Electronic Journal of Statistics
Vol. 12 (2018) 2806–2872
ISSN: 1935-7524
https://doi.org/10.1214/18-EJS1474

Mass volume curves and anomaly ranking

Stephan Clémençon

LTCI, Télécom ParisTech, Université Paris Saclay, 75013, Paris, France
e-mail: stephan.clemencon@telecom-paristech.fr

and

Albert Thomas

LTCI, Télécom ParisTech, Université Paris Saclay, 75013, Paris, France
Airbus Group Innovations, 92150, Suresnes, France
e-mail: albert.thomas@centraliens.net

Abstract: This paper aims at formulating the issue of ranking multivariate unlabeled observations depending on their degree of abnormality as an unsupervised statistical learning task. In the 1-d situation, this problem is usually tackled by means of tail estimation techniques: univariate observations are viewed as all the more 'abnormal' as they are located far in the tail(s) of the underlying probability distribution. It would be desirable as well to dispose of a scalar valued 'scoring' function allowing for comparing the degree of abnormality of multivariate observations. Here we formulate the issue of scoring anomalies as a M-estimation problem by means of a novel functional performance criterion, referred to as the *Mass Volume* curve (MV curve in short), whose optimal elements are strictly increasing transforms of the density almost everywhere on the support of the density. We first study the statistical estimation of the MV curve of a given scoring function and we provide a strategy to build confidence regions using a smoothed bootstrap approach. Optimization of this functional criterion over the set of piecewise constant scoring functions is next tackled. This boils down to estimating a sequence of empirical minimum volume sets whose levels are chosen adaptively from the data, so as to adjust to the variations of the optimal MV curve, while controlling the bias of its approximation by a stepwise curve. Generalization bounds are then established for the difference in sup norm between the MV curve of the empirical scoring function thus obtained and the optimal MV curve.

Keywords and phrases: Anomaly ranking, unsupervised learning, bootstrap, M-estimation.

Received April 2017.

1. Introduction

In a wide variety of applications, ranging from the monitoring of aircraft engines in aeronautics to non destructive control quality in the industry through fraud detection, network intrusion surveillance or system management in data centers (see for instance (Viswanathan et al., 2012)), *anomaly detection* is of crucial importance. In most common situations, anomalies correspond to 'rare'

observations and must be automatically detected based on an unlabeled dataset. In practice, the very purpose of anomaly detection techniques is to rank observations by degree of abnormality/novelty, rather than simply assigning them a binary label, 'abnormal' *vs* 'normal'. In the case of univariate observations, abnormal values are generally those which are extremes, *i.e.* 'too large' or 'too small' in regard to central quantities such as the mean or the median, and anomaly detection may then derive from standard tail distribution analysis: the farther in the tail the observation lies, the more 'abnormal' it is considered. In contrast, it is far from easy to formulate the issue in a multivariate context. In the present paper, motivated by applications such as those aforementioned, we place ourselves in the situation where (unlabeled) observations take their values in a possibly very high-dimensional space, $\mathcal{X} \subset \mathbb{R}^d$ with $d \geq 1$ say, making approaches based on nonparametric density estimation or multivariate (heavy-) tail modeling hardly feasible, if not unfeasible, due to the curse of dimensionality. In this framework, a variety of statistical techniques, relying on the concept of *minimum volume set* investigated in the seminal contributions of Einmahl and Mason (1992) and Polonik (1997) (see Section 2), have been developed in order to split the feature space $\mathcal{X}$ into two halves and decide whether observations should be considered as 'normal' (namely when lying in the *minimum volume* set $\Omega_\alpha \subset \mathcal{X}$ estimated on the basis of the dataset available) or not (when lying in the complementary set $\mathcal{X} \setminus \Omega_\alpha$) with a given confidence level $1 - \alpha \in (0, 1)$. One may also refer to (Scott and Nowak, 2006) and to (Koltchinskii, 1997) for closely related notions. The problem considered here is of different nature, the goal pursued is not to assign to all possible observations a label 'normal' *vs* 'abnormal', but to rank them according to their level of 'abnormality'. The most natural way to define a preorder on the feature space $\mathcal{X}$ is to transport the natural order on the real line through some (measurable) *scoring function* $s : \mathcal{X} \to \mathbb{R}_+$: the 'smaller' the score $s(x)$, the more likely the observation x is viewed as 'abnormal'. This problem shall be here referred to as *anomaly scoring*. It can be somehow related to the literature dedicated to *statistical depth functions* in nonparametric statistics and operations research, see (Zuo and Serfling, 2000) and the references therein. Such parametric functions are generally proposed *ad hoc* in order to define a 'center' for the probability distribution of interest and a notion of distance to the latter, *e.g.* the concept of 'center-outward ordering' induced by *halfspace depth* in (Tukey, 1975) or (Donoho and Gasko, 1992). The angle embraced in this paper is quite different, its objective is indeed twofold: 1) propose a performance criterion for the anomaly scoring problem so as to formulate it in terms of M-estimation 2) investigate the accuracy of scoring rules which optimize empirical estimates of the criterion thus tailored.

Due to the global nature of the ranking problem, the criterion we promote is *functional*, just like the *Receiver Operating Characteristic* (ROC) and *Precision-Recall* curves in the supervised ranking setup (*i.e.* when a binary label, *e.g.* 'normal' *vs* 'abnormal', is assigned to the sampling data), and shall be referred to as the *Mass Volume curve* (MV curve in abbreviated form). The latter induces a partial preorder on the set of scoring functions: the collection of optimal elements is defined as the set of scoring functions whose MV curve is minimum

everywhere. Such optimal scoring functions are proved to coincide to strictly increasing transforms of the underlying probability density almost everywhere on the support of this underlying probability density (see Section 3 for the exact definition). In the unsupervised setting, MV curve analysis is shown to play a role quite comparable to that of ROC curve analysis for supervised anomaly detection. The issue of estimating the MV curve of a given scoring function based on sampling data is then tackled and a smooth bootstrap method for constructing confidence regions is analyzed. A statistical methodology to build a nearly optimal scoring function is next described, which works as follows: first, a piecewise constant approximant of the optimal MV curve is estimated by solving a few minimum volume set estimation problems where confidence levels are chosen adaptively from the data to adjust to the variations of the optimal MV curve; second, a piecewise constant scoring function is built based on the sequence of estimated minimum volume sets. The MV curve of the scoring rule thus produced can be related to a stepwise approximant of the (unknown) optimal MV curve, which permits to establish the generalization ability of the algorithm through rate bounds in terms of sup norm in the MV space.

The rest of the article is structured as follows. Section 2 describes the mathematical framework, sets out the main notations and recalls the crucial notions related to anomaly/novelty detection on which the analysis carried out in the paper relies. Section 3 first provides an informal description of the anomaly scoring problem and then introduces the MV curve criterion, dedicated to evaluate the performance of any scoring function. The set of optimal elements is described and statistical results related to the estimation of the MV curve of a given scoring function are stated in Section 4. Statistical learning of an anomaly scoring function is then formulated as a functional M-estimation problem in Section 5, while Section 6 is devoted to the study of a specific algorithm for the design of nearly optimal anomaly scoring functions. Technical proofs are deferred to the Appendix section.

We finally point out that a very preliminary version of this work has been presented in the conference AISTATS 2013 (see (Clémençon and Jakubowicz, 2013)). The present article gives a better characterization of the optimal scoring functions and investigates much deeper the statistical assessment of the performance of a given scoring function. In particular, it shows how to construct confidence regions for the MV curve of a scoring function in a computationally feasible fashion, using the (smooth) bootstrap methodology for which we state a consistency result. This consistency result gives a rate of convergence which promotes the use of a smooth bootstrap approach, rather than a naive bootstrap technique. Additionally, in Section 6, the confidence levels of the minimum volume sets to be estimated are chosen in a data-driven way (instead of considering a regular subdivision of the interval $[0, 1]$ fixed in advance), giving to the statistical estimation and learning procedures proposed in this paper crucial adaptivity properties, with respect to the (unknown) shape of the optimal MV curve. Eventually, we give an example showing that the nature of the problem tackled here is very different than that of density estimation and we also give a simpler formula of the derivative of the optimal MV curve (that of the

underlying density) compared to the one originally given in (Clémençon and Jakubowicz, 2013).

2. Background and preliminaries

As a first go, we start off with describing the mathematical setup and recalling key concepts in anomaly detection involved in the subsequent analysis.

2.1. Framework and notations

Here and throughout, we suppose that we observe independent and identically distributed realizations $X_1, \ldots, X_n$ of an unknown continuous probability distribution $F(dx)$, copies of a generic random variable X, taking their values in a (possibly very high dimensional) feature space $\mathcal{X} \subset \mathbb{R}^d$, with $d \geq 1$. The density of the random variable X with respect to $\lambda(dx)$, the Lebesgue measure on $\mathbb{R}^d$, is denoted by f, its support $\{x \in \mathcal{X},\ f(x) > 0\}$ by $\mathrm{supp}(f)$ and the indicator function of any event $\mathcal{E}$ by $\mathbb{I}\{\mathcal{E}\}$. For any set $\mathcal{Z} \subset \mathcal{X}$, its complementary is denoted by $\overline{\mathcal{Z}} = \mathcal{X} \setminus \mathcal{Z}$. The sup norm of any real valued function $g : \mathbb{R}^d \to \mathbb{R}$ is denoted by $\|g\|_\infty$ and the Dirac mass at any point a by δ_a. The notation $O_{\mathbb{P}}(1)$ is used to mean boundedness in probability. The quantile function $H^\dagger$ of any real valued random variable Z with cumulative distribution function $H(\cdot) = \mathbb{P}(Z \leq \cdot)$ is defined by $H^\dagger(\alpha) = \inf\{t \in \mathbb{R} : H(t) \geq \alpha\}$ for all $\alpha \in (0, 1)$. For any real valued random variable Z, the generalized inverse G^{-1} of the decreasing function $G(\cdot) = \mathbb{P}(Z \geq \cdot)$ is defined by $G^{-1}(\alpha) = \inf\{t \in \mathbb{R} : G(t) \leq \alpha\}$ for all $\alpha \in (0, 1)$. For any function $s : \mathcal{X} \to \mathbb{R}$, F_s denotes the cumulative distribution function of the random variable $s(X)$. In addition, for any $\alpha \in (0, 1)$, $Q(s, \alpha) = F_s^\dagger(1 - \alpha)$ denotes the quantile at level $1 - \alpha$ of the distribution of $s(X)$. We also set $Q^*(\alpha) = Q(f, \alpha)$ for all $\alpha \in (0, 1)$. Finally, constants in inequalities are denoted by either C or c and may vary at each occurrence.

A natural way of defining preorders on $\mathcal{X}$ is to map its elements onto $\mathbb{R}_+$ and use the natural order on the real half-line.

Definition 1. (Scoring function) *A scoring function is any measurable function $s : \mathcal{X} \to \mathbb{R}_+$ that is integrable with respect to the Lebesgue measure.*

The set of all scoring functions is denoted by $\mathcal{S}$ and we denote the level sets of any scoring function $s \in \mathcal{S}$ by:

$$\Omega_{s,t} = \{x \in \mathcal{X} :\ \ s(x) \geq t\},\ \ t \in [-\infty,\ +\infty]\,.$$

Observe that the family is decreasing (for the inclusion, as t increases from $-\infty$ to $+\infty$):

$$\forall (t, t') \in \mathbb{R}^2,\ \ t \geq t' \Rightarrow \Omega_{s,t} \subset \Omega_{s,t'}$$

and that $\lim_{t \to +\infty} \Omega_{s,t} = \emptyset$ and $\lim_{t \to -\infty} \Omega_{s,t} = \mathcal{X}$[1].

[1]Recall that a sequence $(A_n)_{n \geq 1}$ of subsets of an ensemble E is said to converge if and only if $\limsup A_n = \liminf A_n$. In such a case, one defines its limit, denoted by $\lim A_n$ as $\limsup A_n = \liminf A_n$.

The following quantities shall also be used in the sequel. For any scoring function s and threshold level $t \geq 0$, define:

$$\alpha_s(t) = \mathbb{P}\{s(X) \geq t\} \text{ and } \lambda_s(t) = \lambda\left(\{x \in \mathcal{X} : \ \ s(x) \geq t\}\right) .$$

The quantity $\alpha_s(t)$ is referred to as the *mass* of the level set $\Omega_{s,t}$, while $\lambda_s(t)$ is generally termed the *volume* (with respect to the Lebesgue measure).

Remark 1. *The integrability of a scoring function (see Definition 1) implies that the volumes are finite on $\mathbb{R}_+^*$: for any $s \in \mathcal{S}$,*

$$\forall t > 0, \quad \lambda_s(t) = \int_{\mathcal{X}} \mathbb{I}\{x, s(x) \geq t\} dx \leq \int_{\mathcal{X}} \frac{s(x)}{t} dx < +\infty .$$

Incidentally, we point out that, in the specific case $s = f$, the set $\Omega_{f,t} = \{x \in \mathcal{X}, f(x) \geq t\}$ is the *density contour cluster* of the density function f at level t, see (Polonik, 1995) for instance. Such a set is also referred to as a *density level set*. Observe in addition that, using the terminology introduced in (Liu, Parelius and Singh, 1999), $\Omega_{s,t}$ is the *region enclosed by the contour of depth* t when s is a depth function.

2.2. Minimum volume sets

The notion of minimum volume sets has been introduced and investigated in the seminal contributions of Einmahl and Mason (1992) and Polonik (1997) in order to describe regions where a multivariate random variable X takes its values with highest/smallest probability. Let $\alpha \in (0, 1)$, a minimum volume set Ω_α^* of mass at least α is any solution of the constrained minimization problem

$$\min_{\Omega} \lambda(\Omega) \text{ subject to } \mathbb{P}\{X \in \Omega\} \geq \alpha , \tag{1}$$

where the minimum is taken over all measurable subsets Ω of $\mathcal{X}$. Application of this concept includes in particular novelty/anomaly detection: for large values of α, abnormal observations (outliers) are those which belong to the complementary set $\mathcal{X} \setminus \Omega_\alpha^*$. In the continuous setting, it can be shown that there exists a threshold value equal to $Q^*(\alpha)$ such that the level set $\Omega_{f,Q^*(\alpha)}$ is a solution of the constrained optimization problem above. The (generalized) quantile function is then defined by:

$$\forall \alpha \in (0, 1), \ \ \lambda^*(\alpha) \stackrel{def}{=} \lambda(\Omega_\alpha^*) .$$

The definition of λ^* can be extended to the interval $[0, 1]$ by setting $\lambda^*(0) = 0$ and $\lambda^*(1) = \lambda(\text{supp}(f)) \leq \infty$. The following assumptions shall be used in the subsequent analysis.

(A1) The density f is bounded: $\|f\|_\infty < +\infty$.
(A2) The density f has no flat parts, *i.e.* for any $t > 0$,

$$\mathbb{P}\{f(X) = t\} = 0 .$$

Under the assumptions above, for any $\alpha \in (0,1)$, there exists a unique minimum volume set Ω_α^* (equal to $\Omega_{f,Q^*(\alpha)}$ up to subsets of null λ-measure), whose mass is equal to α exactly. Additionally, the mapping λ^* is continuous on $(0,1)$ and uniformly continuous on $[0, 1-\varepsilon]$ for all $\varepsilon \in (0,1)$ (when the support of $F(dx)$ is compact, uniform continuity holds on the whole interval $[0,1]$).

From a statistical perspective, estimates $\widehat{\Omega}_\alpha$ of minimum volume sets are built by replacing the unknown probability distribution F by its empirical version $\widehat{F} = (1/n)\sum_{i=1}^n \delta_{X_i}$ and restricting optimization to a collection $\mathcal{A}$ of borelian subsets of $\mathcal{X}$, supposed rich enough to include all density level sets (or reasonable approximations of the latter). In (Polonik, 1997), functional limit results are derived for the generalized empirical quantile process $\{\lambda(\widehat{\Omega}_\alpha^*) - \lambda^*(\alpha)\}$ under certain assumptions for the class $\mathcal{A}$ (stipulating in particular that $\mathcal{A}$ is a Glivenko-Cantelli class for $F(dx)$). In (Scott and Nowak, 2006), it is proposed to replace the level α by $\alpha - \phi_n$ where ϕ_n plays the role of tolerance parameter. The latter should be chosen of the same order as the supremum $\sup_{\Omega\in\mathcal{A}} |\widehat{F}(\Omega) - F(\Omega)|$ roughly, complexity of the class $\mathcal{A}$ being controlled by the VC dimension or by means of the concept of Rademacher averages, so as to establish rate bounds at $n < +\infty$ fixed.

Alternatively, so-termed *plug-in* techniques, consisting in computing first an estimate $\widehat{f}$ of the density f and considering next level sets $\Omega_{\widehat{f},t}$ of the resulting estimator have been investigated in several papers, among which (Cavalier, 1997; Tsybakov, 1997; Rigollet and Vert, 2009; Cadre, 2006; Cadre, Pelletier and Pudlo, 2013). Such an approach however yields significant computational issues even for moderate values of the dimension, inherent to the curse of dimensionality phenomenon.

3. Ranking anomalies

In this section, the issue of scoring observations depending on their level of novelty/abnormality is first described in an informal manner and next formulated quantitatively, as a functional optimization problem, by means of a novel concept, termed the *Mass Volume curve*.

3.1. Overall objective

The problem considered in this article is to learn a scoring function s, based on training data $X_1, \ldots, X_n$, so as to describe the extremal behavior of the (high-dimensional) random vector X by that of the univariate variable $s(X)$, which can be summarized by its tail behavior near 0: hopefully, the smaller the score $s(x)$, the more abnormal/rare the observation x should be considered. Hence, an optimal scoring function should naturally permit to rank observations x by increasing order of magnitude of $f(x)$ (see Section 3.2 for the precise definition of the set of optimal scoring functions). The preorder on $\mathcal{X}$ induced by such an optimal scoring function could then be used to rank observations by their

degree of abnormality: for any optimal scoring function s, an observation x' is more abnormal than an observation x if and only if $s(x') < s(x)$.

3.2. A functional criterion: The mass volume curve

We now introduce the concept of Mass Volume curve and shows that it is a natural criterion to evaluate the accuracy of decision rules in regard to anomaly scoring.

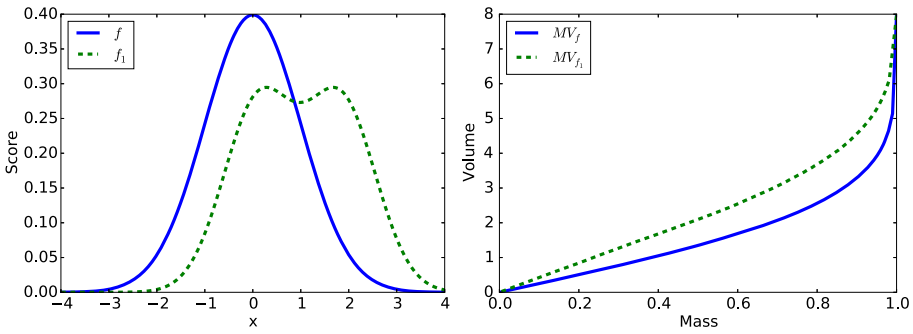
Definition 2. (True Mass Volume curve) *Let $s \in \mathcal{S}$. Its Mass Volume curve (*MV *curve in abbreviated form) with respect to the probability distribution of the random variable X is the plot of the function*

$$\mathrm{MV}_s : \quad \alpha \in (0,1) \mapsto \mathrm{MV}_s(\alpha) \stackrel{def}{=} \lambda_s \circ \alpha_s^{-1}(\alpha)\,.$$

If the scoring function s is upper bounded, $\alpha_s^{-1}(0)$ exists and MV_s is defined on $[0,1)$.

Remark 2. (Parametric Mass Volume Curve) *If α_s is invertible, and therefore α_s^{-1} is the ordinary inverse of α_s, the Mass Volume curve of the scoring function s can also be defined as the parametric curve:*

$$t \in \mathbb{R}_+ \mapsto (\alpha_s(t), \lambda_s(t)) \in [0,1) \times [0, +\infty)\,.$$



(a) Scoring functions (b) Mass Volume curves

Fig 1. *Scoring functions and their associated Mass Volume curves when considering a truncated Gaussian distribution with density f.*

Remark 3. (Connections to ROC/concentration analysis) *We point out that the curve $\alpha \in (0,1) \mapsto 1 - \lambda_s \circ \alpha_s^{-1}(1-\alpha)$ resembles a receiver operating characteristic (ROC) curve of the test/diagnostic function s (see (Egan, 1975)), except that the distribution under the alternative hypothesis is not a probability measure but the image of Lebesgue measure on $\mathcal{X}$ by the function s, while the distribution under the null hypothesis is the probability distribution of the random variable $s(X)$. Hence, the curvature of the* MV *graph somehow*

measures the extent to which the spread of the distribution of $s(X)$ *differs from that of a uniform distribution. Observe also that, in the case where the support of* $F(dx)$ *coincides with the unit square* $[0,1]^d$*, the* MV *curve of any scoring function* $s \in \mathcal{S}$ *corresponds to the concentration function* $(GF^{-1})_{\mathcal{C}_\mathcal{S}}$ *introduced in (Polonik, 1999), where* $G(dx)$ *is the uniform probability distribution on* $[0,1]^d$ *and* $\mathcal{C}_\mathcal{S} = \{\Omega_{s,t} : \ t \geq 0\}$*, to compare the concentration of* $F(dx)$ *with that of* $G(dx)$ *with respect to the class of sublevel sets of the function* s*. Notice finally that, when* s *is a depth function,* MV_s *is a scale curve, as defined in (Liu, Parelius and Singh, 1999).*

This functional criterion induces a partial order over the set of all scoring functions. Let s_1 and s_2 be two scoring functions on $\mathcal{X}$, the ordering provided by s_1 is better than that induced by s_2 when

$$\forall \alpha \in (0,1), \ \ \mathrm{MV}_{s_1}(\alpha) \leq \mathrm{MV}_{s_2}(\alpha)\,.$$

Typical MV curves are illustrated in Fig. 1. A desirable MV curve increases slowly and rises near 1, just like the lowest curve of Fig. 1. This corresponds to the situation where the distribution of the random variable X is much concentrated around its modes and the highest values (respectively, the lowest values) taken by s are located near the modes of $F(dx)$ (respectively, in the tail region of $F(dx)$). The MV curve of the scoring function s is then close to the right lower corner of the Mass Volume space. We point out that, in certain situations, some parts of the MV curve may be of interest solely, corresponding to large values of α when focus is on extremal observations (the tail region of the random variable X) and to small values of α when modes of the underlying distribution are investigated.

We define the set of optimal scoring functions $\mathcal{S}^*$ as follows. We recall here that for any set $\mathcal{Z} \subset \mathcal{X}$, its complementary is denoted by $\overline{\mathcal{Z}} = \mathcal{X} \setminus \mathcal{Z}$.

Definition 3. *The set of optimal scoring functions* $\mathcal{S}^*$ *is the set of scoring functions* $s \in \mathcal{S}$ *such that there exist* $\mathcal{Z} \subset \mathcal{X}$ *and a function* $T : Im f \to \mathbb{R}^+$ *such that*

(i) $\lambda(\overline{\mathcal{Z}} \cap \mathrm{supp}(f)) = 0$,
(ii) For all $x \in \mathcal{Z} \cap \mathrm{supp}(f)$, $s(x) = T \circ f(x)$,
(iii) $T_{|f(\mathcal{Z} \cap \mathrm{supp}(f))}$ *is strictly increasing,*
(iv) For λ*-almost all* $x \in \overline{\mathrm{supp}(f)}$, $F_s(s(x)) = 0$.

One can note that, as expected, the density f and strictly increasing transforms of the density belong to $\mathcal{S}^*$. More generally and roughly speaking, an optimal scoring function $s \in \mathcal{S}^*$ is a strictly increasing transform of the density λ-almost everywhere on the support of the density. Note also that on the one hand, if $a_s = \sup\{t, F_s(t) = 0\}$, thanks to (i), (ii) and (iii), we have $F_s(s(x)) = F_f(f(x)) > 0$ for almost all $x \in \mathrm{supp}(f)$ (because $\mathrm{supp}(f) = \{x, F_f(f(x)) > 0\}$ up to subsets of null λ measure as shown in Lemma 1). Therefore $s \geq a_s$ λ-almost everywhere on $\mathrm{supp}(f)$. On the other hand, thanks to (iv), $F_s(s(x)) = 0$ for almost all $x \in \overline{\mathrm{supp}(f)}$ and therefore $s \leq a_s$ λ-almost everywhere on $\overline{\mathrm{supp}(f)}$.

Thus the values of s on $\overline{\text{supp}(f)}$ are λ-almost everywhere lower than the values of s on $\text{supp}(f)$, where s is almost everywhere a strictly increasing transform of the density f.

The result below shows that optimal scoring functions are those whose MV curves are minimum everywhere.

Proposition 1. (Optimal MV curve) *Let assumptions **(A1)** and **(A2)** be fulfilled. The elements of the class $\mathcal{S}^*$ have the same* MV *curve, equal to* MV_f*, and provide the best possible ordering of $\mathcal{X}$'s elements in regard to the* MV *curve criterion:*

$$s^* \in \mathcal{S}^* \iff \forall s \in \mathcal{S} \text{ and } \forall \alpha \in (0,1),\ \text{MV}_{s^*}(\alpha) \leq \text{MV}_s(\alpha)\,. \tag{2}$$

In addition, we have: $\forall s \in \mathcal{S}$ and $\forall \alpha \in (0,1)$,

$$\text{MV}_s(\alpha) - \text{MV}^*(\alpha) \leq \lambda\left(\Omega^*_\alpha \Delta \Omega_{s,Q(s,\alpha)}\right)\,, \tag{3}$$

where $\text{MV}^*(\alpha) = \text{MV}_f(\alpha)$ for all $\alpha \in (0,1)$ and where Δ denotes the symmetric difference.

From now on we will thus use $\text{MV}^* = \text{MV}_f$ to denote the MV curve of elements of $\mathcal{S}^*$. The proof of Proposition 1 can be found in Appendix A. Incidentally, notice that, equipped with the notations introduced in 2.2, we have $\lambda^*(\alpha) = \text{MV}^*(\alpha)$ for all $\alpha \in (0,1)$. We also point out that bound (3) reveals that the pointwise difference between the optimal MV curve and that of a scoring function candidate s is controlled by the error made in recovering the specific minimum volume set Ω^*_α through $\Omega_{s,Q(s,\alpha)}$.

Remark 4. *In the framework we develop, anomaly scoring boils down to recovering the decreasing collection of all level sets of the density function f, $\{\Omega^*_\alpha : \alpha \in (0,1)\}$, without necessarily disposing of the corresponding levels. Indeed, one may check that any scoring function of the form*

$$s^*(x) = \int_0^1 \mathbb{I}\{x \in \Omega^*_\alpha\} d\mu(\alpha)\,, \tag{4}$$

where $\mu(d\alpha)$ is an arbitrary finite positive measure dominating the Lebesgue measure on $(0,1)$, belongs to $\mathcal{S}^*$. Observe that $F_f \circ f$ corresponds to the case where μ is chosen to be the Lebesgue measure on $(0,1)$ in Eq. (4). The anomaly scoring problem can be thus cast as an overlaid collection of minimum volume set estimation problems. This observation shall turn out to be very useful when designing practical statistical learning strategies in Section 6.

We point out that the optimal MV curve provides a measure of the mass concentration of the random variable X: the lower the curve MV^*, the more concentrated the distribution $f(x)\lambda(dx)$. Indeed, the *excess mass functional* introduced in (Muller and Sawitzki, 1991) can be expressed as $t \geq 0 \mapsto E(t) = \alpha_f(t) - t\lambda_f(t)$. Hence, we have $E(Q^*(\alpha)) = \alpha - Q^*(\alpha)MV^*(\alpha)$ for all $\alpha \in (0,1)$. One may refer to (Polonik, 1995) for results related to the statistical estimation of density contour clusters and applications to multimodality testing using the excess mass functional.

Example 1. (Univariate Gaussian distribution) *Let X be random variable with Gaussian distribution $\mathcal{N}(0,1)$, we have $\mathrm{MV}^*(\alpha) = 2\Phi^{-1}((1+\alpha)/2)$, where Φ is the cumulative distribution function of X.*

Example 2. (Multivariate Gaussian distribution) *Let $X \in \mathbb{R}^d$ be a multivariate random variable with Gaussian distribution $\mathcal{N}(0, \mathbf{\Sigma})$. We assume that Σ is a diagonal matrix and we denote by $a_1, \ldots, a_d \in \mathbb{R}^{+*}$ its diagonal coefficients. The density level set $\{x, f(x) \geq t\}$ is exactly the set $\{x, x^T\Sigma^{-1}x \leq c_t\}$ with $c_t > 0$. It is well known that $X^T\Sigma^{-1}X \sim \chi^2_d$, the χ^2 distribution with d degrees of freedom. Thus $\alpha_f^{-1}(\alpha) = c_t = \chi^2_d(\alpha)$, $\chi^2_d(\alpha)$ denoting the quantile of order α of the χ^2_d distribution. The set $\{x, x^T\Sigma^{-1}x \leq \chi^2_d(\alpha)\}$ is exactly the ellipsoid with semi-principal axes of length $a_1\sqrt{\chi^2_d(\alpha)}, \ldots, a_d\sqrt{\chi^2_d(\alpha)}$. The formula of the volume of an ellipsoid gives $\mathrm{MV}^*(\alpha) = \pi^{d/2}/\Gamma(d/2+1)(\chi^2_d(\alpha))^{d/2}\prod_{i=1}^d a_i$ where Γ is the gamma function.*

The following result reveals that the optimal MV curve is convex and provides a closed analytical form for its derivative. The proof is given in Appendix A.

Proposition 2. (Convexity and derivative) *Suppose that assumptions **(A1)** and **(A2)** are satisfied. Then, as f is bounded, the optimal curve MV^* is defined on $[0,1)$. Furthermore,*

(i) *$\alpha \in [0,1) \mapsto \mathrm{MV}^*(\alpha)$ is convex.*

(ii) *Let F_f be the cumulative distribution function of the random variable $f(X)$ and let $a = \lim_{\alpha\to 0^+} F_f^\dagger(\alpha) = \sup\{t, F_f(t) = 0\} \geq -\infty$. If F_f is invertible on $(a, F_f^\dagger(1)]$ then MV^* is differentiable on $[0,1)$ and*

$$\forall \alpha \in [0,1), \quad \mathrm{MV}^{*\prime}(\alpha) = \frac{1}{Q^*(\alpha)}\,.$$

The convexity of MV^* might be explained as follows: when considering density level sets with increasing probabilities α, their volumes $\lambda(\Omega^*_\alpha)$ increase as well, but they increase more and more quickly since the mass of the distribution becomes less and less concentrated.

Elementary properties of MV curves are summarized in the following proposition.

Proposition 3. (Properties of MV curves) *For any $s \in \mathcal{S}$, the following assertions hold true.*

(i) **Invariance.** *For any strictly increasing function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$, we have $\mathrm{MV}_s = \mathrm{MV}_{\psi\circ s}$.*

(ii) **Monotonicity.** *The mapping $\alpha \in (0,1) \mapsto \mathrm{MV}_s(\alpha)$ is increasing.*

Assertion (i) may be proved using the following property of the quantile function (see *e.g.* Property 2.3 in (Embrechts and Hofert, 2013)): for any cumulative distribution function H, any $z \in \mathbb{R}$ and any $\alpha \in (0,1)$, $z \geq H^\dagger(\alpha)$ if and only if $H(z) \geq \alpha$. Assertion (ii) derives from the fact that the quantile function $F_s^\dagger$ is increasing. Details are left to the reader.

4. Statistical estimation

In practice, MV curves are unknown, just like the probability distribution of the random variable X, and must be estimated based on the observed sample $X_1, \ldots, X_n$. Replacing the mass of each level set by its statistical counterpart in Definition 2 leads to define the notion of *empirical* MV *curve*. We set, for all $t \geq 0$,

$$\widehat{\alpha}_s(t) \stackrel{def}{=} \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{s(X_i) \geq t\} = \int_{\mathcal{X}} \mathbb{I}\{u \geq t\}\widehat{F}_s(du)\,, \tag{5}$$

where $\widehat{F}_s = (1/n)\sum_{i\leq n} \delta_{s(X_i)}$ denotes the empirical distribution of the $s(X_i), 1 \leq i \leq n$. Notice that $\widehat{\alpha}_s$ takes its values in the set $\{k/n : \ k = 0, \ldots, n\}$.

Definition 4. (Empirical MV curve) *Let $s \in \mathcal{S}$. By definition, the empirical* MV *curve of s is the graph of the (piecewise constant) function*

$$\widehat{\mathrm{MV}}_s : \alpha \in [0,1) \mapsto \lambda_s \circ \widehat{\alpha}_s^{-1}(\alpha)\,.$$

Remark 5. (On volume estimation). *Except for very specific choices of the scoring function s (e.g. when s is piecewise constant and the volumes of the subsets of $\mathcal{X}$ on which s is constant can be explicitly computed), no closed analytic form for the volume $\lambda_s(t)$ is available in general and Monte-Carlo procedures should be used to estimate it (which may be practically challenging in a high-dimensional setting). Computation of volumes in high dimensions is an active topic of research (Lovász and Vempala, 2006) and for simplicity, this is not taken into account in the subsequent analysis.*

In order to obtain a smoothed version $\widetilde{\alpha}_s(t)$, a typical strategy consists in replacing the empirical distribution estimate $\widehat{F}_s$ involved in (5) by the continuous distribution $\widetilde{F}_s$ with density $\widetilde{f}_s(t) = (1/n)\sum_{i=1}^{n} K_h(s(X_i) - t)$, with $K_h(t) = h^{-1}K(t/h)$ where $K \geq 0$ is a regularizing Parzen-Rosenblatt kernel (*i.e.* a bounded square integrable function such that $\int K(v)dv = 1$) and $h > 0$ is the smoothing bandwidth (see for instance (Wand and Jones, 1994)).

