



**HAL**  
open science

# A Donsker and Glivenko-Cantelli theorem for random measures linked to extreme value theory

B Bobbia, C Dombry, Davit Varron

► **To cite this version:**

B Bobbia, C Dombry, Davit Varron. A Donsker and Glivenko-Cantelli theorem for random measures linked to extreme value theory. 2021. hal-03402380

**HAL Id: hal-03402380**

**<https://hal.science/hal-03402380>**

Preprint submitted on 25 Oct 2021

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

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

# A Donsker and Glivenko–Cantelli theorem for random measures linked to extreme value theory

B. Bobbia, C. Dombry, D. Varron

October 25, 2021

**Abstract:** We consider a class of random point measures that share properties with empirical measures when conditioned to another exogenous random phenomenon. We investigate the validity of some Glivenko-Cantelli and Donsker theorems for such random measures. In this setup, we prove that the usual conditions on uniform entropy numbers are strong enough to derive these two theorems. A bootstrap Donsker theorem is also proved. Some applications of these results are also presented in the framework of extreme value theory and nearest-neighbor rules.

## 1 Introduction and results

Let  $(\mathfrak{X}, \mathcal{X})$  be a measurable space and let  $\mathfrak{M}$  be the set of all probability measures on  $(\mathfrak{X}, \mathcal{X})$ . Endow  $\mathfrak{M}$  with the  $\sigma$ -algebra

$$\mathcal{M} := \sigma\left(\{Q \rightarrow Q(f), f \text{ real valued, bounded, Borel measurable on } (\mathfrak{X}, \mathcal{X})\}\right),$$

where we use the generic notation

$$Q(f) := \int_{\mathfrak{X}} f dQ.$$

We shall also use that notation for random probability measures, i.e.  $\mathcal{M}$ -measurable random variables. Now let  $\mathcal{F}$  be a class of real Borel functions on  $(\mathfrak{X}, \mathcal{X})$ . Empirical processes indexed by  $\mathcal{F}$  have been intensively investigated during the past decades. This theory is now well known for independent, identically distributed (i.i.d.) samples. One of the main strengths of empirical processes is that several statistics can be expressed as a function  $\Phi$  of the empirical measure, where  $\Phi$  is explicit and sufficiently regular (continuous or Hadamard differentiable) on a suitable subset of the space  $\ell^\infty(\mathcal{F})$  of bounded, real valued,  $\mathcal{F}$ -indexed trajectories. However, several statistical problems involve statistics

that are not based on all of the  $n$  observations, but rather on a subsample of observations pertaining to a set of small probability. In extreme value theory, one takes into account only the observations exceeding a high threshold, while regression/density estimation at point  $x \in \mathbb{R}^d$  focuses on observations falling in a neighborhood of  $x$ . This class of empirical processes - commonly called local empirical processes - has been intensively investigated in the literature: see [10, 11, 12, 24, 13, 14, 15, 25, 26, 9, 16, 31, 41, 40, 42]. All of these can be expressed as the centered and rescaled versions of

$$\mathbb{P}_n^{loc} := \frac{1}{\sum_{i=1}^n \mathbb{1}_{A_n}(X_i)} \sum_{i=1}^n \mathbb{1}_{A_n}(X_i) \delta_{X_i}, \quad (1)$$

(with convention  $0/0 := 0$ ), where the  $X_i$  are i.i.d and where  $A_n$  is a sequence of (measurable) sets fulfilling  $p_n := \mathbb{P}(X_1 \in A_n) \rightarrow 0$ . Einmahl and Mason [26] noticed the following representation - as equalities in laws on  $(\mathfrak{M}, \mathcal{M})$  - which is a direct consequence of Khinchin's equality (see, e.g., [33, p. 307, assertion (14.6)])

$$\mathbb{P}_n^{loc} \stackrel{d}{=} \frac{1}{N_n} \sum_{i=1}^{N_n} \delta_{X_{i,n}}. \quad (2)$$

Here  $N_n$  is binomial  $(n, p_n)$ , independent of the i.i.d. sample  $(X_{i,n})_{i \in [1, n]}$ , which itself has common law  $\mathbf{P}_{0,n} := \mathbb{P}(X_1 \in \cdot \mid X_1 \in A_n)$ . Using (2), Einmahl and Mason could derive several asymptotic results by borrowing arguments from Sheehy and Wellner [35]. These latter had previously studied the case of empirical processes where the sample law changes with  $n$ . In all these cited works, the notion of *uniform entropy numbers* always interacted very smoothly with these modified versions of the empirical measure.

In practice, however, the set  $A_n$  is not deterministic but data-driven, which undermines any direct utility of  $\mathbb{P}_n^{loc}$ . In extreme value theory, the threshold is often taken  $k_n$ -largest order statistic, while in regression/density estimation, the bandwidth can be chosen by a plug-in or a nearest neighbor rule. So far, the best way to tackle the additional randomness brought by a data-driven  $A_n$  has been to prove asymptotic results for local empirical processes that hold "uniformly in bandwidth" (see, e.g., [6, 7, 17, 19, 20, 21, 28, 29, 31, 37, 38] and the references therein): the asymptotic random fluctuations of  $A_n$  are handled by incorporating all those possible fluctuations into a class of functions  $\tilde{\mathcal{F}} \supset \mathcal{F}$  that is still small enough to derive asymptotic results. For example,  $\tilde{\mathcal{F}}$  can be taken as all the elements  $f \circ \rho$ , where  $f \in \mathcal{F}$  and  $\rho \in \mathcal{I}$ , given a class  $\mathcal{I}$  of small perturbations of the identity (see, e.g., the set of conditions (G) and (F) in [32, p. 1394], the set of assumptions (F) in [20, p. 108], or the set of assumptions (HK) in [38, p. 1345]). While this method did prove to have practical applications, it may seem unsatisfactory to require structural conditions upon a overclass  $\tilde{\mathcal{F}}$  rather than  $\mathcal{F}$  itself. In this article, we exhibit a class of random measures which may be more accurately tailored to fit in the framework of data driven bandwidth/threshold selection. Roughly speaking, we make the assumption that, for fixed  $n$ , the conditional laws of  $\mathbb{P}_n^{loc}$  (implicitly on  $(\mathfrak{M}, \mathcal{M})$ ) given another random phenomenon  $\mathbf{U}_n$  are those of classical empirical measures. We

then assume that, as  $n \rightarrow \infty$ , the randomness of  $\mathbf{U}_n$  is asymptotically controlled. As an example,  $\mathbf{U}_n$  may capture the randomness of  $A_n$  in (1) when the latter is data driven, and we may assume that  $\mathbf{U}_n$  asymptotically behaves deterministically. We will show that, for this class of random measures, both a Glivenko-Cantelli and a Donsker theorem hold under classical conditions upon uniform entropy numbers that rise from classic abstract empirical theory. Then we make use of those results for statistical applications. The next subsection is devoted to a rigorous description of our model.

