



HAL
open science

Estimation of the covariate conditional tail expectation : a depth-based level set approach

Elisabeth Armaut, Roland Diel, Thomas Laloë

► To cite this version:

Elisabeth Armaut, Roland Diel, Thomas Laloë. Estimation of the covariate conditional tail expectation : a depth-based level set approach. 2021. hal-03336080v1

HAL Id: hal-03336080

<https://hal.science/hal-03336080v1>

Preprint submitted on 6 Sep 2021 (v1), last revised 5 May 2022 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimation of the covariate conditional tail expectation : a depth-based level set approach

Elisabeth Armaut * Roland Diel † Thomas Laloë ‡

Université de Nice Sophia-Antipoli, Laboratoire J-A Dieudonné, Parc Valrose, 06108 Nice Cedex 02, France

September 6, 2021

Abstract. The aim of this paper is to study the asymptotic behavior of a particular multivariate risk measure, the Covariate-Conditional-Tail-Expectation (CCTE), based on a multivariate statistical depth function. Depth functions have become increasingly powerful tools in nonparametric inference for multivariate data, as they measure a degree of centrality of a point with respect to a distribution. A multivariate risks scenario is then represented by a depth-based lower level set of the risk factors, meaning that we consider a non-compact setting. 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 $\mathcal{L}_D(\alpha) := \{z \in \mathbb{R}^d : D(z, P) \leq \alpha\}$. First, we present a *plug-in* approach in order to estimate $\mathcal{L}_D(\alpha)$. In a second part, we provide a consistent estimator of our CCTE for a general depth function with a rate of convergence, and we consider the particular case of the *Mahalanobis* depth. A simulation study complements the performances of our estimator.

Key-Words. *Plug-in* estimation, multivariate depth function, *Mahalanobis* depth, risk theory.

*armaut@unice.fr

†diel@unice.fr

‡lalo@unice.fr

Contents

1	Introduction	3
2	Notations and definitions	5
2.1	General Notations	5
2.2	Depth functions	6
3	Main results	8
3.1	General Covariate-Conditional-Tail-Expectation consistency .	8
3.2	Consistency of general depth-based level sets in terms of the Hausdorff distance	10
3.3	<i>MHD</i> -depth based Covariate-Conditional-Tail-Expectation consistency	12
4	Simulations and illustrations	14
5	Proofs of Section 3	18

1 Introduction

Risk theory is a branch of statistics which mainly focuses on unlikely events in the aim of managing the degree of uncertainty of such events and/or the associated costs. For instance, in the context of risk management in financial institutions such as banks or insurance companies, adverse consequences may occur and usually mean potential large losses on a portfolio of assets (Eberlein et al. [11]). In hydrology, risk could represent dam floods (and/or failures) and the associated risk factors could be the amount of rainfall, water flow... Broadly speaking, a risk measure can be viewed as a mapping from a set of real-valued random variables to \mathbb{R}^d , $d \geq 1$, 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. One of the most studied risk measure in the univariate risk theory is the Conditional-Tail-Expectation (CTE) (Denuit et al. [7]). 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 :

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

where

$$\text{VaR}(\alpha) := \inf \{t \in \mathbb{R} : F_X(t) \geq \alpha\}$$

is the well-known *Value at Risk* which corresponds to the univariate quantile of order α of X . Thus, the CTE is nothing but the mathematical description of an average loss in the worst $100(1 - \alpha)\%$ risk scenario.

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 univariate case, a wide panel of univariate quantiles has been reviewed in the literature. When it comes to multivariate risks, 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 (Belzunce et al. [2], Dehaan and Huang [6], Cousin and Di Bernardino [4]), or on a depth function (Zuo and Serfling [16]). In this way, several generalizations in higher dimension of the CTE emerged in the literature. In particular, one can mention the one proposed by Di Bernardino et al. [8] : given a random vector $\mathbf{X} \in \mathbb{R}^d$, $\mathbf{X} = (X_1, \dots, X_d) \in \mathbb{R}^d$, $d \geq 1$, with multivariate distribution function $F_{\mathbf{X}} : \mathbb{R}^d \rightarrow [0, 1]$, a generalization of the CTE in higher dimension is defined by

$$\text{CTE}_\alpha(\mathbf{X}) = \mathbb{E}[\mathbf{X}|\mathbf{X} \in \mathcal{L}_{F_{\mathbf{X}}}(\alpha)] \in \mathbb{R}^d, \quad (1.2)$$

where

$$\mathcal{L}_{F_{\mathbf{X}}}(\alpha) := \{x \in \mathbb{R}^d : F_{\mathbf{X}}(x) \geq \alpha\}, \quad \alpha \in (0, 1),$$

is the α -upper level set of $F_{\mathbf{X}}$, which is one generalization of the univariate quantile region $[\text{VaR}_{\alpha}(X), +\infty)$ in dimension $d \geq 1$.

Another interesting problem is to study the behavior of an expected cost $Y \in \mathbb{R}$ 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. To address this, Di Bernardino et al. [9] 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) defined by :

$$\text{CCTE}_{\alpha}(Y, \mathbf{X}) := \mathbb{E}[Y | \mathbf{X} \in \mathcal{L}_{F_{\mathbf{X}}}(\alpha)], \quad \alpha \in (0, 1). \quad (1.3)$$

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. [14] 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, consisting in replacing the distribution function by a depth function (see Zuo and Serfling [16]).

Roughly speaking, a depth function is a mapping

$$D : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}_+$$

which provides a $P_{\mathbf{X}}$ -based *center-outward* ordering of points in \mathbb{R}^d , where $\mathcal{P}(\mathbb{R}^d)$ denotes the set of all probability measures on \mathbb{R}^d . Thus, in order to deal with risk regions, we will consider the *lower*-level sets of a depth function and propose a depth-based CCTE defined by :

$$\text{CCTE}_{D,\alpha}(Y, \mathbf{X}) := \mathbb{E}[Y | \mathbf{X} \in \mathcal{L}_D(\alpha)], \quad \alpha > 0, \quad (1.4)$$

where $\mathcal{L}_D(\alpha) = \{x \in \mathbb{R}^d : D(x, P_{\mathbf{X}}) \leq \alpha\}$ is the α -depth based lower level set.

The paper is organized as follows. In Section 2, we introduce some notations, and tools and the mathematical definition of a depth function. Section 3 is devoted to our main results : in Section 3.1, a construction and consistency and convergence rates of an estimator of our CCTE_D are given in a general setting, in Section 3.2 we study the general asymptotic behavior of our estimator of the level set $\mathcal{L}_D(\alpha)$, and in Section 3.3 we provide consistency results and convergence rates of the CCTE_D in the particular case of *Mahalanobis* depth. Illustrations and simulations are presented in Section 4. Finally, proofs are postponed to section 5.

2 Notations and definitions

This section is dedicated to introducing some useful notations and tools.