4.1. Consistency and asymptotic normality

The theorem below reveals that, under mild assumptions, the empirical MV curve is a consistent and asymptotically Gaussian estimate of the MV curve, uniformly over any subinterval of $[0, 1)$. It involves the assumptions listed below.

(A3) The scoring function s is bounded: $\|s\|_\infty < +\infty$.

(A4) The random variable $s(X)$ has a continuous cumulative distribution function F_s. Let $a = \sup\{t \in \mathbb{R},\ F_s(t) = 0\} \geq -\infty$ and $b = \inf\{t \in \mathbb{R},\ F_s(t) = 1\} \leq +\infty$. The distribution function F_s is twice differentiable on (a, b) and

$$\forall t \in (a, b), \quad F_s'(t) = f_s(t) > 0\,. \tag{6}$$

(A5) There exists $c > 0$ such that

$$\sup_{t\in(a,b)} F_s(t)(1-F_s(t))\left|\frac{f_s'(t)}{f_s^2(t)}\right| \leq c < +\infty\,.$$

(A6) f_s has a limit $A > 0$ when t tends towards b from the left:

$$\lim_{t\to b^-} f_s(t) = A < \infty\,.$$

(A7) The mapping λ_s is of class $\mathcal{C}^2$.

Assumption **(A3)** implies that $F_s^\dagger(1)$ exists and $F_s^\dagger(1) = b$. Note also that if s is not bounded, one can always consider $s' = \arctan\circ s$. Indeed s' takes its values in $[0, \pi/2)$ and as arctan is strictly increasing $\mathrm{MV}_s = \mathrm{MV}_{s'}$ thanks to Proposition 3. Assumptions **(A4)** and **(A5)** are common assumptions for the strong approximation of the quantile process (see for instance (Csörgő and Révész, 1978)). Condition (6) and assumption **(A5)** are respectively equivalent to $f_s(F_s^\dagger(\alpha)) > 0$ for all $\alpha \in (0,1)$ and

$$\sup_{\alpha\in(0,1)} \alpha(1-\alpha)\left|\frac{f_s'(F_s^\dagger(\alpha))}{f_s^2(F_s^\dagger(\alpha))}\right| \leq c < +\infty\,.$$

Eventually, assumption **(A6)** is equivalent to: f_s has a limit $A > 0$ when x tends towards b. The density f_s can therefore be extended by continuity to the interval $(a, b]$. As $A > 0$, (6) can then be replaced by $f_s(t) > 0$ for all $t \in (a, b]$ which also gives $f_s(F_s^\dagger(\alpha)) > 0$ for all $\alpha \in (0, 1]$. One can also note that F_s is thus invertible on $(a, b]$ with inverse $F_s^\dagger$.

In order to state part (ii) of the following theorem, we assume that the probability space is rich enough in the sense that an infinite sequence of Brownian bridges can be defined on it.

Theorem 1. *Let $\varepsilon \in (0, 1]$ and $s \in \mathcal{S}$. Assume that assumptions **(A3)**–**(A7)** are fulfilled. The following assertions hold true.*

(i) (Consistency) *With probability one, we have uniformly over $[0, 1-\varepsilon]$:*

$$\lim_{n\to+\infty} \widehat{\mathrm{MV}}_s(\alpha) = \mathrm{MV}_s(\alpha)\,.$$

(ii) (Strong approximation) *There exists a sequence of Brownian bridges $\{B_n(\alpha), \alpha \in [0,1]\}_{n\geq 1}$ such that we almost surely have, uniformly over the compact interval $[0, 1-\varepsilon]$: as $n \to \infty$,*

$$\sqrt{n}\left(\widehat{\mathrm{MV}}_s(\alpha) - \mathrm{MV}_s(\alpha)\right) = Z_n(\alpha) + O(n^{-1/2}\log n)\,,$$

where

$$Z_n(\alpha) = \frac{\lambda_s'(\alpha_s^{-1}(\alpha))}{f_s(\alpha_s^{-1}(\alpha))} B_n(\alpha),\ \text{for } \alpha \in [0, 1-\varepsilon]\,.$$

The technical proof is given in Appendix A. One can note from the proof of assertion (i) that this assertion does not require assumption **(A5)** to be satisfied. The proof of assertion (ii) relies on standard strong approximation results for the quantile process (Csörgő and Révész, 1978, 1981; Csörgő, 1983). Assertion (ii) means that the fluctuation process $\{\sqrt{n}(\widehat{\mathrm{MV}}_s(\alpha) - \mathrm{MV}_s(\alpha)), \alpha \in [0, 1-\varepsilon]\}$ converges in the space of càd-làg functions on $[0, 1-\varepsilon]$ equipped with the sup norm, to the law of a Gaussian stochastic process $\{Z_1(\alpha), \alpha \in [0, 1-\varepsilon]\}$.

Remark 6. *Assumption **(A6)** is required in order to state the results of assertions (i) and (ii) on the interval $[0, 1-\varepsilon]$. However this assumption is restrictive as it implies that f_s is discontinuous at b since $A > 0$ whereas $f_s(t) = 0$ for all $t > b$. Instead of assumption **(A6)**, if one assumes that f_s is decreasing on an interval to the left of b, then the result of assertion (ii) can be obtained on the interval $(0, 1-\varepsilon]$ with the same rate of convergence if $c < 2$ and with the same rate of convergence up to $\log n$ and $\log\log n$ factors if $c \geq 2$. In this case, assertion (i) also holds on $(0, 1-\varepsilon]$ (see for instance (Csörgő and Révész, 1978)).*

Remark 7. (Asymptotic normality) *It results from assertion (ii) in Theorem 1 that, for any $\alpha \in (0,1)$, the pointwise estimator $\widehat{\mathrm{MV}}_s(\alpha)$ is asymptotically Gaussian under assumptions **(A3)**–**(A7)**. For all $\alpha \in (0,1)$ $\sqrt{n}(\widehat{\mathrm{MV}}_s(\alpha) - \mathrm{MV}_s(\alpha))$ converges in distribution towards $\mathcal{N}(0, \sigma_s^2)$, as $n \to +\infty$, with $\sigma_s^2 = \alpha(1-\alpha)(\lambda_s'(\alpha_s^{-1}(\alpha))/f_s(\alpha_s^{-1}(\alpha)))^2$.*

4.2. Confidence regions in the mass volume space

The true MV curve of a given scoring function is unknown in practice and its performance must be statistically assessed based on a data sample. Beyond consistency of the empirical curve in sup norm and the asymptotic normality of the fluctuation process, we now tackle the question of constructing confidence bands in the MV space.

Definition 5. *Based on a sample $\mathcal{D}_n = (X_1, \ldots, X_n)$, a (random) confidence region for the* MV *curve of a given scoring function $s \in \mathcal{S}$ at confidence level $\eta \in (0,1)$ is any borelian set $\mathcal{R}_\eta \subset [0,1] \times \mathbb{R}_+$ of the* MV *space (possibly depending on $\mathcal{D}_n$) that covers the curve MV_s with probability larger than $1-\eta$:*

$$\mathbb{P}\{\mathrm{MV}_s \in \mathcal{R}_\eta\} \geq 1 - \eta\,.$$

In practice, confidence regions shall be of the form of balls in the Skorohod's space $\mathbb{D}([0, 1-\varepsilon])$ of càd-làg functions on $[0, 1-\varepsilon]$ with respect to the sup norm and with the estimate $\widehat{\mathrm{MV}}_s$ introduced in Definition 4 as center for some fixed $\varepsilon \in (0,1)$:

$$\mathcal{B}(\widehat{\mathrm{MV}}_s, \nu) = \left\{ g \in \mathbb{D}([0, 1-\varepsilon]) : \sup_{\alpha \in [0, 1-\varepsilon]} |g(\alpha) - \widehat{\mathrm{MV}}_s(\alpha)| \leq \nu \right\}.$$

Constructing confidence regions based on the approximation stated in Theorem 1 would require to know the density f_s. Hence a bootstrap approach (Efron, 1979) should be preferred. Following in the footsteps of Silverman and Young (1987), it is recommended to implement a smoothed bootstrap procedure. The asymptotic validity of such a resampling method derives from a strong approximation result similar to the one of Theorem 1. Let r_n be the fluctuation process defined for all $\alpha \in [0, 1-\varepsilon]$ by

$$r_n(\alpha) = \sqrt{n}(\widehat{\mathrm{MV}}_s(\alpha) - \mathrm{MV}_s(\alpha)) \,.$$

The bootstrap approach suggests to consider, as an estimate of the law of the fluctuation process r_n, the conditional law given the original sample $\mathcal{D}_n = (X_1, \ldots, X_n)$ of the *naive bootstrapped fluctuation process*

$$r_n^{Boot} = \{\sqrt{n}(\widehat{\mathrm{MV}}_s^{Boot}(\alpha) - \widehat{\mathrm{MV}}_s(\alpha))\}_{\alpha\in[0,1)} \,, \tag{7}$$

where, given $\mathcal{D}_n$, $\widehat{\mathrm{MV}}_s^{Boot}$ is the empirical MV curve of the scoring function s based on a sample of *i.i.d.* random variables with distribution the empirical distribution $\widehat{F}_s = (1/n)\sum_{i\leq n} \delta_{s(X_i)}$. The difficulty is twofold. First, the target is a distribution on a *path space*, namely a subspace of the Skorohod's space $\mathbb{D}([0, 1-\varepsilon])$ equipped with the sup norm. Second, r_n is a functional of the quantile process $\{\widehat{F}_s^{\dagger}(\alpha)\}_{\alpha\in[\varepsilon,1]}$. The *naive bootstrap*, which consists in resampling from the raw empirical distribution $\widehat{F}_s$, generally provides bad approximations of the distribution of empirical quantiles: the rate of convergence for a given quantile is indeed of order $O_{\mathbb{P}}(n^{-1/4})$ (Falk and Reiss, 1989) whereas the rate of the Gaussian approximation is $O(n^{-1/2}\log n)$ (see (25) in Appendix A). The same phenomenon may be naturally observed for MV curves. In a similar manner to what is usually recommended for empirical quantiles, a *smoothed version* of the bootstrap algorithm shall be implemented in order to improve the approximation rate of the distribution of $\sup_{\alpha\in[0,1-\varepsilon]} |r_n(\alpha)|$, namely, to resample the data from a smoothed version $\widetilde{F}_s$ of the empirical distribution $\widehat{F}_s$. We thus consider the *smooth boostrapped fluctuation process*

$$r_n^* = \{\sqrt{n}(\mathrm{MV}_s^{Boot}(\alpha) - \widetilde{\mathrm{MV}}_s(\alpha))\}_{\alpha\in[0,1)} \,, \tag{8}$$

where, given $\mathcal{D}_n$, $\mathrm{MV}_s^{Boot} = \lambda_s \circ (\alpha_s^{Boot})^{-1}$ is the empirical MV curve of the scoring function s based on a sample of *i.i.d.* random variables with distribution $\widetilde{F}_s$ and where $\widetilde{\mathrm{MV}}_s = \lambda_s \circ \widetilde{\alpha}_s^{-1}$ is the smooth version of the empirical MV curve, $\widetilde{\alpha}_s^{-1}$ being the generalized inverse of $\widetilde{\alpha}_s$. The algorithm for building a confidence band at level $1-\eta$ in the MV space from sampling data $\mathcal{D}_n = \{X_i : \ i = 1, \ldots, n\}$ is described in Algorithm 1.

Before turning to the theoretical analysis of this algorithm, its description calls a few comments. From a computational perspective, the smoothed bootstrap distribution $\mathbb{P}^*(\sup_{\alpha\in[\varepsilon,1-\varepsilon]} |r_n^*(\alpha)| \leq \cdot)$ must be approximated in its turn, by means of a Monte-Carlo approximation scheme. Based on the N bootstrap fluctuation processes obtained, $r_n^{*(j)}$ with $j = 1, \ldots, N$, the radius ν_η then

Algorithm 1 Smoothed MV curve bootstrap

1: Based on the sample $\mathcal{D}_n$, compute the smoothed version $\{\widetilde{\alpha}_s^{-1}(\alpha), \alpha \in [0,1)\} = \{\widetilde{F}_s^{\dagger}(1-\alpha), \alpha \in [0,1)\}$ of the empirical estimate $\{\widehat{\alpha}_s^{-1}(\alpha), \alpha \in [0,1)\}$.
2: Plot the smooth MV curve estimate, *i.e.*, the graph of the mapping

$$\alpha \in [0,1) \mapsto \widetilde{\mathrm{MV}}_s(\alpha) = \lambda_s \circ \widetilde{\alpha}_s^{-1}(\alpha)\,.$$

3: Draw a bootstrap sample $\mathcal{D}_n^{Boot} \sim \widetilde{F}_s \mid \mathcal{D}_n$.
4: Based on $\mathcal{D}_n^{Boot}$, compute the smoothed bootstrap version $\{(\alpha_s^{Boot})^{-1}(\alpha), \alpha \in [0,1)\}$ of the empirical estimate $\{\widehat{\alpha}_s^{-1}(\alpha), \alpha \in [0,1)\}$.
5: Plot the bootstrap MV curve, *i.e.* the graph of the mapping

$$\alpha \in [0,1) \mapsto \mathrm{MV}_s^{Boot}(\alpha) = \lambda_s \circ (\alpha_s^{Boot})^{-1}(\alpha)\,.$$

6: Get the *bootstrap confidence bands at level* $1-\eta$ defined by the ball of center $\widehat{\mathrm{MV}}_s$ and radius $\nu_\eta/\sqrt{n}$ in $\mathbb{D}([\varepsilon, 1-\varepsilon])$, where ν_η is defined by

$$\mathbb{P}^*\Big(\sup_{\alpha\in[\varepsilon,1-\varepsilon]} |r_n^*(\alpha)| \leq \nu_\eta\Big) = 1-\eta\,,$$

denoting by $\mathbb{P}^*(.)$ the conditional probability given the original data $\mathcal{D}_n$.

coincides with the empirical quantile at level $1-\eta$ of the statistical population $\{\sup_{\alpha\in[\varepsilon,1-\varepsilon]} |r_n^{*(j)}(\alpha)| : j = 1, \ldots, N\}$. Concerning the number of bootstrap replications, picking $N = n$ does not modify the rate of convergence. However, choosing N of magnitude comparable to n so that $(1+N)(1-\eta)$ is an integer may be more appropriate: the $(1-\eta)$-quantile of the approximate bootstrap distribution is then uniquely defined and this does not impact the rate of convergence neither (Hall, 1986).

4.2.1. Bootstrap consistency

The next theorem reveals the asymptotic validity of the bootstrap estimate proposed above where we assume that the smoothed version $\widetilde{F}_s$ of the distribution function $\widehat{F}_s$ is computed at step 1 of Algorithm 1 using a kernel K_{h_n}. It requires the following assumptions.

(B1) The density f_s is bounded and of class $\mathcal{C}^3$.
(B2) The bandwidth h_n decreases towards 0 as $n \to +\infty$ in a way that $nh_n \to +\infty$, $nh_n/\log(h_n^{-1}) \to +\infty$ and $\log(h_n^{-1})/\log\log n \to +\infty$.
(B3) The kernel K has finite support and satisfies the following conditions: $\int K(y)dy = 1$, $\int yK(y)dy = 0$ and $\int y^2K(y)dy < +\infty$.
(B4) The kernel K is such that $\|K\|_2 < +\infty$ and is of the form $K(y) = \Phi_1(P_1(y))$, P_1 being a polynomial and Φ_1 a bounded real function of bounded variation.

(B5) The kernel K is differentiable with derivative K' such that $\|K'\|_2 < +\infty$ and of the form $K'(y) = \Phi_2(P_2(y))$, P_2 being a polynomial and Φ_2 a bounded real function of bounded variation.

(B6) The bandwidth h_n is such that $\log(h_n^{-1})/(nh_n^3)$ tends to 0 as $n \to +\infty$.

Assumptions **(B1)** and **(B3)** are needed to control the biases $\sup_{\mathbb{R}} |\mathbb{E}[\widetilde{f}_s] - f_s|$ and $\sup_{\mathbb{R}} |\mathbb{E}[\widetilde{f}_s{}'] - f_s'|$. Assumption **(B2)** on the bandwidth h_n and assumption **(B4)** (respectively assumption **(B5)**) are needed to control $\sup_{\mathbb{R}} |\widetilde{f}_s - \mathbb{E}[\widetilde{f}_s]|$ (respectively $\sup_{\mathbb{R}} |\widetilde{f}_s{}' - \mathbb{E}[\widetilde{f}_s']|$) thanks to the result of Giné and Guillou (2002). Assumption **(B6)** ensures that $\sup_{\mathbb{R}} |\widetilde{f}_s{}' - f_s'|$ tends to 0 almost surely as $n \to \infty$. Assumptions **(B3)**–**(B5)** are fulfilled, for instance, by the biweight kernel defined for all $t \in \mathbb{R}$ by:

$$K(t) = \frac{15}{16}(1 - t^2)^2 \cdot \mathbb{I}\{|t| \leq 1\}\,. \tag{9}$$

Theorem 2. *(*Asymptotic validity*) Let $\varepsilon \in (0,1)$ and let $\mathbb{P}^*(.)$ denote the conditional probability given the original data $\mathcal{D}_n$. Suppose that assumptions **(A3)**–**(A5)**, **(A7)** and **(B1)**–**(B6)** are fulfilled. Then, the distribution estimate $\mathbb{P}^*(\sup_{\alpha\in[\varepsilon,1-\varepsilon]} |r_n^*(\alpha)| \leq \cdot)$ given by Algorithm 1 is such that we almost surely have*

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}^* \left(\sup_{\alpha\in[\varepsilon,1-\varepsilon]} |r_n^*(\alpha)| \leq t \right) - \mathbb{P} \left(\sup_{\alpha\in[\varepsilon,1-\varepsilon]} |r_n(\alpha)| \leq t \right) \right| = O(w_n).$$

where $w_n = \sqrt{\log(h_n^{-1})/nh_n} + h_n^2$.

The proof is given in Appendix A. Note that under assumption **(B2)** w_n tends to 0 as $n \to +\infty$. The primary tuning parameters of Algorithm 1 concern the bandwidth h_n. The optimal bandwidth is obtained by minimizing the term w_n with respect to h_n. This leads to $h_n \sim (\log n/n)^{1/5}$ and an approximation error of order $O((\log n/n)^{2/5})$ for the bootstrap estimate (see (Stute, 1982) for details on the derivation of the optimal bandwidth). Notice that assumptions **(B2)** and **(B6)** are fulfilled by such a bandwidth.

Although the rate of the bootstrap estimate is slower than that of the Gaussian approximation, the smoothed bootstrap method remains very appealing from a computational perspective: it is indeed very difficult to build confidence bands from simulated Brownian bridges in practice. Finally, as said above, it should be noticed that a non smoothed bootstrap of the MV curve would lead to worse approximation rates, of the order $O_{\mathbb{P}}(n^{-1/4})$ namely, see (Falk and Reiss, 1989).

Remark 8. *In the result of Theorem 2 we consider the supremum over $\alpha \in [\varepsilon, 1-\varepsilon]$ instead of the supremum over $\alpha \in [0, 1-\varepsilon]$ for several reasons. One of the reasons is that to obtain a rate of convergence for the bootstrap approximation we need to have the boundedness of the density of the supremum of the absolute value of the Gaussian process Z_1. This is almost immediate from the result of Pitt and Tran (1979) if the supremum is over $[\varepsilon, 1-\varepsilon]$ (see the proof of Theorem*

2 in Appendix A for more details). This is more complicated if the supremum is over $[0, 1-\varepsilon]$. *Using the result of Lifshits (1987), a closer look at the Gaussian process* Z_1 *might lead to the boundedness of the density of the supremum of* $|Z_1|$ *over* $[0, 1-\varepsilon]$. *Another reason is that several arguments in the proof also use the fact that* $\inf_{[\varepsilon,1-\varepsilon]} f_s \circ \alpha_s^{-1} > 0$. *However we do not have* $\inf_{[0,1-\varepsilon]} f_s \circ \alpha_s^{-1} > 0$ *as under the assumptions of Theorem 2* f_s *is continuous on* $\mathbb{R}$ *and therefore* $f_s(\alpha_s^{-1}(0)) = 0$.

4.2.2. Illustrative numerical experiments

Let $x \in \mathbb{R} \mapsto \mathcal{N}(\mu, \Sigma)(x) \in \mathbb{R}$ denote the density of a Gaussian distribution with mean μ and covariance Σ. We consider a two-dimensional Gaussian mixture whose density is given by:

$$\forall x \in \mathbb{R}, \quad f(x) = 0.5\mathcal{N}(\mu_1, \Sigma_1)(x) + 0.5\mathcal{N}(\mu_2, \Sigma_2)(x) \tag{10}$$

where $\mu_1 = (0, 0)$, $\mu_2 = (-1, -1)$,

$$\Sigma_1 = \begin{pmatrix} 2 & 2 \\ 2 & 4 \end{pmatrix} \text{ and } \Sigma_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

Density level sets of such a distribution are shown in Fig. 3(a). We draw a sample of size $n = 500$ from this two-dimensional Gaussian mixture and compute $\widehat{\mathrm{MV}}_f$. We then apply Algorithm 1 using the biweight kernel defined in (9) with a bandwidth $h = 0.005$ to obtain a 90% confidence band. We take $\varepsilon = 0.05$ and $N = n$ bootstrap replications to approximate $\mathbb{P}^*(\sup_{\alpha\in[\varepsilon,1-\varepsilon]} |r_n^*(\alpha)| \leq \cdot)$. All the volumes required to compute the MV curves are estimated using Monte-Carlo integration, drawing uniformly 1,000,000 samples in the hypercube enclosing the data. Fig. 2 shows the 90% confidence curves.

In this section, statistical estimation of the true MV curve of a given scoring function s has been investigated, as well as the problem of building confidence regions in the Mass Volume space. In all the rest of the paper, focus is on statistical learning of a nearly optimal scoring function s based on a training sample $X_1, \ldots, X_n$ with respect to the MV curve criterion.

5. A M-estimation approach to anomaly scoring

Now we are are equipped with the concept of Mass Volume curve, the anomaly scoring task can be formulated as the building of a scoring function s, based on the training set $X_1, \ldots, X_n$, such that MV_s is as close as possible to the optimum MV^*. Due to the functional nature of the criterion performance, there are many ways of measuring how close the MV curve of a scoring function candidate and the optimal one are. The L_p-distances, for $1 \leq p \leq +\infty$, provide a relevant collection of risk measures. Let $\varepsilon \in (0, 1)$ be fixed (take $\varepsilon = 0$ if

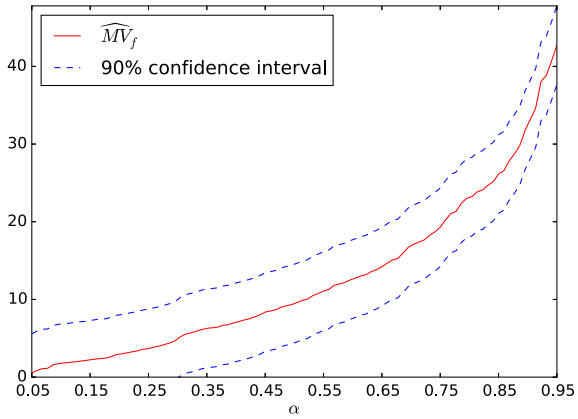


FIG 2. *Empirical MV curve and* 90% *confidence interval curves obtained with the smooth bootstrap approach. Note that it seems that the* 90% *confidence interval curves get closer to the empirical* MV *curve as* α *increases but this is just an optical illusion.*

$\lambda(\mathrm{supp}(f)) < +\infty$) and consider the losses related to the L_1-distance and that related to the sup norm:

$$\begin{aligned} d_1(s,f) &= \int_0^{1-\varepsilon} |\mathrm{MV}_s(\alpha) - \mathrm{MV}^*(\alpha)| d\alpha \,, \\ d_\infty(s,f) &= \sup_{\alpha\in[0,1-\varepsilon]} \{\mathrm{MV}_s(\alpha) - \mathrm{MV}^*(\alpha)\} \,. \end{aligned}$$

Observe that, by virtue of Proposition 1, the 'excess-risk' decomposition applies in the L_1 case and the learning problem can be directly tackled through standard M-estimation arguments:

$$d_1(s,f) = \int_0^{1-\varepsilon} \mathrm{MV}_s(\alpha) d\alpha - \int_0^{1-\varepsilon} \mathrm{MV}^*(\alpha) d\alpha \,.$$

Hence, possible learning techniques could be based on the minimization, over a set $\mathcal{S}_0 \subset \mathcal{S}$ of candidates, of empirical counterparts of the area under the MV curve, such as $\int_0^{1-\varepsilon} \widehat{\mathrm{MV}}_s(\alpha) d\alpha$. In contrast, the approach cannot be straightforwardly extended to the sup norm situation. A possible strategy is to combine M-estimation with approximation methods so as to 'discretize' the functional optimization task. This strategy can be implemented as follows. First, we replace the unknown target curve MV^* by an approximation that can be described by a finite number of scalar parameters, by a piecewise constant approximant MV^*_σ whose breakpoints are given by a subdivision $\sigma : \ 0 < \alpha_1 < \cdots < \alpha_K = 1 - \varepsilon$ precisely, and that is itself a MV curve, namely the MV curve of the piecewise constant scoring function

$$s^*_\sigma(x) = \sum_{k=1}^{K} (K - k + 1) \cdot \mathbb{I}\{x \in \Omega^*_{\alpha_k} \setminus \Omega^*_{\alpha_{k-1}}\} \tag{11}$$

which is indeed a piecewise constant approximant of MV^* related to the meshgrid σ (see (12)). Then, the L_∞-risk can be decomposed as the sum of two terms

$$d_\infty(s,f) \leq \sup_{\alpha\in[0,1-\varepsilon]} |\mathrm{MV}_s(\alpha) - \mathrm{MV}^*_\sigma(\alpha)| + \sup_{\alpha\in[0,1-\varepsilon]} \{\mathrm{MV}^*_\sigma(\alpha) - \mathrm{MV}^*(\alpha)\},$$

the second term on the right-hand side being viewed as the *bias* of the statistical method. Restricting optimization to the first term on the right-hand side of the L_∞-risk decomposition, the problem thus boils down to recovering the bilevel sets $\mathcal{R}^*_k = \Omega^*_{\alpha_k} \setminus \Omega^*_{\alpha_{k-1}}$ for $k = 1,\ldots,K$ as we obviously have

$$s^*_\sigma \in \operatorname*{argmin}_{s\in\mathcal{S}} \sup_{\alpha\in[0,1-\varepsilon]} |\mathrm{MV}_s(\alpha) - \mathrm{MV}^*_\sigma(\alpha)|.$$

This simple observation paves the way for designing scoring strategies relying on the estimation of a finite number of minimum volume sets.

In the next section, we describe a learning algorithm for anomaly ranking, that can be viewed to a certain extent as a statistical version of an adaptive approximation method by piecewise constants introduced in (DeVore, 1987) (see also Section 3.3 in (DeVore, 1998)), to build a piecewise constant estimate of the optimal curve MV^* and a nearly optimal piecewise constant scoring function, mimicking s^*_σ. The subdivision σ is entirely learnt from the data, in order to produce accurate estimates of MV^* in an adaptive fashion: looking at Fig. 1, an ideal meshgrid should be loose where MV^* is nearly flat or grows very slowly ('near' 0) and refined when it exhibits high degrees of variability (as one gets closer to 1).

Before describing and analyzing a prototypal approach to MV curve optimization, a few remarks are in order.

Remark 9. (Connections with supervised ranking) *Based on the observation made in Remark 3, one may see that, in the specific case where the support of $F(dx)$ coincides with the unit square $[0,1]^d$, the* MV *curve of any scoring function $s(x)$ corresponds to the reflection about the first diagonal of the* ROC *curve of $s(x)$ when the 'negative distribution' is the uniform distribution on $[0,1]^d$ and the 'positive distribution' is $F(dx)$. As shown in (Clémençon and Robbiano, 2014), this permits to turn unsupervised ranking into supervised ranking in the compact support situation and to exploit supervised ranking algorithms combined with random sampling to solve the* MV *curve minimization problem. A similar idea had been proposed in (Steinwart, Hush and Scovel, 2005) to turn anomaly detection into supervised binary classification.*

Remark 10. (Plug-in) *As the density f is an optimal scoring function, a natural strategy would be to estimate first the unknown density function f by means of (non-) parametric techniques and next use the resulting estimator as a scoring function. Beyond the computational difficulties one would be confronted to for large or even moderate values of the dimension, we point out that the goal pursued in this paper is by nature very different from density estimation: the local properties of the density function are useless here, only the ordering of the*

possible observations $x \in \mathcal{X}$ it induces is of importance (see Proposition 1). One may also show that a candidate f_1 can be a better approximation of the density f than a candidate f_2 for the L_q loss say, but a worse approximation for the MV *curve criterion (see Example 3).*

Example 3. *Let f be the density of the truncated normal distribution with support $[0,1]$, mean equal to 0.5 and variance equal to 0.0225. Let's consider $f_1 : x \in \mathbb{R} \mapsto f(x - 0.05) \in \mathbb{R}$ and f_2 defined for all $x \in \mathbb{R}$ as*

$$f_2(x) = \begin{cases} f(x) + \|f_1 - f\|_{L_2} + 2 \text{ if } x \in [0,1] \\ 0 \text{ otherwise}. \end{cases}$$

We thus have $\|f_2 - f\|_{L_2} = \|f_1 - f\|_{L_2} + 2 > \|f_1 - f\|_{L_2}$. Hence f_1 is a better approximation of f with respect to the L_2 distance. However $\mathrm{MV}_{f_2} < \mathrm{MV}_{f_1}$ and f_2 is a better scoring function than f_1 with respect to the MV *curve criterion. Indeed one can first show that $\mathrm{MV}_{f_2} = \mathrm{MV}_f$. This thus gives $\mathrm{MV}_{f_2} = \mathrm{MV}^* \leq \mathrm{MV}_{f_1}$. Now let $\alpha \in (0,1)$, $\Omega_{f_1,Q(f_1,\alpha)} = \{x, f_1(x) \geq Q(f_1,\alpha)\}$ is of the form $[0.05 - x_\alpha, 0.05 + x_\alpha]$ with $x_\alpha > 0$ and $F(\Omega_{s,Q(f_1,\alpha)}) \geq \alpha$. The set $\Omega_{f,Q^*(\alpha)}$ is of the form $[-x^*_\alpha, x^*_\alpha]$ with $x^*_\alpha > 0$ and $F(\Omega_{f,Q^*(\alpha)}) = \alpha$. If $\mathrm{MV}_{f_1}(\alpha) = \mathrm{MV}^*(\alpha)$, i.e., $\lambda(\Omega_{f_1,Q(f_1,\alpha)}) = \lambda(\Omega_{f,Q^*(\alpha)})$, then on the one hand $x_\alpha = x^*_\alpha$ and on the other hand $\Omega_{f_1,Q(f_1,\alpha)} = \Omega_{f,Q^*(\alpha)}$ λ almost everywhere by the uniqueness of the solution of (1) which is impossible. Eventually, $\mathrm{MV}_{f_2} < \mathrm{MV}_{f_1}$. One can also observe that in contrary to f_1, f_2 preserves the order induced by f.*

6. The A-Rank algorithm

Now that the anomaly scoring problem has been rigorously formulated, we propose a statistical method to solve it and establish learning rates for the sup norm loss.

6.1. Piecewise constant scoring functions

We focus on scoring functions of the simplest form, piecewise constant functions. Let $K \geq 1$ and consider a partition $\mathcal{W}$ of the feature space $\mathcal{X}$ in K pairwise disjoint subsets of finite Lebesgue measure: $\mathcal{C}_1, \ldots, \mathcal{C}_K$ and the subset $\mathcal{X} \backslash \cup_{k=1}^K \mathcal{C}_k$. When $\lambda(\mathrm{supp}(f))$ is finite, one may suppose $\mathcal{X} \setminus \cup_{k=1}^K \mathcal{C}_k$ of finite Lebesgue measure. Then, define the piecewise constant scoring function given by:

$$\forall x \in \mathcal{X}, \quad s_{\mathcal{W}}(x) = \sum_{k=1}^{K} (K - k + 1) \cdot \mathbb{I}\{x \in \mathcal{C}_k\}\,.$$

Its piecewise constant MV curve is given by: $\forall \alpha \in [0, F(\cup_{k=1}^K \mathcal{C}_k))$,

$$\mathrm{MV}_{s_{\mathcal{W}}}(\alpha) = \sum_{k=0}^{K-1} \lambda_{k+1} \cdot \mathbb{I}\{\alpha \in [\alpha_k,\ \alpha_{k+1})\}\,, \tag{12}$$

where $\alpha_0 = 0$, $\alpha_k = F(\cup_{j=1}^k \mathcal{C}_j)$ and $\lambda_k = \lambda(\cup_{j=1}^k \mathcal{C}_j)$ for all $k \in \{1, \ldots, K\}$. If $\lambda(\mathrm{supp}(f))$ is finite, $\mathrm{MV}_{s_{\mathcal{W}}}$ can be defined on the whole interval $[0,1]$.