## 1.1 The probabilistic model

Denote by  $\mathfrak{M}_0 \in \mathcal{M}$  the set of all probability measures with finite support. Let  $(E, d)$  be a separable metric space endowed with its Borel  $\sigma$ -algebra. Now consider a  $E$ -indexed family of probability measures  $\{\mathbf{P}_u, u \in E\}$  on  $(\mathfrak{X}, \mathcal{X})$  defining a Markov transition kernel, i.e: for each bounded Borel function  $f$  the map  $u \rightarrow \mathbf{P}_u(f)$  is Borel. For each  $n \geq 1$ , we shall consider a couple of random variables  $(\mathbb{P}_n^{(n)}, \mathbf{U}_n)$  from a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  to  $\mathfrak{M}_0 \times E$  for which we make the following assumption:

### Conditional empirical measure (CEM) assumption

There is a sequence of integers  $(k_n)_{n \geq 1}$  such that, for all  $n \geq 1$  and  $\mathbb{P}$ -almost all  $u \in E$ , the law of  $\mathbb{P}_n^{(n)}$  given  $\mathbf{U}_n = u$  is equal to that of the empirical measure

$$\mathbf{P}_{k_n}^{\mathbf{P}_u} := \frac{1}{k_n} \sum_{i=1}^{k_n} \delta_{X_i^{(u)}},$$

where  $k_n$  is an integer and  $(X_1^{(u)}, \dots, X_{k_n}^{(u)})$  is i.i.d with law  $\mathbf{P}_u$ .

Note that "equality in laws" here refers to the  $\sigma$ -algebra  $\mathcal{M}$  of  $\mathfrak{M}$ . The two statistical applications we have in mind here are as follows - see Subsection 2.1 for more details.

**Example 1:** Let  $Z_i := (X_i, Y_i)_{i \in \llbracket 1, n \rrbracket}$  be i.i.d with law on  $\mathbb{R}^d \times \mathbb{R}$  (which may include the case  $d = 1$  and  $X_i = Y_i$ ). Then several statistics in extreme value theory can be expressed through the empirical measure based on the concomitant statistic  $(X_i, i \in \mathcal{I})$  of the  $k_n$  elements in  $(X_1, \dots, X_n)$  that correspond to the  $k_n$  largest observations among  $(Y_1, \dots, Y_n)$ . If ties among the  $Y_i$  are broken evenly, then such empirical measures satisfy the CEM assumption.

**Example 2:** Take the same setup as Example 1, but now consider the empirical measure based on observations  $X_i$  that belong to a data-driven neighborhood of some point  $x \in \mathbb{R}^d$  built through a "nearest neighbor" rule: given a distance  $\rho$ , write  $D_i := \rho(X_i, x)$  and consider the set of  $k_n$  indices  $\mathcal{I}$  corresponding to the  $k_n$  smallest values among  $(D_1, \dots, D_n)$ . Then, if ties are broken evenly, the empirical measure based on the  $(Z_i, i \in \mathcal{I})$  satisfies the CEM assumption.

We also assume that  $\mathbf{U}_n \xrightarrow{d} \mathbf{U}$  in  $(E, d)$  for some random variable  $\mathbf{U}$  with support  $E_0$ . For example, if  $E_0 = \{u_0\}$  (i.e.  $\mathbf{U} \equiv u_0$  is deterministic) we basically suppose that, for large  $n$ , the law of  $\mathbb{P}_n^{(n)}$  is "almost that of"  $\mathbf{P}_{k_n}^{\mathbf{P}_{u_0}}$  - up to a continuity condition

upon  $u \rightarrow \mathbf{P}_u$  - see (14) below. When  $\mathbf{U}$  is non deterministic, then the latter heuristic approximation holds under a mixture upon  $u_0$  according to the law of  $\mathbf{U}$ .

**Remark :** A straightforward way to build a random measure that satisfies (CEM) is to take a sample  $(X_1, \dots, X_{k_n})$  for which the law given  $\mathbf{U}_n = u$  is  $\mathbf{P}_u^{\otimes k_n}$ , then define

$$\mathbb{P}_n^{(n)} := \frac{1}{k_n} \sum_{i=1}^{k_n} \delta_{X_i}. \quad (3)$$

However, we would like to point out the fact that the two above mentioned examples *cannot* be expressed with (3). In both examples the set  $\mathcal{I} \subset \llbracket 1, n \rrbracket$  is not constantly equal to  $\llbracket 1, \dots, k_n \rrbracket$ . It is rather a random set of indices with fixed cardinality  $k_n$ . This illustrates the flexibility of the (CEM) framework in statistics: it encompasses more objects than those of type (3) as it doesn't require any explicit indexing of the subsample that is used to build the empirical measure.