2.1 General Notations

Let \emptyset be the empty set, $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$ be the set of positive integers, and $\mathcal{P} := \mathcal{P}(\mathbb{R}^d)$ be the set of all probability measures on \mathbb{R}^d , $d \geq 1$. When dealing with random variables, we assume that they are defined on a common underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Given a r.v \mathbf{X} with distribution $P_{\mathbf{X}} := P \in \mathcal{P}$ and an i.i.d sample $S_n := (\mathbf{X}_i)_{1 \leq i \leq n}$ of size $n \in \mathbb{N}^*$ with distribution P and independent of \mathbf{X} , we denote by $P_n := \sum_{i=1}^n \delta_{\mathbf{X}_i}$ the empirical measure based on this finite sample. For notational convenience, we denote by \mathbb{E}_P the mathematical expectation under P , and by $\mathbb{E}_{S_n}[Z] := \mathbb{E}[Z | \mathbf{X}_1, \dots, \mathbf{X}_n]$ the conditional expectation of Z knowing $\mathbf{X}_1, \dots, \mathbf{X}_n$. Moreover, denoting $\Phi(A) := P_{\mathbf{X}}[A] = P[A]$ for any Borel-set $A \subset \mathbb{R}^d$, we denote by $P_{S_n}[A] := \Phi[A(X_1, \dots, X_n)]$, where $A := A(X_1, \dots, X_n)$ is a subset of \mathbb{R}^d which depends on the data X_1, \dots, X_n . Thus, $A(X_1, \dots, X_n)$ is a random subset, so that $P_{S_n}[A]$ is a r.v. Furthermore, for any real number $q > 0$, let $\mathbb{L}^q(\Omega) := \mathbb{L}^q(\Omega, \mathcal{A}, \mathbb{P})$ denote the vector space of real-valued random variables U for which $\mathbb{E}[|U|^q] < +\infty$.

Unambiguously, we denote by $\|\cdot\|$ the Euclidean norm in \mathbb{R}^d and by $\|\|\cdot\|\|$ the matrix norm induced by the Euclidean norm in \mathbb{R}^d , i.e for any $d \times d$ real matrix M ,

$$\|\|M\|\| := \sup_{x \neq 0} \frac{\|Mx\|}{\|x\|} = \sup_{\|x\|=1} \|Mx\|.$$

For $p \in [1, +\infty]$, we also denote by

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

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

the $\mathbb{L}^p(\mathbb{R}^d, \lambda_d)$ norm of f w.r.t the Lebesgue measure 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,

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

Now, given two sequences of real numbers $(u_n)_{n \geq 1}$ and $(v_n)_{n \geq 1}$, we recall that $u_n = O_n(v_n)$ 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 $(X_n)_{n \in (\mathbb{N}^*)^r}$, $r \geq 1$, 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 stochastic dominance

$$\mathbf{X}_n = \mathcal{O}_{\mathbb{P}, n}(\mathbf{u}_n) \stackrel{\text{def}}{\iff}$$

$$\forall \varepsilon > 0, \exists M_\varepsilon > 0, \exists N_\varepsilon \geq 1, \forall n := (n_1, \dots, n_r) \in (\mathbb{N}^*)^r,$$

$$\min_{1 \leq i \leq r} n_i \geq N_\varepsilon \implies \mathbb{P}[X_n \geq M_\varepsilon \cdot \mathbf{u}_n] \leq \varepsilon.$$

2.2 Depth functions

In this section, we formally introduce the definition of a statistical multivariate depth function as in Zuo and Serfling [16] (Definition 2.1 in [16]). Let us begin with some common multivariate symmetry notions which have been widely used in the literature (the interested reader can refer to Zuo and Serfling [16], Liu [12], and Beran and Millar [3]). In what follows, we review some standard symmetry notions.

- C -symmetry : a random vector $\mathbf{X} \in \mathbb{R}^d$ is *centrally-symmetric* (or C -symmetric) about $\theta \in \mathbb{R}^d$ if $\mathbf{X} - \theta \stackrel{d}{=} \theta - \mathbf{X}$.

- *A*-symmetry (Liu, 1990) : \mathbf{X} is said to be *angularly-symmetric* (or *A*-symmetric) about θ , if $(\mathbf{X} - \theta)/\|\mathbf{X} - \theta\|$ is centrally symmetric about the origin.

- *H*-symmetry (Zuo and Serfling, 2000) : \mathbf{X} is said to be *halfspace-symmetric* (or *H*-symmetric) about θ if $P_{\mathbf{X}}[H] \geq 1/2$ for every closed halfspace H containing θ .

Note that, it is easily established that *C*-symmetry \Rightarrow *A*-symmetry \Rightarrow *H*-symmetry. In a natural terminology, θ is called the *center* of the distribution $P_{\mathbf{X}}$.

Definition 2.1 (Zuo and Serfling [16]). *A statistical depth function is a mapping $D : \mathbb{R}^d \times \mathcal{P} \rightarrow \mathbb{R}$ which is bounded, non negative, measurable in its first argument and satisfying :*

(D1) Affine invariance : for any $P_{\mathbf{X}} \in \mathcal{P}$, $b \in \mathbb{R}^d$, and any invertible size d matrix A , $D(Ax + b, P_{A\mathbf{X}+b}) = D(x, P_{\mathbf{X}})$

(D2) Maximality at center : for any $P_{\mathbf{X}} \in \mathcal{P}$ having a unique center $\theta \in \mathbb{R}^d$ (for one of the symmetry notions previously presented), $D(\theta, P_{\mathbf{X}}) = \sup_{x \in \mathbb{R}^d} D(x, P_{\mathbf{X}})$

(D3) Monotonicity relative to deepest point : for any $P_{\mathbf{X}}$ having deepest point θ i.e. $D(\theta, P_{\mathbf{X}}) = \sup_{x \in \mathbb{R}^d} D(x, P_{\mathbf{X}})$, $D(x, P_{\mathbf{X}}) \leq D(\alpha x + (1 - \alpha)\theta, P_{\mathbf{X}})$ holds for $\alpha \in [0, 1]$

(D4) Vanishing at infinity : $D(x, P_{\mathbf{X}}) \rightarrow 0$ as $\|x\| \rightarrow \infty$, for each $P_{\mathbf{X}} \in \mathcal{P}$.

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. As far as property **(D2)** is concerned, for a distribution having a unique "center" i.e., the point of symmetry with respect to some notion of multivariate symmetry, the depth function should attain its maximum value at this center. 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. Last but not least, property **(D4)** implies that the depth of a point x approaches zero as $\|x\|$ approaches infinity. Note that **(D3)** and **(D4)** mean that the upper level sets

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

are bounded and starshaped about the point of maximum depth.

3 Main results

In this section, 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 probability under $P \in \mathcal{P}$ of the volume of the symmetric difference is provided.

3.1 General Covariate-Conditional-Tail-Expectation consistency

Fix a depth function $D : \mathbb{R}^d \times \mathcal{P} \rightarrow \mathbb{R}$ and a distribution $P \in \mathcal{P}$. We denote

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

Consider a couple (Y, \mathbf{X}) s.t. Y is a real random variable which is dependent on a random vector $\mathbf{X} \in \mathbb{R}^d$ with distribution P . In Definition 3.1, we formally define our CCTE_D and propose an estimator of the latter. For $n_1, n_2 \geq 1$, let

$$\begin{aligned} \tilde{S}_{n_1} &:= (\tilde{\mathbf{X}}_i)_{i=1, \dots, n_1} \text{ be an i.i.d } n_1\text{-sample from } P, \text{ and} \\ S_{n_2} &:= ((Y_i, \mathbf{X}_i))_{i=1, \dots, n_2} \text{ be an i.i.d } n_2\text{-sample from } P_{(Y, \mathbf{X})}, \end{aligned}$$

s.t. \tilde{S}_{n_1} and S_{n_2} are independent. Now, we define the α -lower level set of D and its plug-in estimator based on \tilde{S}_{n_1} by

$$\begin{aligned} \mathcal{L}_D(\alpha) &= \mathcal{L}_D(\alpha, P) := \{x \in \mathbb{R}^d : D(x, P) \leq \alpha\}, \text{ and} \\ \mathcal{L}_{n_1}(\alpha) &:= \mathcal{L}_D(\alpha, \tilde{P}_{n_1}) = \left\{x \in \mathbb{R}^d : D_{n_1}(x) := D(x, \tilde{P}_{n_1}) \leq \alpha\right\}, \end{aligned}$$

where \tilde{P}_{n_1} is the empirical measure based on the sample $\tilde{S}_{n_1} := (\tilde{\mathbf{X}}_i)_{1 \leq i \leq n_1}$. Finally, we provide the definition of our CCTE_D and its associated estimator.

