However, since our results are presented with absolute value, we work here with the RMAE for which we provide $\mathbb{L}^1$-estimation as well. As a matter of fact, the $O(\sqrt{n})$ convergence observed in our simulations is stronger than just the result of Corollary 3.10 since in the latter rates are obtained in probability, while here we performed $\mathbb{L}^1$-estimation (see Figure 5).

For the sake of comparison, we performed the same previous estimations for the projection and halfspace depths. A similar procedure to the one of Section 4.1 was followed when giving an approximated "theoretical" $\mathrm{CCTE}_{D_{HS}}$ (resp. $\mathrm{CCTE}_{D_{PJ}}$) using a 100000-sample for the level set and 400-sample for the mean (since both of these depths are computationally demanding). This is summarized in Tables 9 to 12 in the appendix. Further, the $\mathrm{RMAE}_{D_{HS}}$ (resp. $\mathrm{RMAE}_{D_{PJ}}$) based on halfspace (resp. projection) depth was computed with $N_{mc} = 400$ for each different sample size n enabling the estimation of the level set. Regarding the overall CCTE estimations, from Figure 5, the critical regime seems to be again $\sqrt{n}$. Recall this is a stronger result than our Corollary 3.10 for MHD, but it just arises from the simulations for D_{PJ} and D_{HS} since we have not derived explicit rates for both of these depths. This can be one explanation to the RMAE behavior for the halspace depth in this particular setup (Figure 5). Remark further for low levels α ($\alpha = 0.1$, $\alpha = 0.15$ and $\alpha = 0.01$ for MHD, D_{PJ} and D_{HS} resp.) and any sample sizes n ($n \leq 50000$ for MHD and $n \leq 5000$ for D_{PJ} and D_{HS}), the value of RMAE is relatively high. This may be explained by the fact that for small values of α there is fewer data to observe so that it becomes more difficult to estimate the mean $\widehat{\text{CCTE}}$ as well as the α-level set. Indeed, for low levels α the constant $A = 2/m_{\nabla}$ can be large since m_{∇} approaches zero (see Theorem 6.1 and Remark 6.4 in Section 3.1) meaning that the constant bounding the error (RMAE) becomes large, thus we will need a "large" dataset to get a reasonable estimation.

4.3. Application

In this section we present a real world environmental application. The data set, which was downloaded from *RTE France*[1] and *Meteo France*[2] consists of three-hourly power consumptions and temperatures for the region of *Nice* (France) over a three-year period (2019-2021), giving rise to 4377 observations. To be more precise, observations are of the form $(y, \boldsymbol{x}) \in \mathbb{R} \times \mathbb{R}^2$, where we consider bivariate risk factors $\boldsymbol{x} := (t_1, t_2)$ with t_1 being the temperature at a given hour H and t_2 is that six hours before. To each $\boldsymbol{x}$ is associated a cost variable y representing the power consumption (in MWh) at hour H (see Figure 6). This kind of problem with lagged dataset has been intensively studied for decades in the short-term load forecasting literature. Most commonly, the aim is to forecast one to five days ahead electric hourly load, based on weather input variables such as one or two days-lagged temperature, humidity... For example, Dudek [27] and Oreshkin et al. [58] proposed models for short and mid-term electricity load forecasts respectively (i.e from hours to days ahead resp.) based on lags of electricity load. This has become an essential task in the scheduling of accurate electricity supplies and the management of power system.Indeed, it is possible to forecast temperatures several days in advance. Then, knowing the behavior of the cost, the electricity consumption, given the predictable covariables, the temperatures, is useful to adapt the electricity production (by starting up a power plant, for example). In practice, the main variable that has been included is the air temperature, since it has been known that the demand rises on cold days because of the use of electric space- and water-heating devices, and on hot days, because of air conditioning. This phenomenon can be observed in Figure 6 as the function that relates the bivariate temperature dataset to the load is clearly non-linear and rather U-shaped. For instance, Khotanzad et al. [45], Chow and Leung [10], Hippert et al. [41] and Elias et al. [32] used neural networks with input vector containing the time-lagged desired forecasting variable "Electric Load" and the time-lagged exogenous variables among which one can find (hourly or daily) lagged-temperatures.

In the above setting, the empirical CCTE at level α can be interpreted as an average power consumption knowing that temperatures are "far away" from the usual ones (e.g. considering low values of α). This can be interesting from electric utility companies point of view such as *EDF* for instance (which is the leading French-multinational-electric utility company). Indeed, if potential great heat

[1] https://www.rte-france.com/
[2] https://meteofrance.com/

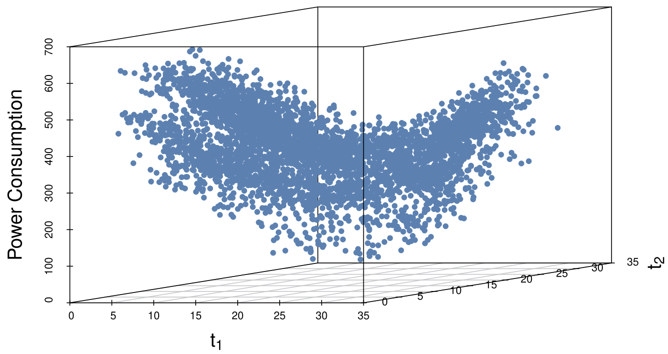

FIG 6. *Illustration of the dataset.*

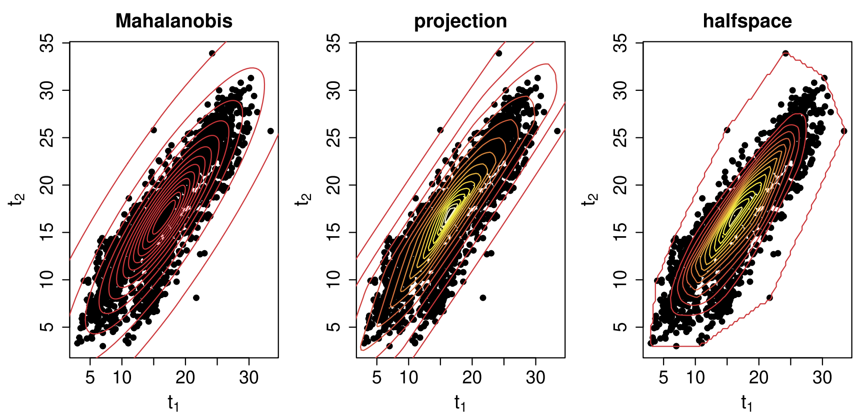

FIG 7. *Depths contours based on the sample of temperatures $(\boldsymbol{x}_\xi)_\xi$ for Mahalanobis, projection and halfspace depth resp.*

or/and cold waves are expected in the next decade, then *EDF* could forecast future average electricity investments for the considered period. In what follows, we illustrate the behavior of $\widehat{\mathrm{CCTE}}_\alpha$ as a function of α for the three different depths D presented in Figure 7 since they capture more or less the features of the risk factors dataset. Observe that low (resp. high) values of t_1 and t_2 corresponds to winter periods (resp. summer periods); the other configurations might correspond to periods in between.

In our analysis, we split the global dataset $\{(y_\xi, \boldsymbol{x}_\xi)_\xi\}$ into two samples of size $n_1 = 1800$ and $n_2 = 4377 - n_1$ to estimate the level set and the mean respectively in the definition of $\widehat{\mathrm{CCTE}}$. Note that in the first n_1-sample, we only need the values of the $(\boldsymbol{x}_\xi)'s$. However, in the second n_2-sample we removed the missing values of $(y_\xi)'s$ (i.e. NA's, there were 26 of them) but kept the associated $(\boldsymbol{x}_\xi)'s$ to add them to the first n_1-sample which is useful in the estimation of the level sets. So that now, $n_1 = 1826$ and $n_2 = 2551$. Observe that the dataset

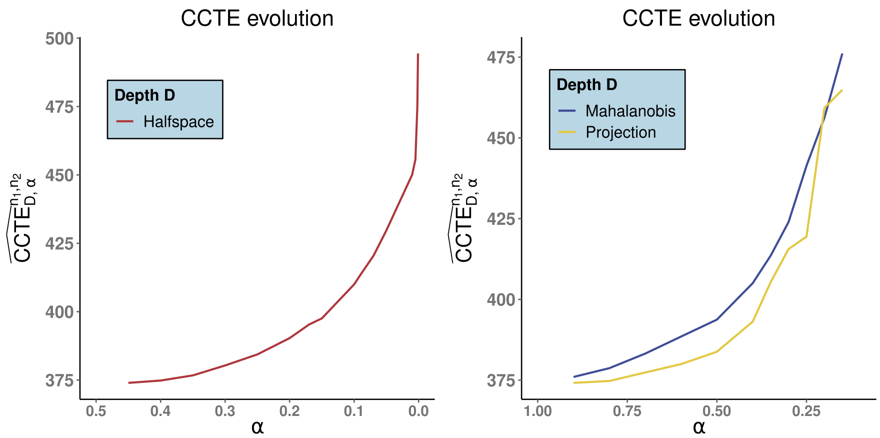

Fig 8. *Empirical* CCTE *behavior w.r.t.* α *(x-axis must be read from left to right according to the decreasing order of* α*).*

above exhibits a dependence structure as a process indexed by time, so that the assumption of i.i.d observations may be misleading. However, since we consider both of the previous split datasets as data clouds in $\mathbb{R}^2$ and $\mathbb{R}^3$ resp., it is not irrelevant to consider that they are realisations of i.i.d r.v. on the data cloud (particularly since we took as many seasons as within each year). As one can expect, $\widehat{\text{CCTE}}_{\alpha}^{n_1,n_2}$ increases as α decreases (Figure 8), which means when less central and hence "riskier" data are observed the cost is more and more significant. In other words, low values of α corresponding to all configurations of significant temperatures (t_1, t_2) (low at time H and low 6 hours before H, low at H and high 6 hours before H, etc..., see Figure 7) result naturally in higher power consumptions. Here, depths are grouped together according to the minimum achievable value of α. For the halfspace depth, the minimum level we could go for was $\alpha = 0.001$, while $\alpha = 0.14$ was the minimum level taken for Mahalanobis and projection depths. This may be explained by the fact that Mahalanobis and projection depths are based on mean, covariance and median calculations (resp.) so that one needs larger datasets to be able to reach possible lower values of α. Whereas halfspace depth is based on proportions (observed in halfspaces) which allow for very low values of α even if the sample size is not "too large".

5. Conclusion and perspectives

In this paper we provided convergence results for the plug-in estimator of the level sets of a given multivariate depth function in terms of the pseudo-metric under an unknown distribution. We derived a rate of convergence for Mahalanobis depth. In this setting we propose and estimate a new multivariate risk measure $\text{CCTE}_{D,\alpha}(\text{Y}, \mathbf{X})$. Comparing our CCTE with existing risk measures in terms of classical properties (monotonicity, translation invariance, homogeneity, subadditivity), behavior with respect to different risk scenarios is still an interesting and open question. Another interesting topic, which is in preparation, would be to derive possible rates of convergence when the level α is no more fixed and depends on the sample size with the aim of exploring extreme depth-regions, this, while relying on the broad extreme value literature.

6. Proofs

In this section, we denote by

$$\|f\|_p := \left(\int_{\mathbb{R}^d} |f(x)|^p dx\right)^{\frac{1}{p}} \quad \text{for } p < +\infty, \text{ and}$$
$$\|f\|_\infty := \text{ess sup}_{x\in\mathbb{R}^d}|f(x)| \quad \text{for } p = +\infty,$$

the $\mathbb{L}^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$ norm of f w.r.t the Lebesgue measure λ_d on $\mathbb{R}^d$. We recall that for A and B non-empty compact sets in $(\mathbb{R}^d, \|\cdot\|)$, the Hausdorff distance between A and B is defined by

$$d_H(A, B) = \sup\left(\sup_{a\in A} \text{dist}(a, B), \sup_{b\in B} \text{dist}(b, A)\right),$$

where $\text{dist}(x, A) := \inf_{a\in A} \|x - a\|$.

6.1. Preliminary results

The following result is a slight modification of Proposition 3.1 in the Ph.D. thesis of Rodríguez-Casal [62] adapted to depth functions, and an adapted version of Theorem 2 in Cuevas et al. [16] to depths in which we weaken the assumption of continuity of the empirical depth function by an assumption of upper-semicontinuity (u.s.c.). This is interesting since several existing depth functions are u.s.c but not continuous. For instance, for each distribution, the halfspace depth is u.s.c. but not continuous. The simplicial depth is u.s.c., and continuous only for absolutely continuous distributions, particularly, the empirical version of the simplicial depth is u.s.c. but not continuous. Theorem 6.1 is rather a technical result of which the proof can be found thereafter. We recall Assumption (**R**) as stated in Section 3.1.