## 1.2 Indexing by a class of functions

For each  $n \geq 1$ , denote by  $Q_n$  the law of  $\mathbf{U}_n$ . Let  $\mathcal{F}$  be a class of real valued, Borel functions on  $(\mathfrak{X}, \mathcal{X})$ . An elementary calculus shows that, if  $f \in \mathcal{F}$  is bounded, then one has

$$\begin{aligned} \left| \mathbb{P}_n^{(n)}(f) - \mathbf{P}_{\mathbf{U}_n}(f) \right| &\xrightarrow{d} 0, \text{ as } n \rightarrow \infty, \text{ where} \\ \mathbb{P}(\mathbf{P}_{\mathbf{U}_n}(f) \leq t) &:= \int_E \mathbb{1}_{\{\mathbf{P}_u(f) \leq t\}} dQ_n(u), \quad t \in \mathbb{R}. \end{aligned}$$

In view of these convergences, it is natural to investigate whether

$$\left\| \mathbb{P}_n^{(n)} - \mathbf{P}_{\mathbf{U}_n} \right\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} \left| \mathbb{P}_n^{(n)}(f) - \mathbf{P}_{\mathbf{U}_n}(f) \right| \xrightarrow{d} 0, \quad (4)$$

as  $n \rightarrow \infty$ , which is a Glivenko-Cantelli type result. Handling the possible non (Borel) measurability of the latter map is usual in empirical processes theory. The now standard approach of Hoffman-Jørgensen (see [30]) is to consider outer expectations/probabilities on  $(\Omega, \mathcal{A}, \mathbb{P})$  when measurability fails to hold. However outer integrals do not interact easily with conditioning (e.g. upon  $\mathbf{U}_n$ ). Consequently, we shall take a special care in this paper to avoid outer expectations by only using measurable functions/sets in our proofs. These measurability properties will be derived from the following standard assumption upon  $\mathcal{F}$ , for some countable  $\check{\mathcal{F}} \subset \mathcal{F}$ :

$$\mathcal{F} \text{ is pointwise measurable with countable separant } \check{\mathcal{F}}, \quad (5)$$

as defined in [36, p. 116]. This assumption automatically induces the measurability of the following envelope function  $F$  defined by

$$F(x) := \sup_{f \in \mathcal{F}} |f(x)| \vee 1 = \sup_{f \in \check{\mathcal{F}}} |f(x)| \vee 1, \quad x \in \mathfrak{X}, \quad (6)$$

which we assume  $\mathbf{P}_u$ -integrable for each  $u \in E$ . In addition, a simple verification - see Lemma 3.2 below - shows that the involved objects in (4) are Borel measurable. Our integrability assumptions upon  $F$  permits us to see each  $\mathbf{P}_{\mathbf{U}_n}$  and  $\mathbb{P}_n^{(n)}$  as random elements taking values in  $\ell^\infty(\mathcal{F})$ , the space of all bounded functions from  $\mathcal{F}$  to  $\mathbb{R}$ .

### 1.3 A Donsker and a Glivenko-Cantelli theorem

Write generically  $\|f\|_{Q,r}^r := Q(|f|^r)$  for  $r > 0$  and a probability measure  $Q$ , and write, for fixed  $\epsilon > 0$  the associated *covering number*  $N(\epsilon, \mathcal{F}, \|\cdot\|_{Q,r})$  defined as the (possibly infinite) minimal number of closed  $\|\cdot\|_{Q,r}$ -balls with radius  $\epsilon$  needed to cover  $\mathcal{F}$ . The *uniform entropy integral* of  $\mathcal{F}$  with respect to  $F$  is commonly defined as

$$J_{\mathcal{F},F}(\eta) := \int_0^\eta \sqrt{\sup_{Q \in \mathfrak{M}} \log \left( 2N(\epsilon \|F\|_{Q,2}, \mathcal{F}, \|\cdot\|_{Q,2}) \right)} d\epsilon, \quad (7)$$

for  $\eta \in ]0, \infty]$ . Our first result can be stated as follows, denoting by  $\mathcal{F}_M$  the truncated class

$$\mathcal{F}_M := \left\{ f \mathbf{1}_{\{F \leq M\}}, f \in \mathcal{F} \right\}, \quad (8)$$

and " $\limsup_{n \rightarrow \infty} \mathbb{P}$ " standing for "lim sup in probability".

**Theorem 1 (a Glivenko-Cantelli theorem)** *Assume that  $F$  is  $\mathbf{P}_u$  integrable for each  $u \in E$ , and that*

$$\lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \mathbf{P}_{\mathbf{U}_n} \left( F \mathbf{1}_{\{F > M\}} \right) = 0. \quad (9)$$

*Assume that, for each  $\epsilon > 0$  and  $M > 0$  we have*

$$\log \left( N(\epsilon, \mathcal{F}_M, \|\cdot\|_{\mathbb{P}_n^{(n)},1}) \right) = o_{\mathbb{P}}(k_n). \quad (10)$$

*Then we have*

$$\left\| \mathbb{P}_n^{(n)} - \mathbf{P}_{\mathbf{U}_n} \right\|_{\mathcal{F}} \xrightarrow{d} 0.$$

The implicit measurabilities in (9) and (10) are consequences of (5) and Lemma 3.3 below.

Let us now focus on a possible weak convergence in  $(\ell^\infty(\mathcal{F}), \|\cdot\|_{\mathcal{F}})$  of  $\mathbb{G}_n^{(n)}$ , where

$$\mathbb{G}_n^{(n)}(f) := \sqrt{k_n} \left( \mathbb{P}_n^{(n)}(f) - \mathbf{P}_{\mathbf{U}_n}(f) \right), f \in \mathcal{F}. \quad (11)$$

For a given probability measure  $Q$ , we will write  $\mathbb{G}_Q$  for the  $Q$ -Brownian bridge indexed by  $\mathcal{F}$  (see [36, p. 82]), which induces a tight probability measure on  $\ell^\infty(\mathcal{F})$  as soon as  $J_{\mathcal{F},F}(\infty) < \infty$ . A straightforward analysis of the finite dimensional convergence of such processes shows that the only possible limit of  $\mathbb{G}_n^{(n)}$  must be  $\mathbb{G}_{\mathbf{P}_{\mathbf{U}}}$ , the "mixture of the laws of  $\mathbb{G}_Q$  with  $Q \rightsquigarrow \mathbf{P}_{\mathbf{U}}$ ". A rigorous definition of such a mixture can be found in [39], and it is not guaranteed that  $\mathbb{G}_{\mathbf{P}_{\mathbf{U}}}$  induces tight probability measure on  $\ell^\infty(\mathcal{F})$  - even if each  $\mathbb{G}_Q$  does. This tightness however holds under the conditions of the following Donsker-type result.