6.2. Adaptive approximation of the optimal MV curve

Using ideas developed in (DeVore, 1987) (see also Section 3.3 in (DeVore, 1998)), we propose to design an adaptive approximation scheme instead of a subdivision fixed in advance. In such a procedure, the subdivision is progressively refined by adding new breakpoints, as further information about the local variation of the target MV^* is gained: the subdivision will be coarse where the optimal MV^* is almost flat and fine where it grows rapidly.

We restrict ourselves only to dyadic subdivisions with breakpoints $\alpha_{j,k} = k(1-\varepsilon)/2^j$, with $j \in \mathbb{N}$ and $k \in \{0, \ldots, 2^j\}$ and to partitions of the interval $[0, 1-\varepsilon]$ produced by recursive dyadic partitioning: any dyadic subinterval $I_{j,k} = [\alpha_{j,k}, \alpha_{j,k+1})$ is possibly split into two halves, producing two siblings $I_{j+1,2k}$ and $I_{j+1,2k+1}$, depending on the local properties of the target MV^*. This adaptive estimation algorithm can be viewed as a top-down search strategy through a binary tree structure.

A binary tree $\mathcal{T}$ is a tree where all internal nodes have exactly two children. The root of the tree $\mathcal{T}$ represents the interval $[0, 1-\varepsilon]$ and each node of depth j a subinterval $I_{j,k}$. The two children resulting of a split of the node $I_{j,k}$ are the nodes $I_{j+1,2k}$ and $I_{j+1,2k+1}$. A terminal node or a leaf is a node without children. In the sequel, $I_{j,k}$ will denote interchangeably the interval and the related node. Equipped with this flexible tree structure, $\mathcal{T}$ define a possibly very heterogeneous subdivision $\sigma_{\mathcal{T}}$ of $[0, 1-\varepsilon]$

The approximant associated with the the binary tree $\mathcal{T}$ is given by: $\forall \alpha \in [0, 1-\varepsilon]$,

$$\mathrm{MV}_{\sigma_{\mathcal{T}}}(\alpha) = \sum_{\substack{j,k \\ I_{j,k} \text{ leaf}}} \mathrm{MV}^*(\alpha_{j,k+1}) \cdot \mathbb{I}\{\alpha \in I_{j,k}\}$$

where the sum is over all the j, k such that $I_{j,k}$ is a leaf. In order to decide whether a subinterval $I = [\alpha_1, \alpha_2] \subset [0, 1-\varepsilon]$ should be split or not, we use the local error on the subinterval I defined by

$$\mathcal{E}(I) = \mathrm{MV}^*(\alpha_2) - \mathrm{MV}^*(\alpha_1)\,.$$

The local error $\mathcal{E}(I)$ provides a simple way of estimating the variability of the nondecreasing function MV^* on the interval I. This measure is nonnegative and additive: for any siblings I_1 and I_2 of the same subinterval I,

$$\mathcal{E}(I) = \mathcal{E}(I_1) + \mathcal{E}(I_2)\,.$$

From (DeVore, 1987), we know that $\mathcal{E}$ controls the approximation rate of MV^* by a constant on any interval I in the sense that:

$$\inf_{c \in [0,\infty)} \|\mathrm{MV}^*(.) - c\|_I \leq \mathcal{E}(I)\,,$$

where for any function $g : \mathbb{R} \to \mathbb{R}$, $\|g\|_I$ denotes the sup norm over the interval I.

Given a tolerance $\tau > 0$, the general concept of the algorithm generating a binary tree $\mathcal{T}$ and leading to its associated approximant is as follows: if $\mathcal{E}(I) > \tau$,

the interval I is split in two siblings, otherwise I is not split and becomes a leaf. If the algorithm ends we thus have:

$$\sup_{\alpha\in[0,1-\varepsilon]} \{\mathrm{MV}^*_{\sigma_{\mathcal{T}}}(\alpha) - \mathrm{MV}^*(\alpha)\} \leq \tau.$$

An illustration is given in Fig. 3: Fig. 3(a) shows the piecewise constant approximant of MV^* as well as the corresponding piecewise scoring function $s_{\mathcal{T}}(x)$ and Fig. 3(b) shows the related binary tree.

6.3. Empirical adaptive estimation of the optimal MV curve

The adaptive approximation procedure described above can be turned itself into an adaptive estimation technique. As the optimal curve MV^* is unknown, the local error $\mathcal{E}(I)$ on a subinterval I is estimated by its empirical counterpart

$$\widehat{\mathcal{E}}(I) = \lambda(\widehat{\Omega}_{\alpha_2}) - \lambda(\widehat{\Omega}_{\alpha_1}),$$

where $\widehat{\Omega}_\alpha$ is an estimation of the minimum volume set Ω^*_α. Let $\mathcal{G}$ be a class of measurable subsets of the feature space $\mathcal{X}$ of finite Lebesgue measure and $\alpha \in (0,1)$. The empirical minimum volume set $\widehat{\Omega}_\alpha$ related to the class $\mathcal{G}$ and level α is the solution of the optimization problem:

$$\inf_{\Omega\in\mathcal{G}} \lambda(\Omega) \text{ subject to } \widehat{F}(\Omega) \geq \alpha - \phi, \tag{13}$$

where ϕ is a penalty term related to the complexity of the class, referred to as the *penalty parameter*. The accuracy of the solution depends on the choices for the class $\mathcal{G}$ and for the penalty parameter (see (Scott and Nowak, 2006)). In this paper, we do not address the issue of designing algorithms for empirical minimum volume set estimation, we refer to Section 7 of (Scott and Nowak, 2006) for a description of off-the-shelf methods documented in the literature, including partition-based techniques.

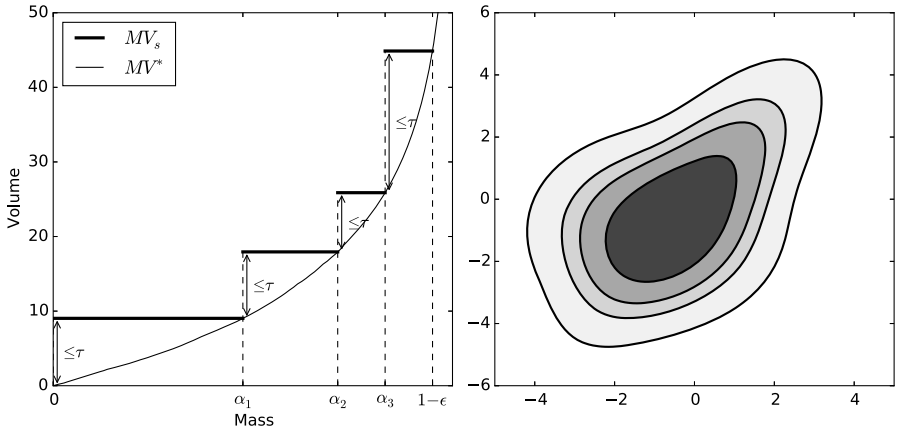
As $\mathcal{E}(I)$, the empirical local error $\widehat{\mathcal{E}}(I)$ is nonnegative and additive:

$$\widehat{\mathcal{E}}(I) = \widehat{\mathcal{E}}(I_1) + \widehat{\mathcal{E}}(I_2)$$

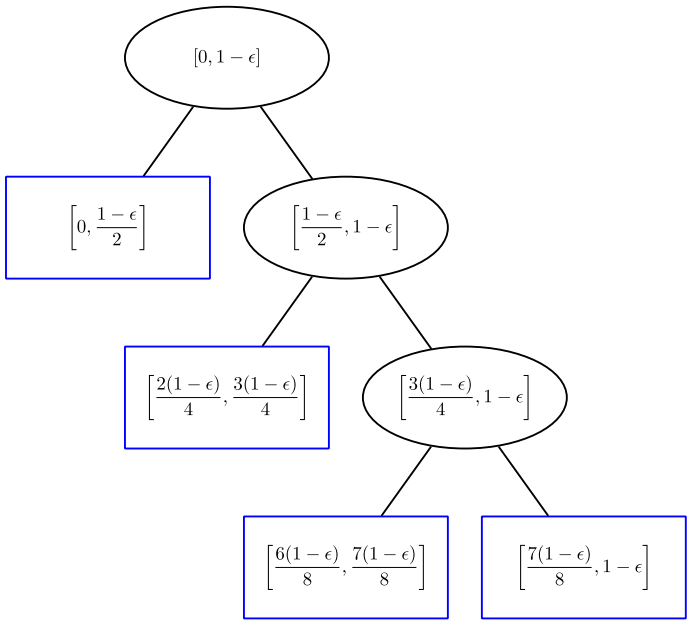
for any siblings I_1 and I_2 of the same subinterval.

Splitting criterion The empirical local error $\widehat{\mathcal{E}}$ will serve as a splitting criterion. Given a tolerance $\tau > 0$, if $\widehat{\mathcal{E}}(I) > \tau$, the interval I is split in two subintervals I_1 and I_2. If $\widehat{\mathcal{E}}(I) \leq \tau$, the interval I is not split and is a leaf of the tree $\mathcal{T}$.

Stopping criterion One can observe that the estimation algorithm does not necessarily terminate for any given tolerance $\tau > 0$. Indeed, as $\widehat{F}(\Omega) \in \{k/n : k = 0,\ldots,n\}$ for any $\Omega \in \mathcal{G}$, the function $\widehat{\lambda} : \alpha \in [0,1) \mapsto \lambda(\widehat{\Omega}_\alpha) \in \mathbb{R}$ is a piecewise constant function with breakpoints $\{k/n + \phi : k = 0,\ldots,n-1\}$. This follows by observing that, for any $\alpha \in (k/n+\phi, (k+1)/n+\phi]$, $0 \leq k \leq n-1$,



(a) Piecewise constant adaptive approximant of MV^* (on the left) and associated piecewise constant scoring function (on the right).



(b) Associated binary tree

FIG 3. *Illustration of the output of the adaptive approximation by piecewise constants of* MV^* *for a tolerance* $\tau = 20$ *and where* $\mathrm{MV}^* = \mathrm{MV}_f$, f *being the density given in* (10). *Note that this algorithm is run assuming that we know* f *and therefore that we have access to* MV^* *and the minimum volume sets* $\{\Omega^*_\alpha,\ \alpha \in (0,1)\}$.

the solution of the empirical optimization problem (13) is $\widehat{\Omega}_{(k+1)/n+\phi}$. Therefore if τ is strictly lower than the minimum of the amplitude of the jumps of the piecewise constant function $\widehat{\lambda}$, the algorithm does not stop.

However, the fact that the function $\widehat{\lambda}$ is a piecewise constant function tells us that the estimation algorithm should stop when the level $j_{\max} = \lfloor \log n / \log 2 \rfloor + 1$ is reached, where $\lfloor \cdot \rfloor$ denotes the floor part function. Indeed if $j = j_{\max}$, the interval $I_{j,k} = [\alpha_1, \alpha_2]$ is such that $\alpha_2 - \alpha_1 < 1/n$. This either means that α_1 and α_2 belong to the same interval $(k/n + \phi, (k+1)/n + \phi]$ or that $\alpha_1 \in (k/n + \phi, (k+1)/n + \phi]$ and $\alpha_2 \in ((k+1)/n + \phi, (k+2)/n + \phi]$. In the former case $\lambda(\widehat{\Omega}_{\alpha_2}) = \lambda(\widehat{\Omega}_{\alpha_1})$ and $\widehat{\mathcal{E}}(I_{j,k}) = 0 \leq \tau$. In the latter case, $\widehat{\mathcal{E}}(I_{j,k})$ is equal to the amplitude of the jump of $\widehat{\lambda}$ at $(k+1)/n + \phi$. If this amplitude is lower than τ, $I_{j,k}$ is a leaf. If this amplitude is strictly greater than τ, we should normally split $I_{j,k}$. However this split will not improve the error $\widehat{\mathcal{E}}$ as $\widehat{\mathcal{E}}$ cannot be lower than the amplitude of the jump of $\widehat{\lambda}$. We therefore add a condition $j < j_{\max}$ to the algorithm. This ensures that the algorithm terminates while not changing the output when τ is greater that the maximum of the amplitudes of the jumps of $\widehat{\lambda}$. Note however that we do not necessarily have $\widehat{\mathcal{E}}(I_{j,k}) \leq \tau$ for all j, k such that $I_{j,k}$ is a leaf.

Algorithm 2 Adaptive estimation of the optimal curve MV^*

Inputs: training set, penalty ϕ, tolerance τ
Create a binary tree $\mathcal{T}$ with root node $I_{0,0} = [0, 1 - \varepsilon]$
Create an empty list of nodes $\mathcal{L}$
Add the node $I_{0,0}$ to the list $\mathcal{L}$
while $\mathcal{L}$ is not empty **do**
    Get first element of $\mathcal{S}$: node $I_{j,k} = [\alpha_1, \alpha_2]$
    if $j < j_{\max}$ **then**
        Compute $\widehat{\mathcal{E}}(I_{j,k}) = \lambda(\widehat{\Omega}_{\alpha_2}) - \lambda(\widehat{\Omega}_{\alpha_1})$ where $\widehat{\Omega}_\alpha$ is the solution of (13)
        if $\widehat{\mathcal{E}}(I_{j,k}) > \tau$ **then**
            Split $I_{j,k}$ in two siblings $I_{j+1,2k}$ and $I_{j+1,2k+1}$
            Add the node $I_{j+1,2k}$ to $\mathcal{L}$
            Add the node $I_{j+1,2k+1}$ to $\mathcal{L}$
        end if
    end if
end while
Output: Let $\widehat{\sigma}_\tau$ denote the collection of the dyadic levels $\alpha_{j,k}$ corresponding to the leafs of the tree $I_{j,k}$.

$$\widehat{\mathrm{MV}^*}(\alpha) = \sum_{\substack{j,k \\ I_{j,k} \text{ leaf}}} \lambda(\widehat{\Omega}_{\alpha_{j,k+1}}) \cdot \mathbb{I}\{\alpha \in I_{j,k}\} = \sum_{\substack{j,k \\ \alpha_{j,k} \in \widehat{\sigma}_\tau}} \lambda(\widehat{\Omega}_{\alpha_{j,k+1}}) \cdot \mathbb{I}\{\alpha \in I_{j,k}\}\,.$$

The adaptive estimation method is summarized in Algorithm 2. We denote by $\hat{\sigma}_\tau : 0 = \alpha_0 < \cdots < \alpha_{\widehat{K}} = 1 - \varepsilon$ the subdivision output by the algorithm such that the empirical piecewise constant estimator of MV^* (and of $\mathrm{MV}^*_{\hat{\sigma}_\tau}$) is

given by

$$\forall \alpha \in [0, 1-\varepsilon],\ \widehat{\mathrm{MV}^*}(\alpha) = \sum_{k=0}^{\widehat{K}-1} \lambda(\widehat{\Omega}_{\alpha_{k+1}}) \cdot \mathbb{I}\{\alpha \in [\alpha_k, \alpha_{k+1})\}\,.$$

Its bias incorporates in particular that caused by the approximation/discretization stage inherent to the method.

6.4. *The anomaly ranking algorithm* A-Rank

The goal pursued is actually to build a scoring function $\hat{s}(x)$ whose MV curve is asymptotically close to the empirical estimate $\widehat{\mathrm{MV}^*}$. In this respect, one should pay attention to the fact that $\widehat{\mathrm{MV}^*}$ is not a MV curve in general. Indeed, the sequence of estimated minimum volume sets $\widehat{\Omega}_{\alpha_k}, k \in \{0, \ldots, \widehat{K}\}$ sorted by increasing order of their mass α_k, is not necessarily increasing (for the inclusion), in contrast to the true minimum level sets $\Omega^*_{\alpha_k}$. This explains the monotonicity step of the following algorithm.

The A-Rank algorithm is implemented in two stages. Stage 1 consists in running Algorithm 2 to find the breakpoints $\alpha_0 = 0 < \alpha_1 < \cdots < \alpha_{\widehat{K}} = 1-\varepsilon < 1$ and the corresponding empirical minimum volume sets $\widehat{\Omega}_{\alpha_0}, \ldots, \widehat{\Omega}_{\alpha_{\widehat{K}}}$. Stage 2 consists in the monotonicity step before overlaying the estimated sets. The statistical learning method is described at length in Algorithm 3.

Before analyzing the statistical performance of the A-rank algorithm, a few remarks are in order.

Notice first that we could alternatively build a monotone sequence of subsets $(\widetilde{\Omega}_k)_{0 \leq k \leq \widehat{K}}$, recursively through: $\widetilde{\Omega}_{\widehat{K}} = \widehat{\Omega}_{\widehat{K}}$ and $\widetilde{\Omega}_k = \widehat{\Omega}_k \cap \widetilde{\Omega}_{k+1}$ for $k \in \{\widehat{K}-1, \ldots, 0\}$. The results established in the sequel straightforwardly extend to this construction.

One obtains as a byproduct of the algorithm a rough piecewise constant estimator of the (optimal scoring) function $(F_f \circ f)(x)$, namely $\sum_{k=1}^{K} (\alpha_{k+1} - \alpha_k) \cdot \mathbb{I}\{x \in \widetilde{\Omega}_k\}$, the Riemann sum approximating the integral (4) when μ is the Lebesgue measure.

6.5. *Performance bounds for the* A-Rank *algorithm*

We now prove a result describing the performance of the scoring function produced by the algorithm proposed in the previous section to solve the anomaly scoring problem, as formulated in Section 5. The following assumptions are required.

(C1) $\forall \alpha \in [0, 1-\varepsilon]$, $\Omega^*_\alpha \in \mathcal{G}$.

(C2) Let $(\varepsilon_i)_{i \geq 1}$ be a Rademacher chaos independent from the X_i's. The Rademacher average given by

$$\mathcal{R}_n = \mathbb{E}\left[\sup_{\Omega \in \mathcal{G}} \frac{1}{n} \left| \sum_{i=1}^{n} \varepsilon_i \cdot \mathbb{I}\{X_i \in \Omega\} \right| \right], \tag{14}$$

Algorithm 3 A-RANK

Inputs: training set, penalty Φ, tolerance τ
Get $(\alpha_k, \widehat{\Omega}_{\alpha_k})_{0\leq k\leq \widehat{K}}$ from Algorithm 2
Monotonicity step: $\widetilde{\Omega}_{\alpha_0} = \widehat{\Omega}_{\alpha_0}$
for $k \in \{1, \ldots, \widehat{K}\}$ **do**

$$\widetilde{\Omega}_{\alpha_k} = \widehat{\Omega}_{\alpha_k} \cup \widetilde{\Omega}_{\alpha_{k-1}}$$

end for
Output:

$$\hat{s}(x) = \sum_{k=0}^{\widehat{K}} (\widehat{K} - k + 1) \cdot \mathbb{I}\{x \in \widetilde{\Omega}_{\alpha_k} \setminus \widetilde{\Omega}_{\alpha_{k-1}}\}$$

where the expectation is over all the random variables, is such that for all $n \geq 1$, $\mathcal{R}_n \leq Cn^{-1/2}$, where $C > 0$ is a constant.

Assumption **(C2)** is very general, it is satisfied in particular when the class $\mathcal{G}$ is of finite VC dimension, see (Koltchinskii, 2006). We recall here that for $\delta \in (0, 1)$ we have (Scott and Nowak, 2006):

$$F\Big(\Big\{\sup_{\Omega\in\mathcal{G}} |\widehat{F}(\Omega) - F(\Omega)| > \phi_n(\delta)\Big\}\Big) \leq \delta \tag{15}$$

with $\forall n \geq 1$,

$$\phi_n(\delta) = 2\mathcal{R}_n + \sqrt{\frac{\log(1/\delta)}{2n}}\,. \tag{16}$$

Therefore assumption **(C2)** says that the rate of uniform convergence of true to empirical probabilities is of the order $O_{\mathbb{P}}(n^{-1/2})$. Assumption **(C1)** is in contrast very restrictive, it could be however relaxed at the price of a much more technical analysis, involving the study of the bias in empirical minimum volume set estimation under adequate assumptions on the smoothness of the boundaries of the Ω^*_α's, like in Tsybakov (1997) for instance (see Remark 11 below).

We state a rate of convergence for Algorithm 2. The following assumption shall be required.

(C3) The function MV^* is differentiable on $[0, 1)$ with derivative $\mathrm{MV}^{*\prime}(\alpha) = 1/Q^*(\alpha)$ for all $\alpha \in [0, 1)$. Futhermore the derivative $\mathrm{MV}^{*\prime}$ belongs to the space $L \log L$ of Borel functions $g : (0, 1) \to \mathbb{R}$ such that:

$$\|g\|_{L\log L} \stackrel{def}{=} \int_0^1 (1 + \log|g(\alpha)|)|g(\alpha)|d\alpha < +\infty\,.$$

As explained in (DeVore, 1987), the space $L \log L$ contains all spaces L_p, $p > 1$ but is strictly contained in L_1.

Theorem 3. *Let $\delta \in (0,1)$. Suppose that assumptions **(A1)**,**(A2)** and **(C1)**–**(C3)** hold true. Take $\tau = 5\phi_n(\delta)/Q^*(1-\varepsilon)$ with $\phi_n(\delta)$ as in (16). Then we have with probability at least $1-\delta$:*

$$\forall n \geq 1, \quad \sup_{\alpha \in [0,1-\varepsilon]} \left\{ \widehat{\mathrm{MV}^*}(\alpha) - \mathrm{MV}^*(\alpha) \right\} \leq \frac{1}{Q^*(1-\varepsilon)} \left(11\phi_n(\delta) + \frac{1}{n} \right) .$$

In addition, there exists a constant $C > 0$ such that with probability at least $1-\delta$ the number of terminal nodes $\mathrm{card}(\hat{\sigma}_\tau)$ output by Algorithm 2 is such that

$$\mathrm{card}(\hat{\sigma}_\tau) \leq C Q^*(1-\varepsilon) \frac{\|\mathrm{MV}^{*\prime}\|_{L \log L}}{\phi_n(\delta)} .$$

The proof of this result can be found in Appendix A. Its argument essentially combines the analysis of the approximation scheme described in Section 6.2 and the generalization bounds for empirical minimum volume sets established in in (Scott and Nowak, 2006), see Theorem 3 and Lemma 19 therein. Here, the rate obtained is not proved to be optimal in the minimax sense, lower bounds will be investigated in a future work.

Corollary 1. *Suppose that assumptions of Theorem 3 are satisfied. Given assumption **(C2)**, we have with probability at least $1-\delta$, $\forall n \geq 1$,*

$$\sup_{\alpha \in [0,1-\varepsilon]} \left\{ \widehat{\mathrm{MV}^*}(\alpha) - \mathrm{MV}^*(\alpha) \right\} \leq \frac{1}{Q^*(1-\varepsilon)} \left(11 \cdot \left(\frac{2C}{\sqrt{n}} + \sqrt{\frac{\log(1/\delta)}{2n}} \right) + \frac{1}{n} \right) .$$

To derive the rate of convergence of the MV curve of the estimated scoring function $\hat{s}$ towards MV^* we need the following assumption on the density f:

(C4) There exist constants $\gamma \geq 0$ and $C > 0$ such that for all $t > 0$ sufficiently small

$$\sup_{q>0} \lambda(\{x, |f(x) - q| \leqslant t\}) \leqslant C t^\gamma .$$

This condition (stated with $\mathbb{P}$ instead of λ) was first introduced in (Polonik, 1995) and used in (Polonik, 1997) to derive rates of convergence of $F(\widehat{\Omega}_\alpha \Delta \Omega^*_\alpha)$. It is also used in (Rigollet and Vert, 2009) to obtain fast rates for plug-in estimators of density level sets. The exponent γ controls the slope of the function f around any level $q > 0$. It is related to Tsybakov's margin assumption stated for the binary classification framework. The relation between Tsybakov's margin assumption and the γ-exponent assumption is given in (Steinwart, Hush and Scovel, 2005).

Theorem 4. *Let $\delta \in (0,1)$. Suppose that assumptions **(A1)**,**(A2)** and **(C1)**–**(C4)** hold true. Then, with probability at least $1-\delta$,*

$$\sup_{\alpha \in [0,1-\varepsilon]} \left\{ \mathrm{MV}_{\hat{s}}(\alpha) - \mathrm{MV}^*(\alpha) \right\} = O\left(n^{-\frac{\gamma}{4(1+\gamma)}} \right) .$$

Notice that the convergence rate is always slower than $n^{-1/4}$. Furthermore, because of assumption **(C1)**, the convergence rate only depends on the dimension through the parameter γ (refer to (Polonik, 1995) for examples of this parameter). However, assumption **(C1)**, made here for simplicity's sake, is very restrictive and relaxing this assumption is discussed in Remark 11.

Remark 11. (Bias analysis) *Whereas the bias resulting from the discretization of the anomaly ranking problem discussed in Section 6.2 is taken into account in the present rate bound analysis, that related to the approximation of the minimum volume sets Ω_α^* has been neglected here for simplicity's sake (assumption **(C1)**). We point out that, under smoothness assumptions on the density sublevel sets, it is possible to control such a bias term. Indeed, consider for instance the case where the boundary $\partial\Omega_\alpha^*$ is of finite perimeter $per(\partial\Omega_\alpha^*) < \infty$ (note that this is the case as soon as $f(x)$ is of bounded variation, the boundary being then $\partial\Omega_\alpha^* = \{x \in \mathcal{X} : \ f(x) = t_\alpha^*\}$ by virtue of f's continuity). In this case, if $\mathcal{G} = \mathcal{G}_j$ is the collection of all subsets of $\mathbb{R}^d$ obtained by binding together an arbitrary number of hypercubes of side length 2^{-j}, cartesian products of intervals of the form $[k/2^j,\ (k+1)/2^j)$ for $k \in \mathbb{Z}$, the bias term inherent to the estimation of the minimum volume set Ω_α^* is bounded by $per(\partial\Omega_\alpha^*)2^{-jd}$, up to a multiplicative constant (see Proposition 9.7 in (Mallat, 1990)). Smaller bounds can be naturally established under more restrictive assumptions involving a regularity parameter θ of $\partial\Omega_\alpha^*$, such as its box dimension. Although the optimal choice for j would then depend on θ, a standard fashion of nearly achieving the optimal rate of convergence is to perform model selection (see Section 4 in (Scott and Nowak, 2006)).*

The general approach described above can be extended in various ways. The class $\mathcal{G}$ over which minimum volume set estimation is performed (and the penalty parameter as well) could vary depending on the mass target α. Additionally assumption **(C1)** could be relaxed as discussed in Remark 11. In such a case, if $\Omega_\alpha^{\mathcal{G}}$ denotes the solution of the minimum volume set optimization problem (1) over the class $\mathcal{G}$, the error term $\lambda(\widehat{\Omega}_\alpha) - \lambda(\Omega_\alpha^*)$ could be decomposed into the sum of a stochastic error and an approximation error:

$$\left(\lambda(\widehat{\Omega}_\alpha) - \lambda(\Omega_\alpha^{\mathcal{G}})\right) + \left(\lambda(\Omega_\alpha^{\mathcal{G}}) - \lambda(\Omega_\alpha^*)\right).$$

The choice of the class $\mathcal{G}$ should then balanced the two errors. On the one hand, if the class $\mathcal{G}$ is small, *e.g.* of finite VC dimension, then the stochastic error can be controlled thanks to assumption **(C2)** but the approximation error may be very large for some distributions. On the other hand, if the class $\mathcal{G}$ is large, *e.g.* of infinite VC dimension, then the stochastic error will be very large for some distributions. A model selection approach could thus be incorporated, so as to select an adequate class $\mathcal{G}$ (refer to Section 4 in (Scott and Nowak, 2006)).

In the present analysis, we have restricted our attention to a truncated part of the MV space, corresponding to mass levels in the interval $[0, 1-\varepsilon]$ and avoiding thus out-of-sample tail regions (in the case where $\mathrm{supp}(f)$ is of infinite Lebesgue measure). An interesting direction for further research could consist

in investigating the accuracy of anomaly scoring algorithms when letting $\varepsilon = \varepsilon_n$ slowly decay to zero as $n \to +\infty$, under adequate assumptions on the asymptotic behavior of MV^*.

Finally, we emphasize that the A-Rank algorithm is prototypal, the essential objective of its description/analysis in this paper is to provide an insight into the nature of the anomaly scoring issue, viewed here as a continuum of minimum volume set estimation problems. Beyond the possible improvements mentioned above (*e.g.* target mass level dependent classes $\mathcal{G}$), the main limitation for the practical implementation of such an approach arises from the apparent lack of minimum volume set estimation techniques that can be scaled to high-dimensional settings documented in the literature (except the dyadic recursive partitioning method considered in (Scott and Nowak, 2006), see Section 6.3 therein). However, as observed in (Steinwart, Hush and Scovel, 2005) (see also (Clémençon and Robbiano, 2014)), supervised methods combined with sampling procedures can be used for this purpose in moderate dimensions. Refer also to Remark 9.

7. Conclusion

Motivated by a wide variety of applications including health monitoring of complex infrastructures or fraud detection for instance, we have formulated the issue of learning how to rank observations in the same order as that induced by the density function, which we called *anomaly scoring* here. For this problem, much less ambitious than estimation of the local values taken by the density, a functional performance criterion, the Mass Volume curve namely, is proposed. What MV curve analysis achieves for unsupervised anomaly detection is quite akin to what ROC curve analysis accomplishes in the supervised setting.

The statistical estimation of MV curves has been investigated from an asymptotic perspective and we have provided a strategy, where the feature space is overlaid with a few well-chosen empirical minimum volume sets, to build a scoring function with statistical guarantees in terms of rate of convergence for the sup norm in the MV space.

Acknowledgements

We are grateful to François Portier and Prof. Patrice Bertail for very helpful discussions. We also thank an anomymous reviewer for raising an important question which led to an improvement of Proposition 1.

Appendix A: Proofs

A.1. Properties of the MV curve

To prove Proposition 1 we need the following lemma.

Lemma 1. *The support of the density* $\mathrm{supp}(f) = \{x, f(x) > 0\}$ *is equal to the set* $\mathcal{Z}_f = \{x, F_f(f(x)) > 0\}$ *up to subsets of null* λ *measure, i.e.*

$$\lambda(\mathrm{supp}(f) \Delta \mathcal{Z}_f) = 0\,.$$

The proof is deferred to Appendix B.

Proof of Proposition 1. We start by proving that all elements of $\mathcal{S}^*$ have the same MV curve equal to MV_f. One can first show that for all $s \in \mathcal{S}$ and all $\alpha \in (0,1)$, $\alpha_s^{-1}(\alpha) = F_s^\dagger(1-\alpha) = Q(s,\alpha)$. Let $s^* \in \mathcal{S}^*$ and let $\alpha \in (0,1)$,

$$\begin{aligned}\mathrm{MV}_{s^*}(\alpha) &= \lambda\{x, s^*(x) \geq F_{s^*}^\dagger(1-\alpha)\} = \lambda\{x, F_{s^*}(s^*(x)) \geq 1-\alpha\} \\ &= \lambda\{x \in \mathrm{supp}(f), F_{s^*}(s^*(x)) \geq 1-\alpha\} \\ &\quad + \lambda\{x \in \overline{\mathrm{supp}(f)}, F_{s^*}(s^*(x)) \geq 1-\alpha\}\end{aligned}$$

where the second equality holds because for any x and any $\alpha \in (0,1)$, $s^*(x) \geq F_{s^*}^\dagger(1-\alpha)$ if and only if $F_{s^*}(s^*(x)) \geq 1-\alpha$ (see *e.g.* Property 2.3 in (Embrechts and Hofert, 2013)). Thanks to assertion (iv) in the definition of $\mathcal{S}^*$, for $x \in \overline{\mathrm{supp}(f)}$, $F_{s^*}(s^*(x)) = 0$ and as $1-\alpha > 0$, the term of the last line of the previous equation is equal to 0. Using the notations introduced in the definition of $\mathcal{S}^*$, the term $\lambda\{x \in \mathrm{supp}(f), F_{s^*}(s^*(x)) \geq 1-\alpha\}$ can be decomposed as follows:

$$\lambda\{x \in \mathrm{supp}(f) \cap \overline{\mathcal{Z}}, F_{s^*}(s^*(x)) \geq 1-\alpha\} + \lambda\{x \in \mathrm{supp}(f) \cap \mathcal{Z}, F_{s^*}(s^*(x)) \geq 1-\alpha\}\,.$$

The first term is lower than $\lambda(\mathrm{supp}(f) \cap \overline{\mathcal{Z}}) = 0$ (thanks to assertion (i) in the definition of $\mathcal{S}^*$) and is therefore equal to 0. We deal with the second term as follows. Let $x \in \mathrm{supp}(f) \cap \mathcal{Z}$,

$$\begin{aligned}F_{s^*}(s^*(x)) &= \mathbb{P}(s^*(X) \leq s^*(x)) \\ &= \mathbb{P}(X \in \mathrm{supp}(f), s^*(X) \leq s^*(x)) + \mathbb{P}(X \in \overline{\mathrm{supp}(f)}, s^*(X) \leq s^*(x))\,.\end{aligned}$$

The second term is lower than $\mathbb{P}(X \in \overline{\mathrm{supp}(f)}) = 0$ and is therefore equal to 0. Thanks again to assertion (i), the first term is equal to

$$\mathbb{P}(X \in \mathrm{supp}(f) \cap \mathcal{Z}, s^*(X) \leq s^*(x)) = \mathbb{P}(X \in \mathrm{supp}(f) \cap \mathcal{Z}, f(X) \leq f(x))$$

where we also used assertions (ii) and (iii). As done above for s^*, this last term can be shown to be equal to $F_f(f(x))$. Therefore,

$$\lambda\{x \in \mathrm{supp}(f) \cap \mathcal{Z}, F_{s^*}(s^*(x)) \geq 1-\alpha\} = \lambda\{x \in \mathrm{supp}(f) \cap \mathcal{Z}, F_f(f(x)) \geq 1-\alpha\}$$

and as done above for s^*, this can be shown to be equal to $\mathrm{MV}_f(\alpha)$.