Definition 3.1 (Depth-based Covariate-Conditional-Tail-Expectation). *Let $\mathbf{X} \in \mathbb{R}^d$ be a random vector with distribution $P \in \mathcal{P}$ 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-Conditionale-Tail-Expectation at level α is defined by :*

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

(ii) Its estimator based on the sample S_{n_2} is given by :

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

with the convention $0/0 = 0$.

Our first result, namely Theorem 3.2, 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:

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

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

where $A \Delta B = (A \setminus B) \cup (B \setminus A)$ is the symmetric difference between A and B .

In the spirit of Di Bernardino et al. [9], Theorem 3.2 states that, under some conditions, the CCTE_D estimator is consistent with at most a convergence rate $O(\sqrt{n_2})$. Remark that in Theorem 3.2, the r -th moment of Y is only involved in the rate (v_{n_1}) . Note that in our setting, the boundaries of the depth-based level sets at hand are compact, contrary to the non-compact setting studied in Di Bernardino et al. [9].

Theorem 3.2. *Let $\alpha > 0$ and $P \in \mathcal{P}$. 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*

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

Furthermore, following the approach of Di Bernardino et al. [9], Assumption **(H0)** can be replaced by Assumption **(H1)** and one can derive a similar result to Theorem 3.2 (c.f. Corollary 3.3) :

(H1): (i) there exists an increasing sequence of positive real numbers $(v_{n_1})_{n_1 \geq 1}$ s.t.

$$\lambda_d(\mathcal{L}_{n_1}(\alpha) \Delta \mathcal{L}_D(\alpha)) = \mathcal{O}_{P,n_1}(v_{n_1}^{-1}), \text{ and}$$

(ii) P is absolutely continuous with density function $f \in \mathbb{L}^p(\mathbb{R}^d, \lambda_d)$ for some $p \in (1, +\infty]$.

Corollary 3.3. *Let $\alpha > 0$ and $P \in \mathcal{P}$. Assume that $P[\mathcal{L}_D(\alpha)] > 0$, and that there exists $r \in [2, +\infty]$ s.t. $Y \in \mathbb{L}^r(\Omega)$. Let $(v_{n_1})_{n_1}$ satisfy **(H1)**, then*

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

Proof. It is sufficient to show that under the assumptions of Corollary 3.3, assumption **(H0)** of Theorem 3.2 is satisfied by the sequence $(v_{n_1}^{1-\frac{1}{p}})_{n_1}$. When $p \in (1, +\infty)$, it holds almost-surely

$$\begin{aligned} v_{n_1}^{1-\frac{1}{p}} P_{\widehat{\mathcal{S}}_{n_1}}[\mathcal{L}_{n_1}(\alpha) \Delta \mathcal{L}_D(\alpha)] &= v_{n_1}^{1-\frac{1}{p}} \int \mathbf{1}_{\mathbf{x} \in \mathcal{L}_{n_1}(\alpha) \Delta \mathcal{L}_D(\alpha)} f(\mathbf{x}) d\mathbf{x} \\ &\leq v_{n_1}^{1-\frac{1}{p}} \lambda_d(\mathcal{L}_D(\alpha) \Delta \mathcal{L}_{n_1}(\alpha))^{1-\frac{1}{p}} \|f\|_{p,\lambda} \quad (\text{H\"older}) \\ &\leq \underbrace{(v_{n_1} \lambda_d(\mathcal{L}_D(\alpha) \Delta \mathcal{L}_{n_1}(\alpha))^{1-\frac{1}{p}})}_{\mathcal{O}_{P,n_1}(1) \text{ (H1)(i)}} \underbrace{\|f\|_{p,\lambda}}_{< \infty \text{ (H1)(ii)}}. \end{aligned}$$

When $p = +\infty$, the result is trivially valid by bounding f by its essential supremum. \square

3.2 Consistency of general depth-based level sets in terms of the Hausdorff distance

In this section, the problem of interest is to study the conditions under which assumption **(H1)(i)** is satisfied : that would provide a rate of convergence for the general CCTE_D . This means studying the rate of convergence of the volume of the symmetric difference between \mathcal{L}_n and \mathcal{L}_D . It happens that, by controlling the Hausdorff distance between the respective boundaries $\partial \mathcal{L}_D(\alpha)$ and $\partial \mathcal{L}_n(\alpha)$ of those two sets, one can control the volume of the symmetric difference. Remark first that the Hausdorff distance between those sets is asymptotically well defined since the boundaries of the level sets are compact and non empty (c.f. Remark 3.4 together with Remark 3.7).

Remark 3.4. *On one hand, if $\alpha \in (0, \alpha_{\max}(P))$, the sets $\{x : D(x, P) \geq \alpha\}$ and $\{x : D(x, P) < \alpha\}$ are both non-empty, and since D is vanishing at infinity, the empirical level set $\{x : D(x, P_n) < \alpha\}$, $n \geq 1$, is non empty. Thus, $\mathcal{L}_D(\alpha)$ and its boundary $\partial \mathcal{L}_D(\alpha)$ are non-empty. On the other hand, if the empirical depth a.s. (almost-surely) converges pointwise to its true version on \mathbb{R}^d then (cf. Theorem 4.1 in Dyckerhoff [10]), for any $\alpha \in (0, \alpha_{\max}(P))$ and a.s. for any n large enough, the upper-level set $\{x : D(x, P_n) \geq \alpha\}$ is non-empty as well. Thus, a.s, for n large enough, $\partial \mathcal{L}_n(\alpha)$ is non-empty.*

Proposition 3.5 is a slight modification of the Proposition 3.1 in the Ph.D. thesis of Rodríguez-Casal [13] adapted to depth functions. We introduce the following assumption **(L)** which 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 $\alpha > 0$.

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

Proposition 3.5. *Let $D : \mathbb{R}^d \times \mathcal{P} \rightarrow \mathbb{R}_+$ be a multivariate depth function. Let $\alpha > 0, 0 < \varepsilon < \alpha$ and $P \in \mathcal{P}$ be fixed. Denoting $D(x) := D(x, P)$, assume that*

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

(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 .*

*Then D satisfies Assumption **(L)**, with $A = \frac{2}{m_\nabla}$.*

The following result is an adapted version of Theorem 2 in Cuevas et al. [5] to depth functions, where we weaken the assumption of continuity of the empirical depth function by an assumption of upper-semicontinuity.