Assumption (**R**). *Let $D : \mathbb{R}^d \times \mathscr{P} \to \mathbb{R}_+$ be a depth as in Definition 2.1. Fix $P \in \mathscr{P}$, $\alpha \in (0, \alpha_{\max}(P))$ and $0 < \varepsilon < \alpha$. Denoting $D(x) := D(x, P)$, we assume the following:*

(i) the function $x \mapsto D(x)$ is continuous on $\mathbb{R}^d$ and of class $\mathscr{C}^2$ on the set $\mathcal{K}_\varepsilon(\alpha) := D^{-1}([\alpha - \varepsilon, \alpha + \varepsilon])$,

(ii) $m_\nabla := m_\nabla(\alpha, \varepsilon, P) := \inf_{x\in\mathcal{K}_\varepsilon(\alpha)} \|(\nabla D)_x\| > 0$, where $(\nabla D)_x$ is the gradient of $D(\cdot)$ at x,

(iii) $\|D_n - D\|_{\infty,\mathbb{R}^d} \xrightarrow[n\to\infty]{\mathbb{P}\text{-}a.s.} 0$.

Theorem 6.1. *Suppose that Assumption (**R**) is satisfied. Then, it holds that*

$$d_H(\partial\mathcal{L}(\alpha), \partial\mathcal{L}_n(\alpha)) = \underset{n\to\infty}{O}(\|D_n - D\|_{\infty,\mathbb{R}^d}),\ \mathbb{P}\text{-a.s.}.$$

Remark 6.2. We should mention that in Theorem 6.1 only two main properties of a depth are sufficient for the previous result to hold, namely **(D2) upper semi-continuity** and **(D4) vanishing at infinity.**

Remark 6.3. On one hand, if $\alpha \in (0, \alpha_{\max}(P))$, then $\varnothing \subsetneq \mathcal{L}_D(\alpha) \subsetneq \mathbb{R}^d$, so that $\partial\mathcal{L}_D(\alpha)$ is non-empty. On the other hand, if the empirical depth D_n converges pointwise $\mathbb{P}$-a.s. (almost-surely) to D on $\mathbb{R}^d$ then (cf. Theorem 4.1 in Dyckerhoff [30]), for any $\alpha \in (0, \alpha_{\max}(P))$ and $\mathbb{P}$-a.s. for any n large enough, the upper-level set $\{x : D(x, P_n) \geq \alpha\}$ is non-empty as well as the set $\{x : D(x, P_n) < \alpha\}$. Thus, $\mathbb{P}$-a.s., for n large enough, $\partial\mathcal{L}_n(\alpha)$ is non-empty. Property **(D4) vanishing at infinity** guarantees that the upper level sets are bounded. As a consequence, the Hausdorff distance is well defined since $\partial\mathcal{L}_n(\alpha)$ and $\partial\mathcal{L}_D(\alpha)$ are closed by definition and bounded as they are included in compact sets ($\{x : D_n(x) \geq \alpha\}$ and $\{x : D(x) \geq \alpha\}$ resp.).

Most of the depth functions commonly encountered in the multivariate depth literature satisfy assumption (*iii*) of Theorem 6.1 (e.g. halfspace, simplicial, projection, Mahalanobis depths...). Continuity of the depth function at x is obtained in most cases (for instance the simplicial depth for absolutely continuous distributions as said before, the L^2-depth (Zuo & Serfling, 2000), and Mahalanobis depth). However, assumptions (*i*) and (*ii*) are not always satisfied by classical depths, but they are satisfied by Mahalanobis and L^2 depths for instance (see Example 2.3 in [70]). Since Mahalanobis depth is infinitely differentiable in its first argument, under some conditions it will satisfy all three assumptions of Theorem 6.1; this will be discussed hereafter (Proposition 6.6).

Remark 6.4. In the first part of the proof of Theorem 6.1, we show that $\tilde{\alpha} \mapsto \{D = \tilde{\alpha}\}$ is locally A-Lipschitz w.r.t the Hausdorff distance in a neighborhood of the level α, with $A = 2/m_\nabla$, $m_\nabla > 0$. This means assuming that $D(x)$ has a non-zero gradient around the α-level set. One scenario where zero gradient occurs is the depth level set $\{x : D(x) = \alpha\}$ such that the depth function $D(x)$ is flat at the level α. While studying strict monotonicity of depths, Dyckerhoff [30] provided examples of distributions for which the halfspace depth is flat around some level (see Supplement to [30]) but is not smooth however. Figure 9 below illustrates a special case where D_n converges pointwise from "above" to D with a plateau. Now under assumptions of Theorem 6.1, D is continuous on $\mathcal{K}_\varepsilon(\alpha)$ and $m_\nabla > 0$, there is no plateau in the graph of D for each level β such that $|\alpha - \beta| \leq \gamma$. This condition of non-zero gradient is commonly assumed in the literature, mostly for density level sets [5, 48, 54, 53, 51]. Notice further when the level α approaches zero, the constant A can be large so that the estimation of the level set becomes hard. In this case, we will need a "large" dataset to get a reasonable estimation.

Remark 6.5. Among the four properties of a depth function, property **(D4)** (vanishing at infinity) guarantees that the set $\mathcal{K}_\varepsilon(\alpha)$ is compact in $\mathbb{R}^d$. Indeed,

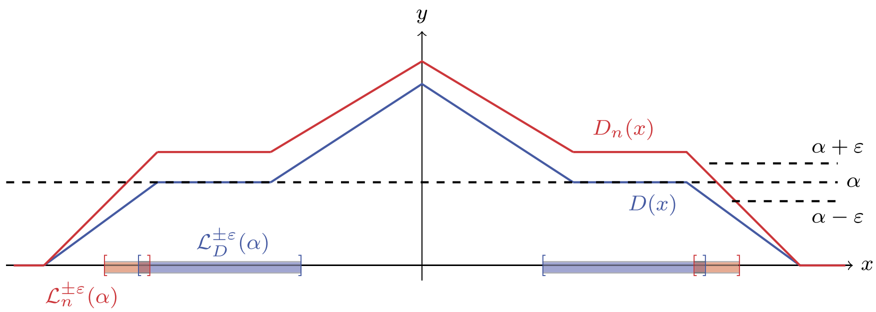

FIG 9. *Example of a flat univariate depth around the α-level set. Here, $\mathcal{L}_D^{\pm\varepsilon}(\alpha) = D^{-1}([\alpha\pm\varepsilon])$ and $\mathcal{L}_n^{\pm\varepsilon}(\alpha) = D_n^{-1}([\alpha\pm\varepsilon])$.*

as D satisfies **(D4)**, assumption $0 < \varepsilon < \alpha$ implies $\mathcal{K}_\varepsilon(\alpha)$ is bounded, moreover under assumption (i) of Theorem 6.1, D is continuous on $\mathcal{K}_\varepsilon(\alpha)$ so that $\mathcal{K}_\varepsilon(\alpha)$ is a closed set. By denoting $(HD)_x$ the Hessian matrix of D at x, one can note that $M_H := \sup_{x\in\mathcal{K}_\varepsilon(\alpha)} |||(HD)_x||| < \infty$, as a supremum of a continuous mapping on a compact set.

In Proposition 6.6 below, we provide the particular version of Theorem 6.1 associated to MHD depth.

Proposition 6.6. *Let $\boldsymbol{X} \in \mathbb{R}^d$ be a random vector from $P \in \mathscr{P}$ s.t. $\mathbb{E}_P[\| \boldsymbol{X} \|^2] < \infty$ and $\Sigma_{\boldsymbol{X}}$ invertible. It holds*

$$d_H(\partial\mathcal{L}_{MHD}(\alpha), \partial\mathcal{L}_n(\alpha)) = \underset{n\to\infty}{O}(\|MHD_n - MHD\|_{\infty,\mathbb{R}^d}),\ \mathbb{P}\text{-a.s.}.$$

Proof of Theorem 6.1. The proof of the theorem is divided into two main parts.

Part I. We show that under assumptions (i) and (ii) of Theorem 6.1, the following assumption

(**L**): $\exists\ A > 0,\ \exists\ \gamma > 0,\ \forall\ \beta > 0, |\alpha-\beta| \le \gamma \Rightarrow d_H(\{D=\alpha\}, \{D=\beta\}) \le A|\alpha-\beta|$,

is satisfied with $A = 2/m_\nabla$. It characterizes the locally Lipschitz behavior of the mapping $\tilde{\alpha} \mapsto \{D = \tilde{\alpha}\}$ w.r.t the Hausdorff distance in a neighborhood of the fixed level $\alpha > 0$.

Under assumptions of Theorem 6.1, $K_\varepsilon(\alpha) := D^{-1}([\alpha-\varepsilon, \alpha+\varepsilon])$ is compact and

$$M_H := \sup_{x\in K_\varepsilon(\alpha)} |||HD_x||| < \infty$$

(cf. Remark 6.5). We state the following useful lemma.

Lemma 6.7. *Under the assumptions of Theorem 6.1, there exist $N := N_\varepsilon \ge 1$, some points $x_i := x_{i,\varepsilon} \in \mathcal{K}_{\frac{\varepsilon}{2}}(\alpha)$, and some positive real numbers $r_i := r_{x_i} \in \mathbb{R}_+^*$, $1 \le i \le N$, s.t.*

$$\mathcal{K}_{\frac{\varepsilon}{2}}(\alpha) \subset \bigcup_{i=1}^N \mathrm{B}\left(x_i, \frac{r_i}{2}\right) \subset \bigcup_{i=1}^N \mathrm{B}(x_i, r_i) \subset \mathcal{K}_\varepsilon(\alpha).$$

Proof of Lemma 6.7. Since $\varepsilon > 0$, then $K_{\frac{\varepsilon}{2}}(\alpha)$ is a subset of the interior of $\mathcal{K}_\varepsilon(\alpha)$. Thus, for any $x \in K_{\frac{\varepsilon}{2}}(\alpha)$, there exists $r_x := r_x(\varepsilon) > 0$ s.t.

$$\mathrm{B}(x, r_x) \subset \mathcal{K}_\varepsilon(\alpha),$$

that is,

$$K_{\frac{\varepsilon}{2}}(\alpha) \subset \bigcup_{x \in K_{\frac{\varepsilon}{2}}(\alpha)} \mathrm{B}\left(x, \frac{r_x}{2}\right) \subset \bigcup_{x \in K_{\frac{\varepsilon}{2}}(\alpha)} \mathrm{B}(x, r_x) \subset \mathcal{K}_\varepsilon(\alpha).$$

Since we have an open cover of the compact set $K_{\frac{\varepsilon}{2}}(\alpha)$, then the latter has a finite sub-cover. In other words, there exist $N \geq 1$, some points $x_i := x_{i,\varepsilon} \in \mathcal{K}_{\frac{\varepsilon}{2}}(\alpha)$, and $r_i := r_{x_i} \in \mathbb{R}_+^*$, $1 \leq i \leq N$, s.t.

$$\mathcal{K}_{\frac{\varepsilon}{2}}(\alpha) \subset \bigcup_{i=1}^N \mathrm{B}\left(x_i, \tfrac{r_i}{2}\right) \subset \bigcup_{i=1}^N \mathrm{B}(x_i, r_i) \subset \mathcal{K}_\varepsilon(\alpha).$$

Hence the result. □

Let $0 < \gamma \leq \varepsilon/2$ and $x \in \mathcal{K}_\gamma(\alpha)$. For $\lambda \in \mathbb{R}$, define

$$y_\lambda := y_{\lambda,x} = x + \lambda \frac{(\nabla D)_x}{\|(\nabla D)_x\|},$$

with $\|(\nabla D)_x\| \geq m_\nabla > 0$, since $\mathcal{K}_\gamma(\alpha) \subset \mathcal{K}_\varepsilon(\alpha)$. In what follows, we take

$$\|y_\lambda - x\| = |\lambda| < \min_{1 \leq i \leq N} \frac{r_i}{2}.$$

It holds $[y_\lambda, x] \subset \mathcal{K}_\varepsilon(\alpha)$. Indeed, $x \in \mathcal{K}_\gamma(\alpha)$ so that Lemma 6.7 applies, namely, there exists $1 \leq i_0 \leq N$ s.t. $x \in \mathrm{B}(x_{i_0}, r_{i_0}/2)$, and for all $z \in [y_\lambda, x]$,

$$\begin{aligned}
\|z - x_{i_0}\| &\leq \|z - x\| + \|x - x_{i_0}\| \\
&\leq \|y_\lambda - x\| + \|x - x_{i_0}\| \\
&= |\lambda| + \|x - x_{i_0}\| \\
&< \min_{1 \leq i \leq N} \frac{r_i}{2} + \frac{r_{i_0}}{2} \\
&\leq r_{i_0}.
\end{aligned}$$