**Theorem 2 (a Donsker theorem)** Assume that  $J_{\mathcal{F},F}(\infty) < \infty$ , assume that  $F^2$  is  $\mathbf{P}_u$  integrable for each  $u \in E$ , and that

$$\lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \mathbf{P}_{\mathbf{U}_n} \left( F^2 \mathbb{1}_{\{F > M\}} \right) = 0. \quad (12)$$

Write

$$\mathcal{G} := \mathcal{F} \cup \left\{ f^2, f \in \mathcal{F} \right\} \cup \left\{ (f_1 - f_2)^2, (f_1, f_2) \in \mathcal{F}^2 \right\}, \quad (13)$$

and assume that

$$\begin{aligned} \text{The map } H : (E, d) &\mapsto (\ell^\infty(\mathcal{G}), \|\cdot\|_{\mathcal{G}}) \\ u &\rightarrow \{f \rightarrow \mathbf{P}_u(f)\} \end{aligned} \quad (14)$$

is continuous at each  $u \in E_0$ .

Then  $\mathbb{G}_{\mathbf{P}_U}$  induces a tight probability measure on  $\ell^\infty(\mathcal{F})$  and we have

$$\mathbb{G}_n^{(n)} \xrightarrow{d} \mathbb{G}_{\mathbf{P}_U}. \quad (15)$$

**Remark 1:** In both theorems the centering parameter is  $\mathbf{P}_{\mathbf{U}_n}$ . For statistical application, the presence of this centering parameter can be interpreted as a "bias". We use quotes here since  $\mathbf{P}_{\mathbf{U}_n}$  is not deterministic. A natural question is then: In which conditions can these "biases" vanish as  $n \rightarrow \infty$ ? If we assume that  $\mathbf{U}_n \rightarrow_{\mathbb{P}} \mathbf{U}$  and that (14) holds with  $\mathcal{F}$  instead of  $\mathcal{G}$ , then  $\|\mathbf{P}_{\mathbf{U}_n} - \mathbf{P}_U\|_{\mathcal{F}} \xrightarrow{d} 0$ , which strengthens Theorem 1 to its unbiased version. Now if, in a particular statistical application, one can establish that

$$\sqrt{k_n} \left\| \mathbf{P}_{\mathbf{U}_n} - \mathbf{P}_U \right\|_{\mathcal{F}} \xrightarrow{d} 0, \quad (16)$$

by analytical arguments, then Theorem 2 can be strengthened to its unbiased version as well, namely

$$\sqrt{k_n} \left( \mathbb{P}_n^{(n)} - \mathbf{P}_U \right) \xrightarrow{d} \mathbb{G}_{\mathbf{P}_U}. \quad (17)$$

**Remark 2:** It turns out that, in all the statistical applications we are considering (see Section 2) we have  $\mathbf{U} \equiv u_0$ , and  $E_0 := \{u_0\}$ , hence the limit is not a mixture of gaussian processes but rather a Gaussian process. In this particular case, conditions (12) and (14) boil down to

$$\lim_{M \rightarrow \infty} \limsup_{u \rightarrow u_0} \mathbf{P}_u \left( F^2 \mathbb{1}_{\{F \geq M\}} \right) = 0, \text{ and} \quad (18)$$

$$\lim_{u \rightarrow u_0} \left\| \mathbf{P}_u - \mathbf{P}_{u_0} \right\|_{\mathcal{G}} = 0. \quad (19)$$

Those are extremely similar to conditions (i) and (ii) of Sheehy and Wellner [35, Theorem 3.1]. The reader may notice that we heavily borrow from their methods to achieve our proofs.

**Remark 3:** In empirical processes theory, the natural underlying probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  is the canonical product space supporting an i.i.d sequence, and each outer probability/expectation is taken over  $(\Omega, \mathcal{A}, \mathbb{P})$ . The main reason for this choice is that

it makes each  $X_i$  a perfect map, which have several nice properties with respect to outer expectations - see [36, Chapter 1]. However, as this space is not suitable for the (CEM) model, our analysis will take place on the space that is the most natural for (CEM) i.e. the space defined at the beginning of Section 1.1. Transporting the usual techniques of empirical processes in this new setup is possible because we take a particular care on measurability, making outer expectations *true expectations*, hence not depending anymore upon the initial choice of probability space - see Section 3.1 for more details.

## 1.4 A bootstrap theorem

In this section we focus on the use of the bootstrap in order to approximate the limit law in Theorem 2. Let us first point out that the law of  $\mathbf{U}_n$  cannot in general be approximated by a bootstrap procedure. In most statistical applications,  $\mathbf{U}_n$  is not even observable. We will hence restrict ourselves to the case where  $\mathbf{U}_n \xrightarrow{d} u_0$  for some deterministic  $u_0 \in E$ . Another point to notice is that the setup of the CEM assumption does not permit a simple way to make a convex combination of random weights with point masses at random locations, due to the lack of proper indexing of the atoms - a drawback absent of the statistical applications of Section 2 though. We will avoid this inconvenience by focusing on Efron's bootstrap, which has the advantage of being interpreted as  $k$  mutually independent draws from an observed discrete measure. We will only treat the case when  $k = k_n$  for simplicity. The two above mentioned difficulties make our result formulated differently (yet equivalently) from the usual Donsker bootstrap theorems for empirical processes.