Theorem 3.6. *Let $D : \mathbb{R}^d \times \mathcal{P} \rightarrow \mathbb{R}_+$ be a depth function, $P \in \mathcal{P}$ and $\alpha \in (0, \alpha_{\max}(P))$. Denote by P_n an estimator of P , $n \geq 1$. Suppose that $x \in \mathbb{R}^d \mapsto D(x, P) := D(x)$ is a continuous function, and $x \in \mathbb{R}^d \mapsto D_n(x) := D(x, P_n)$ is upper semi-continuous \mathbb{P} -almost surely for any $n \geq 1$, and that*

$$\|D_n - D\|_{\infty, \mathbb{R}^d} \xrightarrow[n \rightarrow \infty]{a.s.} 0.$$

Under the same assumptions as in Proposition 3.5, it holds that

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

Remark 3.7. *Recall that, for n large enough, $\partial\mathcal{L}_n(\alpha) \neq \emptyset$ \mathbb{P} -a.s., so that the Hausdorff distance is well-defined for large n (see Remark 3.4 as well). Indeed, in Theorem 3.6, one can underline two main properties of a depth function, namely : upper semi-continuity which is equivalent to having closed depth-based upper level sets, and Property **(D4)** (vanishing at infinity) which guarantees that the upper level sets are bounded. As a consequence, the Hausdorff distance is well defined since $\partial\mathcal{L}_n(\alpha)$ is closed by definition and is bounded as it is included in the compact α -upper level set ($\alpha > 0$). The same applies for $\partial\mathcal{L}(\alpha)$ (or immediately by continuity of D). What is more, $\alpha \in (0, \alpha_{\max}(P))$ implies $\partial\mathcal{L}(\alpha) \neq \emptyset$ (cf. Remark 3.4).*

3.3 MHD-depth based Covariate-Conditional-Tail-Expectation consistency

The *Mahalanobis* depth function (Example 2.5 in Zuo and Serfling [16]) is a depth function in the sense of Definition 2.1 (see Definition 3.8), and is smooth as a function of x (which implies the upper-semicontinuity property in the empirical case as well).

In order to study the rate of convergence of the CCTE_D based on *MHD*, we check here Assumption **(H1)(i)**. According to Section 3.2, the problem reduces to studying the rate of convergence of $\|D_n - D\|_\infty$, in probability, when $D = \text{MHD}$ (c.f. Section 3.2, Theorem 3.6).

Definition 3.8 (*Mahalanobis depth*, Zuo and Serfling [16]). Let $\mathbf{X} \in \mathbb{R}^d$ be a random vector with distribution $P \in \mathcal{P}$. The Mahalanobis depth is defined by

$$\text{MHD}(x, P) = \begin{cases} (1 + d_{\Sigma_{\mathbf{X}}}^2(x, \mu_{\mathbf{X}}))^{-1} & \text{if } \mathbb{E}_P[\|\mathbf{X}\|^2] < +\infty \\ 0 & \text{if } \mathbb{E}_P[\|\mathbf{X}\|^2] = +\infty \end{cases}$$

where $\mu_{\mathbf{X}} = \mathbb{E}_P[\mathbf{X}]$ is the mean vector of \mathbf{X} and $\Sigma_{\mathbf{X}}$ is its covariance matrix (which is assumed to be invertible) and

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

is the Mahalanobis distance.

Remark 3.9. Note that, the above definition of *MHD* depth is introduced as such in order to highlight the fact that, it is restricted to distributions with second moment while still remaining a depth function in the sense of Definition 2.1. Furthermore, for a fixed distribution P , the function $x \in \mathbb{R}^d \mapsto \text{MHD}(x, P)$ is infinitely differentiable, concave, and has $x = \mu_{\mathbf{X}}$ as unique critical point, thus $\mu_{\mathbf{X}} = \arg \max_{x \in \mathbb{R}^d} \text{MHD}(x, P)$. And $\alpha_{\max}(P) :=$

$$\max_{x \in \mathbb{R}^d} \text{MHD}(x, P) = 1.$$

In Zuo and Serfling [16], one can also find the following result which underlines the properties of *MHD* as a depth function (Theorem 2.10, Zuo and Serfling [16]).

Proposition 3.10 (Zuo and Serfling [16]). Let $\mathbf{X} \in \mathbb{R}^d$ be a random vector with distribution $P \in \mathcal{P}$. Assume \mathbf{X} is symmetric (for some notion of symmetry) about $\theta \in \mathbb{R}^d$. If $\mu_{\mathbf{X}} = \theta$, then *MHD* is a statistical depth function in the sense of Definition 2.1.

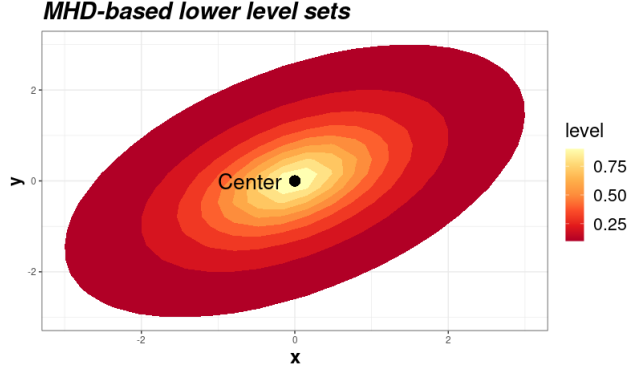


Figure 1: Theoretical lower-level sets based on $MHD(\cdot, P)$, with P the law of a Gaussian vector in \mathbb{R}^2 .

A natural estimator of MHD is given by

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

where $\hat{\mu}_n$ is the empirical mean vector and $\hat{\Sigma}_n$ is the empirical covariance matrix.

In Proposition 3.11, we provide the particular version of Proposition 3.5 associated to MHD and MHD_n depths.

Proposition 3.11. *Let $\mathbf{X} \in \mathbb{R}^d$ be a random vector from $P \in \mathcal{P}$ s.t. $\mathbb{E}_P[\|\mathbf{X}\|^2] < \infty$. Consider $D(x) := MHD(x) := MHD(x, P)$ and fix $\alpha \in (0, 1)$ and $0 < \varepsilon < \alpha \wedge (1 - \alpha)$. Denoting $\mathcal{K}_\varepsilon(\alpha) := D^{-1}([\alpha \pm \varepsilon])$, it holds*

- (i) $m_\nabla := \inf_{x \in \mathcal{K}_\varepsilon(\alpha)} \|(\nabla D)_x\| > 0$, and
- (ii) $d_H(\partial \mathcal{L}_{MHD}(\alpha), \partial \mathcal{L}_n(\alpha)) = O_{n \rightarrow \infty}(\|MHD_n - MHD\|_{\infty, \mathbb{R}^d})$, \mathbb{P} -a.s.

Now, Theorem 3.12 is a useful result in which we provide the rate of convergence (in probability) of MHD_n to its population version MHD uniformly on \mathbb{R}^d .

Theorem 3.12. *Let $\mathbf{X} = (X^{(1)}, \dots, X^{(d)})$ be a random vector with distribution $P \in \mathcal{P}$ satisfying $\mathbb{E}_P[|X^{(i)}|^4] < \infty$ for all $1 \leq i \leq d$. Then, it holds that*

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

Finally, in Theorem 3.13, we derive the specific rate of convergence for the CCTE based on MHD -depth.