We now show (3) and $\Longrightarrow$ in (2). Let $s \in \mathcal{S}$, $\alpha \in (0,1)$. We have,

$$\mathrm{MV}_s(\alpha) - \mathrm{MV}^*(\alpha) = \lambda(\Omega_{s,Q(s,\alpha)}) - \lambda(\Omega_\alpha^*) \leq \lambda(\Omega_\alpha^* \Delta \Omega_{s,Q(s,\alpha)})\,.$$

Now let $s^* \in \mathcal{S}^*$. As shown above, we have $\mathrm{MV}_{s^*} = \mathrm{MV}_f = \mathrm{MV}^*$. Furthermore, $F(\Omega_{s,Q(s,\alpha)}) = \mathbb{P}(s(X) \geq F_s^\dagger(1-\alpha)) = 1 - \mathbb{P}(s(X) < F_s^\dagger(1-\alpha))$ and one can show

that $\mathbb{P}(s(X) < F_s^\dagger(1-\alpha)) \leq 1-\alpha$. We therefore have $\mathbb{P}(\Omega_{s,Q(s,\alpha)}) \geq \alpha$. As Ω_α^* minimizes λ over all sets with mass at least α, we have $\mathrm{MV}_s(\alpha) - \mathrm{MV}^*(\alpha) \geq 0$.

Finally we prove $\Longleftarrow$ in (2). On the one hand, for $s = f$ we have $\mathrm{MV}_{s^*} \leq \mathrm{MV}_f$. On the other hand, as $f \in \mathcal{S}^*$, thanks to $\Longrightarrow$ in (2) (which has been proven above), for all scoring function s, $\mathrm{MV}_f \leq \mathrm{MV}_s$. Therefore $\mathrm{MV}_f \leq \mathrm{MV}_{s^*}$. We thus have $\mathrm{MV}_f = \mathrm{MV}_{s^*}$ on $(0,1)$. From now on, we replace s^* by s for the sake of clarity. We thus want to prove that $\mathrm{MV}_f = \mathrm{MV}_s$ on $(0,1)$ implies $s \in \mathcal{S}^*$. The proof is decomposed in three steps.

First step We first show that for λ-almost all $x \in \mathcal{X}$,

$$\mathbb{I}\{x \in \Omega_\alpha^*\} = \mathbb{I}\{x \in \Omega_{s,Q(s,\alpha)}\} \text{ for all } \alpha \in (0,1)$$

and

$$F_f(f(x)) = F_s(s(x))\,.$$

Let $\alpha \in (0,1)$ be fixed. We have $\lambda(\Omega_\alpha^*) = \lambda(\Omega_{s,Q(s,\alpha)})$. Since $\mathbb{P}(X \in \Omega_{s,Q(s,\alpha)}) \geq \alpha$, by uniqueness of the solution of the minimum volume set optimization problem up to subsets of null λ measure, we have $\lambda(\Omega_\alpha^* \Delta \Omega_{s,Q(s,\alpha)}) = 0$ which is equivalent to: for λ-almost all $x \in \mathcal{X}$

$$\mathbb{I}\{x \in \Omega_\alpha^*\} = \mathbb{I}\{x \in \Omega_{s,Q(s,\alpha)}\}$$

i.e. there exists $\mathcal{Z}_\alpha \subset \mathcal{X}$ such that $\lambda(\overline{\mathcal{Z}_\alpha}) = 0$ and for all $x \in \mathcal{Z}_\alpha$, $\mathbb{I}\{x \in \Omega_\alpha^*\} = \mathbb{I}\{x \in \Omega_{s,Q(s,\alpha)}\}$. Note that $\overline{\mathcal{Z}_\alpha}$ depends on α and we cannot consider to take the union over all $\alpha \in (0,1)$ of the $\overline{\mathcal{Z}_\alpha}$ because this union would not necessarily be of null λ measure (as we have uncountably many α). However considering the union over the rationals of $(0,1)$ is sufficient. Indeed, let $\mathcal{Z}_\mathbb{Q}$ denote such a union:

$$\overline{\mathcal{Z}_\mathbb{Q}} = \bigcup_{\alpha \in (0,1)\cap\mathbb{Q}} \overline{\mathcal{Z}_\alpha}\,.$$

As it is a union over a countable set of α, $\lambda(\overline{\mathcal{Z}_\mathbb{Q}}) = 0$ and for all $x \in \mathcal{Z}_\mathbb{Q} = \cap_{\alpha\in(0,1)\cap\mathbb{Q}} \mathcal{Z}_\alpha$ and all $\alpha \in (0,1)\cap\mathbb{Q}$,

$$\mathbb{I}\{x \in \Omega_\alpha^*\} = \mathbb{I}\{x \in \Omega_{s,Q(s,\alpha)}\}\,.$$

Let $x \in \mathcal{Z}_\mathbb{Q}$ and let $\alpha \in (0,1)\cap\mathbb{R}\setminus\mathbb{Q}$. We know that there exists a decreasing sequence $(\alpha_m)_{m\geq 0}$ of elements of $(0,1)\cap\mathbb{Q}$ such that α_m converges towards α when m tends to infinity. Now, thanks to a property of the quantile function (see e.g. assertion (5) of Property 2.3 in (Embrechts, 2013)), we have, for all $\alpha' \in (0,1)$,

$$\Omega_{\alpha'}^* = \{x', f(x') \geq F_f^\dagger(1-\alpha')\} = \{x', F_f(f(x')) \geq 1-\alpha'\}\,.$$

Thus the sequence $(\Omega_{\alpha_m}^*)_{m\geq 0}$ is decreasing (with respect to set inclusion) and

$$\lim_{m\to+\infty} \mathbb{I}\{x \in \Omega_{\alpha_m}^*\} = \mathbb{I}\{x \in \bigcap_{m\geq 0} \Omega_{\alpha_m}^*\} = \mathbb{I}\{x \in \Omega_\alpha^*\}$$

as one can show that

$$\bigcap_{m\geq 0} \Omega^*_{\alpha_m} = \Omega^*_\alpha \,.$$

Similarly,

$$\lim_{m\to+\infty} \mathbb{I}\{x \in \Omega_{Q(s,\alpha_m)}\} = \mathbb{I}\{x \in \Omega_{Q(s,\alpha)}\} \,.$$

For all $m \geq 0$, $\mathbb{I}\{x \in \Omega^*_{\alpha_m}\} = \mathbb{I}\{x \in \Omega_{Q(s,\alpha_m)}\}$ (because $\alpha_m \in (0,1) \cap \mathbb{Q}$) and thus the two limits are equal. Hence the first result.

Now this also implies that for $x \in \mathcal{Z}_\mathbb{Q}$,

$$\int_0^1 \mathbb{I}\{x \in \Omega^*_\alpha\}\mathrm{d}\alpha = \int_0^1 \mathbb{I}\{x \in \Omega_{s,Q(s,\alpha)}\}\mathrm{d}\alpha \,.$$

If U denotes a random variable with uniform distribution on $(0,1)$, the left-hand side can be rewritten as

$$\mathbb{E}_U[\mathbb{I}\{f(x) \geq F_f^\dagger(1-U)\}] = \mathbb{E}_X[\mathbb{I}\{f(x) \geq f(X)\}] = F_f(f(x)) \,.$$

Similarly, the right-hand side is equal to $F_s(s(x))$ and therefore for all $x \in \mathcal{Z}_\mathbb{Q}$, $F_f(f(x)) = F_s(s(x))$.

Second step We can now prove assertions (i), (ii), (iii) and (iv). We will first show that for almost all $x \in \mathrm{supp}(f)$, there exists $\alpha' \in (0,1]$ such that $f(x) = F_f^\dagger(\alpha')$.

One may first observe that $\lambda(\overline{\mathcal{Z}_f} \cap \mathrm{supp}(f)) = 0$ where $\mathcal{Z}_f = \{x, F_f(f(x)) > 0\}$. Indeed, $\mathcal{Z}_f \subset \mathrm{supp}(f)$ (if $f(x) = 0$ then $F_f(f(x)) = 0$) and therefore $\lambda(\overline{\mathcal{Z}_f} \cap \mathrm{supp}(f)) = \lambda(\mathrm{supp}(f) \setminus \mathcal{Z}_f)$ which is equal to 0 since, thanks to Lemma 1, $\lambda(\mathrm{supp}(f) \Delta \mathcal{Z}_f) = 0$.

Let $x \in \mathrm{supp}(f) \cap \mathcal{Z}_f$. If $F_f^\dagger$ is continuous at $\alpha' = F_f(f(x)) \in (0,1]$ then $F_f^\dagger(\alpha') = F_f^\dagger(F_f(f(x))) = f(x)$. Now, since $F_f^\dagger$ is increasing, $F_f^\dagger$ has at most countably many discontinuities and each discontinuity m corresponds to a jump of $F_f^\dagger$ between two values t_1^m and t_2^m. We also know that this jump of $F_f^\dagger$ corresponds to a flat part of F_f between t_1^m and t_2^m. However if F_f has a flat part between t_1^m and t_2^m then $\mathbb{P}(t_1^m \leq f(X) \leq t_2^m) = F_f(t_2^m) - F_f(t_1^m) = 0$ (recall that F_f is continuous as f has no flat parts). This gives

$$\int_{\{u, t_1^m \leq f(u) \leq t_2^m\}} f(u)\mathrm{d}u = 0 \,.$$

Thus $f = 0$ λ-almost everywhere on $\{u, t_1^m \leq f(u) \leq t_2^m\}$. This implies that $\lambda(\{x, t_1^m \leq f(x) \leq t_2^m\} \cap \mathrm{supp}(f)) = 0$. Let $\overline{\mathcal{Z}_0}$ be the union of such sets over $m \geq 0$:

$$\overline{\mathcal{Z}_0} = \bigcup_{m\geq 0} \{x, t_1^m \leq f(x) \leq t_2^m\}$$

and let $\mathcal{Z}_1 = \mathcal{Z}_0 \cap \mathcal{Z}_f$. We have $\lambda(\overline{\mathcal{Z}_1} \cap \mathrm{supp}(f)) = 0$ and for all $x \in \mathcal{Z}_1 \cap \mathrm{supp}(f)$, there exists $\alpha' \in (0,1]$ such that $f(x) = F_f^\dagger(\alpha')$ because $F_f^\dagger$ is continuous at $\alpha' =$

$F_f(f(x))$ (otherwise this would imply that x belongs to a region corresponding to a flat part of F_f, *i.e.* one of the sets $\{x, t_1^m \leq f(x) \leq t_2^m\}$).

Similarly, we can show that there exists a set $\mathcal{Z}_2$ such that $\lambda(\overline{\mathcal{Z}_2} \cap \text{supp}(f)) = 0$ and for all $x \in \mathcal{Z}_2 \cap \text{supp}(f)$, there exists $\alpha' \in (0,1]$ such that $s(x) = F_s^\dagger(\alpha')$. The only difference with f is that we here have to consider $\mathcal{Z}_\mathbb{Q}$ so that $F_s(s(x)) = F_f(f(x))$.

Let $\mathcal{Z} = \mathcal{Z}_\mathbb{Q} \cap \mathcal{Z}_1 \cap \mathcal{Z}_2$. We have $\lambda(\overline{\mathcal{Z}} \cap \text{supp}(f)) = 0$. Let $x_1, x_2 \in \mathcal{Z} \cap \text{supp}(f)$ such that $f(x_1) > f(x_2)$ and let $\Omega_1 = \{x \in \mathcal{Z}, f(x) \geq f(x_1)\}$. There exists $\alpha_1 \in [0,1)$ such that $F_f^\dagger(1-\alpha_1) = f(x_1)$. Let's first consider the case where $\alpha_1 \in (0,1)$. We have

$$\Omega_1 = \{x \in \mathcal{Z}, f(x) \geq F_f^\dagger(1-\alpha_1)\} = \{x \in \mathcal{Z}, s(x) \geq F_s^\dagger(1-\alpha_1)\}\,. \tag{17}$$

As $x_1 \in \Omega_1$, $x_1 \in \{x \in \mathcal{Z}, s(x) \geq F_s^\dagger(1-\alpha_1)\}$ and $s(x_1) \geq F_s^\dagger(1-\alpha_1)$. Analogously, as $x_2 \notin \Omega_1$, $x_2 \notin \{x \in \mathcal{Z}, s(x) \geq F_s^\dagger(1-\alpha_1)\}$ and $s(x_2) < F_s^\dagger(1-\alpha_1)$. Thus $s(x_1) > s(x_2)$. Let's now consider the case $\alpha_1 = 0$. We do not know that (17) still holds. However we can always assume without loss of generality that s is bounded. Indeed, if s is not bounded, consider the function $s' = \arctan \circ s$. As arctan is strictly increasing, thanks to Proposition 3, $\text{MV}_s = \text{MV}_{s'}$. Besides, as s takes its values in $\mathbb{R}^+$, s' takes its values in $[0, \pi/2)$ and is bounded. Finally, if we show that $s' = T \circ f$ with T strictly increasing then we also have that s is a strictly increasing transform of f. Then as s is bounded, $F_s^\dagger(1-\alpha_1)$ is defined at $\alpha_1 = 0$ and we can replace $\text{MV}_s = \text{MV}^*$ on $(0,1)$ by $\text{MV}_s = \text{MV}^*$ on $[0,1)$ in the proposition. We can then show that (17) holds for $\alpha_1 = 0$.

Similarly, $s(x_1) > s(x_2)$ implies $f(x_1) > f(x_2)$. Therefore for all $x_1, x_2 \in \mathcal{Z} \cap \text{supp}(f)$,

$$s(x_1) > s(x_2) \iff f(x_1) > f(x_2)\,.$$

and hence the existence of T. One can for instance take $T : f(x) \in f(\mathcal{Z} \cap \text{supp}(f)) \mapsto s(x) \in s(\mathcal{Z} \cap \text{supp}(f))$.

Third step Finally, assertion (iv) derives from the fact that for $x \in \mathcal{Z}_\mathbb{Q} \cap \overline{\text{supp}(f)}$ we have $F_s(s(x)) = F_f(f(x)) = 0$. □

Proof of Proposition 2. To prove the convexity of MV^* we show that the slopes are increasing. Let $\alpha_1 < \alpha_2 < \alpha_3$ be in $[0,1)$,

$$\begin{aligned}
\text{MV}^*(\alpha_2) - \text{MV}^*(\alpha_1) &= \lambda(\Omega^*_{\alpha_2}) - \lambda(\Omega^*_{\alpha_1}) \\
&= \lambda\{f \geq F_f^\dagger(1-\alpha_2)\} - \lambda\{f \geq F_f^\dagger(1-\alpha_1)\} \\
&= \lambda\{x, F_f(f(x)) \geq 1-\alpha_2\} - \lambda\{x, F_f(f(x)) \geq 1-\alpha_1\} \\
&= \lambda\{x, 1-\alpha_2 \leq F_f(f(x)) < 1-\alpha_1\} \\
&= \int_\mathcal{X} \mathbb{I}\{x, 1-\alpha_2 \leq F_f(f(x)) < 1-\alpha_1\} dx \\
&= \int_\mathcal{X} \mathbb{I}\{x, 1-\alpha_2 \leq F_f(f(x)) < 1-\alpha_1\} \frac{1}{f(x)} f(x) dx
\end{aligned}$$

where the third equality holds because for any x and any $\alpha \in [0,1)$, $f(x) \geq F_f^\dagger(1-\alpha)$ if and only if $F_f(f(x)) \geq 1-\alpha$ (see *e.g.* Property 2.3 in (Embrechts and Hofert, 2013)). Now as $x \in \{x, F_f^\dagger(1-\alpha_2) \leq f(x) < F_f^\dagger(1-\alpha_1)\}$, we obtain

$$\begin{aligned}\mathrm{MV}^*(\alpha_2) - \mathrm{MV}^*(\alpha_1) &\leq \frac{1}{F_f^\dagger(1-\alpha_2)} \mathbb{P}(F_f(f(X)) \in [1-\alpha_2, 1-\alpha_1)) \\ &= \frac{\alpha_2 - \alpha_1}{F_f^\dagger(1-\alpha_2)}\end{aligned}$$

because the distribution of $F_f(f(X))$ is the uniform distribution on $(0,1)$ (as F_f is continuous since f has no flat parts). Similarly, one can show that $\mathrm{MV}^*(\alpha_3) - \mathrm{MV}^*(\alpha_2) \geq (\alpha_3 - \alpha_2)/F_f^\dagger(1-\alpha_2)$. Hence, we finally have

$$\frac{\mathrm{MV}^*(\alpha_2) - \mathrm{MV}^*(\alpha_1)}{\alpha_2 - \alpha_1} \leq \frac{1}{F_f^\dagger(1-\alpha_2)} \leq \frac{\mathrm{MV}^*(\alpha_3) - \mathrm{MV}^*(\alpha_2)}{\alpha_3 - \alpha_2}\,.$$

We now prove the differentiability of MV^* and the formula of its derivative. First as F_f is invertible on $(a, F_f^\dagger(1)]$, $F_f^\dagger$ is the ordinary inverse of F_f on $(0,1]$. Let $\alpha \in (0,1)$ and $h > 0$. Proceeding similarly as above we can show that

$$\frac{1}{F_f^\dagger(1-\alpha)} \leq \frac{\mathrm{MV}^*(\alpha+h) - \mathrm{MV}^*(\alpha)}{h} \leq \frac{1}{F_f^\dagger(1-(\alpha+h))}\,.$$

Therefore as $F_f^\dagger$ is continuous (because $F_f^\dagger$ is the ordinary inverse of F_f),

$$\lim_{h \to 0^+} \frac{\mathrm{MV}^*(\alpha+h) - \mathrm{MV}^*(\alpha)}{h} = \frac{1}{F_f^\dagger(1-\alpha)}\,.$$

Analogously, we also have

$$\lim_{h \to 0^+} \frac{\mathrm{MV}^*(\alpha-h) - \mathrm{MV}^*(\alpha)}{-h} = \frac{1}{F_f^\dagger(1-\alpha)}$$

which implies that $\mathrm{MV}'^*(\alpha) = 1/F_f^\dagger(1-\alpha)$. For $\alpha = 0$, we have $\mathrm{MV}^*(0) = 0$ and proceeding as for $\alpha \in (0,1)$, if $h > 0$,

$$\mathrm{MV}^*(h) = \int_{\mathcal{X}} \mathbb{I}\{x, F_f(f(x)) \geq 1-h\} \frac{1}{f(x)} f(x) dx$$

and

$$\frac{h}{F_f^\dagger(1)} \leq \mathrm{MV}^*(h) \leq \frac{h}{F_f^\dagger(1-\alpha)}\,.$$

Therefore $\lim_{h \to 0^+} \mathrm{MV}^*(h)/h = 1/F_f^\dagger(1)$. Thus, as $Q^*(\alpha) = F_f^\dagger(1-\alpha)$ for all $\alpha \in [0,1)$ we obtain the formula of the derivative of MV^*. □

A.2. Statistical estimation of the MV curve

A.2.1. Strong approximation: Proof of Theorem 1

To prove Theorem 1 we need the following lemma.

Lemma 2. *Under assumptions (A3)–(A4) and (A6) there exists a constant $C > 0$ such that we almost surely have for n large enough,*

$$\sup_{\alpha' \in [\varepsilon, 1]} |\widehat{F}_s^{\dagger}(\alpha') - F_s^{\dagger}(\alpha')|^2 \leq C \frac{\log n}{n} .$$

Proof of Lemma 2. Let $U_i = F_s(s(X_i))$ for all $i \in \{1, \ldots, n\}$. As F_s is continuous, the random variables $U_1, \ldots, U_n$ are *i.i.d.* with uniform distribution on $(0, 1)$. Their empirical cumulative distribution function is given by $\widehat{U}(\alpha) = 1/n \sum_{i=1}^{n} \mathbb{I}\{U_i \leq \alpha\}$ for all $\alpha \in [0, 1]$. Furthermore we have for all $\alpha \in (0, 1]$, $F_s(\widehat{F}_s^{\dagger}(\alpha)) = \widehat{U}^{\dagger}(\alpha)$. We also define $\widehat{U}^{\dagger}(0) = 0$. Let U_n denote the uniform empirical process defined for all $\alpha \in [0, 1]$ by $U_n(\alpha) = \sqrt{n}(\widehat{U}(\alpha) - \alpha)$ and u_n denote the uniform empirical quantile process defined for all $\alpha \in [0, 1]$ by $u_n(\alpha) = \sqrt{n}(\widehat{U}^{\dagger}(\alpha) - \alpha)$. Applying the Dvoretzky-Kiefer-Wolfowitz (DKW) inequality (Massart, 1990) gives for all $z > 0$ and all $n \geq 1$,

$$\mathbb{P}\Big(\sup_{\alpha \in [0,1]} |U_n(\alpha)| \geq z \Big) \leq 2 \exp(-2z^2) . \tag{18}$$

As $\sup_{0 \leq \alpha \leq 1} |U_n(\alpha)| = \sup_{0 \leq \alpha \leq 1} |u_n(\alpha)|$ for all ω in the sample space (see equation (1.4.5) in (Csörgő, 1983)), the DKW inequality (18) holds when replacing U_n by u_n. Taking $z = \sqrt{\log n}$, we obtain for all $n \geq 1$,

$$\mathbb{P}\Big(\sup_{\alpha \in [0,1]} \big|\widehat{U}^{\dagger}(\alpha) - \alpha\big|^2 \geq \frac{\log n}{n} \Big) \leq \frac{2}{n^2} . \tag{19}$$

As $\sum_{n=1}^{\infty} 1/n^2 < \infty$, by the Borel-Cantelli lemma, we almost surely have for n large enough,

$$\sup_{\alpha \in [0,1]} \big|\widehat{U}^{\dagger}(\alpha) - \alpha\big|^2 \leq \frac{\log n}{n} . \tag{20}$$

Using the mean value theorem we can write for $\alpha' \in [\varepsilon, 1]$,

$$\begin{aligned} |\widehat{F}_s^{\dagger}(\alpha') - F_s^{\dagger}(\alpha')| &= |F_s^{\dagger}(\widehat{U}^{\dagger}(\alpha'))) - F_s^{\dagger}(\alpha')| \\ &= \left| \frac{1}{f_s(F_s^{\dagger}(\xi))} \right| \cdot |\widehat{U}^{\dagger}(\alpha') - \alpha'| \end{aligned}$$

where ξ is between α' and $\widehat{U}^{\dagger}(\alpha')$. As $\widehat{U}^{\dagger}$ is increasing we have for all $\alpha' \in [\varepsilon, 1]$, $\widehat{U}^{\dagger}(\varepsilon) \leq \widehat{U}^{\dagger}(\alpha') \leq 1$. From (20) we know that there exists a constant $c_1 > 0$ such that we almost surely have for n large enough, $\widehat{U}^{\dagger}(\varepsilon) \geq \varepsilon - c_1 > 0$. Therefore, almost surely and for n large enough,

$$f_s(F_s^{\dagger}(\xi)) \geq \inf_{y' \in [\varepsilon - c_1, 1]} f_s(F_s^{\dagger}(y')) = \inf_{y \in [0, 1 - \varepsilon + c_1]} f_s(\alpha_s^{-1}(y)) > 0 .$$

Note that the last infimum is strictly positive because as $f_s \circ \alpha_s^{-1}$ is continuous, the infimum is reached at a point $y_0 \in [0, 1-\varepsilon+c_1] \subset [0,1)$ and as $f_s \circ \alpha_s^{-1} > 0$ on $[0,1)$ (assumptions **(A4)** and **(A6)**), $f_s(\alpha_s^{-1}(y_0)) > 0$.

This eventually leads to: almost surely and for n large enough,

$$\sup_{\alpha' \in [\varepsilon,1]} |\widehat{F}_s^\dagger(\alpha') - F_s^\dagger(\alpha')|^2 \leq \frac{1}{\left(\inf_{[0,1-\varepsilon+c_1]} f_s \circ \alpha_s^{-1}\right)^2} \frac{\log n}{n}\,. \qquad \square$$

Proof of assertion (i) **of Theorem 1.**

Proof. Assume that assumptions **(A3)**–**(A4)** and **(A6)**–**(A7)** are fulfilled. Using the mean value theorem we can write for all $\alpha \in [0, 1-\varepsilon]$

$$\widehat{\mathrm{MV}}_s(\alpha) - \mathrm{MV}_s(\alpha) = \lambda_s(\widehat{\alpha}_s^{-1}(\alpha)) - \lambda_s(\alpha_s^{-1}(\alpha)) = \lambda_s'(\xi)(\widehat{\alpha}_s^{-1}(\alpha) - \alpha_s^{-1}(\alpha))$$

where ξ is between $\widehat{\alpha}_s^{-1}(\alpha)$ and $\alpha_s^{-1}(\alpha)$. Therefore,

$$\sup_{\alpha \in [0,1-\varepsilon]} |\widehat{\mathrm{MV}}_s(\alpha) - \mathrm{MV}_s(\alpha)| = \sup_{\alpha \in [0,1-\varepsilon]} |\lambda_s'(\xi)| \cdot |\widehat{\alpha}_s^{-1}(\alpha) - \alpha_s^{-1}(\alpha)|\,.$$

Now for all $\alpha \in [0, 1-\varepsilon]$, $\alpha_s^{-1}(\alpha)$ and $\widehat{\alpha}_s^{-1}(\alpha)$ belong to the support of f_s, *i.e.*, $[0, \|s\|_\infty]$. This is true for $\widehat{\alpha}_s^{-1}(\alpha)$ because it is equal to one of the $s(X_i)$, $1 \leq i \leq n$. Thus $\xi \in [0, \|s\|_\infty]$. As λ_s' is continuous, we have

$$\sup_{\alpha \in [0,1-\varepsilon]} |\lambda_s'(\xi)| \leq \sup_{[0,\|s\|_\infty]} |\lambda_s'| < +\infty$$

and

$$\begin{aligned}\sup_{\alpha \in [0,1-\varepsilon]} |\widehat{\mathrm{MV}}_s(\alpha) - \mathrm{MV}_s(\alpha)| &\leq \sup_{[0,\|s\|_\infty]} |\lambda_s'| \cdot \sup_{\alpha \in [0,1-\varepsilon]} |\widehat{\alpha}_s^{-1}(\alpha) - \alpha_s^{-1}(\alpha)| \\ &= \sup_{[0,\|s\|_\infty]} |\lambda_s'| \cdot \sup_{\alpha' \in [\varepsilon,1]} |\widehat{F}_s^\dagger(\alpha') - F_s^\dagger(\alpha')|\,.\end{aligned}$$

From Lemma 2 we know that there exists a constant $C > 0$ such that we almost surely have for n large enough,

$$\sup_{\alpha' \in [\varepsilon,1]} |\widehat{F}_s^\dagger(\alpha') - F_s^\dagger(\alpha')| \leq \sqrt{C \frac{\log n}{n}}\,.$$

Therefore $\sup_{\alpha' \in [\varepsilon,1]} |\widehat{F}_s^\dagger(\alpha') - F_s^\dagger(\alpha')|$ converges almost surely to 0 and so is $\sup_{\alpha \in [0,1-\varepsilon]} |\widehat{\mathrm{MV}}_s(\alpha) - \mathrm{MV}_s(\alpha)|$. $\square$

Proof of assertion (ii) **of Theorem 1.**

Proof. For all $n \geq 1$, let $v_n = n^{-1/2} \log n$. Assume that assumptions **(A3)**–**(A7)** are fulfilled. By virtue of Theorem 3.1.2 in (Csörgő, 1983), there exists a sequence of Brownian bridges $\{B_n^1(\alpha), \alpha \in [0,1]\}_{n \geq 1}$ such that, we almost surely have:

$$\sup_{0 \leq \alpha' \leq 1} |u_n(\alpha') - B_n^1(\alpha')| = O(v_n) \tag{21}$$

where u_n is the uniform quantile process as defined in the proof of Lemma 2. From (3.3) of Theorem 3 in (Csörgő and Révész, 1978), we also have almost surely, for n large enough,

$$\sup_{a_n \leq \alpha' \leq 1-a_n} |f_s\big(F_s^\dagger(\alpha')\big)\sqrt{n}\big(\widehat{F}_s^\dagger(\alpha') - F_s^\dagger(\alpha')\big) - u_n(\alpha')| \leq C_1 n^{-1/2}\log\log n$$

where $C_1 = 40c10^c$ and $a_n = 25n^{-1}\log\log n$. Hence for n large enough we have $a_n \leq \varepsilon$ and therefore we almost surely have

$$\sup_{\varepsilon \leq \alpha' \leq 1-a_n} |f_s\big(F_s^\dagger(\alpha')\big)\sqrt{n}\big(\widehat{F}_s^\dagger(\alpha') - F_s^\dagger(\alpha')\big) - u_n(\alpha')| = O(n^{-1/2}\log\log n) \quad (22)$$

Now from the proof of theorem 3.2.1 in (Csörgő, 1983), we also almost surely have,

$$\sup_{1-a_n \leq \alpha' \leq 1} |f_s\big(F_s^\dagger(\alpha')\big)\sqrt{n}\big(\widehat{F}_s^\dagger(\alpha') - F_s^\dagger(\alpha')\big) - u_n(\alpha')| = O(n^{-1/2}\log\log n)\,. \quad (23)$$

Combining (21), (22) and (23), we almost surely have:

$$\sup_{\varepsilon \leq \alpha' \leq 1} |f_s\big(F_s^\dagger(\alpha')\big)\sqrt{n}\big(\widehat{F}_s^\dagger(\alpha') - F_s^\dagger(\alpha')\big) - B_n^1(\alpha')| = O(v_n)\,.$$

As one can show that $F_s^\dagger(\alpha) = \alpha_s^{-1}(1-\alpha)$ and $\widehat{F}_s^\dagger(\alpha) = \widehat{\alpha}_s^{-1}(1-\alpha)$ for all $\alpha \in (0,1]$ we almost surely have

$$\sup_{\varepsilon \leq \alpha' \leq 1} |f_s\big(\alpha_s^{-1}(1-\alpha')\big)\sqrt{n}\big(\widehat{\alpha}_s^{-1}(1-\alpha') - \alpha_s^{-1}(1-\alpha')\big) - B_n^1(\alpha')| = O(v_n)\,.$$

With the change of variable $\alpha = 1-\alpha'$ this leads to

$$\sup_{0 \leq \alpha \leq 1-\varepsilon} |f_s\big(\alpha_s^{-1}(\alpha)\big)\sqrt{n}\big(\widehat{\alpha}_s^{-1}(\alpha) - \alpha_s^{-1}(\alpha)\big) - B_n(\alpha)| = O(v_n)$$

where $\{B_n(\alpha), \alpha \in [0,1]\}_{n\geq 1}$ is the sequence of stochastic processes defined by $\{B_n(\alpha), \alpha \in [0,1]\}_{n\geq 1} = \{B_n^1(1-\alpha), \alpha \in [0,1]\}_{n\geq 1}$. Notice that for all $n \geq 1$, B_n is a Brownian Bridge as it has the same distribution as B_n^1. Thus there exists a constant C_2 independent of α such that almost surely for n large enough and for all $\alpha \in [0, 1-\varepsilon]$,

$$|f_s\big(\alpha_s^{-1}(\alpha)\big)\sqrt{n}\big(\widehat{\alpha}_s^{-1}(\alpha) - \alpha_s^{-1}(\alpha)\big) - B_n(\alpha)| \leq C_2 v_n$$

Hence, dividing by $f_s\big(\alpha_s^{-1}(\alpha)\big)$ which is strictly positive for all $\alpha \in [0, 1-\varepsilon]$ (assumptions **(A4)** and **(A6)**), we almost surely have, for n large enough and for all $\alpha \in [0, 1-\varepsilon]$,

$$\left|\sqrt{n}\big(\widehat{\alpha}_s^{-1}(\alpha) - \alpha_s^{-1}(\alpha)\big) - \frac{B_n(\alpha)}{f_s\big(\alpha_s^{-1}(\alpha)\big)}\right| \leq \frac{C_2}{\inf_{[0,1-\varepsilon]} f_s \circ \alpha_s^{-1}} v_n \quad (24)$$

where $\inf_{0\leq\alpha\leq 1-\varepsilon} f_s(\alpha_s^{-1}(\alpha))$ is strictly positive by the same argument as the one given in the proof of Lemma 2. Now using a Taylor expansion of λ_s which is of class $\mathcal{C}^2$ (assumption **(A7)**) we can write for all $\alpha \in [0, 1-\varepsilon]$:

$$\lambda_s(\widehat{\alpha}_s^{-1}(\alpha)) = \lambda_s(\alpha_s^{-1}(\alpha)) + \lambda_s'(\alpha_s^{-1}(\alpha))(\widehat{\alpha}_s^{-1}(\alpha) - \alpha_s^{-1}(\alpha)) + R_2(\alpha)$$

where $R_2(\alpha) = \lambda_s''(\xi)(\widehat{\alpha}_s^{-1}(\alpha) - \alpha_s^{-1}(\alpha))^2/2$ is the remainder of the Taylor expansion, ξ being between $\alpha_s^{-1}(\alpha)$ and $\widehat{\alpha}_s^{-1}(\alpha)$. We thus have

$$\begin{aligned}
&\sqrt{n}(\widehat{\mathrm{MV}}_s(\alpha) - \mathrm{MV}_s(\alpha)) - \frac{\lambda_s'(\alpha_s^{-1}(\alpha))}{f_s(\alpha_s^{-1}(\alpha))} B_n(\alpha) \\
&\quad = \lambda_s'(\alpha_s^{-1}(\alpha)) \left(\sqrt{n}(\widehat{\alpha}_s^{-1}(\alpha) - \alpha_s^{-1}(\alpha)) - \frac{B_n(\alpha)}{f_s(\alpha_s^{-1}(\alpha))} \right) + \sqrt{n} R_2(\alpha) \\
&\quad = C_n(\alpha) + D_n(\alpha)
\end{aligned}$$

For the first term we almost surely have, for all $\alpha \in [0, 1-\varepsilon]$ and for n large enough,

$$|C_n(\alpha)| \leq |\lambda_s'(\alpha_s^{-1}(\alpha))| \frac{C_2}{\inf_{[0,1-\varepsilon]} f_s \circ \alpha_s^{-1}} v_n \,.$$

For all $\alpha \in [0, 1-\varepsilon]$, $\alpha_s^{-1}(\alpha) \in [0, \|s\|_\infty]$. As λ_s' is continuous, $|\lambda_s'(\alpha_s^{-1}(\alpha))|$ is bounded by $\sup_{[0,\|s\|_\infty]} |\lambda_s'|$.