For any probability measure  $P$  with finite support and bounded real-valued Lipschitz function  $\phi$  on  $\ell^\infty(\mathcal{F})$  define

$$\Delta_k(P, \phi) := \left| \mathbb{E} \left( \phi \left( \mathbb{G}_{k,P} \right) \right) - \mathbb{E} \left( \phi \left( \mathbb{G}_{\mathbf{P}_{u_0}} \right) \right) \right|, \text{ where} \quad (20)$$

$$\mathbb{G}_{k,P} := \sqrt{k} \left( \frac{1}{k} \sum_{i=1}^k \delta_{X_i} - P \right), \text{ with } (X_1, \dots, X_k) \rightsquigarrow P^{\otimes k}. \quad (21)$$

Note that the first expectation in (20) is well defined since the considered random element  $\mathbb{G}_{k,P}$  of  $\ell^\infty(\mathcal{F})$  takes a finitely many values. Now define

$$Donsker_k(P) := \sup_{\phi \in BL_1(\ell^\infty(\mathcal{F}))} \Delta_k(P, \phi), \quad (22)$$

where, for a metric space  $G$ ,  $BL_1(G)$  stands for the set of all 1-Lipschitz functions on  $G$  that are bounded by 1. Now consider  $Donsker_{k_n}(\mathbb{P}_n^{(n)})$  which is a sequence of (non necessarily measurable) random elements from  $\Omega$  to  $\mathbb{R}^+$ . It quantifies the quality of the (random) approximation between the limit law  $\mathbb{G}_{\mathbf{P}_{u_0}}$  and the bootstrap law of an empirical process built from  $k_n$  i.i.d draws from the observed value of  $\mathbb{P}_n^{(n)}$ .

**Theorem 3 (a bootstrap theorem)** Assume that  $E_0 = \{u_0\}$ . Then, under the conditions of Theorem 2 one has

$$\left( \text{Donsker}_{k_n}(\mathbb{P}_n^{(n)}) \right)^* \xrightarrow{d} 0, \quad (23)$$

where “ $*$ ” stands for the (equivalence class of the) measurable cover function on  $(\Omega, \mathcal{A}, \mathbb{P})$  (see, e.g., [36, p. 6]).

## 2 Some statistical applications

### 2.1 The CEM assumption and thresholds based on ranks

We shall use the standard notation  $(y_{1:n}, \dots, y_{n:n})$  for the order statistics, i.e. the vector of increasingly ordered values of  $(y_1, \dots, y_n)$ . Let  $(X_i, Y_i)_{i \in \llbracket 1, n \rrbracket}$  be an i.i.d sequence taking values in  $\mathfrak{X}' \times \mathbb{R}$ . In this section we are interested in the law of the empirical measure built on the  $X_i$  for which  $Y_i$  exceed  $Y_{n-k_n+1:n}$ . When ties occur in the sample, these latter are supposed to be broken by an extra randomness (uniformly among the tie-breaking possibilities). Such a way to break the ties can be performed by the following algorithm, using the following partitions of  $\llbracket 1, n \rrbracket$ :

$$\mathcal{I}_{>}(y_1, \dots, y_n) := \{i \in \llbracket 1, n \rrbracket, y_i > y_{n-k_n+1:n}\} \quad (24)$$

$$\mathcal{I}_{=}(y_1, \dots, y_n) := \{i \in \llbracket 1, n \rrbracket, y_i = y_{n-k_n+1:n}\}, \quad (25)$$

$$\mathcal{I}_{<}(y_1, \dots, y_n) := \{i \in \llbracket 1, n \rrbracket, y_i < y_{n-k_n+1:n}\}. \quad (26)$$

#### Tie-breaking algorithm 1

1. If  $\mathcal{I}_{=}(y_1, \dots, y_n)$  contains at least two indices, then generate  $\mathcal{I}'_{=}(y_1, \dots, y_n)$  by selecting uniformly at random a subset of  $\mathcal{I}_{=}(y_1, \dots, y_n) := \{i \in \llbracket 1, n \rrbracket, y_i = y_{n-k_n+1:n}\}$  with cardinality  $k_n - \#\mathcal{I}_{>}(y_1, \dots, y_n)$ .
2. Set  $\mathcal{I}_1(y_1, \dots, y_n) := \mathcal{I}_{>}(y_1, \dots, y_n) \cup \mathcal{I}'_{=}(y_1, \dots, y_n)$ .

This algorithm clearly picks out a subset of  $k_n$  indices  $i$  fulfilling  $y_i \geq y_{n-k_n+1:n}$ , by breaking ties for  $y_{n-k_n+1:n}$  uniformly at random.

The statistical applications of this section are all based on the following type of random measure,  $\tilde{\phi} : \mathfrak{X}' \times \mathbb{R} \rightarrow \mathfrak{X}$  denoting a measurable function:

$$\tilde{\mathbb{P}}_n^{(n)} := \frac{1}{k_n} \sum_{i \in \mathcal{I}_1(y_1, \dots, y_n)} \delta_{\tilde{\phi}(x_i, y_{n-k_n:n})}. \quad (27)$$

At first sight, it is unclear that  $\tilde{\mathbb{P}}_n^{(n)}$  satisfies the CEM assumption. To prove this, we will randomly generate each  $\mathcal{I}_1(y_1, \dots, y_n)$  using a specific form (or representation) of that randomness. This alternative algorithm is presented below, and has the upside of

making the CEM assumption clearer to understand. Write  $F$  for the c.d.f of  $Y_1$ , and  $F^\leftarrow$  for the corresponding quantile function, namely

$$F^\leftarrow(u) := \inf\{t \in \mathbb{R}, F(t) \geq u\}, \quad u \in ]0, 1[.$$

For fixed  $y$  we shall write  $\Delta F(y)$  for  $\mathbb{P}(Y_1 = y) = F(y) - F^-(y)$ .

**Tie-breaking algorithm 2:**

1. Generate an i.i.d  $n$ -uple  $(\boldsymbol{\eta}_i)_{i \in \llbracket 1, n \rrbracket}$  uniformly distributed on  $]0, 1[$ .
2. For each  $i \in \llbracket 1, n \rrbracket$ , set  $U_i(y_i, \boldsymbol{\eta}_i) := F^-(y_i) + \boldsymbol{\eta}_i \Delta F(y_i)$ .
3. Consider the vector of (almost surely untied) ranks  $(R_1, \dots, R_n)$  of the vector  $(U_1(y_1, \boldsymbol{\eta}_1), \dots, U_n(y_n, \boldsymbol{\eta}_n))$ .
4. Set  $\mathcal{I}_2(y_1, \dots, y_n) := \{i \in \llbracket 1, n \rrbracket, R_i > n - k_n\}$ .