Theorem 3.13. *Let $P \in \mathcal{P}$, $D(\cdot, P) = MHD(\cdot, P)$ and $\alpha \in (0, 1)$. Assume $P[\mathcal{L}_D(\alpha)] > 0$. Under the assumptions of Theorem 3.12, and Assumption **(H1)(ii)** and assuming moreover that there exists $r \in [2, +\infty]$ s.t. $Y \in \mathbb{L}^r(\Omega)$, it holds that*

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

4 Simulations and illustrations

In this section, we provide an illustration of Theorem 3.13. 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 :

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

where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, $\sigma^2 > 0$, is a gaussian noise. In our simulations we will take $\sigma^2 = 0.005$. Here, we choose the squared euclidian norm which has fourth moment under P , defined by : $\|\mathbf{x}\|^2 = |x_1|^2 + |x_2|^2$ (see Figure 3). Moreover, we consider dependent risk factors X_1 and X_2 via a bivariate Frank Copula with Gumbel marginals with parameter $(\mu, \beta) = (0, 0.25)$ and $(-0.5, 0.25)$ respectively (Figure 2).

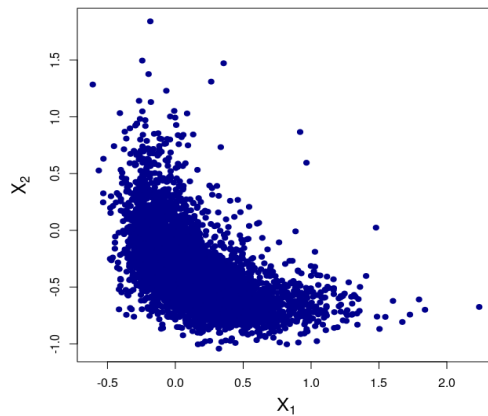


Figure 2: Sample of dependent Gumbel marginals via a Frank Copula.

Note that the above example satisfies the assumptions of Theorem 3.13.

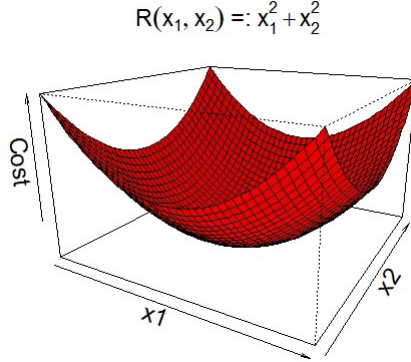


Figure 3: *Monotonic* function : $R(x_1, x_2) = |x_1|^2 + |x_2|^2$.

Here we compare $\widehat{\text{CCTE}}_{MHD,\alpha}^{n_1,n_2}$ with the theoretical $\text{CCTE}_{MHD,\alpha}$ for *Ma-*
halanobis depth. For the sake of simplicity, we take $n_1 = n_2 = n$. As-
suming that $p, r \approx +\infty$ with our sample, we obtain that $|\widehat{\text{CCTE}}_{MHD,\alpha}^n -$
 $\text{CCTE}_{MHD,\alpha}|$ decays to zero at most with a convergence rate $o(\sqrt{n})$.

We provide a deterministic approximation of the true mean vector and
covariance matrix, m and Σ respectively. However, due to the complexity of
the level-sets as domains of integration in the computation of the CCTE_D ,
we perform a Monte Carlo procedure to fix the "true" value of the CCTE_D
based on a sample of size 10^8 (without noise), that is:

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

Recall that Theorem 3.13 illustrates convergence rates in probability, how-
ever, for the sake of computational simplicity, we provide \mathbb{L}^1 -estimation for
the CCTE_D (which implies convergence results in probability). More pre-
cisely, we denote $\overline{\widehat{\text{CCTE}}}_{\alpha}^n := \overline{\widehat{\text{CCTE}}}_{\alpha,MHD}^n$ the mean of the $\widehat{\text{CCTE}}_{MHD,\alpha}^n$
based on 400 simulations. The empirical standard deviation is

$$\hat{\sigma} = \sqrt{\frac{1}{399} \sum_{j=1}^{400} \left(\widehat{\text{CCTE}}_{\alpha,j}^n - \overline{\widehat{\text{CCTE}}}_{\alpha}^n \right)^2},$$

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

RMAE, is defined as follows :

$$\text{RMAE} := \text{RMAE}_{n,\alpha} = \frac{1}{400} \sum_{j=1}^{400} \frac{|\widehat{\text{CCTE}}_{\alpha,j}^n - \text{CCTE}_{MHD,\alpha}(Y, \mathbf{X})|}{|\text{CCTE}_{MHD,\alpha}(Y, \mathbf{X})|}.$$

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 (particularly Theorem 3.13), we work here with the RMAE for which we provide \mathbb{L}^1 -estimation as well.

In Table 1, we provide the above estimations for different values of α and sample size n .

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
	$\hat{\sigma}$	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
	$\hat{\sigma}$	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
	$\hat{\sigma}$	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
	$\hat{\sigma}$	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
	$\hat{\sigma}$	0.0283	0.0118	0.0033	0.002	0.0019
	RMAE	0.0158	0.0106	0.0055	0.0042	0.0041

Table 1: \mathbb{L}^1 -estimation of $\text{CCTE}_{\alpha,MHD}(\mathbf{X}, Y)$ and associated RMAE for bivariate Frank Copulas with Gumbel marginals.

According to the results in Table 1, we observe that the error RMAE_n decreases as the sample size n increases, as one may expect. Besides, remark that for low levels α ($\alpha = 0.1$) and sample sizes n ($n = 100, 1000$), 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 $A = 2/m_{\nabla}$ is large since m_{∇} approaches zero (see the proof of Proposition 3.5 and Theorem 3.6 in Section 5), meaning that the constant

bounding the error (RMAE) becomes large.

In Table 2, we also provide \mathbb{L}^1 -convergence rates : $V_{n,\alpha,\delta} = v_n(\delta) \cdot \text{RMAE}_{n,\alpha}$, with $v_n(\delta) = n^{\frac{1}{2}-\delta}$ for different values of δ (see Figure 4).

α	δ	n				
		100	1000	5000	10000	50000
0.1	0.05	2.9656	2.3641	2.1971	2.2773	2.0529
	0.01	3.5654	3.1164	3.0889	3.2916	3.1646
	0	3.7335	3.3393	3.3635	3.6092	3.5262
	-0.01	3.9094	3.5782	3.6625	3.9574	3.9292
0.2	0.05	1.8749	1.5589	1.5602	1.5217	1.3821
	0.01	2.2542	2.0551	2.1935	2.1996	2.1305
	0	2.3604	2.2021	2.3885	2.4118	2.374
	-0.01	2.4716	2.3595	2.6009	2.6445	2.6452
0.5	0.05	0.9666	0.8522	0.8298	0.7407	0.7124
	0.01	1.1411	1.1234	1.1666	1.0707	1.0982
	0	1.2169	1.2038	1.2703	1.1740	1.2236
	-0.01	1.2742	1.2899	1.3832	1.2872	1.3635
0.8	0.05	0.7577	0.7119	0.6362	0.5989	0.5481
	0.01	0.9111	0.9385	0.8945	0.8656	0.845
	0	0.9539	1.0056	0.9740	0.9491	0.9415
	-0.01	0.9988	1.0775	1.0606	1.0407	1.0491
0.9	0.05	0.7415	0.6949	0.6200	0.5952	0.5281
	0.01	0.8914	0.9160	0.8717	0.8603	0.8141
	0	0.9334	0.9815	0.9492	0.9433	0.9072
	-0.01	0.9774	1.0517	1.0335	1.0344	1.0108

Table 2: Estimated $n^{\frac{1}{2}-\delta} \cdot \text{RMAE}_{n,\alpha}$ based on *MHD* for bivariate Frank Copulas with Gumbel marginals.