Thus, $z \in \mathrm{B}(x_{i_0}, r_{i_0}) \subset \mathcal{K}_\varepsilon(\alpha)$ (cf. Lemma 6.7). Since $|\lambda| < \min_{1 \leq i \leq N} r_i/2$, using a Taylor expansion on the line $[x, y_\lambda] \subset \mathcal{K}_\varepsilon(\alpha)$, it holds

$$D(y_\lambda) = D(x) + \langle (\nabla D)_x, y_\lambda - x \rangle + \tfrac{1}{2} \langle y_\lambda - x, (HD)_{\overline{x}}(y_\lambda - x) \rangle, \; \overline{x} \in [x, y_\lambda],$$

then,

$$D(y_\lambda) = D(x) + \lambda \|(\nabla D)_x\| + \frac{\lambda^2}{2\|(\nabla D)_x\|^2} \langle (\nabla D)_x, (HD)_{\overline{x}} (\nabla D)_x \rangle .$$

Using Cauchy-Schwarz inequality, it holds

$$|D(y_\lambda) - D(x) - \lambda\|(\nabla D)_x\|| \leq \frac{\lambda^2}{2\|(\nabla D)_x\|^2}\|(\nabla D)_x\| |\!|\!|(HD)_{\overline{x}}|\!|\!| \cdot \|(\nabla D)_x\| = \frac{\lambda^2}{2}|\!|\!|(HD)_{\overline{x}}|\!|\!|.$$

Since $\overline{x} \in \mathcal{K}_\varepsilon(\alpha)$, then $|\!|\!|(HD)_{\overline{x}}|\!|\!| \leq \sup_{x \in \mathcal{K}_\varepsilon(\alpha)} |\!|\!|(HD)_{\overline{x}}|\!|\!| = M_H < \infty$. For any $|\lambda| < \min_{1\leq i \leq N} r_i/2$, we obtain

$$D(x) + \lambda\|(\nabla D)_x\| - \frac{\lambda^2}{2}M_H \leq D(y_\lambda) \leq D(x) + \lambda\|(\nabla D)_x\| + \frac{\lambda^2}{2}M_H. \tag{6.1}$$

If $0 < \lambda < \min_{1\leq i \leq N} r_i/2$, then with the above inequality, we have

$$D(y_\lambda) \geq D(x) + \lambda \inf_{x \in \mathcal{K}_\varepsilon(\alpha)} \|(\nabla D)_x\| - \frac{\lambda^2}{2}M_H = D(x) + \lambda\left(m_\nabla - \lambda\frac{M_H}{2}\right) \tag{6.2}$$

Suppose now $M_H > 0$ (the case $M_H = 0$ is trivial). That way, if $0 < \lambda < \frac{m_\nabla}{M_H} \wedge \min_{1\leq i \leq N} r_i/2$, using (6.2),

$$D(y_\lambda) \geq D(x) + \lambda\tfrac{m_\nabla}{2}.$$

Similarly, using the right hand side of inequality (6.1), for any $0 < \lambda < \frac{m_\nabla}{M_H} \wedge \min_{1\leq i \leq N} \frac{r_i}{2}$,

$$D(y_{-\lambda}) \leq D(x) - \lambda\frac{m_\nabla}{2}.$$

To sum up, for any $0 < \gamma \leq \varepsilon/2$, $x \in \mathcal{K}_\gamma(\alpha)$ and $0 < \lambda < \frac{m_\nabla}{M_H} \wedge \min_{1\leq i \leq N} \frac{r_i}{2}$, it holds

$$D(y_\lambda) \geq D(x) + \lambda\frac{m_\nabla}{2}, \tag{6.3}$$

$$D(y_{-\lambda}) \leq D(x) - \lambda\frac{m_\nabla}{2}. \tag{6.4}$$

Choose $\gamma := \left[\frac{m_\nabla}{4}\left(\frac{m_\nabla}{M_H} \wedge \min_{1\leq i \leq N} \frac{r_i}{2}\right)\right] \wedge \frac{\varepsilon}{2} > 0$. Now we show:

$$\text{if } |\alpha - \beta| \leq \gamma, \text{ then } d_H(\{D = \alpha\}, \{D = \beta\}) \leq \tfrac{2}{m_\nabla}|\alpha - \beta|.$$

Assume $|\alpha - \beta| \leq \gamma$.

Let β be s.t. $0 < \beta - \alpha \leq \gamma$. In this case, $\beta = \alpha + \eta$ with $0 < \eta \leq \gamma$. First, we have to find an upper bound for $\sup_{x \in \{D = \beta\}} \mathrm{dist}(x, \{D = \alpha\})$. Let $x \in \{D = \beta\}$, i.e. $D(x) = \beta = \alpha + \eta$. Since $0 < \eta \leq \gamma$, $0 < D(x) - \alpha \leq \gamma$, i.e. $x \in \mathcal{K}_\gamma(\alpha)$. Choose $\lambda := \frac{2\eta}{m_\nabla} \in \left(0, \frac{m_\nabla}{M_H} \wedge \min_{1\leq i \leq N} \frac{r_i}{2}\right)$ so that with (6.4),

$$D(y_{-\lambda}) \leq D(x) - \lambda\frac{m_\nabla}{2} = D(x) - \eta = \alpha < D(x).$$

From the above inequality and the continuity property of $z \mapsto D(z)$ on $\mathcal{K}_\varepsilon(\alpha) \supset [y_{-\lambda}, x]$, there exists a point $y \in [y_{-\lambda}, x]$ s.t. $D(y) = \alpha$. Moreover,

$$\|x - y\| \leq \|x - y_{-\lambda}\| = |\lambda| = \frac{2\eta}{m_\nabla} = \frac{2}{m_\nabla}(\beta - \alpha).$$

As a consequence, for all $x \in \{D = \beta\}$,

$$\operatorname{dist}(x, \{D = \alpha\}) \leq \|x - y\| \leq \frac{2}{m_\nabla}|\beta - \alpha|.$$

So

$$\sup_{x \in \{D=\beta\}} \operatorname{dist}(x, \{D = \alpha\}) \leq \frac{2}{m_\nabla}|\beta - \alpha|.$$

In order to get an upper bound for $\sup_{x\in\{D=\alpha\}} \operatorname{dist}(x, \{D = \beta\})$, we use the inequality (6.3) by proceeding in a similar way.

The proof in the case $0 > \beta - \alpha > -\gamma$ is completely analogous.

Part II. In this part, we show that the Hausdorff distance between the respective boundaries of $\mathcal{L}_n(\alpha)$ and $\mathcal{L}_D(\alpha)$ is $O(\|D_n - D\|_\infty)$ up to a constant depending on A.

Let $\alpha \in (0, \alpha_{\max}(P))$.

Step 1 : we need to find an upper bound for $\sup_{x\in\partial\mathcal{L}(\alpha)} d(x, \partial\mathcal{L}_n(\alpha))$.

Let $x \in \partial\mathcal{L}(\alpha)$. Denote $\varepsilon_n = 2\|D_n - D\|_\infty$. Under the assumptions of Theorem 6.1 $\varepsilon_n \to 0$ $\mathbb{P}$-a.s., so that $\mathbb{P}$-a.s., there exists an integer $n_0 := n_0(\omega) \geq 1$ (independent of x), s.t. for all $n \geq n_0$, $\varepsilon_n \leq \gamma$. Taking $\beta = \alpha + \varepsilon_n$, it holds

$$\mathbb{P}\text{-a.s., for all } n \geq n_0,\ d_H(\partial\mathcal{L}(\alpha + \varepsilon_n), \partial\mathcal{L}(\alpha)) \leq A\varepsilon_n.$$

Thus, from the above inequality and using the continuity property of D, $\mathbb{P}$-a.s., for all $n \geq n_0$, there exists $u_n := u_{x,\varepsilon_n} \in \partial\mathcal{L}(\alpha + \varepsilon_n)$ i.e. $D(u_n) = \alpha + \varepsilon_n$, and $l_n := l_{x,\varepsilon_n} \in \partial\mathcal{L}(\alpha - \varepsilon_n)$ i.e. $D(l_n) = \alpha - \varepsilon_n$, s.t.

$$\|u_n - x\| \leq A\varepsilon_n \text{ and } \|l_n - x\| \leq A\varepsilon_n.$$

Let us assume $\|D_n - D\|_\infty > 0$ (the case $\|D_n - D\|_\infty = 0$ is trivial). In this case,

$$D_n(u_n) = D_n(u_n) + \alpha + \varepsilon_n - D(u_n) \geq \alpha + \varepsilon_n - \|D_n - D\|_\infty = \alpha + \|D_n - D\|_\infty > \alpha.$$

Similarly, we have $D_n(l_n) < \alpha$. So $\mathbb{P}$-a.s., for all $n \geq n_0$,

$$D_n(l_n) < \alpha < D_n(u_n).$$

For the sake of simplicity, we denote here $\mathcal{L}_n := \{x : D_n(x) \leq \alpha\}$. Then, almost surely, for all $n \geq n_0$, $\mathcal{L}_n$ is non-empty (since it contains l_n). And by definition, $l_n \in \mathcal{L}_n \subset \overline{\mathcal{L}_n}$. Denoting by $\mathcal{L}_n^c$ the complementary of $\mathcal{L}_n$ in $\mathbb{R}^d$, it holds $u_n \in \mathcal{L}_n^c \subset \overline{\mathcal{L}_n^c} = (\mathring{\mathcal{L}}_n)^c$, that is, $u_n \notin \mathring{\mathcal{L}}_n$. Then,

$$\mathbb{P}\text{-a.s., for all } n \geq n_0, \text{ there exists } z_n \in [l_n, u_n] \cap \partial\mathcal{L}_n.$$

Thus, $\mathbb{P}$-a.s., for all $n \geq n_0$,

$$\begin{aligned}
\operatorname{dist}(x, \partial\mathcal{L}_n(\alpha)) &\leq \|x - z_n\| \\
&\leq \|x - u_n\| + \|u_n - z_n\| \\
&\leq \|u_n - x\| + \|u_n - l_n\|
\end{aligned}$$

$$\leq \|u_n - x\| + \|u_n - x\| + \|x - l_n\|$$
$$\leq 3A\varepsilon_n$$
$$= 6A\|D_n - D\|_\infty.$$

Since the previous inequality holds for all $x \in \partial\mathcal{L}(\alpha)$, we have, $\mathbb{P}$-a.s., for all $n \geq n_0$,

$$\sup_{x\in\partial\mathcal{L}(\alpha)} d(x, \partial\mathcal{L}_n(\alpha)) \leq 6A\|D_n - D\|_\infty.$$

Step 2 : Let us find an upper bound for $\sup_{x\in\partial\mathcal{L}_n(\alpha)} d(x, \partial\mathcal{L}(\alpha))$.

Let $x_n \in \partial\mathcal{L}_n(\alpha) := \partial\mathcal{L}_n = \overline{\{D_n \leq \alpha\}} \cap \overline{\{D_n > \alpha\}} \subset \overline{\{D_n \geq \alpha\}} = \{D_n \geq \alpha\}$, since D_n is a.s. upper-semicontinuous, so that the upper level set based on D_n is closed. Then, $D_n(x_n) \geq \alpha$. Furthermore, since $x_n \in \overline{\{D_n \leq \alpha\}}$, there exists ℓ_n "close" enough to x_n s.t. $D_n(\ell_n) \leq \alpha$, and s.t. by continuity of D, $|D(x_n) - D(\ell_n)| \leq \varepsilon_n/2$. On the one hand,

$$D(x_n) = D_n(x_n) - D_n(x_n) + D(x_n) \geq \alpha - \varepsilon_n/2 \geq \alpha - \varepsilon_n,$$

on the other hand,

$$D(x_n) = D_n(\ell_n) - D_n(\ell_n) + D(\ell_n) - D(\ell_n) + D(x_n)$$
$$\leq \alpha + \varepsilon_n/2 + \varepsilon_n/2,$$

so,

$$|D(x_n) - \alpha| \leq \varepsilon_n.$$

Recall that, a.s. for all $n \geq n_0$, $\varepsilon_n \leq \gamma$. Then, using property **(L)** with $\beta = D(x_n)$, we can write

$$\mathrm{dist}(x_n, \partial\mathcal{L}(\alpha)) \leq d_H(\partial\mathcal{L}(D(x_n)), \partial\mathcal{L}(\alpha)) \leq A|D(x_n) - \alpha|$$
$$\leq 2A\|D_n - D\|_\infty.$$

Now we deduce that, a.s. for n large enough,

$$\sup_{x\in\partial\mathcal{L}_n(\alpha)} d(x, \partial\mathcal{L}(\alpha)) \leq 2A\|D_n - D\|_\infty.$$