We treat the second term as follows. For all $\alpha \in [0, 1-\varepsilon]$, $\alpha_s^{-1}(\alpha)$ and $\widehat{\alpha}_s^{-1}(\alpha)$ are in $[0, \|s\|_\infty]$. Therefore $\xi \in [0, \|s\|_\infty]$ and as λ_s'' is continuous, $|\lambda_s''(\xi)|$ is bounded by $\sup_{[0,\|s\|_\infty]} |\lambda_s''|$. Thus,

$$\begin{aligned}
\sup_{\alpha\in[0,1-\varepsilon]} |D_n(\alpha)| &\leq \sup_{[0,\|s\|_\infty]} |\lambda_s''| \frac{\sqrt{n}}{2} \sup_{\alpha\in[0,1-\varepsilon]} |\widehat{\alpha}_s^{-1}(\alpha) - \alpha_s^{-1}(\alpha)|^2 \\
&\leq \sup_{[0,\|s\|_\infty]} |\lambda_s''| \frac{C}{2} \frac{\log n}{\sqrt{n}} \quad \text{(thanks to Lemma 2)}
\end{aligned}$$

where the last inequality holds almost surely and for n large enough and where $C > 0$ is the constant of Lemma 2.

Eventually, combining the bounds on $\sup_{[0,1-\varepsilon]} |C_n|$ and $\sup_{[0,1-\varepsilon]} |D_n|$ we almost surely have

$$\sup_{\alpha\in[0,1-\varepsilon]} \left| \sqrt{n}(\widehat{\mathrm{MV}}_s(\alpha) - \mathrm{MV}_s(\alpha)) - \frac{\lambda_s'(\alpha_s^{-1}(\alpha))}{f_s(\alpha_s^{-1}(\alpha))} B_n(\alpha) \right| = O(v_n) \,.$$

This concludes the proof. □

A.2.2. Bootstrap consistency: Proof of Theorem 2

Recall that for all $n \geq 1$, $v_n = n^{-1/2} \log n$ and $w_n = \sqrt{\log(h_n^{-1})/(nh_n)} + h_n^2$. Let $y_s(\alpha)$ and $\widetilde{y}_s(\alpha)$ be respectively defined for all $\alpha \in [\varepsilon, 1-\varepsilon]$ as $y_s(\alpha) = \lambda_s'(\alpha_s^{-1}(\alpha))/f_s(\alpha_s^{-1}(\alpha))$ and $\widetilde{y}_s(\alpha) = \lambda_s'(\widetilde{\alpha}_s^{-1}(\alpha))/\widetilde{f}_s(\widetilde{\alpha}_s^{-1}(\alpha))$.

Sketch of proof The argument is based on a strong approximation type inequality for r_n (see Lemma 3) and r_n^* (see Lemma 4). These inequalities can then be used to obtain a result on the rate of convergence of the cumulative distribution function of $\sup_{[\varepsilon,1-\varepsilon]} |r_n|$ (respectively, the conditional cumulative distribution function of $\sup_{[\varepsilon,1-\varepsilon]} |r_n^*|$) towards the supremum of a Gaussian process which depends on y_s (respectively the smoothed empirical version $\widetilde{y}_s$ of y_s). We finally obtain the result using the rate of the strong uniform convergence of $\widetilde{y}_s/y_s$ towards 1 on $[\varepsilon, 1-\varepsilon]$ (see Lemma 5). The technical proofs of Lemma 3, Lemma 4 and Lemma 5 are deferred to Appendix B for the sake of clarity of the proof of the main result.

We first need the following strong approximation result.

Lemma 3. *Let $\varepsilon \in (0,1)$. There exists a positive constant C such that*

$$\mathbb{P}\left(\sup_{\alpha\in[\varepsilon,1-\varepsilon]} |r_n(\alpha) - Z_n(\alpha)| > Cv_n \right) = O(v_n)$$

We also need the counterpart of Lemma 3 conditionally on the data set $\mathcal{D}_n$.

Lemma 4. *Let $\varepsilon \in (0,1)$. Under assumptions **(A3)**–**(A5)**, **(A7)** and **(B1)**–**(B6)**, there exists a constant C such that we $\mathbb{P}$-almost surely have,*

$$\mathbb{P}^*\left(\sup_{\alpha\in[\varepsilon,1-\varepsilon]} |r_n^*(\alpha) - Z_n^*(\alpha)| > Cv_n \right) = O(v_n)$$

where

$$\forall \alpha \in [\varepsilon, 1-\varepsilon], \quad Z_n^*(\alpha) = \frac{\lambda_s'(\widetilde{\alpha}_s^{-1}(\alpha))}{\widetilde{f}_s(\widetilde{\alpha}_s^{-1}(\alpha))} B_n^*(\alpha)$$

and where for each $n \geq 1$ the distribution of B_n^* conditionally to $\mathcal{D}_n$ is the one of a Brownian bridge.

Eventually, to control the distance between the distribution of $\sup_{[\varepsilon,1-\varepsilon]} |Z_n|$ and the conditional distribution of $\sup_{[\varepsilon,1-\varepsilon]} |Z_n^*|$ we need the following lemma.

Lemma 5. *Under assumptions **(A3)** and **(B1)**–**(B4)** there exists a constant $C > 0$ such that we almost surely have, for n large enough,*

$$\sup_{\alpha\in[\varepsilon,1-\varepsilon]} \left| \frac{\widetilde{y}_s(\alpha)}{y_s(\alpha)} - 1 \right| \leq Cw_n \,.$$

We can now proof the main result.

Proof of Theorem 2. Let $\varepsilon \in (0,1)$. Let $C > 0$ be the constant of Lemma 3. If $\sup_{\alpha\in[\varepsilon,1-\varepsilon]} |r_n(\alpha) - Z_n(\alpha)| \leq Cv_n$ then

$$\left| \sup_{\alpha\in[\varepsilon,1-\varepsilon]} |r_n(\alpha)| - \sup_{\alpha\in[0,1-\varepsilon]} |Z_n(\alpha)| \right| \leq Cv_n \,.$$

Thus thanks to the result of Lemma 3

$$\mathbb{P}\left(\left|\sup_{[\varepsilon,1-\varepsilon]}|r_n| - \sup_{[\varepsilon,1-\varepsilon]}|Z_n|\right| > Cv_n\right) \leq \mathbb{P}\left(\sup_{[\varepsilon,1-\varepsilon]}|r_n - Z_n| > Cv_n\right) = O(v_n)\,.$$

Therefore using a result of Sargan and Mikhail (1971), we have for all $t \in \mathbb{R}$,

$$\begin{aligned}
&\left|\mathbb{P}\Big(\sup_{[\varepsilon,1-\varepsilon]}|r_n| \leq t\Big) - \mathbb{P}\Big(\sup_{[\varepsilon,1-\varepsilon]}|Z_n| \leq t\Big)\right| \\
&\leq \mathbb{P}\Big(\Big|\sup_{[\varepsilon,1-\varepsilon]}|r_n| - \sup_{[\varepsilon,1-\varepsilon]}|Z_n|\Big| > Cv_n\Big) + \mathbb{P}\Big(\Big|\sup_{[\varepsilon,1-\varepsilon]}|Z_n| - t\Big| < Cv_n\Big) \\
&= \mathbb{P}\Big(\Big|\sup_{[\varepsilon,1-\varepsilon]}|Z_n| - t\Big| < Cv_n\Big) + O(v_n)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}\Big(\Big|\sup_{[\varepsilon,1-\varepsilon]}|Z_n| - t\Big| < Cv_n\Big) &= \mathbb{P}\Big(t - Cv_n < \sup_{[\varepsilon,1-\varepsilon]}|Z_1| < t + Cv_n\Big) \\
&= \int_{t-Cv_n}^{t+Cv_n} \phi_{\bar{Z}}(x)dx \leq 2DCv_n
\end{aligned}$$

where $\phi_{\bar{Z}}$ denotes the density of $\sup_{[\varepsilon,1-\varepsilon]}|Z_1|$ and where for the first equality we use the fact that for all $n \geq 1$ the random variable $Z_n(\alpha)$ is equal in distribution to $Z_1(\alpha)$ and for the last inequality the result from (Pitt and Tran, 1979) stating that $\sup_{\alpha\in[\varepsilon,1-\varepsilon]}|Z_1(\alpha)|$ has a density bounded by a constant $D > 0$ as the supremum of a Gaussian process $\{Z_1(\alpha), \varepsilon \leq \alpha \leq 1-\varepsilon\}$ such that for all $\alpha \in [\varepsilon, 1-\varepsilon]$, $\inf_{\alpha\in[\varepsilon,1-\varepsilon]}\mathrm{Var}[Z_1(\alpha)] > 0$ and $\mathbb{P}(\sup_{\alpha\in[\varepsilon,1-\varepsilon]}|Z_1(\alpha)| < \infty) = 1$. Indeed, as B_1 is a Brownian bridge it is almost surely continuous on $[0,1]$ and as y_s is continuous on $[\varepsilon, 1-\varepsilon]$, Z_1 is almost surely continuous on $[\varepsilon, 1-\varepsilon]$ and therefore almost surely $\sup_{\alpha\in[\varepsilon,1-\varepsilon]}|Z_1(\alpha)| < \infty$. Furthermore, $\alpha \in [\varepsilon, 1-\varepsilon] \mapsto \mathrm{Var}[Z_1(\alpha)] = \alpha(1-\alpha)y_s^2(\alpha)$ is continuous and thanks to Lemma 10, for all $\alpha \in [\varepsilon, 1-\varepsilon]$, $\mathrm{Var}[Z_1(\alpha)] > 0$. Therefore $\inf_{\alpha\in[\varepsilon,1-\varepsilon]}\mathrm{Var}[Z_1(\alpha)]$ is attained and is strictly positive.

Therefore,

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\Big(\sup_{\alpha\in[\varepsilon,1-\varepsilon]}|r_n(\alpha)| \leq t\Big) - \mathbb{P}\Big(\sup_{\alpha\in[\varepsilon,1-\varepsilon]}|Z_1(\alpha)| \leq t\Big)\right| = O(v_n) \tag{25}$$

Reasoning similarly as above, thanks to the result of Lemma 4, there exists a constant C_1 such that, $\mathbb{P}$-almost surely, as $n \to \infty$,

$$\mathbb{P}^*\left(\left|\sup_{\alpha\in[\varepsilon,1-\varepsilon]}|r_n^*(\alpha)| - \sup_{\alpha\in[0,1-\varepsilon]}|Z_n^*(\alpha)|\right| > C_1v_n\right) = O(v_n)\,.$$

Therefore using the result of (Sargan and Mikhail, 1971), we $\mathbb{P}$-almost surely have, for all $t \in \mathbb{R}$,

$$\left|\mathbb{P}^*\Big(\sup_{[\varepsilon,1-\varepsilon]}|r_n^*| \leq t\Big) - \mathbb{P}^*\Big(\sup_{[\varepsilon,1-\varepsilon]}|Z_n^*| \leq t\Big)\right|$$

$$
\begin{aligned}
&= \mathbb{P}^*\Big(t - C_1 v_n < \sup_{[\varepsilon,1-\varepsilon]} |Z_1^*| < t + C_1 v_n\Big) + O(v_n) \\
&= \mathbb{P}^*\Big(\sup_{[\varepsilon,1-\varepsilon]} |Z_1^*| < t + C_1 v_n\Big) - \mathbb{P}\Big(\sup_{[\varepsilon,1-\varepsilon]} |Z_1| < t + C_1 v_n\Big) \\
&\quad + \mathbb{P}\Big(\sup_{[\varepsilon,1-\varepsilon]} |Z_1| < t + C_1 v_n\Big) - \mathbb{P}\Big(\sup_{[\varepsilon,1-\varepsilon]} |Z_1| < t - C_1 v_n\Big) \\
&\quad + \mathbb{P}\Big(\sup_{[\varepsilon,1-\varepsilon]} |Z_1| < t - C_1 v_n\Big) - \mathbb{P}^*\Big(\sup_{[\varepsilon,1-\varepsilon]} |Z_1^*| < t - C_1 v_n\Big) + O(v_n) \\
&\leq 2 \sup_{t\in\mathbb{R}} \Big|\mathbb{P}^*\Big(\sup_{[\varepsilon,1-\varepsilon]} |Z_1^*| \leq t\Big) - \mathbb{P}\Big(\sup_{[\varepsilon,1-\varepsilon]} |Z_1| \leq t\Big)\Big| \\
&\quad + \mathbb{P}\Big(t - C_1 v_n < \sup_{[\varepsilon,1-\varepsilon]} |Z_1| < t + C_1 v_n\Big) + O(v_n) \\
&\leq 2 \sup_{t\in\mathbb{R}} \Big|\mathbb{P}^*\Big(\sup_{[\varepsilon,1-\varepsilon]} |Z_1^*| \leq t\Big) - \mathbb{P}\Big(\sup_{[\varepsilon,1-\varepsilon]} |Z_1| \leq t\Big)\Big| + 2DC_1 v_n + O(v_n)\,.
\end{aligned}
$$

We now need the following result.

Lemma 6. *The exists a constant $C_2 > 0$ such that we $\mathbb{P}$-almost surely have for n large enough,*

$$
\sup_{t\in\mathbb{R}} \Big|\mathbb{P}^*\Big(\sup_{\alpha\in[\varepsilon,1-\varepsilon]} |Z_1^*(\alpha)| \leq t\Big) - \mathbb{P}\Big(\sup_{\alpha\in[\varepsilon,1-\varepsilon]} |Z_1(\alpha)| \leq t\Big)\Big| \leq C_2 w_n
$$

Proof of Lemma 6. Let $t \in \mathbb{R}$. If $t < 0$ both probabilities are equal to 0 as the two random variables involved are positive. Hence we have (for all ω in the sample space)

$$
\sup_{t<0} \Big|\mathbb{P}^*\Big(\sup_{\alpha\in[\varepsilon,1-\varepsilon]} |Z_1^*(\alpha)| \leq t\Big) - \mathbb{P}\Big(\sup_{\alpha\in[\varepsilon,1-\varepsilon]} |Z_1(\alpha)| \leq t\Big)\Big| = 0 \leq C_2 w_n \tag{26}
$$

Let $t \geq 0$. Thanks to Lemma 5 there exists a constant $C_3 > 0$ independent of α such that we almost surely have for n large enough:

$$
\forall \alpha \in [\varepsilon, 1-\varepsilon], 1 - C_3 w_n \leq \left|\frac{\widetilde{y}_s(\alpha)}{y_s(\alpha)}\right| \leq 1 + C_3 w_n\,. \tag{27}
$$

Furthermore, we can assume, given n large enough, that there exists a constant $C_4 > 0$ such that $1 - C_3 w_n > C_4$ as w_n tends towards 0. Observe that

$$
\begin{aligned}
\mathbb{P}^*\Big(\sup_{\alpha\in[\varepsilon,1-\varepsilon]} |Z_1^*(\alpha)| \leq t\Big) &= \mathbb{P}^*\Big(\sup_{\alpha\in[\varepsilon,1-\varepsilon]} \left|\frac{\widetilde{y}_s(\alpha)}{y_s(\alpha)} y_s(\alpha) B_1^*(\alpha)\right| \leq t\Big) \\
&= \mathbb{P}^*\left(\forall \alpha \in [\varepsilon, 1-\varepsilon], \left|\frac{\widetilde{y}_s(\alpha)}{y_s(\alpha)}\right| |y_s(\alpha) B_1^*(\alpha)| \leq t\right)\,.
\end{aligned}
$$

and decompose the last term as follows

$$
\begin{aligned}
&\mathbb{P}^*\left(\left\{\forall \alpha \in [\varepsilon, 1-\varepsilon], \left|\frac{\widetilde{y}_s(\alpha)}{y_s(\alpha)}\right| |y_s(\alpha) B_1^*(\alpha)| \leq t\right\}, \mathcal{Z}\right) \\
&\qquad + \mathbb{P}^*\left(\left\{\forall \alpha \in [\varepsilon, 1-\varepsilon], \left|\frac{\widetilde{y}_s(\alpha)}{y_s(\alpha)}\right| |y_s(\alpha) B_1^*(\alpha)| \leq t\right\}, \overline{\mathcal{Z}}\right)\,.
\end{aligned} \tag{28}
$$

where the event $\mathcal{Z}$ is defined as

$$\mathcal{Z} = \left\{ \forall \alpha \in [\varepsilon, 1-\varepsilon], \left| \frac{\widetilde{y}_s(\alpha)}{y_s(\alpha)} \right| \geq 1 - C_3 w_n \right\}$$

and its complementary is given by

$$\overline{\mathcal{Z}} = \left\{ \exists \alpha \in [\varepsilon, 1-\varepsilon], \left| \frac{\widetilde{y}_s(\alpha)}{y_s(\alpha)} \right| < 1 - C_3 w_n \right\} .$$

The first term of (28) is lower than

$$\mathbb{P}^* \left((1 - C_3 w_n) \sup_{[\varepsilon, 1-\varepsilon]} |y_s B_1^*| \leq t \right) = \mathbb{P} \left(\sup_{[\varepsilon, 1-\varepsilon]} |y_s B_1| \leq \frac{t}{1 - C_3 w_n} \right)$$

where the last equality stands from the fact that the law of B_1^* conditionally on $\mathcal{D}_n$ is equal to the law of B_1. The second term of (28) is lower than

$$\mathbb{P}^* \left(\exists \alpha \in [\varepsilon, 1-\varepsilon], |\widetilde{y}_s(\alpha)/y_s(\alpha)| < 1 - C_3 w_n \right)$$

which is $\mathbb{P}$-almost surely equal to 0 thanks to (27). Note that the fact that this holds $\mathbb{P}$-almost surely is independent of t. Therefore we $\mathbb{P}$-almost surely have: for all $t \geq 0$,

$$\begin{aligned} \mathbb{P}^* \Big(\sup_{\alpha \in [\varepsilon, 1-\varepsilon]} |Z_1^*(\alpha)| \leq t \Big) &\leq \mathbb{P} \left(\sup_{\alpha \in [\varepsilon, 1-\varepsilon]} |y_s(\alpha) B_1(\alpha)| \leq \frac{t}{1 - C_3 w_n} \right) \\ &= \mathbb{P} \left(\sup_{\alpha \in [\varepsilon, 1-\varepsilon]} |Z_1(\alpha)| \leq t \right) + \phi_{\bar{Z}}(\xi) t \frac{C_3 w_n}{1 - C_3 w_n} \end{aligned}$$

where we use a Taylor expansion and where ξ is between t and $t/(1 - C_3 w_n)$.

For n large enough, we have $1 > 1 - C_3 w_n > C_4 > 0$. On the one hand this gives $1 < 1/(1 - C_3 w_n)$ and, as $t \geq 0$, $t < t/(1 - C_3 w_n)$. Therefore, as ξ is between t and $t/(1 - C_3 w_n)$, we have in fact, $t < \xi < t/(1 - C_3 w_n)$. Thus $\phi_{\bar{Z}}(\xi) t \leq \phi_{\bar{Z}}(\xi) \xi$. On the other hand, we also have $1/(1 - C_3 w_n) < 1/C_4$. This eventually gives $\mathbb{P}$-almost surely, for all $t \geq 0$,

$$\mathbb{P}^* \left(\sup_{[\varepsilon, 1-\varepsilon]} |Z_1^*| \leq t \right) \leq \mathbb{P} \left(\sup_{[\varepsilon, 1-\varepsilon]} |Z_1| \leq t \right) + \sup_{x \in \mathbb{R}^+} |\phi_{\bar{Z}}(x) x| \frac{C_3 w_n}{C_4} . \tag{29}$$

A lower bound can be obtained in a similar fashion: $\mathbb{P}$-almost surely, for all $t \geq 0$,

$$\mathbb{P} \left(\sup_{[\varepsilon, 1-\varepsilon]} |Z_1| \leq t \right) \leq \mathbb{P}^* \left(\sup_{[\varepsilon, 1-\varepsilon]} |Z_1^*| \leq t \right) + 2 \sup_{x \in \mathbb{R}^+} |x \phi_{\bar{Z}}(x)| C_3 w_n . \tag{30}$$

Combining the two inequalities (29) and (30) we obtain that there exists a constant $C_5 > 0$ such that we $\mathbb{P}$-almost surely have

$$\sup_{t \geq 0} \left| \mathbb{P}^* \left(\sup_{[\varepsilon, 1-\varepsilon]} |Z_1^*| \leq t \right) - \mathbb{P} \left(\sup_{[\varepsilon, 1-\varepsilon]} |Z_1| \leq t \right) \right| \leq C_5 \sup_{x \in \mathbb{R}^+} |x \phi_{\bar{Z}}(x)| w_n .$$

Let φ be the density of the Gaussian distribution $\mathcal{N}(0,1)$. Using a remark in the paper of (Tsirel'son, 1976, p. 854), there exists $M > 0$ such that for all $x \geq M$,

$$x\phi_{\bar{Z}}(x) \leq \frac{x}{v_1}\varphi\left(\frac{x-a-b}{v_1}\right), \tag{31}$$

where v_1 is such that $v_1^2 = \sup_{\alpha\in[\varepsilon,1-\varepsilon]} \mathrm{Var}[|Z_1(\alpha)|]$, $a > 0$ is arbitrary and $b \geq 0$ only depends on v_1, $\phi_{\bar{Z}}$ and the cumulative distribution function of the Gaussian distribution $\mathcal{N}(0,1)$. Now the function $x \in \mathbb{R}^+ \mapsto x/v_1 \cdot \varphi((x-a-b)/v_1)$ is bounded on $\mathbb{R}^+$ as it is continuous and tends towards 0 when x tends to $+\infty$. Therefore from (31) we have that $\sup_{x\geq M} |x\phi_{\bar{Z}}(x)| < +\infty$. And as from (Pitt and Tran, 1979) $\phi_{\bar{Z}}$ is bounded on $\mathbb{R}$, so is $x \mapsto x\phi_{\bar{Z}}(x)$ on $[0,M]$. Eventually, $\sup_{x\in\mathbb{R}^+} |x\phi_{\bar{Z}}(x)| < +\infty$ and there exists a constant $C_2 > 0$ such that $\mathbb{P}$-almost surely,

$$\sup_{t\geq 0}\left|\mathbb{P}^*\left(\sup_{\alpha\in[\varepsilon,1-\varepsilon]}|Z_1^*(\alpha)| \leq t\right) - \mathbb{P}\left(\sup_{\alpha\in[\varepsilon,1-\varepsilon]}|Z_1(\alpha)| \leq t\right)\right| \leq C_2 w_n \tag{32}$$

Combining (26) and (32) we obtain the result of the Lemma. □

From Lemma 6, we $\mathbb{P}$-almost surely have

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}^*\Big(\sup_{\alpha\in[\varepsilon,1-\varepsilon]}|r_n^*(\alpha)| \leq t\Big) - \mathbb{P}^*\Big(\sup_{\alpha\in[\varepsilon,1-\varepsilon]}|Z_n^*(\alpha)| \leq t\Big)\right| = O(w_n)$$

and thus

$$\begin{aligned}
&\sup_{t\in\mathbb{R}}\left|\mathbb{P}^*\Big(\sup_{\alpha\in[\varepsilon,1-\varepsilon]}|r_n^*(\alpha)| \leq t\Big) - \mathbb{P}\Big(\sup_{\alpha\in[\varepsilon,1-\varepsilon]}|Z_1(\alpha)| \leq t\Big)\right| \\
&\quad\leq \sup_{t\in\mathbb{R}}\left|\mathbb{P}^*\Big(\sup_{\alpha\in[\varepsilon,1-\varepsilon]}|r_n^*(\alpha)| \leq t\Big) - \mathbb{P}^*\Big(\sup_{\alpha\in[\varepsilon,1-\varepsilon]}|Z_n^*(\alpha)| \leq t\Big)\right| \\
&\qquad+ \sup_{t\in\mathbb{R}}\left|\mathbb{P}^*\Big(\sup_{\alpha\in[\varepsilon,1-\varepsilon]}|Z_n^*(\alpha)| \leq t\Big) - \mathbb{P}\Big(\sup_{\alpha\in[\varepsilon,1-\varepsilon]}|Z_1(\alpha)| \leq t\Big)\right| \\
&\quad= O(w_n)
\end{aligned}$$

Finally combining this last inequality with (25) we obtain the result of the theorem. □

A.3. A-Rank algorithm

A.3.1. Proof of Theorem 3

The proof follows closely the one of theorem 2 in (Clémençon and Vayatis, 2009). For clarity, we start off with recalling the following result.

Proposition 4. *(Scott and Nowak (2006)) Suppose that the assumptions of Theorem 3 are fulfilled. For any $\alpha \in (0,1)$, let $\widehat{\Omega}_\alpha \in \mathcal{G}$ be a solution of the constrained optimization problem:*

$$\min_{\Omega \in \mathcal{G}} \lambda(\Omega) \text{ subject to } \widehat{F}(\Omega) \geq \alpha - \phi_n(\delta).$$

Let $\delta \in (0,1)$. With probability at least $1-\delta$: $\forall \alpha \in (0,1)$, $\forall n \geq 1$,

$$\alpha - 2\phi_n(\delta) \leq F(\widehat{\Omega}_\alpha) \leq \alpha \text{ and } \mathrm{MV}^*(\alpha) - \frac{2\phi_n(\delta)}{Q^*(\alpha)} \leq \lambda(\widehat{\Omega}_\alpha) \leq \mathrm{MV}^*(\alpha). \quad (33)$$

The major part of this result is given by Theorem 3 and Lemma 19 in (Scott and Nowak, 2006). Although they prove the result for a given α, it can be extended to all $\alpha \in (0,1)$. The upper bound on $F(\widehat{\Omega}_\alpha)$ can be obtained with the following reasoning. Suppose that we have $F(\widehat{\Omega}_\alpha) > \alpha$, then $F(\widehat{\Omega}_\alpha) \geq \alpha$ and $\lambda(\Omega^*_\alpha) \leq \lambda(\widehat{\Omega}_\alpha)$ because Ω^*_α minimizes λ over all measurable sets Ω such that $F(\Omega) \geq \alpha$. The upper bound on the volume gives $\lambda(\widehat{\Omega}_\alpha) \leq \lambda(\Omega^*_\alpha)$, thus $\lambda(\widehat{\Omega}_\alpha) = \lambda(\Omega^*_\alpha)$. Hence $\widehat{\Omega}$ also minimizes λ over all measurable sets Ω such that $F(\Omega) \geq \alpha$ and by uniqueness of the solution under assumptions **(A1)** and **(A2)**, we have $F(\widehat{\Omega}_\alpha) = F(\Omega^*_\alpha) = \alpha$ which contradicts $F(\widehat{\Omega}_\alpha) > \alpha$.

We now prove the following lemma quantifying the uniform deviation of the empirical local error from the true local error over all dyadic scales.

Lemma 7. *Suppose that assumptions **(C1)** and **(C2)** are satisfied. Let $\delta \in (0,1)$. With probability at least $1-\delta$, we have: $\forall n \geq 1$,*

$$\sup_{\substack{0 \leq j \leq j_{max} \\ 0 \leq k \leq 2^j}} |\widehat{\mathcal{E}}(I_{j,k}) - \mathcal{E}(I_{j,k})| \leq \frac{4\phi_n(\delta)}{Q^*(1-\varepsilon)}.$$

Proof. Let $0 \leq j \leq j_{\max}$, $0 \leq k \leq 2^j - 1$,

$$\begin{aligned} |\widehat{\mathcal{E}}(I_{j,k}) - \mathcal{E}(I_{j,k})| &\leq |\lambda(\widehat{\Omega}_{\alpha_{j,k+1}}) - \lambda(\Omega^*_{\alpha_{j,k+1}})| + |\lambda(\widehat{\Omega}_{\alpha_{j,k}}) - \lambda(\Omega^*_{\alpha_{j,k}})| \\ &\leq 2 \sup_{\substack{0 \leq j \leq j_{\max} \\ 0 \leq k \leq 2^j}} |\lambda(\widehat{\Omega}_{\alpha_{j,k}}) - \lambda(\Omega^*_{\alpha_{j,k}})|. \end{aligned}$$

From proposition 4, with probability at least $1-\delta$ we have: for all $0 \leq j \leq j_{\max}, 0 \leq k \leq 2^j$,

$$|\lambda(\widehat{\Omega}_{\alpha_{j,k}}) - \lambda(\Omega^*_{\alpha_{j,k}})| \leq 2\frac{\phi_n(\delta)}{Q^*(\alpha_{j,k})} \leq \frac{2\phi_n(\delta)}{Q^*(1-\varepsilon)},$$

therefore with probability at least $1-\delta$, for all $0 \leq j \leq j_{\max}$, $0 \leq k \leq 2^j - 1$,

$$|\widehat{\mathcal{E}}(I_{j,k}) - \mathcal{E}(I_{j,k})| \leq \frac{4\phi_n(\delta)}{Q^*(1-\varepsilon)}$$

and the result follows. □

From Lemma 7, with probability at least $1-\delta$, for all $0 \leq j \leq j_{\max}$, $0 \leq k \leq 2^j - 1$,

$$\mathcal{E}(I_{j,k}) - \frac{4\phi_n(\delta)}{Q^*(1-\varepsilon)} \leq \widehat{\mathcal{E}}(I_{j,k}) \leq \mathcal{E}(I_{j,k}) + \frac{4\phi_n(\delta)}{Q^*(1-\varepsilon)}\,. \tag{34}$$

Let $E_{\sigma_\tau}(\mathrm{MV}^*)$ be the piecewise constant approximant of the optimal MV curve built from the same recursive strategy as the one implemented in Algorithm 2 but based on the true local error $\mathcal{E}$ and without the condition $j < j_{\max}$, denoting σ_τ the associated meshgrid. We thus have, for every $\alpha \in [0, 1-\varepsilon]$:

$$E_{\sigma_\tau}(\mathrm{MV}^*)(\alpha) = \sum_{\substack{j,k \\ \alpha_{j,k} \in \sigma_\tau}} \lambda(\Omega^*_{\alpha_{j,k+1}}) \cdot \mathbb{I}\{\alpha \in I_{j,k}\}\,.$$

We also use the notation $E_{\widehat{\sigma}_\tau}$ to denote the function defined for all $\alpha \in [0, 1-\varepsilon]$ as

$$E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\alpha) = \sum_{\substack{j,k \\ \alpha_{j,k} \in \widehat{\sigma}_\tau}} \lambda(\Omega^*_{\alpha_{j,k+1}}) \cdot \mathbb{I}\{\alpha \in I_{j,k}\}$$

where we recall that $\widehat{\sigma}_\tau$ is the meshgrid obtained in Algorithm 2 implemented with the empirical error estimate $\widehat{\mathcal{E}}$.