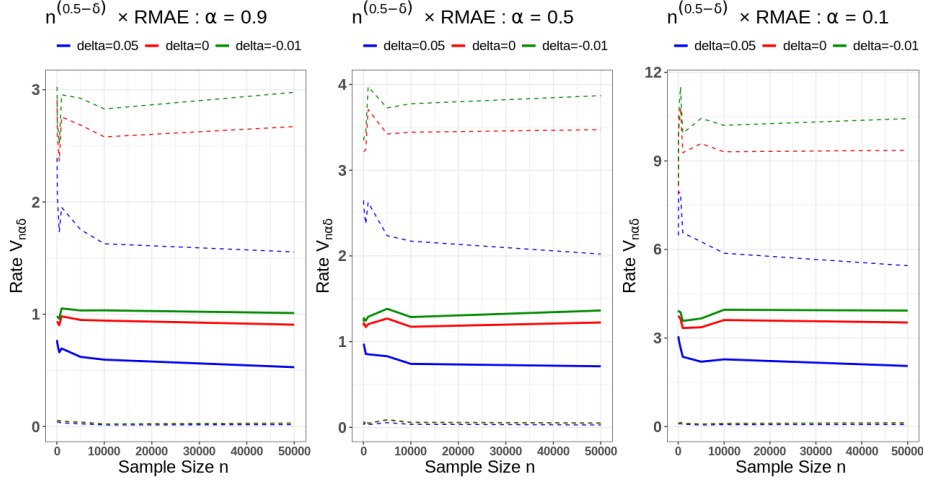


Figure 4: Estimation of the convergence rates $V_{n,\alpha,\delta}$ associated to RMAE, based on MHD for bivariate Frank Copulas with Gumbel marginals.

In Table 2 and Figure 4, the parameter δ allows us to check that the critical regime is $\delta = 0$ which corresponds to the $n^{1/2}$ -convergence rate associated to the RMAE.

5 Proofs of Section 3

Proofs of Section 3.1.

Proof of Theorem 3.2. We can write

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

The proof of Theorem 3.2 is a modified version of the proof of Theorem 5.1 in Di Bernardino et al. [9]. The latter focuses on distribution functions instead of depth functions. Besides, in the proof of Theorem 3.2, 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 3.2 is based on the two following preliminary results, Lemma 5.1 and Lemma 5.2. \square

Lemma 5.1. *Under assumptions of Theorem 3.2, it holds that*

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

Proof of Lemma 5.1. On the event $\left\{ P_{\tilde{S}_{n_1}}[\mathcal{L}_{n_1}(\alpha)] > 0 \right\}$, it holds

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

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

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

so that

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

and

$$\mathbb{P} \left[\left\{ P_{\tilde{\mathcal{S}}_{n_1}}[\mathcal{L}_{n_1}(\alpha)] > 0 \right\} \right] \xrightarrow[n_1 \rightarrow \infty]{} 1. \quad (5.1)$$

On the one hand,

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

so we obtain

$$v_{n_1}^{1-\frac{1}{r}} \left| P_{\tilde{\mathcal{S}}_{n_1}}[\mathcal{L}_{n_1}(\alpha)] - P[\mathcal{L}_D(\alpha)] \right| \xrightarrow[n_1 \rightarrow \infty]{\mathbb{P}} 0. \quad (5.2)$$

On the other hand, using Hölder inequality

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

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

$$v_{n_1}^{1-\frac{1}{r}} \left| \mathbb{E}_{\tilde{\mathcal{S}}_{n_1}}[Y \mathbf{1}_{\mathcal{L}_{n_1}(\alpha)}] - \mathbb{E}[Y \mathbf{1}_{\mathcal{L}_D(\alpha)}] \right| = \mathcal{O}_{P,n_1}(1). \quad (5.3)$$

Since the convergence to zero in probability implies the $\mathcal{O}_P(1)$ result, the lemma follows directly from (5.1), (5.2), and (5.3). The case $r = +\infty$ is analogous. □

Lemma 5.2. *Under assumptions of Theorem 3.2, we obtain*

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

Proof of Lemma 5.2. First, we distinguish the event in which $\frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{1}_{\mathbf{X}_i \in \mathcal{L}_{n_1}(\alpha)} = 0$, then the one in which $\frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{1}_{\mathbf{X}_i \in \mathcal{L}_{n_1}(\alpha)} \neq 0$.

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

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

Since $\varepsilon \in (0, P[\mathcal{L}_D(\alpha)])$ and $P_{\tilde{S}_{n_1}}[\mathcal{L}_{n_1}(\alpha)] \xrightarrow[n_1 \rightarrow \infty]{\mathbb{P}} P[\mathcal{L}_D(\alpha)]$ (see Lemma 5.1), we obtain $\mathbb{P} \left[\frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{1}_{\mathbf{X}_i \in \mathcal{L}_{n_1}(\alpha)} = 0 \right] \xrightarrow[n_1, n_2 \rightarrow \infty]{} 0$. Now on the event $\left\{ \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{1}_{\mathbf{X}_i \in \mathcal{L}_{n_1}(\alpha)} \neq 0 \right\} \cap \left\{ P_{\tilde{S}_{n_1}}[\mathcal{L}_{n_1}(\alpha)] > 0 \right\}$, we can write

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

Let us first clarify the convergence of the denominator terms. Recall that under Assumption **(H0)** of Theorem 3.2,

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

and $\mathbb{P} \left[\left\{ P_{\tilde{S}_{n_1}}[\mathcal{L}_{n_1}(\alpha)] > 0 \right\} \right] \xrightarrow[n_1 \rightarrow \infty]{} 1$ (see the proof of Lemma 5.1). Next, we prove

$$n_2^{\frac{1}{2}} \left(\frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{1}_{\mathbf{X}_i \in \mathcal{L}_{n_1}(\alpha)} - P_{\tilde{S}_{n_1}}[\mathcal{L}_{n_1}(\alpha)] \right) = \mathcal{O}_{P, n_1, n_2}(1), \quad (5.4)$$

so that we obtain

$$\frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{1}_{\mathbf{X}_i \in \mathcal{L}_{n_1}(\alpha)} - P_{\tilde{S}_{n_1}}[\mathcal{L}_{n_1}(\alpha)] \xrightarrow[n_1, n_2 \rightarrow \infty]{\mathbb{P}} 0, \quad (5.5)$$

and

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

Let us prove (5.4). Let $n_1 \geq 1$ 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)

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

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

$$\begin{aligned} \sup_{n_1, n_2 \geq 1} \mathbb{P} \left[\left| \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{1}_{\mathbf{X}_i \in \mathcal{L}_{n_1}(\alpha)} - P_{\tilde{S}_{n_1}}[\mathcal{L}_{n_1}(\alpha)] \right| \geq M_\varepsilon \right] &\leq \frac{1}{n_2 \left(\frac{M_\varepsilon}{n_2^{\frac{1}{2}}} \right)^2} \\ &= \varepsilon, \end{aligned}$$

which means that (5.4) is satisfied. Similarly, we obtain

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

Hence the result. \square

Proofs of Section 3.2.