It is easy to verify that, if  $(Y_1, \dots, Y_n)$  is i.i.d, then

$$(U_1, \dots, U_n) := (U_1(Y_1, \boldsymbol{\eta}_1), \dots, U_n(Y_n, \boldsymbol{\eta}_n)) \quad (28)$$

is itself i.i.d. and uniformly distributed on  $[0, 1]$ , that  $F^\leftarrow(U_i) = Y_i$  almost surely for each  $i \in \llbracket 1, n \rrbracket$  and that the coordinates of  $(Y_1, \dots, Y_n)$  are almost surely ordered exactly as those of  $(U_1, \dots, U_n)$ . We then consider, for given  $k_n \in \llbracket 1, n \rrbracket$ , the empirical measure

$$\mathbb{P}_n^{(n)} := \frac{1}{k_n} \sum_{i \in \mathcal{I}_2(Y_1, \dots, Y_n)} \delta_{\tilde{\phi}(X_i, Y_{n-k_n:n})} = \frac{1}{k_n} \sum_{i \in \mathcal{I}_2(Y_1, \dots, Y_n)} \delta_{\phi(X_i, U_{n-k_n:n})}, \quad (29)$$

with  $\phi(x, u) := \tilde{\phi}(x, F^\leftarrow(u))$ ,  $x \in \mathfrak{X}'$ ,  $u \in [0, 1[$ .

**Proposition 2.1** *The following equality in laws holds on  $\mathcal{M}$ :*

$$\tilde{\mathbb{P}}_n^{(n)} \stackrel{d}{=} \mathbb{P}_n^{(n)}.$$

The proof is postponed to Section 4.

Our next proposition links Theorems 1 and 2 to the applications of the following subsections. Its proof is postponed to §4.2.

**Proposition 2.2** *Let  $(U_1, \dots, U_n)$  be as in (28). Set  $\mathbf{U}_n := U_{n-k_n:n}$ . The law of  $(\mathbb{P}_n^{(n)}, \mathbf{U}_n)$  is that of*

$$\left( \frac{1}{k_n} \sum_{i=n-k_n+1}^n \delta_{\phi(X_i, U_{n-k_n:n})}, U_{n-k_n:n} \right)$$

given  $U_1 < \dots < U_n$ . As a consequence, for each  $n \geq 1$ , the couple  $(\mathbb{P}_n^{(n)}, \mathbf{U}_n)$  satisfies the CEM assumption, with  $\mathbf{P}_u$  equal to the law of  $\phi(X, u)$  given  $U > u$ , for each  $u \in [0, 1[$ .

Note that, if  $Y$  is continuous then there is no point in introducing  $U$  (no tie breaking is needed), and  $\mathbf{P}_u$  is simply the law of  $\phi(X, u)$  given  $Y > F^\leftarrow(u)$ . Models in Extreme Value Theory naturally make assumptions with a conditioning upon  $Y > y$  (rather than  $U(Y, \eta) > u$ ). These assumptions can generically be written of the following form, for a class  $\mathcal{G}$  with envelope  $G$ , and for some function  $\varepsilon(\cdot)$  tending to zero as  $y \uparrow F^\leftarrow(1)$

$$\sup_{g \in \mathcal{G}} \left| \mathbb{E} \left( g \circ \tilde{\phi}(X, y) \mid Y > y \right) - \mathbf{P}_1(g) \right| = O(\varepsilon(y)), \quad (30)$$

$$\lim_{M \rightarrow \infty} \limsup_{y \uparrow F^\leftarrow(1)} \mathbb{E} \left( (G \mathbf{1}_{\{G > M\}}) \circ \tilde{\phi}(X, y) \mid Y > y \right) = 0, \quad (31)$$

where  $F^\leftarrow(1) = \lim_{p \rightarrow 1} F^\leftarrow(p)$ . Our next proposition shows that, under mild conditions, (30) and (31) have a direct implication upon the limit behavior of  $\mathbf{P}_u$  as  $u \rightarrow 1$ . Its proof is postponed to §4.

**Proposition 2.3** *Under the notation of Proposition 2.2, assume that (30) and (31) hold for some probability measure  $\mathbf{P}_1$ , and that*

$$\lim_{y \uparrow F^\leftarrow(1)} \frac{1}{\varepsilon(y)} \frac{\mathbb{P}(Y = y)}{\mathbb{P}(Y > y)} = 0. \quad (32)$$

$$\lim_{y \uparrow F^\leftarrow(1)} \frac{1}{\varepsilon(y)} \frac{\mathbb{E} \left( G \circ \tilde{\phi}(X, y) \mathbf{1}_{\{Y=y\}} \right)}{\mathbb{P}(Y > y)} = 0. \quad (33)$$

Then one has, as  $u \uparrow 1$ :

$$\sup_{g \in \mathcal{G}} \left| \mathbf{P}_u(g) - \mathbf{P}_1(g) \right| = O(\varepsilon(F^\leftarrow(u))), \text{ and} \quad (34)$$

$$\lim_{M \rightarrow \infty} \limsup_{u \uparrow 1} \mathbf{P}_u(G \mathbf{1}_{\{G > M\}}) = 0. \quad (35)$$

In particular both (18) and (19) hold with  $E := [0, 1]$ ,  $E_0 := \{1\}$  and  $\mathbf{P}_u$ ,  $u \in [0, 1]$ .