Hence the result. $\square$

*Proof of Proposition 6.6. **(i)*** The function $MHD(\cdot)$ is infinitely differentiable on $\mathbb{R}^d$, and denoting $\mu = \mu_{\mathbf{X}}$, we can write for any $1 \leq k \leq d$,

$$\frac{\partial MHD(x)}{\partial x_k} = -MHD(x)^2 \frac{\partial}{\partial x_k}\left[\sum_{i,j=1}^{d}(x_i - \mu_i)(\Sigma_{\mathbf{X}}^{-1})_{ij}(x_j - \mu_j)\right]$$
$$= -MHD(x)^2 \cdot 2\left[\sum_{i=1}^{d}(\Sigma_{\mathbf{X}}^{-1})_{ki}(x_i - \mu_i)\right],\ (\Sigma_{\mathbf{X}}^{-1} \text{ is symmetric})$$

$$= -2MHD(x)^2 \left[\Sigma_{\mathbf{X}}^{-1}(x-\mu) \right]_k .$$

So

$$(\nabla MHD)_x = -2MHD(x)^2 \Sigma_{\mathbf{X}}^{-1}(x-\mu).$$

Since $MHD(x) > 0$,

$$(\nabla MHD)_x = 0 \Longleftrightarrow x = \mu = \mu_{\mathbf{X}}.$$

Thus,

$$\|(\nabla MHD)_x\| > 0, \text{ for all } x \neq \mu_{\mathbf{X}}.$$

Now since $x \in \mathbb{R}^d \mapsto \|(\nabla MHD)_x\|$ is continuous and $\mathcal{K}_\varepsilon(\alpha)$ is compact, then there exists $x_0 \in \mathcal{K}_\varepsilon(\alpha)$ in which the infimum m_∇ is attained,

$$m_\nabla = \|(\nabla MHD)_{x_0}\| > 0.$$

The latter inequality is strict since $x_0 \in \mathcal{K}_\varepsilon(\alpha)$, and $\mu_{\mathbf{X}} \notin \mathcal{K}_\varepsilon(\alpha)$ (from the assumption $\varepsilon < 1-\alpha$). Indeed,

$$\mu_{\mathbf{X}} \in \mathcal{K}_\varepsilon(\alpha) \Leftrightarrow |MHD(\mu_{\mathbf{X}}) - \alpha| \leq \varepsilon \Leftrightarrow |1-\alpha| = 1-\alpha \leq \varepsilon.$$

(ii) It remains to prove

$$\|MHD_n - MHD\|_{\infty,\mathbb{R}^d} \xrightarrow[n\to\infty]{\mathbb{P}\text{-a.s.}} 0, \tag{6.5}$$

by recalling that $\alpha_{\max}(P) = 1$ for MHD depth. The result is hence a straight forward application of Theorem 6.1. In order to prove (6.5), one can refer to the computations in the proof of Theorem 3.9 and obtain the desired result knowing that $\hat{\Sigma}_n \xrightarrow[n\to\infty]{\mathbb{P}\text{-a.s.}} \Sigma$, and $\hat{\mu}_n \xrightarrow[n\to\infty]{\mathbb{P}\text{-a.s.}} \mu$. □

6.2. Proofs of Section 3

Proofs of Section 3.1.

Proof of Theorem 3.1. Denote $D(\cdot) := D(\cdot, P)$, where $P \in \mathscr{P}$ is fixed. We want to find an upper bound for $\lambda_d(\mathcal{L}_D(\alpha)\Delta\mathcal{L}_{n_1}(\alpha))$. We introduce:

$$\ell_{n_1} = \ell_{n_1}(\alpha) := d_H(\partial\mathcal{L}_{n_1}(\alpha), \partial\mathcal{L}_D(\alpha)),$$

and the tube around the boundary $\partial\mathcal{L}_D(\alpha)$ of radius ℓ_{n_1} defined by

$$\text{Tube}(\partial\mathcal{L}_D(\alpha), \ell_{n_1}) := \left\{ z \in \mathbb{R}^d : \text{dist}(z, \partial\mathcal{L}_D(\alpha)) \leq \ell_{n_1} \right\}.$$

Since $\partial\mathcal{L}_D(\alpha)$ is a compact submanifold of dimension $d-1$, and ℓ_{n_1} is small enough $\mathbb{P}$-a.s. for large n_1, one can make a first order approximation of the volume of the spherical tube around $\partial\mathcal{L}_D(\alpha)$ of radius ℓ_{n_1} using Weyl's formula (cf. Weyl [69], p. 461):

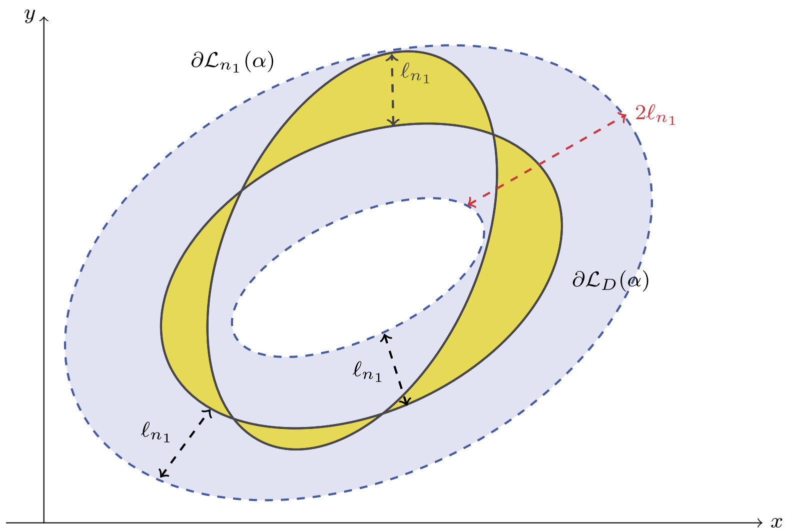

FIG 10. *Illustration of* $\lambda_d(\mathcal{L}_{n_1}(\alpha)\Delta\mathcal{L}_D(\alpha))$ *(yellow) and the tube around* $\mathcal{L}_D(\alpha)$ *of radius* ℓ_{n_1} *(blue), with* $d = 2$.

$$\lambda_d \left[\text{Tube}(\partial\mathcal{L}_D(\alpha), \ell_{n_1})\right] \approx \Omega_1 \ell_{n_1} \cdot k_0(\alpha) \ \mathbb{P}\text{-a.s.},$$

with $\Omega_1 = 2$ being the length of the unit interval $\{t \in \mathbb{R} : |t| \leq 1\}$ and $k_0(\alpha)$ the area of the "surface" $\partial\mathcal{L}_D(\alpha)$. So, $\mathbb{P}$-a.s. for n_1 large enough,

$$\begin{aligned}
\lambda_d(\mathcal{L}_{n_1}(\alpha)\Delta\mathcal{L}_D(\alpha)) &\leq \lambda_d \left[\text{Tube}(\partial\mathcal{L}_D(\alpha), \ell_{n_1})\right] \\
&\approx \Omega_1 \ell_{n_1} \cdot k_0(\alpha) \\
&:= K_d(\alpha)\ell_{n_1} \\
&\leq K_d(\alpha)\|D_{n_1} - D\|_{\infty,\mathbb{R}^d},
\end{aligned}$$

with $K_d(\alpha)$ a constant that may change from a line to another. We specify here that the "$\approx$" sign above means

$$a_n \approx b_n \overset{def}{\Longleftrightarrow} a_n = b_n + \varepsilon_n, \ \varepsilon_n \xrightarrow[n\to\infty]{} 0.$$

In the above, the last inequality is obtained under assumptions of Theorem 6.1. □

Proof of Corollary 3.2. Denote $D(\cdot) := D(\cdot, P)$, where P is absolutely continuous with p-integrable density function f w.r.t the Lebesgue measure on $\mathbb{R}^d$. Under the assumptions of Theorem 3.2, when $p \in (1, +\infty)$, it holds $\mathbb{P}$-almost-surely

$$\begin{aligned}
P_{\tilde{S}_{n_1}}\left(\mathcal{L}_{n_1}(\alpha)\Delta\mathcal{L}_D(\alpha)\right) &= \int \mathbb{1}_{\boldsymbol{x}\in\mathcal{L}_{n_1}(\alpha)\Delta\mathcal{L}_D(\alpha)} f(x)dx \\
&\leq \lambda_d(\mathcal{L}_D(\alpha)\Delta\mathcal{L}_{n_1}(\alpha))^{1-\frac{1}{p}}\|f\|_p \quad \text{(Hölder)}
\end{aligned}$$

$$\leq \lambda_d(\mathcal{L}_D(\alpha)\Delta\mathcal{L}_{n_1}(\alpha))^{1-\frac{1}{p}} \underbrace{\|f\|_p}_{<\infty}.$$

When $p = +\infty$, the above inequality is trivially valid by bounding f by its essential supremum. The proof now follows from Theorem 3.1. □

Proofs of Section 3.2.

Proofs of Section 3.2.1. We first give the proof of Theorem 3.5. Theorem 3.4 will be proved following the guidelines of the proof of Teorem 3.5.

Proof of Theorem 3.5. We can write

$$\begin{aligned}
&\left|\widehat{\mathrm{CCTE}}^{n_1,n_2}_{D,\alpha}(Y,\mathbf{X}) - \mathrm{CCTE}_{D,\alpha}(Y,\mathbf{X})\right| \cdot \mathbb{1}_{P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha))>0} \\
&= \left|\frac{\frac{1}{n_2}\sum_{i=1}^{n_2} Y_i \mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)}}{\frac{1}{n_2}\sum_{i=1}^{n_2} \mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)}} - \mathbb{E}[Y|\,\mathbf{X}\in\mathcal{L}_D(\alpha)]\right| \cdot \mathbb{1}_{P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha))>0} \\
&\leq \left|\frac{\frac{1}{n_2}\sum_{i=1}^{n_2} Y_i \mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)}}{\frac{1}{n_2}\sum_{i=1}^{n_2} \mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)}} - \mathbb{E}_{\tilde{S}_{n_1}}[Y|\,\mathbf{X}\in\mathcal{L}_{n_1}(\alpha)]\right| \cdot \mathbb{1}_{P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha))>0} \\
&\quad + \left|\mathbb{E}_{\tilde{S}_{n_1}}[Y|\,\mathbf{X}\in\mathcal{L}_{n_1}(\alpha)] - \mathbb{E}[Y|\,\mathbf{X}\in\mathcal{L}_D(\alpha)]\right| \cdot \mathbb{1}_{P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha))>0}.
\end{aligned}$$

The proof of Theorem 3.5 is a modified version of the proof of Theorem 5.1 in Di Bernardino et al. [24]. The latter focuses on distribution functions instead of depth functions. Besides, in the proof of Theorem 3.5, we show that $\mathbb{1}_{P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha))>0}$ converges to one in probability.

First recall that, in the following, probability measures involving events which depend on $\mathcal{L}_{n_1}(\alpha)$ are conditional expectations to the sample $\tilde{S}_{n_1}$. For notational convenience, we recall that

$$P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha)) = \mathbb{P}(\mathbf{X}\in\mathcal{L}_{n_1}(\alpha))$$

which is a random variable. Moreover, note that the convergence to zero in probability implies directly the $\mathcal{O}_P(1)$ result.

The proof of Theorem 3.5 is a direct consequence of Lemma 6.8 and Lemma 6.9 below. □

Lemma 6.8. *Under assumptions of Theorem 3.5, it holds that*

$$\left|\mathbb{E}_{\tilde{S}_{n_1}}[Y|\,\boldsymbol{X}\in\mathcal{L}_{n_1}(\alpha)] - \mathbb{E}[Y|\,\boldsymbol{X}\in\mathcal{L}_D(\alpha)]\right| = \mathcal{O}_{P,n_1}\left(v_{n_1}^{-(1-\frac{1}{r})}\right).$$