Remark 5.3. Among the four properties of a depth function, the only necessary property in Proposition 3.5 is **(D4)** (vanishing at infinity). The latter guarantees that the set $\mathcal{K}_\varepsilon(\alpha)$ is compact in \mathbb{R}^d . Indeed, as D satisfies **(D4)**, the assumption $0 < \varepsilon < \alpha$ implies $\mathcal{K}_\varepsilon(\alpha)$ is bounded, moreover under (i), 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.

Proof of Proposition 3.5. For the sake of simplicity, denote by $[\alpha \pm \varepsilon]$ the interval $[\alpha - \varepsilon, \alpha + \varepsilon]$. Under the assumptions of Proposition 3.5, $K_\varepsilon(\alpha) := D^{-1}([\alpha \pm \varepsilon])$ is compact and

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

(cf. Remark 5.3). We state the following useful lemma.

Lemma 5.4. Under the assumptions of Proposition 3.5, there exist $N := N_\varepsilon \geq 1$, some points $x_i := x_{i,\varepsilon} \in \mathcal{K}_{\frac{\varepsilon}{2}}(\alpha)$, and some positive real numbers $r_i := r_{x_i} \in \mathbb{R}_+^*$, $1 \leq i \leq N$, s.t.

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

Proof of Lemma 5.4. 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.

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

that is,

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

Since we have an open cover of the compact set $K_{\frac{\varepsilon}{2}}(\alpha)$, then the latter has a finite 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.

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

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 5.4 applies, namely, there exists $1 \leq i_0 \leq N$ s.t. $x \in \mathcal{B}(x_{i_0}, r_{i_0}/2)$, 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, $z \in \mathcal{B}(x_{i_0}, r_{i_0}) \subset \mathcal{K}_\varepsilon(\alpha)$ (cf. Lemma 5.4). 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

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

then,

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

Using Cauchy-Schwarz inequality, it holds

$$\begin{aligned} |D(y_\lambda) - D(x) - \lambda \|(\nabla D)_x\|| &\leq \frac{\lambda^2}{2 \|(\nabla D)_x\|^2} \|(\nabla D)_x\| \| (HD)_{\bar{x}} \| \cdot \|(\nabla D)_x\| \\ &= \frac{\lambda^2}{2} \| (HD)_{\bar{x}} \|. \end{aligned}$$

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

$$D(x) + \lambda \|(\nabla D)_x\| - \frac{\lambda^2}{2} M_H \leq D(y_\lambda) \leq D(x) + \lambda \|(\nabla D)_x\| + \frac{\lambda^2}{2} M_H. \quad (5.6)$$

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

$$D(y_\lambda) \geq D(x) + \lambda \inf_{x \in \mathcal{K}_\varepsilon(\alpha)} \|(\nabla D)_x\| - \frac{\lambda^2}{2} M_H = D(x) + \lambda \left(m_\nabla - \lambda \frac{M_H}{2} \right) \quad (5.7)$$

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 (5.7),

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

Similarly, using the right hand side of inequality (5.6), 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

$$D(y_\lambda) \geq D(x) + \lambda \frac{m_\nabla}{2}, \quad (5.8)$$

$$D(y_{-\lambda}) \leq D(x) - \lambda \frac{m_\nabla}{2}. \quad (5.9)$$

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 :

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

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 $\sup_{x \in \{D = \beta\}} \text{dist}(x, \{D = \alpha\})$. 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 (5.9),

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

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 $x \in \{D = \beta\}$,

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

So

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

In order to get an upper bound for $\sup_{x \in \{D=\alpha\}} \text{dist}(x, \{D=\beta\})$, we use the inequality (5.8) by proceeding in a similar way.

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

Proof of Theorem 3.6. 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 3.6 $\varepsilon_n \rightarrow 0$ \mathbb{P} -a.s, so that \mathbb{P} -a.s, there exists an integer $n_0 := n_0(\omega) \geq 1$ (independent from 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 $n \geq n_0$,

$$\begin{aligned}
\text{dist}(x, \partial\mathcal{L}_n(\alpha)) &\leq \|x - z_n\| \\
&\leq \|x - u_n\| + \|u_n - z_n\| \\
&\leq \|u_n - x\| + \|u_n - l_n\| \\
&\leq \|u_n - x\| + \|u_n - x\| + \|x - l_n\| \\
&\leq 3A\varepsilon_n \\
&= 6A\|D_n - D\|_\infty.
\end{aligned}$$

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. $D_n(l_n) \leq \alpha$, and s.t. by continuity of D , $|D(x_n) - D(l_n)| \leq \varepsilon_n/2$. 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,

$$\begin{aligned}
D(x_n) &= D_n(l_n) - D_n(l_n) + D(l_n) - D(l_n) + D(x_n) \\
&\leq \alpha + \varepsilon_n/2 + \varepsilon_n/2,
\end{aligned}$$

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

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

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. □

Proofs of Section 3.3.

Proof of Proposition 3.11. (i) The function $MHD(\cdot)$ is infinitely differentiable on \mathbb{R}^d , and denoting $\mu = \mu_{\mathbf{X}}$, we can write for any $1 \leq k \leq d$,

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

So

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

Since $MHD(x) > 0$,

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

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,

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

(ii) It suffices to prove

$$\|MHD_n - MHD\|_{\infty, \mathbb{R}^d} \xrightarrow[n \rightarrow \infty]{a.s.} 0, \quad (5.10)$$

by recalling that $\alpha_{\max}(P) = 1$ for MHD depth. The result is hence a straight forward application of Proposition 3.5 and Theorem 3.6. In order to prove (5.10), one can refer to the computations in the proof of Theorem 3.12 and obtain the desired result knowing that $\hat{\Sigma}_n \xrightarrow[n \rightarrow \infty]{a.s.} \Sigma$, and $\hat{\mu}_n \xrightarrow[n \rightarrow \infty]{a.s.} \mu$. \square

Proof of Theorem 3.12. Let $x \in \mathbb{R}^d$. Denote $\mu := \mu_{\mathbf{X}}$ and $\Sigma := \Sigma_{\mathbf{X}}$, we can write

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

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,

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

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

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

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

$$\begin{aligned} & \frac{\left| \|\hat{\Sigma}_n^{-\frac{1}{2}}(\Sigma^{\frac{1}{2}}y + \mu - \hat{\mu}_n)\|^2 - \|y\|^2 \right|}{1 + \|y\|^2} \\ &\leq \left\| \Sigma^{\frac{1}{2}}\hat{\Sigma}_n^{-1}\Sigma^{\frac{1}{2}} - I_d \right\| + \|\hat{\Sigma}_n^{-\frac{1}{2}}(\mu - \hat{\mu}_n)\|^2 + \left\| \hat{\Sigma}_n^{-1} \right\| \cdot \left\| \Sigma^{\frac{1}{2}} \right\| \|\mu - \hat{\mu}_n\|. \end{aligned}$$

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

$$\begin{aligned} & \|MHD_n - MHD\|_{\infty, \mathbb{R}^d} \\ &\leq \left\| \Sigma^{\frac{1}{2}}\hat{\Sigma}_n^{-1}\Sigma^{\frac{1}{2}} - I_d \right\| + \left\| \hat{\Sigma}_n^{-\frac{1}{2}} \right\|^2 \|\mu - \hat{\mu}_n\| \left(\|\mu - \hat{\mu}_n\| + \left\| \Sigma^{\frac{1}{2}} \right\| \right) \\ &:= A_n(d), \end{aligned} \tag{5.11}$$