## 2.2 Revisiting the peaks over threshold empirical c.d.f

For some  $\gamma \in \mathbb{R}$ , assume that  $Y$  satisfies the conditions of the Pickands-Balkema-de Haan theorem, namely: for fixed  $x$  satisfying  $1 + \gamma x > 0$  and for two normalizing functions  $\tilde{b}(\cdot)$ ,  $\tilde{a}(\cdot)$  one has

$$\lim_{t \rightarrow \infty} t \bar{F}(\tilde{a}(t)x + \tilde{b}(t)) = V(x), \quad (36)$$

where

$$V(x) := (1 + \gamma x)^{-1/\gamma} \mathbf{1}_{\{1 + \gamma x > 0\}} \text{ if } \gamma \neq 0, \text{ and} \\ V(x) := \exp(-x), \text{ for } \gamma = 0.$$

Write  $V_t(x) := t\bar{F}(\tilde{a}(t)x + \tilde{b}(t))$ . In this section we shall work under the stronger framework of *second order assumptions*, namely, for some  $\rho \leq 0$  and for all  $x \geq 0$ :

$$\lim_{t \rightarrow \infty} \frac{V_t(x) - V(x)}{\tilde{A}(t)} = (1 - \gamma x)^{-1-1/\gamma} H_{\gamma, \rho}((1 + \gamma x)^{1/\gamma}), \quad (37)$$

where  $\tilde{A}$  is a  $\rho$ -varying function with eventually constant sign, and

$$H_{\gamma, \rho}(x) := \int_1^x s^{\gamma-1} \int_1^s u^{\rho-1} du ds.$$

Drees, De Haan and Li [22] proved that, as soon as (37) is assumed in the model, a weighted uniform approximation also holds for modifications  $a(\cdot)$ ,  $A(\cdot)$ ,  $b(\cdot)$  of  $\tilde{a}(\cdot)$ ,  $\tilde{A}(\cdot)$  and  $\tilde{b}(\cdot)$ . The exact expression of  $A(\cdot)$  and  $a(\cdot)$  will be of no relevance in this paper. What has to be kept in mind is that  $a(t)/\tilde{a}(t)$  and  $A(t)/\tilde{A}(t)$  both tend to 1 as  $t \rightarrow \infty$ . The expression of  $b(\cdot)$  however needs a special attention:

$$b(t) := U(t) - ca(t)A(t), \quad \text{with } c := \frac{1}{\gamma + \rho} \mathbf{1}_{\{\gamma + \rho \neq 0\}}. \quad (38)$$

**Proposition 2.4 (Corollary from Drees *et. al.* 2006, Proposition 3.2)** *Assume that either  $\gamma \neq 0$  or  $\rho < 0$ . Write  $x^+ := 1/(-\gamma \vee 0)$ . Then for any  $\eta > 0$  one has*

$$\lim_{t \rightarrow \infty} \sup_{x \in [0, x^+]} V(x)^{\rho-1+\eta} \left| \frac{V_t(x) - V(x)}{A(t)} - V(x)^{1+\gamma} K_{\rho, \gamma}(1/V(x)) \right| = 0,$$

where  $K_{\rho, \gamma}(v) = v^{\gamma+\rho}/(\gamma + \rho)$  and  $K_{\rho, \gamma}(v) = \log(v)$  if  $\gamma + \rho = 0$ .

With this proposition at hand the authors could derive a Donsker theorem in weighted topology for the following sequence of "POT empirical functions"

$$\tilde{\mathbb{F}}_{n, k_n}(x) := \frac{1}{k_n} \sum_{i=1}^n \mathbf{1}_{[x, +\infty[} \left( \frac{Y_i - b(\frac{n}{k_n})}{a(\frac{n}{k_n})} \right),$$

under some standard conditions upon  $k_n$  (see [22, Theorem 1]). In this section we will prove a similar result for

$$\mathbb{F}_{n, k_n}(x) := \frac{1}{k_n} \sum_{\substack{i \in \llbracket 1, n \rrbracket \\ i \in \mathcal{I}_1(Y_1, \dots, Y_n)}} \mathbf{1}_{[0, x]} \left( \frac{Y_i - b(1/\bar{F}(Y_{n-k_n:n}))}{a(1/\bar{F}(Y_{n-k_n:n}))} \right), \quad x \in [0, x^+],$$

which is the counterpart of  $\tilde{\mathbb{F}}_{n, k_n}$  where the threshold is data driven and the normalization is deterministic. In the sequel,  $L^{\infty, \beta}([0, x^+])$  shall denote the closed subspace of  $\ell^\infty([0, x^+])$  of all trajectories  $\psi$  satisfying  $\|\psi\|_{\infty, \beta} < \infty$ , where

$$\|\psi\|_{\infty, \beta} := \sup_{x \in [0, x^+]} V(x)^{-\beta} |\psi(x)|.$$

Then  $L^{\infty, \beta}([0, x^+])$  is naturally endowed with the norm  $\|\cdot\|_{\infty, \beta}$ .

**Proposition 2.5** Assume that (37) holds, that  $\gamma \neq 0$  and that  $\sqrt{k_n}A(n/k_n) \rightarrow 0$ . Then for any  $\delta \in ]0, 1/2]$  one has

$$\sqrt{k_n} \left( \mathbb{F}_{n,k_n} - (1 - V) \right) \xrightarrow{d} B(1 - V(\cdot)),$$

in  $L^{\infty, 1/2-\delta}([0, x^+])$ ,  $B$  denoting the standard Brownian bridge.

The proof of Proposition 2.5 is postponed to Section 4.

**Remark:** Note that [22, Theorem 1] implies that, under the conditions of Proposition 2.5 one has  $\sqrt{k_n}(\tilde{\mathbb{F}}_{n,k_n} - V) \xrightarrow{d} W(V(\cdot))$  in  $L^{\infty, 1/2-\delta}([0, x^+])$ , where  $W$  is the standard Brownian motion. Since  $\|B(V(\cdot))\|_{\infty, 1/2-\delta}$  is stochastically lower than  $\|W(V(\cdot))\|_{\infty, 1/2-\delta}$  (use Slepian's Lemma) we can conclude that, while both  $\bar{\mathbb{F}}_{n,k_n}$  and  $\tilde{\mathbb{F}}_{n,k_n}$  are  $\|\cdot\|_{\infty, 1/2-\delta}$ -weak consistent at rate  $k_n^{-1/2}$ ,  $\bar{\mathbb{F}}_{n,k_n}$  does slightly better. This advocates for the use of data driven thresholds based on ordered statistics rather than deterministic ones. To the best of our knowledge, such a phenomenon is new to the literature.