Choosing $\tau = 5\phi_n(\delta)/Q^*(1-\varepsilon)$, we obtain that with probability at least $1-\delta$, the meshgrid $\widehat{\sigma}_\tau$ is finer than $\sigma_{\widetilde{\tau}_1}$ where $\widetilde{\tau}_1 = \tau + (4\phi_n(\delta) + 1/n)/Q^*(1-\varepsilon)$, and coarser than $\sigma_{\widetilde{\tau}_0}$ where $\widetilde{\tau}_0 = \tau - 4\phi_n(\delta)/Q^*(1-\varepsilon)$. Indeed, thanks to (34), $\mathcal{E}(I_{j,k}) > \widetilde{\tau}_1$ implies that with probability $1-\delta$, $\widehat{\mathcal{E}}(I_{j,k}) > \widetilde{\tau}_1 - 4\phi_n/Q^*(1-\varepsilon) > \tau$. It also implies that $j < j_{\max}$. Indeed, if $j \geq j_{\max}$, then $I_{j,k} = [\alpha_1, \alpha_2]$ is such that $\alpha_2 - \alpha_1 < 1/n$ which gives, using the mean value theorem, $\mathcal{E}(I_{j,k}) = \lambda(\Omega_{\alpha_2}) - \lambda(\Omega_{\alpha_1}) < 1/(Q^*(1-\varepsilon)n)$ which contradicts $\mathcal{E}(I_{j,k}) > \widetilde{\tau}_1 > 1/(Q^*(1-\varepsilon)n$. Therefore with probability at least $1-\delta$, splitting $I_{j,k}$ for the true local error and a tolerance $\widetilde{\tau}_1$ implies splitting $I_{j,k}$ in Algorithm 2. Analogously, $\widehat{\mathcal{E}}(I_{j,k}) > \tau$ implies, thanks to (34), that with probability at least $1-\delta$, $\mathcal{E}(I_{j,k}) > \widetilde{\tau}_0$. Therefore with probability at least $1-\delta$, splitting $I_{j,k}$ in Algorithm 2 implies splitting $I_{j,k}$ for the true local error and a tolerance $\widetilde{\tau}_0$.

We now use the following bound

$$\begin{aligned}\sup_{\alpha\in[0,1-\varepsilon]} |\widehat{\mathrm{MV}^*}(\alpha) - \mathrm{MV}^*(\alpha)| \leq{}& \sup_{\alpha\in[0,1-\varepsilon]} |\widehat{\mathrm{MV}^*}(\alpha) - E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\alpha)| \\ &+ \sup_{\alpha\in[0,1-\varepsilon]} |E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\alpha) - \mathrm{MV}^*(\alpha)|\,.\end{aligned}$$

The second term is bounded by $\widetilde{\tau}_1$ with probability at least $1-\delta$ because with probability at least $1-\delta$ the meshgrid $\widehat{\sigma}_\tau$ is finer than $\sigma_{\widetilde{\tau}_1}$ and therefore

$$\begin{aligned}\sup_{\alpha\in[0,1-\varepsilon]} |E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\alpha) - \mathrm{MV}^*(\alpha)| &\leq \sup_{\alpha\in[0,1-\varepsilon]} |E_{\sigma_{\widetilde{\tau}_1}}(\mathrm{MV}^*)(\alpha) - \mathrm{MV}^*(\alpha)| \\ &= \sup_{\substack{\alpha\in I \\ I \text{ interval of } \sigma_{\widetilde{\tau}_1}}} |E_{\sigma_{\widetilde{\tau}_1}}(\mathrm{MV}^*)(\alpha) - \mathrm{MV}^*(\alpha)| \\ &\leq \sup_{\substack{\alpha\in I \\ I \text{ interval of } \sigma_{\widetilde{\tau}_1}}} \mathcal{E}(I) \leq \widetilde{\tau}_1\,.\end{aligned}$$

On the same event, for the first term we have,

$$\sup_{\alpha\in[0,1-\varepsilon]} |\widehat{\mathrm{MV}^*}(\alpha) - E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\alpha)| \leq \sup_{\substack{0\leq j\leq j_{\max}\\ 0\leq k\leq 2^j}} |\lambda(\Omega^*_{\alpha_{j,k+1}}) - \lambda(\widehat{\Omega}_{\alpha_{j,k+1}})\ |$$
$$\leq \frac{2\phi_n(\delta)}{Q^*(1-\varepsilon)}\,.$$

Thus we finally have

$$\sup_{\alpha\in[0,1-\varepsilon]} |\widehat{\mathrm{MV}^*}(\alpha) - \mathrm{MV}^*(\alpha)| \leq \widetilde{\tau}_1 + \frac{2\phi_n(\delta)}{Q^*(1-\varepsilon)} \leq \frac{1}{Q^*(1-\varepsilon)}\left(11\phi_n(\delta) + \frac{1}{n}\right)\,.$$

For the second part of the theorem on the cardinality of the meshgrid $\widehat{\sigma}_\tau$ obtained with Algorithm 2 we use the following lemma (DeVore, 1998).

Lemma 8. *Let* $\mathrm{card}(\sigma_\tau)$ *be the number of terminal nodes obtained when the algorithm is implemented with the true local error* $\mathcal{E}(I)$*. Suppose that assumptions* ***(A1)****,* ***(A2)*** *and* ***(C3)*** *are fulfilled. There exists a universal constant* $C > 0$ *such that, for all* $\tau > 0$*:*

$$\mathrm{card}(\sigma_\tau) \leq \frac{C}{\tau}\|\mathrm{MV}^{*'}\|_{L\log L}\,.$$

Applying this lemma to $\widetilde{\tau}_0 = \phi_n(\delta)/Q^*(1-\varepsilon)$ we obtain that the number of terminal nodes $\mathrm{card}(\sigma_{\widetilde{\tau}_0})$ of $\sigma_{\widetilde{\tau}_0}$ is such that

$$\mathrm{card}(\sigma_{\widetilde{\tau}_0}) \leq CQ^*(1-\varepsilon)\frac{\|\mathrm{MV}^{*'}\|_{L\log L}}{\phi_n(\delta)}\,.$$

As $\widehat{\sigma}_\tau$ is coarser than $\sigma_{\widetilde{\tau}_0}$ with probability at least $1-\delta$, the number of terminal nodes of the former is bounded by the number of terminal nodes of the latter with probability at least $1-\delta$.

A.3.2. Proof of Theorem 4

We use the following notation: $\widetilde{\alpha}_k = F(\widetilde{\Omega}_{\alpha_k})$. We first prove a lemma quantifying the loss coming from the monotonicity step of Algorithm 3.

Lemma 9. *Suppose that the assumptions of Theorem 4 are satisfied. Let* $\delta \in (0,1)$*. For all* $n \geq 1$*, with probability at least* $1-\delta$*, we have:* $\forall k \in \{1,\ldots,\widehat{K}\}$*,*

$$|\widetilde{\alpha}_k - \alpha_k| \leq \max(kC_1\phi_n(\delta)^{\frac{\gamma}{\gamma+1}}, 2\phi_n(\delta)) \tag{35}$$

and

$$|\lambda(\widetilde{\Omega}_{\alpha_k}) - \mathrm{MV}^*(\alpha_k)| \leq \max(kC_2\phi_n(\delta)^{\frac{\gamma}{\gamma+1}}, 2\phi_n(\delta)/Q^*(1-\varepsilon))\,. \tag{36}$$

where $C_1 > 0$ *and* $C_2 > 0$ *are constants depending only on* γ*,* $\|f\|_\infty$ *and the constant* C *of assumption* ***(C4)****.*

Proof. First, from proposition 4, we have with probability at least $1-\delta$, $\forall k \in \{0,\ldots,\widehat{K}\}$,

$$|\alpha_k - F(\widehat{\Omega}_{\alpha_k})| \leq 2\phi_n(\delta) \text{ and } |\lambda(\widehat{\Omega}_{\alpha_k}) - \lambda(\Omega^*_{\alpha_k})| \leq \frac{2\phi_n(\delta)}{Q^*(1-\varepsilon)}\,. \tag{37}$$

For $k=0$, we have $\alpha_0 = 0$ and thus $\widetilde{\Omega}_{\alpha_0} = \widehat{\Omega}_{\alpha_0} = \emptyset$ and the inequalities of the Lemma are trivially satisfied. For $k=1$, we have $\widetilde{\Omega}_{\alpha_1} = \widehat{\Omega}_{\alpha_1}$ and with (37) the inequalities of the Lemma are satisfied. For all $k \in \{0,\ldots,\widehat{K}\}$,

$$F(\widetilde{\Omega}_{\alpha_k}) = F\Big(\bigcup_{j=0}^{k} \widehat{\Omega}_{\alpha_j}\Big) \geq F(\widehat{\Omega}_{\alpha_k}) \geq \alpha_k - 2\phi_n(\delta)$$

and $\widetilde{\alpha}_k - \alpha_k \geq -2\phi_n(\delta)$. The same argument gives

$$\lambda(\widetilde{\Omega}_{\alpha_k}) - \lambda(\Omega^*_{\alpha_k}) \geq -\frac{2\phi_n(\delta)}{Q^*(1-\varepsilon)}\,.$$

We show how to obtain the upper bound for the case $k=2$. We have $\widetilde{\Omega}_{\alpha_2} = \widehat{\Omega}_{\alpha_1} \cup \widehat{\Omega}_{\alpha_2}$ and

$$F(\widetilde{\Omega}_{\alpha_2}) - \alpha_2 = F(\widetilde{\Omega}_{\alpha_2}) - F(\Omega^*_{\alpha_2}) \leq F(\widetilde{\Omega}_{\alpha_2}) - F(\Omega^*_{\alpha_2} \cap \widetilde{\Omega}_{\alpha_2}) = F(\widetilde{\Omega}_{\alpha_2} \setminus \Omega^*_{\alpha_2})\,.$$

Now,

$$\widetilde{\Omega}_{\alpha_2} \setminus \Omega^*_{\alpha_2} = \{\widehat{\Omega}_{\alpha_1} \setminus \Omega^*_{\alpha_2}\} \cup \{\widehat{\Omega}_{\alpha_2} \setminus \Omega^*_{\alpha_2}\} \subset \{\widehat{\Omega}_{\alpha_1} \setminus \Omega^*_{\alpha_1}\} \cup \{\widehat{\Omega}_{\alpha_2} \setminus \Omega^*_{\alpha_2}\}$$

where the last inclusion comes from the fact that $\Omega^*_{\alpha_1} \subset \Omega^*_{\alpha_2}$. Hence,

$$F(\widetilde{\Omega}_{\alpha_2}) - \alpha_2 \leq F(\widehat{\Omega}_{\alpha_1} \setminus \Omega^*_{\alpha_1}) + F(\widehat{\Omega}_{\alpha_2} \setminus \Omega^*_{\alpha_2})$$

and thus

$$F(\widetilde{\Omega}_{\alpha_2}) - \alpha_2 \leq F(\widehat{\Omega}_{\alpha_1} \Delta \Omega^*_{\alpha_1}) + F(\widehat{\Omega}_{\alpha_2} \Delta \Omega^*_{\alpha_2})\,.$$

Using assumption **(C4)**, for any $\alpha \in (0,1)$, if $\widehat{\Omega}_\alpha$ denotes the solution of the optimization problem (13), we can as in Polonik (1997) (see proof of Lemma 3.2) bound $F(\widehat{\Omega}_\alpha \Delta \Omega^*_\alpha)$: with probability at least $1-\delta$,

$$F(\Omega^*_\alpha \Delta \widehat{\Omega}_\alpha) \leq \|f\|_\infty C t^\gamma + 2\frac{\|f\|_\infty}{t}\phi_n(\delta)$$

Minimizing the right-hand side with respect to t gives

$$F(\widehat{\Omega} \Delta \Omega^*_\alpha) \leq C_1 \phi_n(\delta)^{\frac{\gamma}{\gamma+1}}\,,$$

where $C_1 > 0$ is a constant depending only on γ, C and $\|f\|_\infty$. We thus finally have, with probability at least $1-\delta$,

$$\widetilde{\alpha}_2 - \alpha_2 \leq 2C_1 \phi_n(\delta)^{\frac{\gamma}{\gamma+1}}$$

and therefore,

$$|\widetilde{\alpha}_2 - \alpha_2| \leq \max(2C_1\phi_n(\delta)^{\frac{\gamma}{\gamma+1}}, 2\phi_n(\delta))\,.$$

For $k \geq 2$, we can obtain in a similar fashion,

$$F(\widetilde{\Omega}_{\alpha_k}) - \alpha_k \leq \sum_{j=1}^{k} F(\widehat{\Omega}_{\alpha_j} \setminus \Omega^*_{\alpha_j}) \leq \sum_{j=1}^{k} F(\widehat{\Omega}_{\alpha_j} \Delta \Omega^*_{\alpha_j})$$

and derive the following result: with probability at least $1-\delta$, for all $k \geq 2$,

$$|\widetilde{\alpha}_k - \alpha_k| \leq \max(kC_1\phi_n(\delta)^{\frac{\gamma}{\gamma+1}}, 2\phi_n(\delta))\,.$$

For (36), first note that for all $\alpha \in [0, 1-\varepsilon]$, $\lambda(\Omega^*_\alpha \Delta \widehat{\Omega}_\alpha) \leq C_2\phi_n(\delta)^{\frac{\gamma}{\gamma+1}}$ where $C_2 > 0$ depends only on γ and C. Therefore the same procedure leads to: with probability $1-\delta$, for all $k \geq 0$,

$$|\lambda(\widetilde{\Omega}_{\alpha_k}) - \mathrm{MV}^*(\alpha_k)| \leq \max(kC_2\phi_n(\delta)^{\frac{\gamma}{\gamma+1}}, 2\phi_n(\delta)/Q^*(1-\varepsilon))\,. \qquad \square$$

We can now prove Theorem 4. We first write

$$\begin{aligned}\sup_{\alpha\in[0,1-\varepsilon]} \left\{\mathrm{MV}_{\hat{s}}(\alpha) - \mathrm{MV}^*(\alpha)\right\} \leq & \sup_{\alpha\in[0,1-\varepsilon]} \left\{\mathrm{MV}_{\hat{s}}(\alpha) - E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\alpha)\right\} \\ & + \sup_{\alpha\in[0,1-\varepsilon]} \left\{E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\alpha) - \mathrm{MV}^*(\alpha)\right\}.\end{aligned}$$

Taking $\tau = 4\phi_n(\delta)/Q^*(1-\varepsilon) + c_n$, $c_n > 0$ and using the same argument as in the proof of Theorem 3, the second term of the right-hand side is bounded by $\widetilde{\tau}_1 = \tau + (4\phi_n(\delta) + 1/n)/Q^*(1-\varepsilon)$. We now deal with the first term. Let $\alpha \in [0, 1-\varepsilon]$,

$$\mathrm{MV}_{\hat{s}}(\alpha) - E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\alpha) = \sum_{k=0}^{\widehat{K}-1} \lambda(\widetilde{\Omega}_{\alpha_{k+1}}) \cdot \mathbb{I}\{\alpha \in [\widetilde{\alpha}_k, \widetilde{\alpha}_{k+1})\} - E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\alpha)\,.$$

But one can observe that

$$E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\alpha) = \sum_{k=0}^{\widehat{K}-1} E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\alpha) \cdot \mathbb{I}\{\alpha \in [\widetilde{\alpha}_k, \widetilde{\alpha}_{k+1})\}$$

to obtain

$$\begin{aligned}\mathrm{MV}_{\hat{s}}(\alpha) - E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\alpha) &= \sum_{k=0}^{\widehat{K}-1} \lambda(\widetilde{\Omega}_{\alpha_{k+1}}) \cdot \mathbb{I}\{\alpha \in [\widetilde{\alpha}_k, \widetilde{\alpha}_{k+1})\} - E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\alpha) \\ &= \sum_{k=0}^{\widehat{K}-1} (\lambda(\widetilde{\Omega}_{\alpha_{k+1}}) - E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\alpha)) \cdot \mathbb{I}\{\alpha \in [\widetilde{\alpha}_k, \widetilde{\alpha}_{k+1})\}\,.\end{aligned}$$

Now $\alpha \mapsto E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\alpha)$ is increasing, therefore on $[\widetilde{\alpha}_k, \widetilde{\alpha}_{k+1})$, $E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\alpha) \geq E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\widetilde{\alpha}_k)$ and

$$\begin{aligned}
\mathrm{MV}_{\hat{s}}(\alpha) - E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\alpha) &\leq \sum_{k=0}^{\widehat{K}-1} (\lambda(\widetilde{\Omega}_{\alpha_{k+1}}) - E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\widetilde{\alpha}_k)) \cdot \mathbb{I}\{\alpha \in [\widetilde{\alpha}_k, \widetilde{\alpha}_{k+1})\} \\
&\leq \max_{0 \leq k \leq \widehat{K}-1} |\lambda(\widetilde{\Omega}_{\alpha_{k+1}}) - E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\widetilde{\alpha}_k)| \sum_{k=0}^{\widehat{K}-1} \mathbb{I}\{\alpha \in [\widetilde{\alpha}_k, \widetilde{\alpha}_{k+1})\} \\
&= \max_{0 \leq k \leq \widehat{K}-1} |\lambda(\widetilde{\Omega}_{\alpha_{k+1}}) - E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\widetilde{\alpha}_k)| \,.
\end{aligned}$$

We thus have

$$\begin{aligned}
\sup_{[0,1-\varepsilon]} \left\{ \mathrm{MV}_{\hat{s}} - E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*) \right\} &\leq \max_{0 \leq k \leq \widehat{K}-1} |\lambda(\widetilde{\Omega}_{\alpha_{k+1}}) - E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\widetilde{\alpha}_k)| \\
&\leq \max_{0 \leq k \leq \widehat{K}-1} |\lambda(\widetilde{\Omega}_{\alpha_{k+1}}) - E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\alpha_k)| \\
&\quad + \max_{0 \leq k \leq \widehat{K}-1} |E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\alpha_k) - E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\widetilde{\alpha}_k)| \,.
\end{aligned}$$

The first term is equal to $\max_{0 \leq k \leq \widehat{K}-1} |\lambda(\widetilde{\Omega}_{\alpha_{k+1}}) - \mathrm{MV}^*(\alpha_{k+1})|$ and thanks to Lemma 9 we have with probability at least $1-\delta$

$$\max_{0 \leq k \leq \widehat{K}-1} |\lambda(\widetilde{\Omega}_{\alpha_{k+1}}) - \mathrm{MV}^*(\alpha_{k+1})| \leq \max(\widehat{K} C_2 \phi_n(\delta)^{\frac{\gamma}{\gamma+1}}, 2\phi_n(\delta)/Q^*(1-\varepsilon)) \,. \tag{38}$$

For the second term we write

$$\begin{aligned}
&\left| E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\alpha_k) - E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\widetilde{\alpha}_k) \right| \\
&\qquad = \left| \mathrm{MV}^*(\alpha_{k+1}) - \sum_{j=0}^{\widehat{K}-1} \mathrm{MV}^*(\alpha_{j+1}) \cdot \mathbb{I}\{\widetilde{\alpha}_k \in [\alpha_j, \alpha_{j+1})\} \right| \\
&\qquad = \left| \mathrm{MV}^*(\alpha_{k+1}) - \mathrm{MV}^*(\alpha_{j_0+1}) \right| ,
\end{aligned}$$

where $j_0 \in \{0, \ldots, \widehat{K}-1\}$ is such that $\widetilde{\alpha}_k \in [\alpha_{j_0}, \alpha_{j_0+1})$. We bound $|\mathrm{MV}^*(\alpha_{k+1}) - \mathrm{MV}^*(\alpha_{j_0+1})|$ by

$$\left| \mathrm{MV}^*(\alpha_{k+1}) - \mathrm{MV}^*(\alpha_k) \right| + \left| \mathrm{MV}^*(\alpha_k) - \mathrm{MV}^*(\widetilde{\alpha}_k) \right| + \left| \mathrm{MV}^*(\widetilde{\alpha}_k) - \mathrm{MV}^*(\alpha_{j_0+1}) \right| .$$

We bound the first term as follows. Let $I = [\alpha_k, \alpha_{k+1}]$. If the depth $j_{\max}$ has been reached, $\alpha_{k+1} - \alpha_k < 1/n$. Therefore the mean value theorem gives:

$$|\mathrm{MV}^*(\alpha_{k+1}) - \mathrm{MV}^*(\alpha_k)| \leq \frac{1}{n Q^*(1-\varepsilon)} \,.$$

Otherwise, this means that $\widehat{\mathcal{E}}(I) \leq \tau$ and we decompose $|\mathrm{MV}^*(\alpha_{k+1}) - \mathrm{MV}^*(\alpha_k)| = \mathcal{E}(I)$ into

$$\mathcal{E}(I) = \big(\mathcal{E}(I) - \widehat{\mathcal{E}}(I)\big) + \widehat{\mathcal{E}}(I).$$

Using Lemma 9, the right-hand side is bounded by $4\phi_n(\delta)/Q^*(1-\varepsilon)+\tau$. Eventually,

$$|\mathrm{MV}^*(\alpha_{k+1}) - \mathrm{MV}^*(\alpha_k)| \leq \max\left(\frac{4\phi_n(\delta)}{Q^*(1-\varepsilon)} + \tau, \frac{1}{nQ^*(1-\varepsilon)}\right)$$

Using the mean value theorem and Lemma 9, the second term can be bounded by

$$\frac{1}{Q^*(1-\varepsilon)}|\alpha_k - \widetilde{\alpha}_k| \leq \frac{1}{Q^*(1-\varepsilon)}\max(\widehat{K}C_1\phi_n(\delta)^{\frac{\gamma}{\gamma+1}}, 2\phi_n(\delta))\,.$$

As $\widetilde{\alpha}_k \in [\alpha_{j_0}, \alpha_{j_0+1})$ and as MV^* is increasing, the third term is bounded by $|\lambda(\Omega^*_{\alpha_{j_0+1}}) - \lambda(\Omega^*_{\alpha_{j_0}})|$ which is also bounded as the first term by $\max(4\phi_n(\delta)/Q^*(1-\varepsilon)+\tau, 1/(nQ^*(1-\varepsilon)))$. We thus obtain

$$\begin{aligned}\left|E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\alpha_k) - E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\widetilde{\alpha}_k)\right| &\leq 2\max\left(\frac{4\phi_n(\delta)}{Q^*(1-\varepsilon)} + \tau, \frac{1}{nQ^*(1-\varepsilon)}\right)\\ &\quad + \frac{\max(\widehat{K}C_1\phi_n(\delta)^{\frac{\gamma}{\gamma+1}}, 2\phi_n(\delta))}{Q^*(1-\varepsilon)}\,.\end{aligned}$$

and combining this inequality with (38), we have

$$\begin{aligned}\sup_{\alpha\in[0,1-\varepsilon]}\left\{\mathrm{MV}_{\hat{s}}(\alpha) - E_{\widehat{\sigma}_\tau}(\mathrm{MV}^*)(\alpha)\right\} &\leq \max(\widehat{K}C_2\phi_n(\delta)^{\frac{\gamma}{\gamma+1}}, 2\phi_n(\delta)/Q^*(1-\varepsilon))\\ &\quad + 2\max\left(\frac{4\phi_n(\delta)}{Q^*(1-\varepsilon)} + \tau, \frac{1}{nQ^*(1-\varepsilon)}\right)\\ &\quad + \frac{\max(\widehat{K}C_1\phi_n(\delta)^{\frac{\gamma}{\gamma+1}}, 2\phi_n(\delta))}{Q^*(1-\varepsilon)}\,.\end{aligned}$$

Eventually, with probability at least $1-\delta$,

$$\begin{aligned}\sup_{\alpha\in[0,1-\varepsilon]}\left\{\mathrm{MV}_{\hat{s}}(\alpha) - \mathrm{MV}^*(\alpha)\right\} &\leq \max(\widehat{K}C_2\phi_n(\delta)^{\frac{\gamma}{\gamma+1}}, 2\phi_n(\delta)/Q^*(1-\varepsilon))\\ &\quad + 2\max\left(\frac{4\phi_n(\delta)}{Q^*(1-\varepsilon)} + \tau, \frac{1}{nQ^*(1-\varepsilon)}\right)\\ &\quad + \tau + \frac{4\phi_n(\delta)}{Q^*(1-\varepsilon)} + \frac{1}{nQ^*(1-\varepsilon)}\\ &\quad + \frac{\max(\widehat{K}C_1\phi_n(\delta)^{\frac{\gamma}{\gamma+1}}, 2\phi_n(\delta))}{Q^*(1-\varepsilon)}\,.\end{aligned} \tag{39}$$

Using the same argument as in the proof of Theorem 3, on the same event, which holds with probability at least $1-\delta$, $\widehat{\sigma}_\tau$ is coarser than $\sigma_{\widetilde{\tau}_0}$ where $\widetilde{\tau}_0 = \tau - 4\phi_n(\delta)/Q^*(1-\varepsilon) = c_n$ and therefore $\widehat{K}$ is of the order of c_n. Using assumption **(C2)**, for n large enough, $\phi_n(\delta) \leq \phi_n(\delta)^{\gamma/(\gamma+1)}$. Therefore the right-hand side of (39) is of the order of:

$$\frac{1}{c_n}\left(\frac{1}{n}\right)^{\frac{\gamma}{2(\gamma+1)}} + c_n\,.$$

Choosing $c_n = (1/n)^{\gamma/(4(1+\gamma))}$ to balance the two terms of this expression, we finally obtain that with probability at least $1-\delta$,

$$\sup_{\alpha\in[0,1-\varepsilon]} \left\{ \mathrm{MV}_{\hat{s}}(\alpha) - \mathrm{MV}^*(\alpha) \right\} = O\left(n^{-\frac{\gamma}{4(1+\gamma)}} \right) .$$

Appendix B: Technical results

This section contains the proofs of technical results that have been postponed for the sake of clarity.

B.1. Proof of Lemma 1

Proof. We recall here that $\mathcal{Z}_f = \{x, F_f(f(x)) > 0\}$ and $\mathrm{supp}(f) = \{x, f(x) > 0\}$. If $f(x) = 0$ then $F_f(f(x)) = 0$ thus $\overline{\mathrm{supp}(f)} \subset \overline{\mathcal{Z}_f}$ and $\overline{\mathcal{Z}_f} \Delta \overline{\mathrm{supp}(f)} = \overline{\mathcal{Z}_f} \setminus \overline{\mathrm{supp}(f)} = \overline{\mathcal{Z}_f} \cap \mathrm{supp}(f)$. Therefore we have to prove $\lambda(\overline{\mathcal{Z}_f} \cap \mathrm{supp}(f)) = 0$. Let's assume $\lambda(\overline{\mathcal{Z}_f} \cap \mathrm{supp}(f)) > 0$ and let $f^* = \sup_{u \in \overline{\mathcal{Z}_f} \cap \mathrm{supp}(f)} f(u)$. The intersection $\overline{\mathcal{Z}_f} \cap \mathrm{supp}(f)$ is not empty. Otherwise we would have $\lambda(\overline{\mathcal{Z}_f} \cap \mathrm{supp}(f)) = 0$. Therefore there exists $u \in \overline{\mathcal{Z}_f} \cap \mathrm{supp}(f)$ and we have $f^* \geq f(u) > 0$ as $u \in \mathrm{supp}(f)$. We also have $F_f(f^*) = 0$. Indeed, f^* is in the closure of $\{f(u), u \in \overline{\mathcal{Z}_f} \cap \mathrm{supp}(f)\}$ and therefore there exists a sequence $(f_m)_{m\geq 0}$ of elements of $\{f(u), u \in \overline{\mathcal{Z}_f} \cap \mathrm{supp}(f)\}$ which converges towards f^* as m tends to infinity. Thus $F_f(f_m) = 0$ for all $m \geq 0$ and as F_f is continuous, $F_f(f^*) = 0$, i.e.

$$\int_{\{u, 0\leq f(u)\leq f^*\}} f(u)\mathrm{d}u = 0 .$$

This implies that f is equal to 0 λ-almost everywhere on $\{u, 0 \leq f(u) \leq f^*\}$. However, if $u \in \overline{\mathcal{Z}_f} \cap \mathrm{supp}(f)$ then, by definition of f^*, $f(u) \leq f^*$. Therefore $\overline{\mathcal{Z}_f} \cap \mathrm{supp}(f) \subset \{u, 0 \leq f(u) \leq f^*\}$ and as $\lambda(\overline{\mathcal{Z}_f} \cap \mathrm{supp}(f)) > 0$ and $f > 0$ on $\mathrm{supp}(f)$, $\{u, 0 \leq f(u) \leq f^*\}$ contains a subset of non null λ-measure on which $f > 0$. This is a contradiction with the fact that f is equal to 0 λ-almost everywhere on $\{u, 0 \leq f(u) \leq f^*\}$ and thus $\lambda(\overline{\mathcal{Z}_f} \Delta \overline{\mathrm{supp}(f)}) = 0$, which is equivalent to $\lambda(\mathcal{Z}_f \Delta \mathrm{supp}(f)) = 0$. □

B.2. Proof of Lemma 3

Proof of Lemma 3. The proof is based on Remark 3.2.4 in (Csörgő, 1983). Let q_n denote the quantile process defined for all $\alpha \in [\varepsilon, 1-\varepsilon]$ by

$$q_n(\alpha) = \sqrt{n} f_s(F_s^\dagger(\alpha))(\widehat{F}_s^\dagger(\alpha) - F_s^\dagger(\alpha)) .$$

Using the mean value theorem we can write for all $\alpha \in [\varepsilon, 1-\varepsilon]$,

$$\begin{aligned} q_n(\alpha) &= \sqrt{n} f_s(F_s^\dagger(\alpha))\big(F_s^\dagger(\widehat{U}^\dagger(\alpha)) - F_s^\dagger(\alpha)\big) \\ &= \frac{f_s(F_s^\dagger(\alpha))}{f_s(F_s^\dagger(\xi))} \sqrt{n}(\widehat{U}^\dagger(\alpha) - \alpha) \end{aligned}$$

where ξ is between $\widehat{U}^{\dagger}(\alpha)$ and α and where $\widehat{U}^{\dagger}(\alpha)$ is defined as in the proof of Lemma 2. Let u_n be the uniform quantile process defined in Lemma 2. For all $\alpha \in [\varepsilon, 1-\varepsilon]$, $q_n(\alpha) = u_n(\alpha) f_s(F_s^{\dagger}(\alpha))/f_s(F_s^{\dagger}(\xi))$.