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

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

Furthermore, by the multivariate Central Limit theorem, it holds that $n^{\frac{1}{2}}(\hat{\mu}_n - \mu) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \Sigma)$. Thus (by the continuity theorem and Slutsky's lemma), $\left\| \left\| \hat{\Sigma}_n^{-\frac{1}{2}} \right\| \right\|^2 \|\mu - \hat{\mu}_n\| \left(\|\mu - \hat{\mu}_n\| + \left\| \left\| \Sigma^{\frac{1}{2}} \right\| \right\| \right)$ 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:

$$DF_A : H \in \mathcal{S}_d(\mathbb{R}) \mapsto DF_A(H) = -\Sigma^{\frac{1}{2}} A^{-1} H A^{-1} \Sigma^{\frac{1}{2}}. \quad (5.12)$$

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.

$$\sqrt{n}(\hat{\Sigma}_n - \Sigma) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, M^*), \quad (5.13)$$

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 $X \in B(A, r) \subset \mathcal{S}_d^+(\mathbb{R})$, $r > 0$, we can write :

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

Then, \mathbb{P} -almost surely, using (5.12)

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

This, together with (5.13), using the continuity theorem and Slutsky's Lemma, we obtain

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

so $\sqrt{n}R(\hat{\Sigma}_n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0$. In addition, from the continuity of $U \mapsto -\Sigma^{-\frac{1}{2}}U\Sigma^{-\frac{1}{2}}$ and using (5.13), we obtain $-\sqrt{n}\Sigma^{-\frac{1}{2}}(\hat{\Sigma}_n - \Sigma)\Sigma^{-\frac{1}{2}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} -\Sigma^{-\frac{1}{2}}E^*\Sigma^{-\frac{1}{2}}$. Therefore, by continuity of the matrix norm, we deduce that

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

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

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

□

Proof of Corollary 3.13. Denote $D := MHD$. Remark that the upper level sets based on MHD are ellipsoids in \mathbb{R}^d . It is sufficient to prove that under the assumptions of Corollary 3.13, Assumption **(H1)(i)** of Corollary 3.3 is satisfied i.e. $\lambda_d(\mathcal{L}_{n_1}(\alpha)\Delta\mathcal{L}_{n_1}(\alpha)) = \mathcal{O}_P(v_{n_1}^{-1})$ with $v_{n_1} = n_1^{\frac{1}{2}}$. The result is then a straight forward consequence of Corollary 3.3. 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

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

Since $\partial\mathcal{L}_D(\alpha)$ is closed and ℓ_{n_1} is small enough (\mathbb{P} -a.s. for n_1 large enough, according to Proposition 3.11 for MHD), then \mathbb{P} -a.s, (cf. Weyl [15], p. 461)

$$\begin{aligned} \lambda_d(\mathcal{L}_{n_1}(\alpha)\Delta\mathcal{L}_D(\alpha)) &\leq \lambda_d[\text{Tube}(\partial\mathcal{L}_D(\alpha), \ell_{n_1})] \\ &\approx A_{d-1}(\alpha)\ell_{n_1} \quad \text{for large } n_1, \quad (\text{Weyl}) \\ &\leq A_{d-1}(\alpha) \cdot C\|MHD_{n_1} - MHD\|_{\infty, \mathbb{R}^d} \quad \text{for large } n_1, \end{aligned}$$

and the above inequality is obtained by Proposition 3.11, where $A_{d-1}(\alpha)$ is the $(d-1)$ -dimensional volume of $\mathcal{L}_D(\alpha)$.

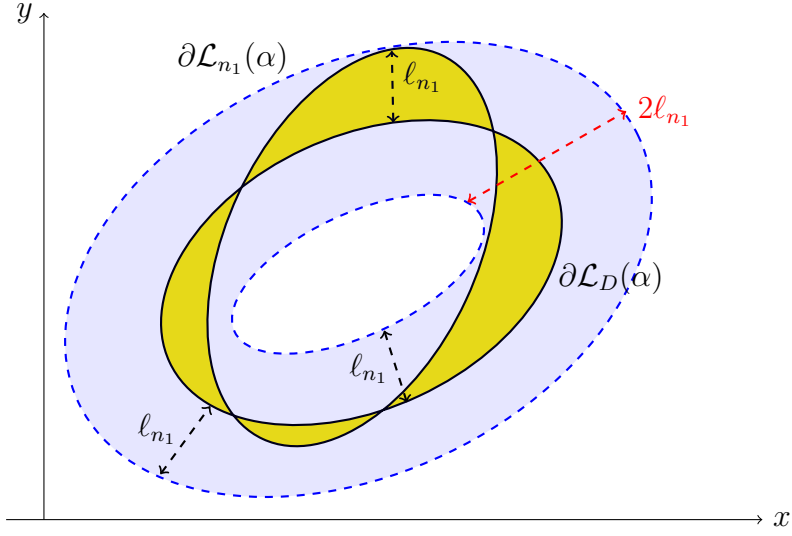


Figure 5: Illustration of $\lambda_d(\mathcal{L}_{n_1}(\alpha)\Delta\mathcal{L}_D(\alpha))$ (yellow), and the tube around $\mathcal{L}_D(\alpha)$ of radius ℓ_{n_1} (blue).

Finally, according to Theorem 3.12, $\|MHD_{n_1} - MHD\|_{\infty, \mathbb{R}^d} = \mathcal{O}_P\left(n^{-\frac{1}{2}}\right)$. Hence,

$$\lambda_d(\mathcal{L}_{n_1}(\alpha)\Delta\mathcal{L}_D(\alpha)) = \mathcal{O}_P\left(n^{-\frac{1}{2}}\right).$$

□

References

- [1] Agresti, A. (2002). *Categorical data analysis*. NJ: John Wiley & Sons, Inc.
- [2] 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.
- [3] Beran, R. J. and Millar, P. W. (1997). *Multivariate Symmetry Models*. In *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics*, pages 13–42. Springer New York.
- [4] Cousin, A. and Di Bernardino, E. (2013). On multivariate extensions of value-at-risk. *Journal of multivariate analysis*, 119:32–46.
- [5] 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.
- [6] Dehaan, L. and Huang, X. (1995). Large quantile estimation in a multivariate setting. *Journal of Multivariate Analysis*, 53(2):247–263.
- [7] Denuit, M., Dhaene, J., Goovaerts, M., and Kaas, R. (2006). *Actuarial theory for dependent risks: measures, orders and models*. John Wiley & Sons.
- [8] 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.
- [9] Di Bernardino, E., Laloë, T., and Servien, R. (2015). Estimating covariate functions associated to multivariate risks: a level set approach. *Metrika, Springer Verlag*, pages 497–526.
- [10] Dyckerhoff, R. (2016). Convergence of depths and depth-trimmed regions. *arXiv preprint arXiv:1611.08721*.
- [11] Eberlein, E., Frey, R., Kalkbrener, M., and Overbeck, L. (2007). Mathematics in financial risk management. *Jahresbericht der DMV*.
- [12] Liu, R. (1990). On a notion of data depth based on random simplices. *The Annals of Statistics*, pages 405–414.

- [13] Rodríguez-Casal, A. (2003). Estimación de conjuntos y sus fronteras. un enfoque geometrico. *PhD thesis, University of Santiago de Compostela.*
- [14] 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.
- [15] Weyl, H. (1939). On the volume of tubes. *American Journal of Mathematics*, Vol. 61:461–472.
- [16] Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *Annals of statistics*, pages 461–482.