Proof of Lemma 6.8. On the event $\left\{P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha))>0\right\}$, it holds

$$v_{n_1}^{1-\frac{1}{r}}\left|\mathbb{E}_{\tilde{S}_{n_1}}[Y|\,\mathbf{X}\in\mathcal{L}_{n_1}(\alpha)] - \mathbb{E}[Y|\,\mathbf{X}\in\mathcal{L}_D(\alpha)]\right|$$

$$= v_{n_1}^{1-\frac{1}{r}} \left| \frac{\mathbb{E}_{\tilde{S}_{n_1}}[Y \mathbb{1}_{\mathbf{X} \in \mathcal{L}_{n_1}(\alpha)}]}{P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha))} - \frac{\mathbb{E}[Y \mathbb{1}_{\mathbf{X} \in \mathcal{L}_D(\alpha)}]}{P(\mathcal{L}_D(\alpha))} \right|$$

$$= \frac{v_{n_1}^{1-\frac{1}{r}}}{P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha)) P(\mathcal{L}_D(\alpha))} \times \Big| P(\mathcal{L}_D(\alpha)) \mathbb{E}_{\tilde{S}_{n_1}}[Y \mathbb{1}_{\mathbf{X} \in \mathcal{L}_{n_1}(\alpha)}] - P(\mathcal{L}_D(\alpha)) \mathbb{E}[Y \mathbb{1}_{\mathbf{X} \in \mathcal{L}_D(\alpha)}] + P(\mathcal{L}_D(\alpha)) \mathbb{E}[Y \mathbb{1}_{\mathbf{X} \in \mathcal{L}_D(\alpha)}] - P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha)) \mathbb{E}[Y \mathbb{1}_{\mathbf{X} \in \mathcal{L}_D(\alpha)}] \Big|$$

$$\leq \frac{v_{n_1}^{1-\frac{1}{r}}}{P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha)) P(\mathcal{L}_D(\alpha))} \times \Big(P(\mathcal{L}_D(\alpha)) \Big| \mathbb{E}_{\tilde{S}_{n_1}}[Y \mathbb{1}_{\mathbf{X} \in \mathcal{L}_{n_1}(\alpha)}] - \mathbb{E}[Y \mathbb{1}_{\mathbf{X} \in \mathcal{L}_D(\alpha)}] \Big| + \Big| P(\mathcal{L}_D(\alpha)) - P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha)) \Big| \left| \mathbb{E}[Y \mathbb{1}_{\mathbf{X} \in \mathcal{L}_D(\alpha)}] \right| \Big)$$

$$\leq \frac{v_{n_1}^{1-\frac{1}{r}}}{P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha)) P(\mathcal{L}_D(\alpha))} \times \Big(\Big| \mathbb{E}_{\tilde{S}_{n_1}}[Y \mathbb{1}_{\mathbf{X} \in \mathcal{L}_{n_1}(\alpha)}] - \mathbb{E}[Y \mathbb{1}_{\mathbf{X} \in \mathcal{L}_D(\alpha)}] \Big| + \mathbb{E}[|Y|] \Big| P(\mathcal{L}_D(\alpha)) - P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha)) \Big| \Big).$$

Under Assumption (**H0**) and since $v_{n_1}^{-1} \to 0$ as $n_1 \to \infty$, it holds that

$$P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha) \Delta \mathcal{L}_D(\alpha)) \xrightarrow[n_1 \to \infty]{\mathbb{P}} 0,$$

so that

$$P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha)) \xrightarrow[n_1 \to \infty]{\mathbb{P}} P(\mathcal{L}_D(\alpha)) > 0,$$

and

$$\mathbb{P}\left(\left\{ P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha)) > 0 \right\} \right) \xrightarrow[n_1 \to \infty]{} 1. \tag{6.6}$$

On the one hand,

$$v_{n_1}^{1-\frac{1}{r}} \left| P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha)) - P(\mathcal{L}_D(\alpha)) \right| \leq v_{n_1}^{1-\frac{1}{r}} P_{\tilde{S}_{n_1}}(\mathcal{L}_D(\alpha) \Delta \mathcal{L}_{n_1}(\alpha)),$$

so we obtain

$$v_{n_1}^{1-\frac{1}{r}} \left| P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha)) - P(\mathcal{L}_D(\alpha)) \right| \xrightarrow[n_1 \to \infty]{\mathbb{P}} 0. \tag{6.7}$$

On the other hand, using Hölder inequality

$$\begin{aligned} v_{n_1}^{1-\frac{1}{r}} \left| \mathbb{E}_{\tilde{S}_{n_1}}[Y \mathbb{1}_{\mathbf{X} \in \mathcal{L}_n(\alpha)}] - \mathbb{E}[Y \mathbb{1}_{\mathbf{X} \in \mathcal{L}_D(\alpha)}] \right| & \\ \leq & v_{n_1}^{1-\frac{1}{r}} \mathbb{E}_{\tilde{S}_{n_1}}\left[|Y| \mathbb{1}_{\mathbf{X} \in \mathcal{L}_{n_1}(\alpha) \Delta \mathcal{L}_D(\alpha)} \right] \\ \leq & v_{n_1}^{1-\frac{1}{r}} \mathbb{E}\left[|Y|^r \right]^{\frac{1}{r}} \mathbb{E}_{\tilde{S}_{n_1}}\left[\mathbb{1}_{\mathbf{X} \in \mathcal{L}_{n_1}(\alpha) \Delta \mathcal{L}_D(\alpha)} \right]^{1-\frac{1}{r}} \end{aligned}$$

$$=v_{n_1}^{1-\frac{1}{r}} P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha)\Delta\mathcal{L}_D(\alpha))^{1-\frac{1}{r}}\|Y\|_{\mathbb{L}^r(\Omega)}.$$

Since $v_{n_1} P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha)\Delta\mathcal{L}_D(\alpha)) = \mathcal{O}_{P,n_1}(1)$, it holds

$$v_{n_1}^{1-\frac{1}{r}}\left|\mathbb{E}_{\tilde{S}_{n_1}}[Y\mathbb{1}_{\mathcal{L}_{n_1}(\alpha)}] - \mathbb{E}[Y\mathbb{1}_{\mathcal{L}_D(\alpha)}]\right| = \mathcal{O}_{P,n_1}(1). \tag{6.8}$$

Since the convergence to zero in probability implies the $\mathcal{O}_P(1)$ result, the lemma follows directly from (6.6), (6.7), and (6.8). The case $r = +\infty$ is analogous. □

Lemma 6.9. *Under assumptions of Theorem 3.5, we obtain*

$$\left|\frac{\frac{1}{n_2}\sum_{i=1}^{n_2} Y_i\mathbb{1}_{\boldsymbol{X}_i\in\mathcal{L}_{n_1}(\alpha)}}{\frac{1}{n_2}\sum_{i=1}^{n_2}\mathbb{1}_{\boldsymbol{X}_i\in\mathcal{L}_{n_1}(\alpha)}} - \mathbb{E}[Y|\,\boldsymbol{X}\in\mathcal{L}_{n_1}(\alpha)]\right| = \mathcal{O}_{P,n_1,n_2}(n_2^{-\frac{1}{2}}).$$

Proof of Lemma 6.9. First, we distinguish the event in which $\frac{1}{n_2}\sum_{i=1}^{n_2}\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)} = 0$, then the one in which $\frac{1}{n_2}\sum_{i=1}^{n_2}\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)} \neq 0$.

Let $0 < \varepsilon < P[\mathcal{L}_D(\alpha)]$. Since the $(\mathrm{Y}_i, \mathbf{X}_i)_{1\leq i\leq n_2}$ are iid so that the $(\mathbf{X}_i)_{1\leq i\leq n_2}$ are iid, it holds that

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{n_2}\sum_{i=1}^{n_2}\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)} = 0\right)\\
&= \mathbb{E}\left[P_{\tilde{S}_{n_1}}\left(\frac{1}{n_2}\sum_{i=1}^{n_2}\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)} = 0\right)\right]\\
&= \mathbb{E}\left[\prod_{i=1}^{n_2} P_{\tilde{S}_{n_1}}(\mathbf{X}_i\notin\mathcal{L}_{n_1}(\alpha))\right]\\
&= \mathbb{E}\left[P_{\tilde{S}_{n_1}}(\mathbf{X}\notin\mathcal{L}_{n_1}(\alpha))^{n_2}\right]\\
&= \mathbb{E}\left[(1-P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha)))^{n_2}\mathbb{1}_{P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha))\geq\varepsilon}\right]\\
&\qquad + \mathbb{E}\left[(1-P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha)))^{n_2}\mathbb{1}_{P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha))<\varepsilon}\right]\\
&\leq (1-\varepsilon)^{n_2} + \mathbb{P}\left(P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha)) < \varepsilon\right).
\end{aligned}$$

Since $\varepsilon \in (0, P(\mathcal{L}_D(\alpha)))$ and $P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha)) \xrightarrow[n_1\to\infty]{\mathbb{P}} P(\mathcal{L}_D(\alpha))$ (see Lemma 6.8), we obtain $\mathbb{P}\left(\frac{1}{n_2}\sum_{i=1}^{n_2}\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)} = 0\right) \xrightarrow[n_1,n_2\to\infty]{} 0$. Now on the event $\left\{\frac{1}{n_2}\sum_{i=1}^{n_2}\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)} \neq 0\right\} \cap \left\{P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha)) > 0\right\}$, we can write

$$\begin{aligned}
&\left|\frac{\frac{1}{n_2}\sum_{i=1}^{n_2} Y_i\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)}}{\frac{1}{n_2}\sum_{i=1}^{n_2}\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)}} - \mathbb{E}_{\tilde{S}_{n_1}}[Y|\,\mathbf{X}\in\mathcal{L}_{n_1}(\alpha)]\right|\\
&= \left|\frac{\sum_{i=1}^{n_2} Y_i\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)}}{\sum_{i=1}^{n_2}\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)}} - \frac{\mathbb{E}_{\tilde{S}_{n_1}}[Y\mathbb{1}_{\mathbf{X}\in\mathcal{L}_{n_1}(\alpha)}]}{P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha))}\right|
\end{aligned}$$

$$
\begin{aligned}
\le & \frac{1}{\frac{1}{n_2}\sum_{i=1}^{n_2}\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)}}\left|\frac{1}{n_2}\sum_{i=1}^{n_2}Y_i\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)}-\mathbb{E}_{\tilde{S}_{n_1}}[Y\mathbb{1}_{\mathbf{X}\in\mathcal{L}_{n_1}(\alpha)}]\right| \\
& +\left|\mathbb{E}_{\tilde{S}_{n_1}}[Y\mathbb{1}_{\mathbf{X}\in\mathcal{L}_{n_1}(\alpha)}]\right|\left|\frac{1}{\frac{1}{n_2}\sum_{i=1}^{n_2}\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)}}-\frac{1}{P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha))}\right| \\
\le & \frac{1}{\frac{1}{n_2}\sum_{i=1}^{n_2}\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)}}\left|\frac{1}{n_2}\sum_{i=1}^{n_2}Y_i\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)}-\mathbb{E}_{\tilde{S}_{n_1}}[Y\mathbb{1}_{\mathbf{X}\in\mathcal{L}_{n_1}(\alpha)}]\right| \\
& +\frac{\mathbb{E}[|Y|]}{P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha))\frac{1}{n_2}\sum_{i=1}^{n_2}\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)}}\left|\frac{1}{n_2}\sum_{i=1}^{n_2}\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)}-P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha))\right|.
\end{aligned}
\tag{$\mathbf{R}$}
$$

Let us first clarify the convergence of the denominator terms. Recall that under Assumption (**H0**) of Theorem 3.5,

$$
P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha)) \xrightarrow[n_1\to\infty]{\mathbb{P}} P(\mathcal{L}_D(\alpha))>0,
$$

and $\mathbb{P}\left(\left\{P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha))>0\right\}\right)\xrightarrow[n_1\to\infty]{} 1$ (see the proof of Lemma 6.8). Next, we prove

$$
n_2^{\frac{1}{2}}\left(\frac{1}{n_2}\sum_{i=1}^{n_2}\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)}-P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha))\right)=\mathcal{O}_{P,n_1,n_2}(1), \tag{6.9}
$$

so that we obtain

$$
\frac{1}{n_2}\sum_{i=1}^{n_2}\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)}-P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha)) \xrightarrow[n_1,n_2\to\infty]{\mathbb{P}} 0, \tag{6.10}
$$

and

$$
\frac{1}{n_2}\sum_{i=1}^{n_2}\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)} \xrightarrow[n_1,n_2\to\infty]{\mathbb{P}} P(\mathcal{L}_D(\alpha))>0.
$$

Let us prove (6.9). Let $n_1\ge 1$ and $\varepsilon>0$. Using Tchebychev inequality, we can write $\mathbb{P}$-a.s. (here the event $\omega\in\Omega$ is one realisation of the sample $\tilde{S}_{n_1}$ and is independent of ε)

$$
\begin{aligned}
P_{\tilde{S}_{n_1}}\left(\left|\frac{1}{n_2}\sum_{i=1}^{n_2}\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)}-P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha))\right|\ge\varepsilon\right) &\le \frac{\mathbb{V}_{\tilde{S}_{n_1}}\left(\frac{1}{n_2}\sum_{i=1}^{n_2}\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)}\right)}{\varepsilon^2} \\
&= \frac{\mathbb{V}_{\tilde{S}_{n_1}}\left(\mathbb{1}_{\mathbf{X}_1\in\mathcal{L}_{n_1}(\alpha)}\right)}{n_2\varepsilon^2} \\
&\le \frac{1}{n_2\varepsilon^2}.
\end{aligned}
$$