Now we know from Theorem 1 in (Csörgő and Révész, 1978) that for all z_1 and for all n,

$$\mathbb{P}\Big(\sup_{\alpha\in[0,1]} |u_n(\alpha) - B_n^1(\alpha)| > \frac{C_1 \log n + z_1}{\sqrt{n}}\Big) \leq C_2 \exp(-C_3 z_1) \tag{40}$$

where C_1, C_2 and C_3 are positive absolute constants and where $\{B_n^1(\alpha), \alpha \in [0,1]\}_{n\geq 1}$ is the same sequence of Brownian Bridges as the one of the proof of assertion (ii) of Theorem 1. For all $\alpha \in [\varepsilon, 1-\varepsilon]$, we can write

$$q_n(\alpha) - B_n^1(\alpha) = \big(u_n(\alpha) - B_n^1(\alpha)\big) + u_n(\alpha)\left(\frac{f_s(F_s^{\dagger}(\alpha))}{f_s(F_s^{\dagger}(\xi))} - 1\right). \tag{41}$$

Using a Taylor expansion of λ_s as in the proof of part (ii) of Theorem 1 we have for all $\alpha \in [\varepsilon, 1-\varepsilon]$,

$$\begin{aligned} r_n(\alpha) - Z_n(\alpha) = \frac{\lambda_s'(\alpha_s^{-1}(\alpha))}{f_s(\alpha_s^{-1}(\alpha))}\Big(\sqrt{n} f_s(\alpha_s^{-1}(\alpha))(\widehat{\alpha}_s^{-1}(\alpha) - \alpha_s^{-1}(\alpha)) - B_n^1(1-\alpha)\Big) \\ + \sqrt{n}\frac{\lambda_s''(\xi_1)}{2}(\widehat{\alpha}_s^{-1}(\alpha) - \alpha_s^{-1}(\alpha))^2 \end{aligned}$$

where ξ_1 is between $\widehat{\alpha}_s^{-1}(\alpha)$ and $\alpha_s^{-1}(\alpha)$. With the change of variable $\alpha' = 1-\alpha$, we obtain for all $\alpha' \in [\varepsilon, 1-\varepsilon]$ that ξ_1 is between $\widehat{F}_s^{\dagger}(\alpha')$ and $F_s^{\dagger}(\alpha')$ and that

$$r_n(\alpha') - Z_n(\alpha') = \frac{\lambda_s'(F_s^{\dagger}(\alpha'))}{f_s(F_s^{\dagger}(\alpha'))}(q_n(\alpha') - B_n^1(\alpha')) + \sqrt{n}\frac{\lambda_s''(\xi_1)}{2}(\widehat{F}_s^{\dagger}(\alpha') - F_s^{\dagger}(\alpha'))^2 .$$

and therefore

$$|r_n(\alpha') - Z_n(\alpha')| \leq \frac{|\lambda_s'(F_s^{\dagger}(\alpha'))|}{f_s(F_s^{\dagger}(\alpha'))}|q_n(\alpha') - B_n^1(\alpha')| + \sqrt{n}\frac{|\lambda_s''(\xi_1)|}{2}|\widehat{F}_s^{\dagger}(\alpha') - F_s^{\dagger}(\alpha')|^2 . \tag{42}$$

To bound the right-hand side of the previous inequality we need the following results. From the DKW inequality (see proof of Lemma 2) we have for all $z_2 > 0$ and for all n,

$$\mathbb{P}(\sup_{\alpha\in[0,1]} |u_n(\alpha)| \geq z_2) \leq 2\exp(-2z_2^2) . \tag{43}$$

From Theorem 1.5.1 in (Csörgő, 1983) we have for all $z_3 > 0$ and for all n,

$$\begin{aligned} \mathbb{P}\left(\sup_{\alpha\in[\varepsilon,1-\varepsilon]}\left|\frac{f_s(F_s^{\dagger}(\alpha))}{f_s(F_s^{\dagger}(\xi))} - 1\right| > z_3\right) \leq 4(\lfloor c \rfloor + 1)\Big(e^{-n\varepsilon h\left((1+z_3)^{1/2(\lfloor c \rfloor + 1)}\right)} \\ + e^{-n\varepsilon h\left((1+z_3)^{-1/2(\lfloor c \rfloor + 1)}\right)}\Big) \end{aligned} \tag{44}$$

where for all $z > 0$, $h(z) = z + \log(1/z) - 1$ and where $x \mapsto \lfloor x \rfloor$ denotes the floor part function. Finally from the proof of Theorem 1.4.3 in (Csörgő, 1983) we have for all $z_4 \geq 1$ and for all n,

$$\mathbb{P}\left(\sup_{\alpha\in[\varepsilon,1-\varepsilon]}\left|\frac{f_s(F_s^\dagger(\alpha))}{f_s(F_s^\dagger(\xi))}\right| > z_4\right) \leq 2(\lfloor c\rfloor+1)\Big(e^{-n\varepsilon h\left(z_4^{1/2(\lfloor c\rfloor+1)}\right)} + e^{-n\varepsilon h\left(z_4^{-1/2(\lfloor c\rfloor+1)}\right)}\Big). \tag{45}$$

Let z_1, z_2 and z_3 be positive and let $z_4 \geq 1$. We consider the four following events:

$$\mathcal{Z}_1 = \left\{\sup_{\alpha\in[0,1]}|u_n(\alpha) - B_n^1(\alpha)| > (C_1\log n + z_1)/\sqrt{n}\right\},$$

$$\mathcal{Z}_2 = \left\{\sup_{\alpha\in[0,1]}|u_n(\alpha)| \geq z_2\right\},$$

$$\mathcal{Z}_3 = \left\{\sup_{\alpha\in[\varepsilon,1-\varepsilon]}\left|f_s(F_s^\dagger(\alpha))/f_s(F_s^\dagger(\xi)) - 1\right| > z_3\right\} \text{ and}$$

$$\mathcal{Z}_4 = \left\{\sup_{\alpha\in[\varepsilon,1-\varepsilon]}\left|f_s(F_s^\dagger(\alpha))/f_s(F_s^\dagger(\xi))\right| > z_4\right\}.$$

Let $\omega \in \overline{\mathcal{Z}_1} \cap \overline{\mathcal{Z}_2} \cap \overline{\mathcal{Z}_3} \cap \overline{\mathcal{Z}_4}$ where $\overline{\mathcal{Z}}$ denotes the complement of a set $\mathcal{Z}$. We first deal with the first term of the right-hand side of (42). From (41) we have for all $\alpha' \in [\varepsilon, 1-\varepsilon]$,

$$\begin{aligned}|q_n(\alpha') - B_n^1(\alpha')| &\leq |u_n(\alpha') - B_n^1(\alpha')| + |u_n(\alpha')|\left|\frac{f_s(F_s^\dagger(\alpha'))}{f_s(F_s^\dagger(\xi))} - 1\right| \\ &\leq \frac{C_1\log n + z_1}{\sqrt{n}} + z_2 z_3\end{aligned}$$

where the second inequality holds because $\omega \in \overline{\mathcal{Z}_1} \cap \overline{\mathcal{Z}_2} \cap \overline{\mathcal{Z}_3}$. Now for all $\alpha' \in [\varepsilon, 1-\varepsilon]$, $F_s^\dagger(\alpha') \in [0, \|s\|_\infty]$. As λ_s' is continuous $|\lambda_s'(F_s^\dagger(\alpha'))| \leq \sup_{[0,\|s\|_\infty]}|\lambda_s'| < \infty$. Furthermore, for all $\alpha' \in [\varepsilon, 1-\varepsilon]$, $f_s(F_s^\dagger(\alpha')) \geq \inf_{\alpha\in[\varepsilon,1-\varepsilon]} f_s(F_s^\dagger(\alpha'))$ where the infimum is strictly positive. Therefore for all $\alpha' \in [\varepsilon, 1-\varepsilon]$,

$$\frac{|\lambda_s'(F_s^\dagger(\alpha'))|}{f_s(F_s^\dagger(\alpha'))}|q_n(\alpha') - B_n^1(\alpha')| \leq \frac{\sup_{[0,\|s\|_\infty]}|\lambda_s'|}{\inf_{[\varepsilon,1-\varepsilon]} f_s \circ F_s^\dagger}\left(\frac{C_1\log n + z_1}{\sqrt{n}} + z_2 z_3\right)$$

We now deal with the second term of the right-hand side of (42). For all $\alpha \in [\varepsilon, 1-\varepsilon]$, $\alpha_s^{-1}(\alpha)$ and $\widehat{\alpha}_s^{-1}(\alpha)$ are in $[0, \|s\|_\infty]$. Therefore $\xi_1 \in [0, \|s\|_\infty]$ and as λ_s'' is continuous, $|\lambda_s''(\xi_1)| \leq \sup_{[0,\|s\|_\infty]}|\lambda_s''|$. Now, as in the proof of Lemma

2, using the mean value theorem we can write for $\alpha' \in [\varepsilon, 1]$,

$$\begin{aligned}|\widehat{F}_s^\dagger(\alpha') - F_s^\dagger(\alpha')| &= \frac{1}{f_s(F_s^\dagger(\xi))} \cdot |\widehat{U}^\dagger(\alpha') - \alpha'| \\ &= \frac{1}{f_s(F_s^\dagger(\alpha'))} \left| \frac{f_s(F_s^\dagger(\alpha'))}{f_s(F_s^\dagger(\xi))} \right| \cdot |u_n(\alpha')| n^{-1/2}\end{aligned}$$

where ξ is the one of (41). As $\omega \in \overline{\mathcal{Z}_2} \cap \overline{\mathcal{Z}_4}$ and bounding $1/f_s(F_s^\dagger(\alpha'))$ by $1/\inf_{[\varepsilon,1-\varepsilon]} f_s \circ F_s^\dagger$ as previously we obtain

$$\sqrt{n}\frac{\lambda_s''(\xi_1)}{2}|\widehat{F}_s^\dagger(\alpha') - F_s^\dagger(\alpha')|^2 \leq \sup_{[0,\|s\|_\infty]} |\lambda_s''| \frac{z_2^2 z_4^2}{2\sqrt{n}(\inf_{[\varepsilon,1-\varepsilon]} f_s \circ F_s^\dagger)^2}\,.$$

We eventually have

$$\begin{aligned}\sup_{\alpha\in[\varepsilon,1-\varepsilon]} |r_n(\alpha) - Z_n(\alpha)| \leq{}& \frac{\sup_{[0,\|s\|_\infty]} |\lambda_s'|}{\inf_{[\varepsilon,1-\varepsilon]} f_s \circ F_s^\dagger} \left(\frac{C_1 \log n + z_1}{\sqrt{n}} + z_2 z_3 \right) \\ &+ \sup_{[0,\|s\|_\infty]} |\lambda_s''| \frac{z_2^2 z_4^2}{2\sqrt{n}(\inf_{[\varepsilon,1-\varepsilon]} f_s \circ F_s^\dagger)^2}\,.\end{aligned}$$

Choosing, $z_1 = c_1 \log n$ with $c_1 > 0$ arbitrary, $z_2 = \sqrt{\log n}$, $z_3 = c_3\sqrt{\log n/n}$ with $c_3 > 0$ arbitrary, and $z_4 = 2$ we obtain that,

$$\begin{aligned}\sup_{\alpha\in[\varepsilon,1-\varepsilon]} |r_n(\alpha) - Z_n(\alpha)| \leq{}& \frac{\sup_{[0,\|s\|_\infty]} |\lambda_s'|}{\inf_{[\varepsilon,1-\varepsilon]} f_s \circ F_s^\dagger} (C_1 + c_1 + c_3) \frac{\log n}{\sqrt{n}} \\ &+ 2 \sup_{[0,\|s\|_\infty]} |\lambda_s''| \frac{\log n}{\sqrt{n}(\inf_{[\varepsilon,1-\varepsilon]} f_s \circ F_s^\dagger)^2} \\ ={}& c_4 v_n\end{aligned}$$

where the constant c_4 is given by

$$c_4 = \frac{\sup_{[0,\|s\|_\infty]} |\lambda_s'|}{\inf_{[\varepsilon,1-\varepsilon]} f_s \circ F_s^\dagger} (C_1 + c_1 + c_3) + \frac{2\sup_{[0,\|s\|_\infty]} |\lambda_s''|}{(\inf_{[\varepsilon,1-\varepsilon]} f_s \circ F_s^\dagger)^2}\,.$$

Let $\mathcal{Z}_5$ be the event defined by

$$\mathcal{Z}_5 = \left\{ \sup_{\alpha\in[\varepsilon,1-\varepsilon]} |r_n(\alpha) - Z_n(\alpha)| > c_4 v_n \right\}.$$

We have shown that $\omega \in \overline{\mathcal{Z}_5}$ therefore

$$\begin{aligned}\mathbb{P}(\mathcal{Z}_5) &\leq \mathbb{P}(\mathcal{Z}_1) + \mathbb{P}(\mathcal{Z}_2) + \mathbb{P}(\mathcal{Z}_3) + \mathbb{P}(\mathcal{Z}_4) \\ &\leq \frac{C_2}{n^{C_3 c_1}} + \frac{2}{n^2} + 4(\lfloor c \rfloor + 1)\Big(e^{-n\varepsilon h\left((1+c_3\sqrt{\log n/n})^{1/2(\lfloor c \rfloor+1)}\right)} \\ &\qquad + e^{-n\varepsilon h\left((1+c_3\sqrt{\log n/n})^{-1/2(\lfloor c \rfloor+1)}\right)} \Big) \\ &\qquad + 2(\lfloor c \rfloor + 1)\Big(e^{-n\varepsilon h\left(2^{1/2(\lfloor c \rfloor+1)}\right)} + e^{-n\varepsilon h\left(2^{-1/2(\lfloor c \rfloor+1)}\right)} \Big)\,.\end{aligned}$$

We can choose c_1 such that $n^{-C_3 c_1} = O(v_n)$. Studying the function h we can also show that $h(2^{1/2(\lfloor c \rfloor+1)})$ and $h(2^{-1/2(\lfloor c \rfloor+1)})$ are strictly positive and if $a = 1/2(\lfloor c \rfloor + 1)$, as $n \to \infty$,

$$\exp(-n\varepsilon h((1 + c_3\sqrt{\log n/n})^a)) = \exp\left(-\varepsilon c_3^2 a^2 \log n/2 + O((\log n)^{3/2}/\sqrt{n})\right).$$

Similarly we can show that

$$\exp(-n\varepsilon h((1 + c_3\sqrt{\log n/n})^{-a})) = \exp\left(-\varepsilon c_3^2 a^2 \log n/2 + O((\log n)^{3/2}/\sqrt{n})\right)$$

and we can choose c_3 such that the right-hand sides of the two previous equations are of order $O(v_n)$. This finally gives the result of the lemma. □

B.3. Proof of Lemma 4

Proof of Lemma 4. To prove this lemma we carefully follow the proof of Lemma 3. From Lemma 11, Lemma 12 and Lemma 14 we know that there exist constants C_1 and C_2 both strictly positive such that for $\mathbb{P}$-almost all $\mathcal{D}_n$, for n large enough, $\sup_{t\in\mathbb{R}} |\widetilde{f}_s(t) - f_s(t)| \leq C_1 w_n$, $\sup_{\alpha\in[\varepsilon,1-\varepsilon]} |\widetilde{F_s}^{\dagger}(\alpha) - F_s^{\dagger}(\alpha)| \leq C_2 w_n$ and $\sup_{t\in\mathbb{R}} |\widetilde{f_s}'(t) - f_s'(t)|$ converges to 0 as n tends to ∞. Let $\mathcal{D}_n$ be such that these assertions are fulfilled. We denote by q_n^* the quantile process given for all $\alpha' \in (0,1)$ by

$$q_n^*(\alpha') = \widetilde{f}_s(\widetilde{F}_s^{\dagger}(\alpha'))\sqrt{n}\left((F_s^{Boot})^{\dagger}(\alpha') - \widetilde{F}_s^{\dagger}(\alpha')\right)$$

where $(F_s^{Boot})^{\dagger}$ is defined on $(0,1)$ as the generalized inverse of the empirical cumulative distribution function F_s^{Boot} which is based on a bootstrap sample $S_1^*, \ldots, S_n^*$ which are *i.i.d.* random variables with distribution $\widetilde{F}_s$.

We have for all $\alpha \in [\varepsilon, 1-\varepsilon]$,

$$q_n^*(\alpha) = \widetilde{f}_s(\widetilde{F}_s^{\dagger}(\alpha))\sqrt{n}\left(\widetilde{F}_s^{\dagger}(\widetilde{U}^{\dagger}(\alpha)) - \widetilde{F}_s^{\dagger}(\alpha)\right)$$

where $\widetilde{U}^{\dagger}(\alpha) = \widetilde{F}_s((F_s^{Boot})^{\dagger}(\alpha))$ for all $\alpha \in (0,1)$. Let u_n^* be the associated uniform quantile process defined for all $\alpha \in (0,1)$ by $u_n^*(\alpha) = \sqrt{n}\left(\widetilde{U}^{\dagger}(\alpha) - \alpha\right)$.

Let $\varepsilon' \in (0,1)$. We first prove the differentiability of $\widetilde{F}_s^{\dagger}$ on $[\varepsilon', 1-\varepsilon']$ for n large enough. Given assumption **(B5)** the kernel K is differentiable on $\mathbb{R}$. Therefore $\widetilde{f}_s$ is differentiable on $\mathbb{R}$. For n large enough and for all $\alpha \in [\varepsilon', 1-\varepsilon']$, $\widetilde{f}_s(\widetilde{F}_s^{\dagger}(\alpha)) = \widetilde{f}_s(\widetilde{\alpha}_s^{-1}(1-\alpha)) > 0$. Indeed, we can write

$$\inf_{\alpha\in[\varepsilon',1-\varepsilon']} \widetilde{f}_s(\widetilde{\alpha_s}^{-1}(\alpha)) \geq \inf_{t\in[\widetilde{\alpha_s}^{-1}(1-\varepsilon'),\widetilde{\alpha_s}^{-1}(\varepsilon')]} \widetilde{f}_s(t).$$

As $\sup_{\alpha\in[\varepsilon',1-\varepsilon']} |\widetilde{F_s}^{\dagger}(\alpha) - F_s^{\dagger}(\alpha)|$ converges to 0, $\widetilde{\alpha}_s^{-1}(\varepsilon')$ and $\widetilde{\alpha}_s^{-1}(1-\varepsilon')$ converge respectively towards $\alpha_s^{-1}(\varepsilon')$ and $\alpha_s^{-1}(1-\varepsilon')$. Therefore there exists a constant $c_1 > 0$, depending only on $\alpha_s^{-1}(1-\varepsilon')$ and $\alpha_s^{-1}(\varepsilon')$, such that for n

large enough, $\widetilde{\alpha_s}^{-1}(1-\varepsilon') \geq \alpha_s^{-1}(1-\varepsilon')-c_1 > a$ and $\widetilde{\alpha_s}^{-1}(\varepsilon') < \alpha_s^{-1}(\varepsilon')+c_1 < b$, *i.e.*, $[\widetilde{\alpha_s}^{-1}(1-\varepsilon'), \widetilde{\alpha_s}^{-1}(\varepsilon')] \subset [\alpha_s^{-1}(1-\varepsilon')-c_1, \alpha_s^{-1}(\varepsilon')+c_1]$. Therefore for n large enough,

$$\inf_{t\in[\widetilde{\alpha_s}^{-1}(1-\varepsilon'),\widetilde{\alpha_s}^{-1}(\varepsilon')]} \widetilde{f}_s(t) \geq \inf_{t\in[\alpha_s^{-1}(1-\varepsilon')-c_1,\alpha_s^{-1}(\varepsilon')+c_1]} \widetilde{f}_s(t)$$

and as $\widetilde{f}_s$ converges uniformly towards f_s, there exists a constant $c_2 > 0$ such that for n large enough,

$$\inf_{t\in[\alpha_s^{-1}(1-\varepsilon')-c_1,\alpha_s^{-1}(\varepsilon')+c_1]} \widetilde{f}_s(t) \geq \inf_{t\in[\alpha_s^{-1}(1-\varepsilon')-c_1,\alpha_s^{-1}(\varepsilon')+c_1]} f_s(t) - c_2 > 0\,.$$

Note that c_2 can be chosen such that the strict positivity holds because the infimum of f_s on $[\alpha_s^{-1}(1-\varepsilon')-c_1, \alpha_s^{-1}(\varepsilon')+c_1]$ is attained and thus strictly positive (assumption **(A4)**). Eventually, combining the three previous inequalities, for n large enough we have: for all $\alpha \in [\varepsilon', 1-\varepsilon']$, $\widetilde{f}_s(\widetilde{\alpha}_s^{-1}(\alpha)) > 0$.

Therefore, $\widetilde{F}_s$ is differentiable on $[\widetilde{\alpha}_s^{-1}(1-\varepsilon'), \widetilde{\alpha}_s^{-1}(\varepsilon')]$ with strictly positive derivative. This gives that $\widetilde{F}_s^{\dagger} : [\varepsilon', 1-\varepsilon'] \to [\widetilde{\alpha}_s^{-1}(1-\varepsilon'), \widetilde{\alpha}_s^{-1}(\varepsilon')]$ is equal to the ordinary inverse of $\widetilde{F}_s$ on $[\varepsilon', 1-\varepsilon']$ and $\widetilde{F}_s^{\dagger}$ is differentiable on $[\varepsilon', 1-\varepsilon']$.

We now need to invoke the same argument as in the beginning of the proof of Lemma 16. This argument says that there exists a constant c_3 such that for n large enough, if $\mathcal{Z}_0$ denotes the event $\mathcal{Z}_0 = \{[\widetilde{U}^{\dagger}(\varepsilon), \widetilde{U}^{\dagger}(1-\varepsilon)] \not\subset [\varepsilon-c_3, 1-\varepsilon+c_3]\}$ then $\mathbb{P}^*(\mathcal{Z}_0) \leq 2/n^2$ (refer to the proof of Lemma 16 for details). Let $\omega \in \overline{\mathcal{Z}}_0$. Assuming n large enough, for all $\alpha \in [\varepsilon, 1-\varepsilon]$, $\widetilde{U}^{\dagger}(\alpha) \in [\varepsilon - c_3, 1-\varepsilon+c_3]$. Furthermore for n large enough, we proved that $\widetilde{F}_s^{\dagger}$ is differentiable on $[\varepsilon', 1-\varepsilon']$ for all $\varepsilon' \in (0,1)$. Therefore we can apply the mean value theorem to write for all $\alpha \in [\varepsilon, 1-\varepsilon]$,

$$q_n^*(\alpha) = \frac{\widetilde{f}_s(\widetilde{F}_s^{\dagger}(\alpha))}{\widetilde{f}_s(\widetilde{F}_s^{\dagger}(\xi))}\sqrt{n}(\widetilde{U}^{\dagger}(\alpha)-\alpha) = \frac{\widetilde{f}_s(\widetilde{F}_s^{\dagger}(\alpha))}{\widetilde{f}_s(\widetilde{F}_s^{\dagger}(\xi))}u_n^*(\alpha)$$

where ξ is between $\widetilde{U}^{\dagger}(\alpha')$ and α'.

Applying Theorem 1 in (Csörgő and Révész, 1978) conditionally on $\mathcal{D}_n$, there exists a sequence of Brownian Bridges $\{B_n^{*1}(\alpha), \alpha \in [0,1]\}_{n\geq 1}$ such that for all z_1 and for all n,

$$\mathbb{P}^*\Big(\sup_{\alpha\in[0,1]} |u_n^*(\alpha) - B_n^{*1}(\alpha)| > \frac{C_3 \log n + z_1}{\sqrt{n}}\Big) \leq C_4 \exp(-C_5 z_1) \tag{46}$$

where C_3, C_4 and C_5 are positive absolute constants. Let $B_n^*(\alpha) = B_n^{*1}(1-\alpha)$ for all $\alpha \in [0,1]$. Reasoning similarly as in the proof of Lemma 3, we have for all $\alpha' \in [\varepsilon, 1-\varepsilon]$,

$$\begin{aligned} |r_n^*(\alpha') - Z_n^*(\alpha')| \leq{}& \frac{|\lambda_s'(\widetilde{F}_s^{\dagger}(\alpha'))|}{\widetilde{f}_s(\widetilde{F}_s^{\dagger}(\alpha'))}|q_n^*(\alpha') - B_n^{*1}(\alpha')| \\ &+ \sqrt{n}\frac{|\lambda_s''(\xi_1)|}{2}|(F_s^{Boot})^{\dagger}(\alpha') - \widetilde{F}_s^{\dagger}(\alpha')|^2 \end{aligned} \tag{47}$$

where ξ_1 is between $(F_s^{Boot})^{\dagger}(\alpha')$ and $\widetilde{F}_s^{\dagger}(\alpha')$.

Still following the proof of Lemma 3 we need the following results. Conditionally on $\mathcal{D}_n$, the DKW inequality gives for all $z_2 > 0$ and for all n,

$$\mathbb{P}^*\Big(\sup_{\alpha\in[0,1]} |u_n^*(\alpha)| \geq z_2\Big) \leq 2\exp(-2z_2^2)\,. \tag{48}$$

Using the result of Lemma 16 there exists a constant $c^* > 0$ such that we have, conditionally on $\mathcal{D}_n$, for all $z_4 \geq 1$ and for n large enough,

$$\begin{aligned}\mathbb{P}^*\left(\sup_{\alpha\in[\varepsilon,1-\varepsilon]}\left|\frac{\widetilde{f}_s(\widetilde{F}_s^\dagger(\alpha))}{\widetilde{f}_s(\widetilde{F}_s^\dagger(\xi))}\right| > z_4\right)\\ \leq \frac{2}{n^2} + 2(\lfloor c^*\rfloor+1)\Big(e^{-n\varepsilon h\left(z_4^{1/2(\lfloor c^*\rfloor+1)}\right)} + e^{-n\varepsilon h\left(z_4^{-1/2(\lfloor c^*\rfloor+1)}\right)}\Big)\end{aligned} \tag{49}$$

where we recall that for all $z > 0$, $h(z) = z + \log(1/z) - 1$. Following the proof of Lemma 16 and with calculations similar as the ones used in the proof of Theorem 1.4.3 in (Csörgő, 1983) it can also be shown that conditionally on $\mathcal{D}_n$ we have for all $z_3 > 0$ and for n large enough,

$$\begin{aligned}\mathbb{P}^*\left(\sup_{\alpha\in[\varepsilon,1-\varepsilon]}\left|\frac{\widetilde{f}_s(\widetilde{F}_s^\dagger(\alpha))}{\widetilde{f}_s(\widetilde{F}_s^\dagger(\xi))} - 1\right| > z_3\right)\\ \leq 4(\lfloor c^*\rfloor+1)\Big(e^{-n\varepsilon h\left((1+z_3)^{1/2(\lfloor c^*\rfloor+1)}\right)} + e^{-n\varepsilon h\left((1+z_3)^{-1/2(\lfloor c^*\rfloor+1)}\right)}\Big) + \frac{2}{n^2}\,.\end{aligned} \tag{50}$$

Now as for all $\alpha' \in [\varepsilon, 1-\varepsilon]$, $\widetilde{F}_s^\dagger(\alpha')$ is in the support of $\widetilde{f}_s$ which has shown in the proof of Lemma 13, is included in an interval of the form $[-c_4, c_4 + \|s\|_\infty]$ where c_4 is a strictly positive constant that depends only on the kernel K. As λ_s' is continuous, $|\lambda_s'(\widetilde{F}_s^\dagger(\alpha'))| \leq \sup_{[-c_4,c_4+\|s\|_\infty]} |\lambda_s'| < \infty$. Furthermore, as proved above, there exist constants c_5 and c_6 both strictly positive such that $[F_s^\dagger(\varepsilon) - c_5, F_s^\dagger(1-\varepsilon) + c_5] \subset (a, b)$ and such that for n large enough,

$$\inf_{\alpha'\in[\varepsilon,1-\varepsilon]} \widetilde{f}_s(\widetilde{F}_s^\dagger(\alpha')) \geq \inf_{t\in[F_s^\dagger(\varepsilon)-c_5, F_s^\dagger(1-\varepsilon)+c_5]} f_s(t) - c_6 > 0\,.$$

For all $\alpha \in [\varepsilon, 1-\varepsilon]$, $(F_s^{Boot})^\dagger(\alpha)$ and $\widetilde{F}_s^\dagger(\alpha)$ are in the support of $\widetilde{f}_s$ which is included in $[-c_4, c_4 + \|s\|_\infty]$. Therefore $\xi_1 \in [-c_4, c_4 + \|s\|_\infty]$ and as λ_s'' is continuous, $|\lambda_s''(\xi_1)| \leq \sup_{[-c_4,c_4+\|s\|_\infty]} |\lambda_s''| < +\infty$. Thus following the steps of the proof of Lemma 3 we have for n large enough,

$$\begin{aligned}\mathbb{P}^*\left(\sup_{\alpha\in[\varepsilon,1-\varepsilon]} |r_n^*(\alpha) - Z_n^*(\alpha)| > Cv_n\right)\\ \leq \frac{C_4}{n^{C_5c_7}} + \frac{8}{n^2} + 4(\lfloor c^*\rfloor+1)\Big(\exp\big(-n\varepsilon h\big((1+c_8\sqrt{\log n/n})^{1/2(\lfloor c^*\rfloor+1)}\big)\big)\\ + \exp\big(-n\varepsilon h\big((1+c_8\sqrt{\log n/n})^{-1/2(\lfloor c^*\rfloor+1)}\big)\big)\Big)\\ + 2(\lfloor c^*\rfloor+1)\Big(\exp\big(-n\varepsilon h\big(2^{1/2(\lfloor c^*\rfloor+1)}\big)\big) + \exp\big(-n\varepsilon h\big(2^{-1/2(\lfloor c^*\rfloor+1)}\big)\big)\Big)\,.\end{aligned}$$

where

$$C = \frac{\sup_{[-c_4, c_4+\|s\|_\infty]} |\lambda_s'|}{\inf_{[F_s^\dagger(\varepsilon)-c_5, F_s^\dagger(1-\varepsilon)+c_5]} f_s - c_6}(C_4 + c_7 + c_8) + \frac{2\sup_{[-c_4, c_4+\|s\|_\infty]} |\lambda_s''|}{(\inf_{[F_s^\dagger(\varepsilon)-c_5, F_s^\dagger(1-\varepsilon)+c_5]} f_s - c_6)^2}.$$

Proceeding analogously as in Lemma 3, we can choose c_7 and c_8 such that the right-hand side of the previous inequality is of order $O(v_n)$. This finally gives the result of the lemma. □

B.4. Proof of Lemma 5

Proof of Lemma 5. Notice first that thanks to Lemma 10 $y_s(\alpha) \neq 0$ for all $\alpha \in (0,1)$. Let $\alpha \in [\varepsilon, 1-\varepsilon]$,

$$\frac{\widetilde{y}_s(\alpha)}{y_s(\alpha)} - 1 = \frac{\widetilde{y}_s(\alpha) - y_s(\alpha)}{y_s(\alpha)}.$$

As the function $|1/y_s(\cdot)|$ is continuous and therefore bounded over $[\varepsilon, 1-\varepsilon]$, we only need to control $\sup_{\alpha\in[\varepsilon,1-\varepsilon]} |\widetilde{y}_s(\alpha) - y_s(\alpha)|$ which can be bounded by the following decomposition

$$|y_s(\alpha) - \widetilde{y}_s(\alpha)| \leq \left| \frac{\lambda_s'(\alpha_s^{-1}(\alpha))}{f_s(\alpha_s^{-1}(\alpha))} - \frac{\lambda_s'(\widetilde{\alpha}_s^{-1}(\alpha))}{f_s(\widetilde{\alpha}_s^{-1}(\alpha))} \right| + \left| \frac{\lambda_s'(\widetilde{\alpha}_s^{-1}(\alpha))}{f_s(\widetilde{\alpha}_s^{-1}(\alpha))} - \frac{\lambda_s'(\widetilde{\alpha}_s^{-1}(\alpha)}{\widetilde{f}_s(\widetilde{\alpha}_s^{-1}(\alpha))} \right|. \tag{51}$$

As λ_s is of class $\mathcal{C}^2$ and f_s is of class $\mathcal{C}^1$ and strictly positive on (a,b), λ_s'/f_s is of class $\mathcal{C}^1$ on (a,b). Thus using the mean value theorem, we can write, for all $\alpha \in [\varepsilon, 1-\varepsilon]$,

$$\left| \frac{\lambda_s'(\alpha_s^{-1}(\alpha))}{f_s(\alpha_s^{-1}(\alpha))} - \frac{\lambda_s'(\widetilde{\alpha}_s^{-1}(\alpha))}{f_s(\widetilde{\alpha}_s^{-1}(\alpha))} \right| \leq \left| \left(\frac{\lambda_s'}{f_s}\right)'(\xi) \right| \cdot |\alpha_s^{-1}(\alpha) - \widetilde{\alpha}_s^{-1}(\alpha)|$$

where ξ is between $\alpha_s^{-1}(\alpha)$ and $\widetilde{\alpha}_s^{-1}(\alpha)$. From Lemma 14 we know that there exists a constant $c_1 > 0$ such that almost surely, for n large enough, $\widetilde{\alpha}_s^{-1}(\alpha) \in [\alpha_s^{-1}(1-\varepsilon)-c_1, \alpha_s^{-1}(\varepsilon)+c_1] \subset (a,b)$. Therefore, as $(\lambda_s'/f_s)'$ is continuous, almost surely and for n large enough,

$$\left| \left(\frac{\lambda_s'}{f_s}\right)'(\xi) \right| \leq \sup_{t\in[\alpha_s^{-1}(1-\varepsilon)-c_1, \alpha_s^{-1}(\varepsilon)+c_1]} \left| \left(\frac{\lambda_s'}{f_s}\right)'(t) \right| < +\infty.$$

Furthermore,

$$\sup_{\alpha\in[\varepsilon,1-\varepsilon]} |\alpha_s^{-1}(\alpha) - \widetilde{\alpha}_s^{-1}(\alpha)| = \sup_{\alpha'\in[\varepsilon,1-\varepsilon]} |F_s^\dagger(\alpha') - \widetilde{F}_s^\dagger(\alpha')|.$$

Therefore, using Lemma 14, there exists a constant $C_1 > 0$ such that we almost surely have, for n large enough,

$$\left| \frac{\lambda_s'(\alpha_s^{-1}(\alpha))}{f_s(\alpha_s^{-1}(\alpha))} - \frac{\lambda_s'(\widetilde{\alpha}_s^{-1}(\alpha))}{f_s(\widetilde{\alpha}_s^{-1}(\alpha))} \right| \leq C_1 w_n \,. \tag{52}$$

For the second term of (51) we write

$$\left| \frac{\lambda_s'(\widetilde{\alpha}_s^{-1}(\alpha))}{f_s(\widetilde{\alpha}_s^{-1}(\alpha))} - \frac{\lambda_s'(\widetilde{\alpha}_s^{-1}(\alpha))}{\widetilde{f}_s(\widetilde{\alpha}_s^{-1}(\alpha))} \right| = |\lambda_s'(\widetilde{\alpha}_s^{-1}(\alpha))| \left| \frac{1}{f_s(\widetilde{\alpha}_s^{-1}(\alpha))} - \frac{1}{\widetilde{f}_s(\widetilde{\alpha}_s^{-1}(\alpha))} \right|$$

and as shown in the proof of Lemma 13, $\widetilde{F}_s$ has a finite support included in an interval of the form $[-c_2, c_2 + \|s\|_\infty]$. Therefore $\widetilde{\alpha}_s^{-1}(\alpha) \in [-c_2, c_2 + \|s\|_\infty]$ and as λ_s' is continuous, it is bounded on this interval. Therefore there exists $C_2 > 0$ such that for all $\alpha \in [\varepsilon, 1-\varepsilon]$, $|\lambda_s'(\widetilde{\alpha}_s^{-1}(\alpha))| \leq C_2$ and

$$\left| \frac{\lambda_s'(\widetilde{\alpha}_s^{-1}(\alpha))}{f_s(\widetilde{\alpha}_s^{-1}(\alpha))} - \frac{\lambda_s'(\widetilde{\alpha}_s^{-1}(\alpha))}{\widetilde{f}_s(\widetilde{\alpha}_s^{-1}(\alpha))} \right| \leq C_2 \frac{\sup_{x \in \mathbb{R}} |\widetilde{f}_s(x) - f_s(x)|}{\widetilde{f}_s(\widetilde{\alpha}_s^{-1}(\alpha)) f_s(\widetilde{\alpha}_s^{-1}(\alpha))} \,. \tag{53}$$