Thus, taking $M_\varepsilon := 1/\varepsilon^{\frac{1}{2}}$ it holds that

$$
\sup_{n_1,n_2\ge 1}\mathbb{P}\left(n_2^{\frac{1}{2}}\left|\frac{1}{n_2}\sum_{i=1}^{n_2}\mathbb{1}_{\mathbf{X}_i\in\mathcal{L}_{n_1}(\alpha)}-P_{\tilde{S}_{n_1}}(\mathcal{L}_{n_1}(\alpha))\right|\ge M_\varepsilon\right)\le\frac{1}{n_2\left(\frac{M_\varepsilon}{n_2^{\frac{1}{2}}}\right)^2}
$$

$$= \varepsilon,$$

which means that (6.9) is satisfied. Similarly, we obtain

$$\frac{1}{n_2}\sum_{i=1}^{n_2} Y_i \mathbb{1}_{\mathbf{X}_i \in \mathcal{L}_{n_1}(\alpha)} - \mathbb{E}_{\tilde{S}_{n_1}}[Y \mathbb{1}_{\mathbf{X} \in \mathcal{L}_{n_1}(\alpha)}] = \mathcal{O}_{P,n_1,n_2}\left(n_2^{-\frac{1}{2}}\right)$$

Hence the result. □

Proof of Theorem 3.4. Let $P \in \mathscr{P}$ and $\mathbf{X}$ a r.v. with law P. Denoting $D(x) := D(x, P)$, remark that

$$\begin{aligned}\mathbb{E}\left[P_{\tilde{S}_{n_1}}\left(\mathcal{L}_D(\alpha)\Delta\mathcal{L}_{n_1}(\alpha)\right)\right] &= \mathbb{E}\left[P_{\tilde{S}_{n_1}}\left(\{x : D(x) \le \alpha < D_{n_1}(x)\}\right)\right] \\ &\quad + \mathbb{E}\left[P_{\tilde{S}_{n_1}}\left(\{x : D_{n_1}(x) \le \alpha < D(x)\}\right)\right].\end{aligned} \tag{6.11}$$

In the first term of the right-hand side of inequality (6.11), assuming $\mathbb{P}(D(\mathbf{X}) = \alpha) = 0$, it holds (using a Fubini-Tonelli type of argument)

$$\begin{aligned}&\mathbb{E}\left[P_{\tilde{S}_{n_1}}\left(\{x : D(x) \le \alpha < D_{n_1}(x)\}\right)\right] \\ &\qquad = \mathbb{E}\left[\int_x \mathbb{1}_{D(x)\le\alpha<D_{n_1}(x)} dP(x)\right] \\ &\qquad = \int_x \mathbb{1}_{D(x)\le\alpha}\mathbb{P}\left(D_{n_1}(x) > \alpha\right) dP(x) \\ &\qquad = \int_x \mathbb{1}_{D(x)<\alpha}\mathbb{P}\left(D_{n_1}(x) > \alpha\right) dP(x) \\ &\qquad \le \int_x \mathbb{1}_{D(x)<\alpha}\mathbb{P}\left(\|D_{n_1} - D\|_{\infty,\mathbb{R}^d} > \alpha - D(x)\right) dP(x).\end{aligned}$$

In an analogous way, it holds

$$\begin{aligned}&\mathbb{E}[P_{\tilde{S}_{n_1}}\left(\{x : D_{n_1}(x) \le \alpha < D(x)\}\right)] \\ &\qquad = \int_x \mathbb{1}_{D(x)>\alpha}\mathbb{P}\left(D_{n_1}(x) \le \alpha\right) dP(x) \\ &\qquad \le \int_x \mathbb{1}_{D(x)>\alpha}\mathbb{P}(\|D_{n_1} - D\|_{\infty,\mathbb{R}^d} > D(x) - \alpha) dP(x).\end{aligned}$$

Since $\|D_{n_1} - D\|_{\infty,\mathbb{R}^d} \xrightarrow[n_1\to\infty]{\mathbb{P}} 0$,

$$\mathbb{1}_{D(x)>\alpha}\mathbb{P}(\|D_{n_1} - D\|_{\infty,\mathbb{R}^d} > D(x) - \alpha) \xrightarrow[n_1\to\infty]{} 0$$

for a.e. x. Thus, using the Dominated Convergence Theorem, it holds

$$\int_x \mathbb{1}_{D(x)>\alpha}\mathbb{P}(\|D_{n_1} - D\|_{\infty,\mathbb{R}^d} > D(x) - \alpha) dP(x) \xrightarrow[n_1\to\infty]{} 0.$$

So that, from equality (6.11) we conclude that

$$\mathbb{E}\left[P_{\tilde{S}_{n_1}}\left(\mathcal{L}_D(\alpha)\Delta\mathcal{L}_{n_1}(\alpha)\right)\right] \xrightarrow[n_1\to\infty]{} 0.$$

The proof now directly follows from the guidelines of the proof of Theorem 3.5. □

Proofs of Section 3.2.2.

Proof of Theorem 3.9. In all of the following, depending on the context, we denote by $\|\cdot\|$ the Euclidean norm in $\mathbb{R}^d$ or the matrix norm induced by the Euclidean norm in $\mathbb{R}^d$.

Let $x \in \mathbb{R}^d$. Denote $\mu := \mu_{\mathbf{X}}$ and $\Sigma := \Sigma_{\mathbf{X}}$. Since $\hat{\Sigma}_n \xrightarrow[n\to\infty]{\mathbb{P}\text{-a.s.}} \Sigma$ and Σ is invertible, then $\hat{\Sigma}_n$ is invertible for large n. Thus, for n large enough, we can write

$$\begin{aligned}
|MHD_n(x) - MHD(x)| &\\
&= \left|\frac{1}{1+\ {}^t(x-\hat{\mu}_n)\hat{\Sigma}_n^{-1}(x-\hat{\mu}_n)} - \frac{1}{1+\ {}^t(x-\mu)\Sigma^{-1}(x-\mu)}\right|\\
&\le \frac{\left|\ {}^t(x-\hat{\mu}_n)\hat{\Sigma}_n^{-1}(x-\hat{\mu}_n) -\ {}^t(x-\mu)\Sigma^{-1}(x-\mu)\right|}{1+\ {}^t(x-\mu)\Sigma^{-1}(x-\mu)}.
\end{aligned}$$

Since Σ is a positive definite symmetric and invertible matrix, we can make the change of variable $y = \Sigma^{-\frac{1}{2}}(x-\mu)$. So that,

$$\|MHD_n - MHD\|_{\infty,\mathbb{R}^d} \le \sup_{y\in\mathbb{R}^d} \frac{\left|\|\hat{\Sigma}_n^{-\frac{1}{2}}(\Sigma^{\frac{1}{2}}y+\mu-\hat{\mu}_n)\|^2 - \|y\|^2\right|}{1+\|y\|^2}.$$

Now, denoting by I_d the identity matrix of size d, and using a triangle inequality then Cauchy-Schwarz inequality, it holds

$$\begin{aligned}
&\frac{\left|\|\hat{\Sigma}_n^{-\frac{1}{2}}(\Sigma^{\frac{1}{2}}y+\mu-\hat{\mu}_n)\|^2 - \|y\|^2\right|}{1+\|y\|^2}\\
&\le \frac{\left|\|\hat{\Sigma}_n^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}y\|^2 - \|y\|^2\right| + \|\hat{\Sigma}_n^{-\frac{1}{2}}(\mu-\hat{\mu}_n)\|^2 + 2\left|\left\langle \hat{\Sigma}_n^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}y, \hat{\Sigma}_n^{-\frac{1}{2}}(\mu-\hat{\mu}_n)\right\rangle\right|}{1+\|y\|^2}\\
&\le \frac{\|\Sigma^{\frac{1}{2}}\hat{\Sigma}_n^{-1}\Sigma^{\frac{1}{2}} - I_d\|\|y\|^2 + \|\hat{\Sigma}_n^{-\frac{1}{2}}(\mu-\hat{\mu}_n)\|^2 + 2\|\hat{\Sigma}_n^{-1}\|\|\Sigma^{\frac{1}{2}}\|\|y\|\|\mu-\hat{\mu}_n\|}{1+\|y\|^2},
\end{aligned}$$

This, together with the fact that $2\|y\|/(1+\|y\|^2) \le 1$, for all $y \in \mathbb{R}^d$,

$$\frac{\left|\|\hat{\Sigma}_n^{-\frac{1}{2}}(\Sigma^{\frac{1}{2}}y+\mu-\hat{\mu}_n)\|^2 - \|y\|^2\right|}{1+\|y\|^2}$$

$$\leq \|\Sigma^{\frac{1}{2}}\hat{\Sigma}_n^{-1}\Sigma^{\frac{1}{2}} - I_d\| + \|\hat{\Sigma}_n^{-\frac{1}{2}}(\mu - \hat{\mu}_n)\|^2 + \|\hat{\Sigma}_n^{-1}\| \cdot \|\Sigma^{\frac{1}{2}}\|\|\mu - \hat{\mu}_n\|.$$

Now since the right hand side of the above inequality is independent of y, for large n we obtain

$$\begin{aligned} &\|MHD_n - MHD\|_{\infty,\mathbb{R}^d} \\ &\leq \|\Sigma^{\frac{1}{2}}\hat{\Sigma}_n^{-1}\Sigma^{\frac{1}{2}} - I_d\| + \|\hat{\Sigma}_n^{-\frac{1}{2}}\|^2\|\mu - \hat{\mu}_n\| \left(\|\mu - \hat{\mu}_n\| + \|\Sigma^{\frac{1}{2}}\|\right) \\ &:= A_n(d), \end{aligned} \tag{6.12}$$

The problem reduces to studying the asymptotic behavior of $A_n(d)$. On one hand, since $\hat{\Sigma}_n \xrightarrow[n\to\infty]{\mathbb{P}\text{-a.s.}} \Sigma$ and $\hat{\mu}_n \xrightarrow[n\to\infty]{\mathbb{P}\text{-a.s.}} \mu$, then by the continuity theorem we obtain

$$\|\hat{\Sigma}_n^{-\frac{1}{2}}\|^2 \left(\|\mu - \hat{\mu}_n\| + \|\Sigma^{\frac{1}{2}}\|\right) \xrightarrow[n\to\infty]{\mathbb{P}} \|\Sigma^{-\frac{1}{2}}\|^2\|\Sigma^{\frac{1}{2}}\| > 0.$$

Furthermore, by the multivariate Central Limit theorem, it holds that $n^{\frac{1}{2}}(\hat{\mu}_n - \mu) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,\Sigma)$. Thus (by the continuity theorem and Slutsky's lemma), $\|\hat{\Sigma}_n^{-\frac{1}{2}}\|^2\|\mu - \hat{\mu}_n\| \left(\|\mu - \hat{\mu}_n\| + \|\Sigma^{\frac{1}{2}}\|\right)$ is $\mathcal{O}_P\left(n^{-\frac{1}{2}}\right)$. On the other hand, to study the first term in $A_n(d)$, we define

$$F : H \in \mathcal{S}_d(\mathbb{R}) \mapsto \Sigma^{\frac{1}{2}} H^{-1} \Sigma^{\frac{1}{2}},$$

where $\mathcal{S}_d(\mathbb{R})$ is the vector space of all symmetric real-valued matrices of size d. We denote $\mathcal{S}_d^+(\mathbb{R})$ the set of all positive definite symmetric matrices which is an open set in $\mathcal{S}_d(\mathbb{R})$. Using classical computations of Fréchet differentiable functions, it holds that for all $A \in \mathcal{S}_d^+(\mathbb{R})$, the differential of F at A is given by:

$$DF_A : H \in \mathcal{S}_d(\mathbb{R}) \mapsto DF_A(H) = -\Sigma^{\frac{1}{2}} A^{-1} H A^{-1} \Sigma^{\frac{1}{2}}. \tag{6.13}$$

By isomorphism, one can see $\hat{\Sigma}_n$ as an element of $\mathbb{R}^{\frac{d(d+1)}{2}}$. Since $\mathbf{X}$ has all of its components in $\mathbb{L}^4$, then a multivariate CLT applies, i.e. there exists $M^* \in \mathcal{S}^+_{\frac{d(d+1)}{2}}(\mathbb{R})$ s.t.

$$\sqrt{n}(\hat{\Sigma}_n - \Sigma) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, M^*), \tag{6.14}$$

and for notational convenience, the gaussian vector $\mathcal{N}(0, M^*)$ could be rearranged in a size d symmetric random matrix which will be denoted by E^*. For the sake of completeness, we resume a proof of the delta method in the words of Agresti [1] (p. 577). Using a first order Taylor expansion, for all $A \in \mathcal{S}_d^+(\mathbb{R})$ and $X \in \mathrm{B}(A, r) \subset \mathcal{S}_d^+(\mathbb{R})$, $r > 0$, we can write:

$$\begin{aligned} &F(X) - F(A) = DF_A(X - A) + R(X), \text{ with} \\ &\frac{R(X)}{\|X - A\|} \xrightarrow[X\to A]{} 0. \end{aligned}$$