We can proceed as in the proof of Lemma 4 to show that there exists constants c_3 and c_4 both strictly positive such that $[\alpha_s^{-1}(1-\varepsilon) - c_3, \alpha_s^{-1}(\varepsilon) + c_3] \subset (a, b)$ and such that almost surely, for n large enough, for all $\alpha \in [\varepsilon, 1-\varepsilon]$,

$$\widetilde{f}_s(\widetilde{\alpha}_s^{-1}(\alpha)) \geq \inf_{t \in [\alpha_s^{-1}(1-\varepsilon) - c_3, \alpha_s^{-1}(\varepsilon) + c_3]} f_s(t) - c_4 > 0 \,.$$

We also almost surely have, for n large enough and for all $\alpha \in [\varepsilon, 1-\varepsilon]$,

$$\begin{aligned} f_s(\widetilde{\alpha}_s^{-1}(\alpha)) &\geq \inf_{\alpha' \in [\varepsilon, 1-\varepsilon]} f_s(\widetilde{\alpha}_s^{-1}(\alpha')) = \inf_{t \in [\widetilde{\alpha}_s^{-1}(1-\varepsilon), \widetilde{\alpha}_s^{-1}(\varepsilon)]} f_s(t) \\ &\geq \inf_{t \in [\alpha_s^{-1}(1-\varepsilon) - c_3, \alpha_s^{-1}(\varepsilon) + c_3]} f_s(t) > 0 \,. \end{aligned}$$

Let $A_\varepsilon = [\alpha_s^{-1}(1-\varepsilon) - c_3, \alpha_s^{-1}(\varepsilon) + c_3]$. We almost surely have, for n large enough,

$$\left| \frac{\lambda_s'(\widetilde{\alpha}_s^{-1}(\alpha))}{f_s(\widetilde{\alpha}_s^{-1}(\alpha))} - \frac{\lambda_s'(\widetilde{\alpha}_s^{-1}(\alpha))}{\widetilde{f}_s(\widetilde{\alpha}_s^{-1}(\alpha))} \right| \leq C_2 \frac{\sup_{t \in \mathbb{R}} |\widetilde{f}_s(t) - f_s(t)|}{\inf_{t \in A_\varepsilon} f_s(t) \cdot (\inf_{t \in A_\varepsilon} f_s(t) - c_4)} \,.$$

Thanks to Lemma 11 and combining the last inequality with (52), we obtain the result of the lemma. □

B.5. Other technical results

Lemma 10. *For all $\alpha \in (0,1)$, $y_s(\alpha) \neq 0$.*

Proof of Lemma 10. First notice that as λ_s is of class $\mathcal{C}^2$ and $F_s' = f_s > 0$ on (a,b), $\mathrm{MV}_s = \lambda_s \circ \alpha_s^{-1}$ is differentiable on $(0,1)$ and for all $\alpha \in (0,1)$,

$$\mathrm{MV}_s'(\alpha) = -\frac{\lambda_s(\alpha_s^{-1}(\alpha))}{f_s(\alpha_s^{-1}(\alpha))} = -y_s(\alpha)\,. \tag{54}$$

Now, following the steps of the proof of the formula of the derivative of MV^* we also have, for all $\alpha \in (0,1)$ and $h > 0$,

$$\begin{aligned}
\frac{\mathrm{MV}_s(\alpha+h) - \mathrm{MV}_s(\alpha)}{h} &= \frac{1}{h}(\lambda\{s \geq F_s^{-1}(1-(\alpha+h))\} - \lambda\{s \geq F_s^{-1}(1-\alpha)\}) \\
&= \frac{1}{h}\lambda\{F_s^{-1}(1-(\alpha+h)) \leq s \leq F_s^{-1}(1-\alpha)\} \\
&= \frac{1}{h}\lambda\{1-(\alpha+h) \leq F_s \circ s \leq 1-\alpha\} \\
&= \frac{1}{h}\int_{\mathcal{X}} \mathbb{I}\{x, 1-(\alpha+h) \leq F_s(s(x)) \leq 1-\alpha\}\frac{f(x)}{f(x)}dx \\
&\geq \frac{1}{h\|f\|_\infty}\mathbb{P}(F_s(s(X)) \in [1-(\alpha+h), 1-\alpha]) \\
&= \frac{1}{\|f\|_\infty}\,.
\end{aligned}$$

where we used the fact that the distribution of $F_s(s(X))$ is the uniform distribution on $(0,1)$. Similarly,

$$\forall \alpha \in (0,1), \quad \frac{\mathrm{MV}_s(\alpha-h) - \mathrm{MV}_s(\alpha)}{-h} \geq \frac{1}{\|f\|_\infty}\,.$$

Therefore, for all $\alpha \in (0,1)$, $\mathrm{MV}_s'(\alpha) \geq 1/\|f\|_\infty > 0$. Using (54), this implies that $y_s(\alpha) \neq 0$ for all $\alpha \in (0,1)$. □

Lemma 11. *Under assumptions **(A3)** and **(B1)–(B4)** there exists a constant $C > 0$, depending only on f_s and the kernel K such that we almost surely have, for n large enough,*

$$\sup_{t\in\mathbb{R}} \left|\widetilde{f}_s(t) - f_s(t)\right| \leq C w_n\,.$$

Note that this result does not require K to have a finite support.

Proof of Lemma 11. We first bound $\sup_{t\in\mathbb{R}} |\widetilde{f}_s(t) - f_s(t)|$ by the sum of a stochastic term and a deterministic term:

$$\sup_{t\in\mathbb{R}} \left|\widetilde{f}_s(t) - f_s(t)\right| \leq \sup_{t\in\mathbb{R}} \left|\widetilde{f}_s(t) - \mathbb{E}[\widetilde{f}_s(t)]\right| + \sup_{t\in\mathbb{R}} \left|\mathbb{E}[\widetilde{f}_s(t)] - f_s(t)\right|\,.$$

Under the stipulated conditions, the first term is controlled thanks to Theorem 2.3 in (Giné and Guillou, 2002): we almost surely have, for n large enough,

$$\sup_{t\in\mathbb{R}} \left|\widetilde{f}_s(t) - \mathbb{E}[\widetilde{f}_s(t)]\right| \leq M^2\|f_s\|_\infty\|K\|_2^2\sqrt{\frac{\log h_n^{-1}}{nh_n}}$$

where M is a constant that depends on the VC characteristics of K.

It can then be shown with classical calculations that the second term is bounded by

$$\sup_{t\in\mathbb{R}}\left|\mathbb{E}[\widetilde{f}_s(t)]-f_s(t)\right|\leq\frac{h_n^2}{2}\|f_s''\|_\infty\int_{-\infty}^{+\infty}y^2K(y)\,.$$

Combining this last inequality with the one on the first term we obtain the result of the lemma. □

Lemma 12. *Under assumptions* ***(A3)***, ***(B1)–(B3)*** *and* ***(B5)*** *there exists a constant $C>0$, depending only on f_s and the kernel K such that we almost surely have, for n large enough,*

$$\sup_{t\in\mathbb{R}}\left|\widetilde{f}_s{}'(t)-f_s'(t)\right|\leq C\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^3}}+h_n^2\right)\,.$$

Therefore, as $h_n\to 0$, if assumption ***(B6)*** *is also fulfilled then we almost surely have,*

$$\sup_{t\in\mathbb{R}}\left|\widetilde{f}_s{}'(t)-f_s'(t)\right|\underset{n\to\infty}{\longrightarrow}0\,.$$

Note that this result does not require K to have a finite support.

Proof of Lemma 12. We proceed as in the proof of Lemma 11. We first bound $\sup_{t\in\mathbb{R}}|\widetilde{f}_s{}'(t)-f_s'(t)|$ by the sum of a stochastic term and a deterministic term:

$$\sup_{t\in\mathbb{R}}\left|\widetilde{f}_s{}'(t)-f_s'(t)\right|\leq\sup_{t\in\mathbb{R}}\left|\widetilde{f}_s{}'(t)-\mathbb{E}[\widetilde{f}_s{}'(t)]\right|+\sup_{t\in\mathbb{R}}\left|\mathbb{E}[\widetilde{f}_s{}'(t)]-f_s'(t)\right|\,.$$

We first deal with the first term. Let $t\in\mathbb{R}$,

$$\begin{aligned}&\widetilde{f}_s{}'(t)-\mathbb{E}[\widetilde{f}_s'(t)]\\&\quad=\frac{1}{h_n}\left(\frac{1}{nh_n}\sum_{i=1}^n K'\left(\frac{t-s(X_i)}{h_n}\right)-\mathbb{E}\left[\frac{1}{nh_n}\sum_{i=1}^n K'\left(\frac{t-s(X_i)}{h_n}\right)\right]\right)\,.\end{aligned}$$

Therefore K' satisfying the conditions required for the result of (Giné and Guillou, 2002) we almost surely have, for n large enough,

$$\sup_{t\in\mathbb{R}}\left|\widetilde{f}_s{}'(t)-\mathbb{E}[\widetilde{f}_s'(t)]\right|\leq M'^2\|f_s\|_\infty\|K'\|_2^2\sqrt{\frac{\log h_n^{-1}}{nh_n^3}}$$

where M' is a constant that depends on the VC characteristics of K'.

We now deal with the second term as follows. Let $t\in\mathbb{R}$,

$$\begin{aligned}\mathbb{E}[\widetilde{f}_s{}'(t)]&=\frac{1}{nh_n^2}\sum_{i=1}^n\mathbb{E}\left[K'\left(\frac{t-s(X_i)}{h_n}\right)\right]=\frac{1}{h_n^2}\mathbb{E}\left[K'\left(\frac{t-s(X_1)}{h_n}\right)\right]\\&=\frac{1}{h_n}\int_{-\infty}^{+\infty}\frac{1}{h_n}K'\left(\frac{t-u}{h_n}\right)f_s(u)du\,.\end{aligned}$$

With an integration by parts we obtain,

$$\mathbb{E}[\widetilde{f}_s{}'(t)] - f_s'(t) = \frac{1}{h_n}\int_{-\infty}^{+\infty} K\left(\frac{t-u}{h_n}\right) f_s'(u)du - f_s'(t)$$

and we can deal with this term as we dealt with $\mathbb{E}[\widetilde{f}_s(t)] - f_s(t)$ in the proof of Lemma 11, to obtain

$$\sup_{t\in\mathbb{R}} \left|\mathbb{E}[\widetilde{f}_s{}'(t)] - f_s'(t)\right| \leq \frac{h_n^2}{2}\|f_s'''\|_\infty \int_{-\infty}^{+\infty} y^2 K(y)\,.$$

Combining this last inequality with the one on the first term we obtain the result of the lemma. □

Lemma 13. *Under the conditions of Lemma 11 there exists a constant $C > 0$ such that we almost surely have for n large enough,*

$$\sup_{t\in\mathbb{R}} \left|\widetilde{F}_s(t) - F_s(t)\right| \leq Cw_n\,.$$

Proof of Lemma 13. Let $t \in \mathbb{R}$,

$$\left|F_s(t) - \widetilde{F}_s(t)\right| = \left|\int_{-\infty}^{t} f_s(u)du - \int_{-\infty}^{t} \widetilde{f}_s(u)du\right| \leq \int_{-\infty}^{t} |f_s(u) - \widetilde{f}_s(u)|du\,.$$

As f_s has a support included in $[0, \|s\|_\infty]$, as K has a finite support and as the sequence $(h_n)_{n\geq 1}$ decreases towards 0, $\widetilde{f}_s$ also has a finite support included in an interval of the form $[-c_1, \|s\|_\infty + c_1]$ where $c_1 > 0$. Thus,

$$\int_{-\infty}^{t} |f_s(u) - \widetilde{f}_s(u)|du \leq (2c_1 + \|s\|_\infty) \sup_{t\in\mathbb{R}} \left|f_s(t) - \widetilde{f}_s(t)\right|\,.$$

Thanks to Lemma 11, we finally obtain the result. □

Lemma 14. *Let $\varepsilon \in (0,1)$. Under the conditions of Lemma 11, there exists a constant $C > 0$ such that we almost surely have for n large enough,*

$$\sup_{y\in[\varepsilon,1-\varepsilon]} \left|\widetilde{F}_s^\dagger(y) - F_s^\dagger(y)\right| \leq Cw_n\,.$$

This lemma only requires $\widetilde{F}_s$ to be continuous.

Proof of Lemma 14. First we recall that $\widetilde{F}_s$ being continuous, for all $y \in (0,1)$, $F_s(F_s^\dagger(y)) = y$. Let $y \in [\varepsilon, 1-\varepsilon]$,

$$\begin{aligned}|\widetilde{F}_s(\widetilde{F}_s^\dagger(y)) - F_s(\widetilde{F}_s^\dagger(y))| &= |y - F_s(\widetilde{F}_s^\dagger(y))| = |F_s(F_s^\dagger(y)) - F_s(\widetilde{F}_s^\dagger(y))| \\ &= |f_s(\xi)| \cdot |F_s^\dagger(y) - \widetilde{F}_s^\dagger(y)|\end{aligned}$$

with ξ between $F_s^\dagger(y)$ and $\widetilde{F}_s^\dagger(y)$. As $F_s^\dagger$ and $\widetilde{F}_s^\dagger$ are increasing, $F_s^\dagger(\varepsilon) \leq F_s^\dagger(y) \leq F_s^\dagger(1-\varepsilon)$ and $\widetilde{F}_s^\dagger(\varepsilon) \leq \widetilde{F}_s^\dagger(y) \leq \widetilde{F}_s^\dagger(1-\varepsilon)$. Now

$$|\varepsilon - F_s(\widetilde{F}_s^{\dagger}(\varepsilon))| = |\widetilde{F}_s(\widetilde{F}_s^{\dagger}(\varepsilon)) - F_s(\widetilde{F}_s^{\dagger}(\varepsilon))| \leq \sup_{t\in\mathbb{R}} |\widetilde{F}_s(t) - F_s(t)|\,.$$

Therefore from Lemma 13 as $w_n \to 0$ when n tends to $+\infty$, $F_s(\widetilde{F}_s^{\dagger}(\varepsilon))$ converges almost surely to ε. And as $F_s^{\dagger}$ is continuous, we almost surely have, $F_s^{\dagger}(F_s(\widetilde{F}_s^{\dagger}(\varepsilon))) \underset{n\to\infty}{\longrightarrow} F_s^{\dagger}(\varepsilon)$, i.e., $\widetilde{F}_s^{\dagger}(\varepsilon) \underset{n\to\infty}{\longrightarrow} F_s^{\dagger}(\varepsilon)$. Similarly, we almost surely have $\widetilde{F}_s^{\dagger}(1-\varepsilon) \underset{n\to\infty}{\longrightarrow} F_s^{\dagger}(1-\varepsilon)$. Therefore there exists a constant $c_1 > 0$ such that we almost surely have, for n large enough, $\widetilde{F}_s^{\dagger}(y) \in [F_s^{\dagger}(\varepsilon) - c_1, F_s^{\dagger}(1-\varepsilon) + c_1] \subset (a, b)$. This implies that ξ is between $F_s^{\dagger}(\varepsilon) - c_1 > a$ and $F_s^{\dagger}(1-\varepsilon) + c_1 < b$ and thus,

$$f_s(\xi) \geq \inf_{t\in[F_s^{\dagger}(\varepsilon)-c_1, F_s^{\dagger}(1-\varepsilon)+c_1]} f_s(t)\,.$$

We thus almost surely have, for n large enough,

$$|\widetilde{F}_s(\widetilde{F}_s^{\dagger}(y)) - F_s(\widetilde{F}_s^{\dagger}(y))| \geq |F_s^{\dagger}(y) - \widetilde{F}_s^{\dagger}(y)| \inf_{t\in[F_s^{\dagger}(\varepsilon)-c_1, F_s^{\dagger}(1-\varepsilon)+c_1]} f_s(t)$$

Now as f_s is strictly positive on (a, b) and continuous, the infimum of the right-hand side is strictly positive. We can thus write almost surely and for n large enough,

$$\begin{aligned} |F_s^{\dagger}(y) - \widetilde{F}_s^{\dagger}(y)| &\leq \frac{1}{\inf_{t\in[F_s^{\dagger}(\varepsilon)-c_1, F_s^{\dagger}(1-\varepsilon)+c_1]} f_s(t)} |\widetilde{F}_s(\widetilde{F}_s^{\dagger}(y)) - F_s(\widetilde{F}_s^{\dagger}(y))| \\ &\leq \frac{C}{\inf_{t\in[F_s^{\dagger}(\varepsilon)-c_1, F_s^{\dagger}(1-\varepsilon)+c_1]} f_s(t)} w_n \end{aligned}$$

where we used Lemma 13 for the last inequality. This concludes the proof. □

Lemma 15. *Let $a', b' \in \mathbb{R}$ with $a < a' < b' < b$. Under assumptions **(A3)**–**(A4)** and **(B1)**–**(B6)**,*

$$\sup_{t\in[a',b']} \left| \widetilde{F}_s(t)(1-\widetilde{F}_s(t)) \frac{|\widetilde{f}_s{}'(t)|}{\widetilde{f}_s{}^2(t)} - F_s(t)(1-F_s(t)) \frac{|f_s'(t)|}{f_s^2(t)} \right| \underset{n\to+\infty}{\longrightarrow} 0$$

Proof of Lemma 15. We have for all $t \in [a', b']$,

$$\begin{aligned} &\left| \widetilde{F}_s(t)(1-\widetilde{F}_s(t)) \frac{|\widetilde{f}_s{}'(t)|}{\widetilde{f}_s{}^2(t)} - F_s(t)(1-F_s(t)) \frac{|f_s'(t)|}{f_s^2(t)} \right| \\ &\quad \leq |\widetilde{F}_s(t)(1-\widetilde{F}_s(t))| \left| \frac{|\widetilde{f}_s{}'(t)|}{\widetilde{f}_s{}^2(t)} - \frac{|f_s'(t)|}{f_s^2(t)} \right| \\ &\qquad + \left| \frac{\widetilde{f}_s{}'(t)}{\widetilde{f}_s{}^2(t)} \right| |\widetilde{F}_s(t)(1-\widetilde{F}_s(t)) - F_s(t)(1-F_s(t))| \\ &\quad \leq \left| \frac{|\widetilde{f}_s{}'(t)|}{\widetilde{f}_s{}^2(t)} - \frac{|f_s'(t)|}{f_s^2(t)} \right| + \left| \frac{\widetilde{f}_s{}'(t)}{\widetilde{f}_s{}^2(t)} \right| |\widetilde{F}_s(t)(1-\widetilde{F}_s(t)) - F_s(t)(1-F_s(t))| \end{aligned}$$

where the third inequality holds because for all t, $\widetilde{F}_s(t) \in [0,1]$. The result can then be derived thanks to Lemma 11, Lemma 12 and Lemma 13. □

Lemma 16. *Let $\mathcal{D}_n$ be such that $\sup_{t\in\mathbb{R}} |\widetilde{f}_s(t) - f_s(t)| \leq C_1 w_n$, $\sup_{\alpha\in[\varepsilon,1-\varepsilon]} |\widetilde{F_s}^{\dagger}(\alpha) - F_s^{\dagger}(\alpha)| \leq C_2 w_n$ and $\sup_{t\in\mathbb{R}} |\widetilde{f}_s{}'(t) - f_s'(t)|$ converges to 0 as n tends to ∞. There exists a constant $c^* > 0$ such that at $\mathcal{D}_n$ fixed, for all $z_4 \geq 1$ and for n large enough,*

$$\mathbb{P}^*\left(\sup_{\alpha\in[\varepsilon,1-\varepsilon]}\left|\frac{\widetilde{f}_s(\widetilde{F}_s^{\dagger}(\alpha))}{\widetilde{f}_s(\widetilde{F}_s^{\dagger}(\xi))}\right| > z_4\right) \leq \frac{2}{n^2} + 2(\lfloor c^*\rfloor+1)\Big(\exp\big(-n\varepsilon h\big(z_4^{1/2(\lfloor c^*\rfloor+1)}\big)\big) + \exp\big(-n\varepsilon h\big(z_4^{-1/2(\lfloor c^*\rfloor+1)}\big)\big)\Big)\,.$$

where ξ is between $\widetilde{U}^{\dagger}(\alpha)$ and α with $\alpha \in [\varepsilon, 1-\varepsilon]$.

Proof of Lemma 16. From the DKW inequality (see proof of Lemma 2), we have for all n,

$$\mathbb{P}^*\left(\sup_{\alpha\in[0,1]} |\widetilde{U}^{\dagger}(\alpha) - \alpha| \geq \sqrt{\frac{\log n}{n}}\right) \leq \frac{2}{n^2}\,.$$

Therefore, if $\omega' \in \{\sup_{\alpha\in[0,1]} |\widetilde{U}^{\dagger}(\alpha) - \alpha| < \sqrt{\log n/n}\}$, we have,

$$|\widetilde{U}^{\dagger}(\varepsilon) - \varepsilon| \leq \sup_{\alpha\in[0,1]} |\widetilde{U}^{\dagger}(\alpha) - \alpha| < \sqrt{\log n/n}$$

and similarly,

$$|\widetilde{U}^{\dagger}(1-\varepsilon) - (1-\varepsilon)| \leq \sup_{\alpha\in[0,1]} |\widetilde{U}^{\dagger}(\alpha) - \alpha| < \sqrt{\log n/n}\,.$$

Thus there exists a constant c_1 such that for n large enough (independent of ω'), $\widetilde{U}^{\dagger}(\varepsilon) \geq \varepsilon - c_1 > 0$ and $\widetilde{U}^{\dagger}(1-\varepsilon) \leq 1-\varepsilon+c_1 < 1$. Now as $\alpha \in [\varepsilon, 1-\varepsilon]$, this implies that for n large enough, $\xi \in [\varepsilon - c_1, 1-\varepsilon+c_1]$. Eventually, for n large enough, $\omega' \in \{\xi \in [\varepsilon - c_1, 1-\varepsilon+c_1]\}$. Let $\mathcal{Z}_1 = \{\xi \notin [\varepsilon - c_1, 1-\varepsilon+c_1]\}$. We thus have for n large enough, $\mathbb{P}^*(\mathcal{Z}_1) \leq \mathbb{P}^*(\sup_{\alpha\in[0,1]} |u_n^*(\alpha)| \geq \sqrt{\log n}) \leq 2/n^2$.

Now, from Lemma 14 and Lemma 15 there exist positive constants c_2 and C_3 such that for n high enough,

$$\begin{aligned}\sup_{\alpha\in[\varepsilon-c_1,1-\varepsilon+c_1]} & \alpha(1-\alpha)\frac{|\widetilde{f}_s{}'(\widetilde{F}_s^{\dagger}(\alpha))}{\widetilde{f}_s^2(\widetilde{F}_s^{\dagger}(\alpha))} \\ &\leq \sup_{t\in[F_s^{\dagger}(\varepsilon-c_1)-c_2,F_s^{\dagger}(1-\varepsilon+c_1)+c_2]} F_s(t)(1-F_s(t))\frac{|f_s'(t)|}{f_s^2(t)} + C_3 \\ &\leq c + C_3\end{aligned}$$

where c_2 is such that

$$[\widetilde{F}_s^{\dagger}(\varepsilon - c_1), \widetilde{F}_s^{\dagger}(1-\varepsilon+c_1)] \subset [F_s^{\dagger}(\varepsilon-c_1) - c_2, F_s^{\dagger}(1-\varepsilon+c_1) + c_2] \subset (0, \|s\|_{\infty})\,.$$

Let $c^* = c + C_3$ and let $\alpha_1, \alpha_2 \in [\varepsilon - c_1, 1 - \varepsilon + c_1]$. Following the proof of Lemma 1.4.1 in (Csörgő, 1983), we have for n high enough,

$$\frac{\widetilde{f}_s(\widetilde{F}_s^\dagger(\alpha_1))}{\widetilde{f}_s(\widetilde{F}_s^\dagger(\alpha_2))} \leq \left(\frac{\max(\alpha_1, \alpha_2)}{\min(\alpha_1, \alpha_2)} \frac{1 - \min(\alpha_1, \alpha_2)}{1 - \max(\alpha_1, \alpha_2)} \right)^{c^*} .$$

Let $z_4 \geq 1$ et let $\mathcal{Z}_2$ be the following event

$$\mathcal{Z}_2 = \left\{ \sup_{\alpha \in [\varepsilon, 1-\varepsilon]} \left(\frac{\max(\widetilde{U}^\dagger(\alpha), \alpha)}{\min(\widetilde{U}^\dagger(\alpha), \alpha)} \frac{1 - \min(\widetilde{U}^\dagger(\alpha), \alpha)}{1 - \max(\widetilde{U}^\dagger(\alpha), \alpha)} \right)^{c^*} \geq z_4 \right\} .$$

Let $\omega' \in \overline{\mathcal{Z}_1} \cap \overline{\mathcal{Z}_2}$. We therefore have

$$\sup_{\alpha \in [\varepsilon, 1-\varepsilon]} \frac{\widetilde{f}_s(\widetilde{F}_s^\dagger(\alpha))}{\widetilde{f}_s(\widetilde{F}_s^\dagger(\xi))} \leq \sup_{\alpha \in [\varepsilon, 1-\varepsilon]} \left(\frac{\max(\widetilde{U}^\dagger(\alpha), \alpha)}{\min(\widetilde{U}^\dagger(\alpha), \alpha)} \frac{1 - \min(\widetilde{U}^\dagger(\alpha), \alpha)}{1 - \max(\widetilde{U}^\dagger(\alpha), \alpha)} \right)^{c^*} \leq z_4 .$$

Therefore $\omega' \in \overline{\mathcal{Z}_3}$ defined as the complement of

$$\mathcal{Z}_3 = \left\{ \sup_{\alpha \in [\varepsilon, 1-\varepsilon]} \frac{\widetilde{f}_s(\widetilde{F}_s^\dagger(\alpha))}{\widetilde{f}_s(\widetilde{F}_s^\dagger(\xi))} \geq z_4 \right\} .$$

Thus $\mathbb{P}^*(\overline{\mathcal{Z}_1} \cap \overline{\mathcal{Z}_2}) \leq \mathbb{P}^*(\overline{\mathcal{Z}_3})$ which implies that for n large enough,

$$\mathbb{P}^* \left(\sup_{\alpha \in [\varepsilon, 1-\varepsilon]} \frac{\widetilde{f}_s(\widetilde{F}_s^\dagger(\alpha))}{\widetilde{f}_s(\widetilde{F}_s^\dagger(\xi))} \geq z_4 \right) = \mathbb{P}^*(\mathcal{Z}_3) \leq \mathbb{P}^*(\mathcal{Z}_1) + \mathbb{P}^*(\mathcal{Z}_2) \leq \frac{2}{n^2} + \mathbb{P}^*(\mathcal{Z}_2) .$$

Now from the proof of Theorem 1.4.3 in (Csörgő, 1983) we can bound $\mathbb{P}^*(\mathcal{S}_2)$ and obtain the result of the lemma. □

References

Cadre, B. (2006). Kernel Estimation Of Density Level Sets. *Journal of Multivariate Analysis* **97** 999–1023. MR2256570

Cadre, B., Pelletier, B. and Pudlo, P. (2013). Estimation of density level sets with a given probability content. *Journal of Nonparametric Statistics* **25** 261–272. MR3039981

Cavalier, L. (1997). Nonparametric Estimation of Regression Level Sets. *Statistics* **29** 131–160. MR1484386

Clémençon, S. and Jakubowicz, J. (2013). Scoring Anomalies: a M-estimation formulation. In *Proceedings of the 16-th International Conference on Artificial Intelligence and Statistics, Scottsdale, USA*.

Clémençon, S. and Robbiano, S. (2014). Anomaly ranking as supervised bipartite ranking. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China* 343–351.

Clémençon, S. and Vayatis, N. (2009). Adaptive Estimation of the Optimal ROC Curve and a Bipartite Ranking Algorithm. In *Algorithmic Learning Theory. Lecture Notes in Computer Science* **5809** 216–231. Springer Berlin Heidelberg. MR2564229

Csörgő, M. (1983). *Quantile Processes with Statistical Applications*. Society for Industrial and Applied Mathematics. MR0745130

Csörgő, M. and Révész, P. (1978). Strong Approximations of the Quantile Process. *The Annals of Statistics* **6** 882–894. MR0501290

Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press. MR0666546

DeVore, R. (1987). A note on adaptive approximation. *Approx. Theory Appl.* **3** 74–78. MR0939183

DeVore, R. A. (1998). Nonlinear Approximation. *Acta Numerica* **7** 51–150. MR1689432

Donoho, D. and Gasko, M. (1992). Breakdown properties of location estimates based on half space depth and projected outlyingness. *The Annals of Statistics* **20** 1803–1827. MR1193313

Efron, B. (1979). Bootstrap methods: another look at the jacknife. *Annals of Statistics* **7** 1–26. MR0515681

Egan, J. P. (1975). *Signal Detection Theory and ROC Analysis*. Academic Press.

Einmahl, J. H. J. and Mason, D. M. (1992). Generalized Quantile Processes. *The Annals of Statistics* **20** 1062–1078. MR1165606

Embrechts, P. and Hofert, M. (2013). A note on generalized inverses. *Mathematical Methods of Operations Research* **77** 423–432. MR3072795

Falk, M. and Reiss, R. (1989). Weak convergence of smoothed and nonsmoothed bootstrap quantile estimates. *Annals of Probability* **17** 362–371. MR0972792

Giné, E. and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. Poincaré (B), Probabilités et Statistiques* **38** 907–921. MR1955344

Hall, P. (1986). On the number of bootstrap simulations required to construct a confidence interval. *Annals of Statistics* **14** 1453–1462. MR0868311

Koltchinskii, V. (1997). M-estimation, convexity and quantiles. *The Annals of Statistics* **25** 435–477. MR1439309

Koltchinskii, V. (2006). Local Rademacher complexities and oracle inequalities in risk minimization (with discussion). *The Annals of Statistics* **34** 2593–2706. MR2329463

Lifshits, M. A. (1987). On the distribution of the maximum of a Gaussian process. *Theory of Probability and its Applications* **31** 125–132. MR0836962

Liu, R. Y., Parelius, J. M. and Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference. *Ann. Statist.* **27** 783–858. MR1724033

Lovász, L. and Vempala, S. (2006). Simulated annealing in convex bodies and an $O(n^4)$ volume algorithm. *Journal of Computer and System Sciences* **72** 392–417. MR2205290

Mallat, S. (1990). *A Wavelet Tour of Signal Processing*. Academic Press. MR1614527

Massart, P. (1990). The Tight Constant in the Dvoretzky-Kiefer-Wolfowitz Inequality. *Ann. Probab.* **18** 1269–1283. MR1062069

Muller, D. W. and Sawitzki, G. (1991). Excess Mass Estimates and Tests for Multimodality. *Journal of the American Statistical Association* **86** 738–746. MR1147099

Pitt, L. D. and Tran, L. T. (1979). Local Sample Path Properties of Gaussian Fields. *Ann. Probab.* **7** 477–493. MR0528325

Polonik, W. (1995). Measuring Mass Concentrations And Estimating Density Contour Clusters – An Excess Mass Approach. *The Annals of Statistics* **23** 855–881. MR1345204

Polonik, W. (1997). Minimum volume sets and generalized quantile processes. *Stochastic Processes and their Applications* **69** 1–24. MR1464172

Polonik, W. (1999). Concentration and goodness-of-fit in higher dimensions: (asymptotically) distribution-free methods. *The Annals of Statistics* **27** 1210–1229. MR1740114

Rigollet, P. and Vert, R. (2009). Fast rates for plug-in estimators of density level sets. *Bernoulli* **14** 1154–1178. MR2597587

Sargan, J. D. and Mikhail, W. M. (1971). A General Approximation to the Distribution of Instrumental Variables Estimates. *Econometrica* **39** 131–169. MR0278439

Scott, C. and Nowak, R. (2006). Learning Minimum Volume Sets. *Journal of Machine Learning Research* **7** 665–704. MR2274383

Silverman, B. and Young, G. (1987). The bootstrap: to smooth or not to smooth. *Biometrika* **7** 469–479. MR0909352

Steinwart, I., Hush, D. and Scovel, C. (2005). A classification framework for anomaly detection. *J. Machine Learning Research* **6** 211–232. MR2249820

Stute, W. (1982). A Law of the Logarithm for Kernel Density Estimators. *The Annals of Probability* **10** 414–422. MR0647513

Tsirel'son, V. S. (1976). The Density of the Distribution of the Maximum of a Gaussian Process. *Theory of Probability & Its Applications* **20** 847–856.

Tsybakov, A. (1997). On nonparametric estimation of density level sets. *Annals of Statistics* **25** 948–969. MR1447735

Tukey, J. (1975). Mathematics and picturing data. (R. D. James, ed.) 523–531. Canadian Math. Congress. MR0426989

Viswanathan, K., Choudur, L., Talwar, V., Wang, C., Macdonald, G. and Satterfield, W. (2012). Ranking Anomalies in Data Centers. In *Network Operations and System Management* (R. D. James, ed.) 79–87. IEEE.

Wand, M. P. and Jones, M. C. (1994). *Kernel Smoothing. Chapman & Hall/CRC Monographs on Statistics & Applied Probability*. Taylor & Francis. MR1319818

Zuo, B. Y. and Serfling, R. (2000). General notions of statistical depth function. *The Annals of Statistics* **28** 461–482. MR1790005