Then, $\mathbb{P}$-a.s., using (6.13)

$$\begin{aligned} F(\hat{\Sigma}_n) - F(\Sigma) &= DF_\Sigma(\hat{\Sigma}_n - \Sigma) + R(\hat{\Sigma}_n), \\ &= -\Sigma^{\frac{1}{2}}\Sigma^{-1}(\hat{\Sigma}_n - \Sigma)\Sigma^{-1}\Sigma^{\frac{1}{2}} + R(\hat{\Sigma}_n) \\ &= -\Sigma^{-\frac{1}{2}}(\hat{\Sigma}_n - \Sigma)\Sigma^{-\frac{1}{2}} + R(\hat{\Sigma}_n). \end{aligned}$$

This, together with (6.14), using the continuity theorem and Slutsky's Lemma, we obtain

$$\sqrt{n}R(\hat{\Sigma}_n) = \sqrt{n}\|\hat{\Sigma}_n - \Sigma\| \frac{R(\hat{\Sigma}_n)}{\|\hat{\Sigma}_n - \Sigma\|} \xrightarrow[n\to\infty]{\mathcal{L}} 0,$$

so $\sqrt{n}R(\hat{\Sigma}_n) \xrightarrow[n\to\infty]{\mathbb{P}} 0$. In addition, from the continuity of $U \mapsto -\Sigma^{-\frac{1}{2}}U\Sigma^{-\frac{1}{2}}$ and using (6.14), we obtain $-\sqrt{n}\Sigma^{-\frac{1}{2}}(\hat{\Sigma}_n - \Sigma)\Sigma^{-\frac{1}{2}} \xrightarrow[n\to\infty]{\mathcal{L}} -\Sigma^{-\frac{1}{2}}E^*\Sigma^{-\frac{1}{2}}$. Therefore, by continuity of the matrix norm, we deduce that

$$\|\Sigma^{\frac{1}{2}}\hat{\Sigma}_n^{-1}\Sigma^{\frac{1}{2}} - I_d\| = \mathcal{O}_P\left(n^{-\frac{1}{2}}\right).$$

To conclude, it holds that $A_n(d) = \mathcal{O}_P\left(n^{-\frac{1}{2}}\right)$ which implies the desired result:

$$\|MHD_n - MHD\|_{\infty,\mathbb{R}^d} = \mathcal{O}_P\left(n^{-\frac{1}{2}}\right). \qquad \square$$

Proof of Corollary 3.10. The proof is a combination of Corollary 3.2, Theorem 3.5 and 3.9. Indeed, under the assumptions of Corollary 3.2 and Theorem 3.9, assumption (**H0**) of Theorem 3.5 is satisfied since

$$\begin{aligned} P_{\tilde{S}_{n_1}}(\mathcal{L}_D(\alpha)\Delta\mathcal{L}_{n_1}(\alpha)) &= \mathcal{O}_{P,n_1}\left(\|D_{n_1} - D\|_{\infty,\mathbb{R}^d}^{1-\frac{1}{p}}\right) && \text{(Corollary 3.2)} \\ &= \mathcal{O}_{P,n_1}\left(v_{n_1}^{-1}\right) && \text{(Theorem 3.9)} \end{aligned}$$

with $v_{n_1} = n_1^{\frac{1}{2}\left(1-\frac{1}{p}\right)}$. The result is then a straightforward consequence of Theorem 3.5. $\square$

Appendix A: Simulation tables

Probability based pseudo-metric.

TABLE 1
Estimated $\sqrt{n}\hat{p}^n_{N,M}(D,\alpha)$ with $D = MHD$.

α \ n	100	1000	5000	10000	50000
0.9	0.3010	0.2627	0.3715	0.3032	0.2263
0.8	0.4339	0.3912	0.4107	0.332	0.317
0.5	0.736	0.7353	0.6793	0.7162	0.7406
0.2	0.3391	0.3284	0.3937	0.474	0.3287
0.1	0.1374	0.1187	0.1355	0.1372	0.0792

TABLE 2
Estimated $\sqrt{n}\hat{p}^n_{N,M}(D,\alpha)$ *with* $D = D_{HS}$.

α \ n	50	100	500	1000	2500	5000
0.4	0.1984	0.1589	0.2147	0.2242	0.305	0.2821
0.1	1.0211	0.9703	0.9559	0.9003	0.927	0.7
0.05	0.8608	0.99	0.9474	0.7729	0.9078	0.9007
0.01	0.7497	0.8647	0.5923	0.4712	0.537	0.5607

TABLE 3
Estimated $\sqrt{n}\hat{p}^n_{N,M}(D,\alpha)$ *with* $D = D_{PJ}$.

α \ n	50	100	500	1000	2500	5000
0.7	0.2025	0.2056	0.2689	0.2948	0.299	0.2936
0.5	0.5802	0.6046	0.6933	0.6399	0.6951	0.6139
0.2	0.4345	0.377	0.3501	0.2829	0.2904	0.3111
0.1	0.0519	0.0491	0.0238	0.0224	0.0353	0.0346

Volume based pseudo-metric.

TABLE 4
Estimated $\sqrt{n}\widehat{\mathrm{Vol}}^n_{N,M}(D,\alpha)$ *with* $D = MHD$.

α \ n	100	1000	5000	10000	50000
0.9	0.1535	0.1975	0.1763	0.2538	0.2113
0.8	0.2674	0.2411	0.1737	0.2268	0.2073
0.5	0.6329	0.6136	0.6459	0.3816	0.0.3202
0.2	1.6834	1.4583	1.4001	1.9035	1.2618
0.1	2.752	2.679	2.5596	3.006	2.0608

TABLE 5
Estimated $\sqrt{n}\widehat{\mathrm{Vol}}^n_{N,M}(D,\alpha)$ *with* $D = D_{HS}$.

α \ n	50	100	500	1000	2500	5000
0.4	0.0942	0.065	0.1565	0.0626	0.2093	0.1416
0.1	0.9253	1.0699	1.0213	0.5536	0.8741	0.7907
0.05	1.601	1.7863	2.1649	1.8983	1.2398	1.8646
0.01	3.5555	4.9957	5.5936	5.0859	4.3425	5.0609

TABLE 6
Estimated $\sqrt{n}\widehat{\mathrm{Vol}}^n_{N,M}(D,\alpha)$ *with* $D = D_{PJ}$.

α \ n	50	100	500	1000	2500	5000
0.7	0.0118	0.0752	0.0569	0.0583	0.2025	0.0111
0.5	0.2237	0.3033	0.1046	0.7741	0.3026	0.315
0.2	2.5936	2.1085	2.1634	1.8257	2.0396	2.0062
0.1	4.2131	4.8096	4.0873	3.6664	4.5821	4.8859

CCTE estimations.

TABLE 7
$\mathbb{L}^1$-estimation of $\mathrm{CCTE}_{\alpha,MHD}(\mathrm{Y}, \boldsymbol{X})$ *and associated* RMAE *for bivariate Frank Copulas with Gumbel marginals.*

n		$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.5$	$\alpha = 0.8$	$\alpha = 0.9$
		CCTE = 1.4237	CCTE = 0.8831	CCTE = 0.4804	CCTE = 0.3831	CCTE = 0.3661
100	Mean	1.2744	0.8568	0.4706	0.3806	0.3636
	$\widehat{\sigma}$	0.6596	0.2678	0.0716	0.045	0.042
	RMAE	0.3733	0.2360	0.1217	0.0954	0.0933
1000	Mean	1.4125	0.8758	0.4801	0.3835	0.3664
	$\widehat{\sigma}$	0.1923	0.0782	0.0233	0.0153	0.0143
	RMAE	0.1056	0.0696	0.0381	0.0318	0.0310
5000	Mean	1.4213	0.8808	0.4804	0.3832	0.3661
	$\widehat{\sigma}$	0.0855	0.0377	0.0106	0.0065	0.0061
	RMAE	0.0476	0.0338	0.018	0.0138	0.0134
10000	Mean	1.4178	0.8807	0.4799	0.3828	0.3659
	$\widehat{\sigma}$	0.0625	0.0271	0.0071	0.0046	0.0043
	RMAE	0.0361	0.0241	0.0117	0.0095	0.0094
50000	Mean	1.4231	0.8829	0.4807	0.3833	0.3663
	$\widehat{\sigma}$	0.0283	0.0118	0.0033	0.002	0.0019
	RMAE	0.0158	0.0106	0.0055	0.0042	0.0041

TABLE 8
Estimated $\sqrt{n} \cdot \mathrm{RMAE}_{n,\alpha}$ *based on* MHD *for bivariate Frank Copulas with Gumbel marginals.*

α \ n	100	1000	5000	10000	50000
0.1	3.7335	3.3393	3.3635	3.6092	3.5262
0.2	2.3604	2.2021	2.3885	2.4118	2.374
0.5	1.2169	1.2038	1.2703	1.1740	1.2236
0.8	0.9539	1.0056	0.9740	0.9491	0.9415
0.9	0.9334	0.9815	0.9492	0.9433	0.9072

Table 9

$\mathbb{L}^1$-estimation of $\mathrm{CCTE}_{\alpha,D_{HS}}(\mathrm{Y},\boldsymbol{X})$ *based on halfspace depth and associated* RMAE *for bivariate Frank Copulas with Gumbel marginals.*

n		$\alpha = 0.01$ CCTE $= 1.0061$	$\alpha = 0.05$ CCTE $= 0.6492$	$\alpha = 0.1$ CCTE $= 0.5249$	$\alpha = 0.4$ CCTE $= 0.3573$
50	Mean	0.7389	0.5794	0.4763	0.3558
	$\widehat{\sigma}$	0.3237	0.161	0.0987	0.0563
	RMAE	0.3507	0.2177	0.167	0.1268
100	Mean	0.7765	0.5756	0.4859	0.3539
	$\widehat{\sigma}$	0.2386	0.1007	0.0689	0.0364
	RMAE	0.2776	0.1615	0.1248	0.0808
500	Mean	0.9098	0.6225	0.5115	0.3552
	$\widehat{\sigma}$	0.1503	0.0583	0.0377	0.0188
	RMAE	0.1473	0.08	0.0621	0.0418
1000	Mean	0.9617	0.639	0.5207	0.3573
	$\widehat{\sigma}$	0.1167	0.0423	0.0277	0.013
	RMAE	0.1006	0.0538	0.0425	0.0287
2500	Mean	0.9817	0.6433	0.5216	0.3569
	$\widehat{\sigma}$	0.0734	0.0262	0.0167	0.0078
	RMAE	0.06	0.0328	0.0256	0.0174
5000	Mean	0.9913	0.645	0.5229	0.3569
	$\widehat{\sigma}$	0.00545	0.019	0.0126	0.0064
	RMAE	0.0453	0.0239	0.0194	0.0143

Table 10

Estimated $\sqrt{n}\cdot\mathrm{RMAE}_{n,\alpha}$ *based on halfspace depth for bivariate Frank Copulas with Gumbel marginals.*

α \ n	50	100	500	1000	2500	5000
0.01	2.4798	2.7761	3.2948	3.1802	3.0022	3.1999
0.05	1.5393	1.6151	1.7886	1.7021	1.6385	1.6933
0.1	1.1812	1.248	1.388	1.3454	1.2801	1.369
0.4	0.8969	0.8083	0.935	0.9071	0.8689	1.0102

TABLE 11
$\mathbb{L}^1$*-estimation of* $\text{CCTE}_{\alpha,D_{PJ}}(\text{Y}, \boldsymbol{X})$ *based on projection depth and associated* RMAE *for bivariate Frank Copulas with Gumbel marginals.*

n		$\alpha = 0.15$ CCTE = 1.396	$\alpha = 0.3$ CCTE = 0.5748	$\alpha = 0.5$ CCTE = 0.3938	$\alpha = 0.8$ CCTE = 0.3541
50	Mean	1.0154	0.5097	0.3744	0.3457
	$\widehat{\sigma}$	0.656	0.1324	0.0681	0.0566
	RMAE	0.445	0.2104	0.1441	0.1282
100	Mean	1.1969	0.5371	0.3855	0.3535
	$\widehat{\sigma}$	0.5513	0.0986	0.0476	0.0399
	RMAE	0.3279	0.1489	0.0982	0.0906
500	Mean	1.3404	0.5648	0.3908	0.3539
	$\widehat{\sigma}$	0.2447	0.0482	0.0214	0.018
	RMAE	0.1441	0.0696	0.0442	0.0407
1000	Mean	1.3776	0.5669	0.3919	0.3538
	$\widehat{\sigma}$	0.1858	0.0363	0.0157	0.0128
	RMAE	0.1065	0.0522	0.0313	0.0285
2500	Mean	1.3826	0.5719	0.3932	0.3545
	$\widehat{\sigma}$	0.1242	0.0218	0.0098	0.0082
	RMAE	0.0697	0.03	0.0198	0.0189
5000	Mean	1.3815	0.5731	0.3936	0.3543
	$\widehat{\sigma}$	0.0822	0.0158	0.0071	0.0057
	RMAE	0.445	0.0222	0.0143	0.013

TABLE 12
Estimated $\sqrt{n} \cdot \text{RMAE}_{n,\alpha}$ *based on projection depth for bivariate Frank Copulas with Gumbel marginals.*

α \ n	50	100	500	1000	2500	5000
0.15	3.1464	3.2794	3.2224	3.367	3.484	3.4328
0.3	1.4877	1.4888	1.5561	1.6503	1.4983	1.57
0.5	1.0191	0.982	0.9878	0.9911	0.9902	1.0126
0.8	0.9067	0.9059	0.9093	0.9011	0.9461	0.9162

Acknowledgments

The authors express their gratitude to two anonymous Referees and Associate Editor for their valuable comments on this article.

References

[1] Agresti, A. (2002). *Categorical data analysis.* NJ: John Wiley & Sons, Inc. MR1914507

[2] Baillo, A., Cuesta-Albertos, J. A., and Cuevas, A. (2001). Convergence rates in nonparametric estimation of level sets. *Statistics & probability letters*, 53(1):27–35. MR1843338

[3] Belzunce, F., Castaño, A., Olvera-Cervantes, A., and Suárez-Llorens, A. (2007). Quantile curves and dependence structure for bivariate distributions. *Computational Statistics & Data Analysis*, 51(10):5112–5129. MR2370711

[4] Brunel, V.-E. (2019). Concentration of the empirical level sets of tukey's halfspace depth. *Probability Theory and Related Fields*, 173(3):1165–1196. MR3936153

[5] Cadre, B. (2006). Kernel estimation of density level sets. *Journal of multivariate analysis*, 97(4):999–1023. MR2256570

[6] Cascos, I. and Molchanov, I. (2007). Multivariate risks and depth-trimmed regions. *Finance and stochastics*, 11(3):373–397. MR2322918

[7] Cavalier, L. (1997). Nonparametric estimation of regression level sets. *Statistics A Journal of Theoretical and Applied Statistics*, 29(2):131–160. MR1484386

[8] Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, 91(434):862–872. MR1395753

[9] Chebana, F. and Ouarda, T. B. (2011). Multivariate quantiles in hydrological frequency analysis. *Environmetrics*, 22(1):63–78. MR2843337

[10] Chow, T. W. and Leung, C.-T. (1996). Neural network based short-term load forecasting using weather compensation. *IEEE transactions on power systems*, 11(4):1736–1742.

[11] Coblenz, M., Dyckerhoff, R., and Grothe, O. (2018a). Confidence regions for multivariate quantiles. *Water*, 10(8):996.

[12] Coblenz, M., Dyckerhoff, R., and Grothe, O. (2018b). Nonparametric estimation of multivariate quantiles. *Environmetrics*, 29(2):e2488. MR3775342

[13] Cousin, A. and Di Bernardino, E. (2013). On multivariate extensions of value-at-risk. *Journal of multivariate analysis*, 119:32–46. MR3061413

[14] Cuevas, A., Febrero, M., and Fraiman, R. (2000). Estimating the number of clusters. *Canadian Journal of Statistics*, 28(2):367–382. MR1792055

[15] Cuevas, A., Febrero, M., and Fraiman, R. (2001). Cluster analysis: a further approach based on density estimation. *Computational Statistics & Data Analysis*, 36(4):441–459. MR1855727

[16] Cuevas, A., González-Manteiga, W., and Rodríguez–Casal, A. (2006). Plug-in estimation of general level sets. *Australian & New Zealand J. Statist.*, 48:7–19. MR2234775

[17] Dai, W. and Genton, M. G. (2019). Directional outlyingness for multivariate functional data. *Computational Statistics & Data Analysis*, 131:50–65. MR3906794

[18] Dehaan, L. and Huang, X. (1995). Large quantile estimation in a multivariate setting. *Journal of Multivariate Analysis*, 53(2):247–263. MR1336056

[19] Denecke, L. and Müller, C. H. (2014). Consistency of the likelihood depth estimator for the correlation coefficient. *Statistical Papers*, 55(1):3–13.

MR3152764

[20] Denuit, M., Dhaene, J., Goovaerts, M., and Kaas, R. (2006). *Actuarial theory for dependent risks: measures, orders and models*. John Wiley & Sons.

[21] Desforges, M., Jacob, P., and Cooper, J. (1998). Applications of probability density estimation to the detection of abnormal conditions in engineering. *Proceedings of the institution of mechanical engineers, part c: Journal of mechanical engineering science*, 212(8):687–703.

[22] Devroye, L. and Wise, G. L. (1980). Detection of abnormal behavior via nonparametric estimation of the support. *SIAM Journal on Applied Mathematics*, 38(3):480–488. MR0579432

[23] Di Bernardino, E., Laloë, T., Maume-Deschamps, V., and Prieur, C. (2013). Plug-in estimation of level sets in a non-compact setting with applications in multivariate risk theory. *ESAIM: Probability and Statistics*, 17. MR3021318

[24] Di Bernardino, E., Laloë, T., and Servien, R. (2015). Estimating covariate functions associated to multivariate risks: a level set approach. *Metrika, Springer Verlag*, 78(5):497–526. MR3355458

[25] Di Bernardino, E. and Rulliere, D. (2013). Distortions of multivariate distribution functions and associated level curves: Applications in multivariate risk theory. *Insurance: Mathematics and Economics*, 53(1):190–205. MR3081472

[26] Donoho, D. L. and Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *The Annals of Statistics*, pages 1803–1827. MR1193313

[27] Dudek, G. (2016). Pattern-based local linear regression models for short-term load forecasting. *Electric power systems research*, 130:139–147.

[28] Dümbgen, L. (1992). Limit theorems for the simplicial depth. *Statistics & Probability Letters*, 14(2):119–128. MR1173409

[29] Dyckerhoff, R. (2004). Data depths satisfying the projection property. *Allgemeines Statistisches Archiv*, 88:163–190. MR2074729

[30] Dyckerhoff, R. (2016). Convergence of depths and depth-trimmed regions. *arXiv preprint arXiv:1611.08721*.

[31] Dyckerhoff, R. and Mosler, K. (2012). Weighted-mean regions of a probability distribution. *Statistics & Probability Letters*, 82(2):318–325. MR2875217

[32] Elias, R., Fang, L., and Wahab, M. (2011). Electricity load forecasting based on weather variables and seasonalities: A neural network approach. In *ICSSSM11*, pages 1–6. IEEE.

[33] Febrero, M., Galeano, P., and González-Manteiga, W. (2008). Outlier detection in functional data by depth measures, with application to identify abnormal nox levels. *Environmetrics: The official journal of the International Environmetrics Society*, 19(4):331–345. MR2440036

[34] Gardes, L. and Girard, S. (2010). Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return levels. *Extremes*, 13(2):177–204. MR2643556

[35] Girard, S., Stupfler, G., and Usseglio-Carleve, A. (2021). Extreme condi-

tional expectile estimation in heavy-tailed heteroscedastic regression models. *The Annals of statistics*, 49(6):3358–3382. MR4352533

[36] Goldstein, L. and Messer, K. (1992). Optimal plug-in estimators for nonparametric functional estimation. *The annals of statistics*, pages 1306–1328. MR1186251

[37] Hartigan, J. (1975). Clustering algorithms wiley. *New York*. MR0405726

[38] He, X. and Wang, G. (1997). Convergence of depth contours for multivariate datasets. *The Annals of Statistics*, pages 495–504. MR1439311

[39] He, Y. and Einmahl, J. H. (2017). Estimation of extreme depth-based quantile regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(2):449–461. MR3611754

[40] Hettmansperger, T. P. and Oja, H. (1994). Affine invariant multivariate multisample sign tests. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(1):235–249. MR1257810

[41] Hippert, H. S., Pedreira, C. E., and Souza, R. C. (2001). Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on power systems*, 16(1):44–55.

[42] Hubert, M., Rousseeuw, P., and Segaert, P. (2017). Multivariate and functional classification using depth and distance. *Advances in Data Analysis and Classification*, 11(3):445–466. MR3688976

[43] Jeankyung, K. (2000). Rate of convergence of depth contours: with application to a multivariate metrically trimmed mean. *Statistics & Probability Letters*, 49(4):393–400. MR1796484

[44] Jörnsten, R. (2004). Clustering and classification based on the l1 data depth. *Journal of Multivariate Analysis*, 90(1):67–89. MR2064937

[45] Khotanzad, A., Hwang, R.-C., Abaye, A., and Maratukulam, D. (1995). An adaptive modular artificial neural network hourly load forecaster and its implementation at electric utilities. *IEEE Transactions on Power Systems*, 10(3):1716–1722.

[46] Korostelev, A. P. and Tsybakov, A. B. (1993). Minimax linewise algorithm for image reconstruction. In *Computer intensive methods in statistics*, pages 113–133. Springer. MR1224456

[47] Koshevoy, G. and Mosler, K. (1997). Zonoid trimming for multivariate distributions. *The Annals of Statistics*, 25(5):1998–2017. MR1474078

[48] Laloë, T. and Servien, R. (2013). Nonparametric estimation of regression level sets using kernel plug-in estimator. *Journal of the Korean Statistical Society*, 42(3):301–311. MR3255389

[49] Liu, R. (1990). On a notion of data depth based on random simplices. *The Annals of Statistics*, pages 405–414. MR1041400

[50] Liu, R. Y. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests. *Journal of the American Statistical Association*, 88(421):252–260. MR1212489

[51] Mammen, E. and Polonik, W. (2013). Confidence regions for level sets. *Journal of Multivariate Analysis*, 122:202–214. MR3189318

[52] Massé, J.-C. (2004). Asymptotics for the tukey depth process, with an application to a multivariate trimmed mean. *Bernoulli*, 10(3):397–419.

MR2061438

[53] Molchanov, I. S. (1991). Empirical estimation of distribution quantiles of random closed sets. *Theory of Probability and Its Applications*, 35:594–600. MR1091220

[54] Molchanov, I. S. (1998). A limit theorem for solutions of inequalities. *Scandinavian Journal of Statistics*, 25(1):235–242. MR1614288

[55] Müller, C. H. (2005). Depth estimators and tests based on the likelihood principle with application to regression. *Journal of Multivariate Analysis*, 95(1):153–181. MR2164127

[56] Müller, D. W. and Sawitzki, G. (1991). Excess mass estimates and tests for multimodality. *Journal of the American Statistical Association*, 86(415):738–746. MR1147099

[57] Oreshkin, B., Dudek, G., Pełka, P., and Turkina, E. (2021a). N-beats neural network for mid-term electricity load forecasting. *Applied Energy*, 293:116918.

[58] Oreshkin, B. N., Dudek, G., Pełka, P., and Turkina, E. (2021b). N-beats neural network for mid-term electricity load forecasting. *Applied Energy*, 293:116918.

[59] Park, C., Huang, J. Z., and Ding, Y. (2010). A computable plug-in estimator of minimum volume sets for novelty detection. *Operations Research*, 58(5):1469–1480. MR2560548

[60] Pokotylo, O., Mozharovskyi, P., Dyckerhoff, R., and Nagy, S. (2019). ddalpha: Depth-based classification and calculation of data depth 1.3. 9.

[61] Polonik, W. (1995). Measuring mass concentrations and estimating density contour clusters-an excess mass approach. *The annals of Statistics*, pages 855–881. MR1345204

[62] Rodríguez-Casal, A. (2003). Estimacíon de conjuntos y sus fronteras. un enfoque geometrico. *PhD thesis, University of Santiago de Compostela*.

[63] Ruts, I. and Rousseeuw, P. J. (1996). Computing depth contours of bivariate point clouds. *Computational statistics & data analysis*, 23(1):153–168.

[64] Scott, C. D. and Nowak, R. D. (2006). Learning minimum volume sets. *The Journal of Machine Learning Research*, 7:665–704. MR2274383

[65] Torres, R., Di Bernardino, E., Laniado, H., and Lillo, R. (2020). On the estimation of extreme directional multivariate quantiles. *Communications in Statistics-Theory and Methods*, 49(22):5504–5534. MR4156014

[66] van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer. MR1385671

[67] Wang, H. J. and Li, D. (2013). Estimation of extreme conditional quantiles through power transformation. *Journal of the American Statistical Association*, 108(503):1062–1074. MR3174684

[68] Wang, H. J., Li, D., and He, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association*, 107(500):1453–1464. MR3036407

[69] Weyl, H. (1939). On the volume of tubes. *American Journal of Mathematics*, Vol. 61:461–472. MR1507388

[70] Zuo, Y. and Serfling, R. (2000a). General notions of statistical depth func-

tion. *Annals of statistics*, pages 461–482. MR1790005
[71] Zuo, Y. and Serfling, R. (2000b). On the performance of some robust nonparametric location measures relative to a general notion of multivariate symmetry. *Journal of Statistical Planning and Inference*, 84(1-2):55–79. MR1747497
[72] Zuo, Y. and Serfling, R. (2000c). Structural properties and convergence results for contours of sample statistical depth functions. *Annals of Statistics*, pages 483–499. MR1790006