### 2.3 Multivariate regular variation

Take  $\mathfrak{X} := \mathbb{R}^d$  endowed with the canonical Euclidean norm  $\|\cdot\|_d$ , and write  $\mathcal{S}_{d-1}$  for its unit sphere. In this section, we make the following assumption, for each Borel set  $A \subset ]0, \infty]$  that is bounded away from zero and for each Borel set  $B \subset \mathcal{S}_{d-1}$

$$\frac{\mathbb{P} \left( \left( \frac{\|X\|_d}{y}, \frac{1}{\|X\|_d} X \right) \in A \times B \right)}{\mathbb{P} \left( \|X\|_d > y \right)} \xrightarrow{y \rightarrow \infty} m(A \times B), \quad (39)$$

where  $m$  is a Borel measure on  $]0, \infty[ \times \mathcal{S}_{d-1}$ . This model encompasses that of *multivariate regular variation* (MRV), i.e. when (39) holds for  $m := \mu_\alpha \otimes \nu$ , with  $\nu$  a (spectral) measure on  $\mathcal{S}_{d-1}$  and  $\mu_\alpha(dx) = \alpha x^{-1-\alpha} dx$  for some  $\alpha > 0$ . Specializing to Borel sets  $A \subset ]1, \infty]$  we see that (39) reads

$$\mathbb{P} \left( \left( \frac{\|X\|_d}{y}, \frac{1}{\|X\|_d} X \right) \in \cdot \mid \|X\|_d > y \right) \xrightarrow{y \rightarrow \infty} Q, \quad (40)$$

where  $Q$  is the restriction of  $m$  to  $]1, \infty[ \times \mathcal{S}_{d-1}$ . In this subsection we propose a goodness of fit test procedure for the multivariate regular variation model, by testing the hypothesis

$\mathbf{H}_0$  : "Q has independent marginals and its first margin is Pareto",

in the model (40). Assumption  $\mathbf{H}_0$  is important, as MRV is a widely used model in finance, economics or communication networks, as pointed out in Einmahl *et. al.* [23]. In this article, the authors proposed a test procedure for  $\mathbf{H}_0$ , for which they proved consistency together with real data studies. Their idea was to interpret  $\mathbf{H}_0$  as the constancy in  $A \subset \mathcal{S}_{d-1}$  of  $\gamma(A)$ , the Pareto index of the radial component  $Y := \|X\|_d$  given

the angular part  $\|X\|_d^{-1}X$  belongs to  $A$ . Their test statistic checks that constancy with respect to a partition  $A_1, \dots, A_m$  of  $\mathcal{S}_{d-1}$  with a chi-square type goodness of fit test statistic  $T_n$ , combined with a "PE" test statistic  $Q_n$  which itself tests the Pareto assumption upon the radial component (see their Theorem 2 for more details). Both  $T_n$  and  $Q_n$  can be seen as functionals of the local empirical measure

$$\frac{1}{k_n} \sum_{i: Y_i > Y_{n-k_n:n}} \delta_{(\log(Y_i/Y_{n-k_n:n}), Y_i^{-1}X_i)}.$$

Note that, if the law of  $Y$  has atoms, then the number of elements in the preceding does not necessarily equal  $k_n$ . This caveat is tackled when considering

$$\tilde{\mathbb{P}}_n^{(n)} := \frac{1}{k_n} \sum_{i \in \mathcal{I}_1(Y_1, \dots, Y_n)} \delta_{(\log(Y_i/Y_{n-k_n:n}), Y_i^{-1}X_i)}, \quad (41)$$

where  $\mathcal{I}_1(Y_1, \dots, Y_n)$  selects  $k_n$  indices by *tie breaking algorithm 1*. This empirical measure falls into representation (27) with the choice of  $\tilde{\phi}(x, y) := (\log(\|x\|_d/y), \|x\|_d^{-1}x)$ , and hence - by Propositions 2.1 and 2.2 - satisfies the CEM assumption with

$$\mathbf{P}_u := \mathbb{P}\left(\left(\log\left(\frac{\|X\|_d}{F^{\leftarrow}(u)}\right), \|X\|_d^{-1}X\right) \in \cdot \mid U \geq u\right). \quad (42)$$

Our proposed test procedure is an alternative to that of Einmahl *et al.* [23]: we test both the independence and Pareto marginal (i.e  $\log(Y)$  is exponential) by Kolmogorov-Smirnov statistics based on  $\tilde{\mathbb{P}}_n^{(n)}$ . For any given  $d-1$ -tuple of angles  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{d-1}) \in [0, 2\pi]^{d-1}$  define the corresponding "lower left orthant for polar coordinates"

$$O_{\boldsymbol{\theta}} := \left\{ \left( \cos \theta'_1, \sin \theta'_1 \cos \theta'_2, \dots, \sin \theta'_1 \sin \theta'_2 \dots \sin \theta'_{d-2} \cos \theta'_{d-1}, \sin \theta'_1 \dots \sin \theta'_{d-1} \right), \right. \\ \left. \theta'_1 \in [0, \theta_1], \dots, \theta'_{d-1} \in [0, \theta_{d-1}] \right\}, \quad (43)$$

and with  $O_{\theta_1} := \{(\cos \theta'_1, \sin \theta'_1), \theta'_1 \in [0, \theta_1]\}$  when  $d = 1$ . Now write

$$\tilde{\mathbb{F}}_n(t, \boldsymbol{\theta}) := \frac{1}{k_n} \sum_{i \in \mathcal{I}_1(Y_1, \dots, Y_n)} \mathbb{1}_{]t, +\infty[}(\log(Y_i/Y_{n-k_n:n})) \mathbb{1}_{O_{\boldsymbol{\theta}}}(Y_i^{-1}X_i) \\ = \tilde{\mathbb{P}}_n^{(n)}(]t, +\infty[ \times O_{\boldsymbol{\theta}}), \quad t \in [0, +\infty[, \boldsymbol{\theta} \in [0, 2\pi]^{d-1}.$$

Note that, for any observed sample the partial function  $\check{\mathbb{F}}_n^{(1)} := \check{\mathbb{F}}_n(\cdot, 2\pi, \dots, 2\pi)$  is a survival function on  $[0, +\infty[$ , while  $\check{\mathbb{F}}_n^{(2)} := \check{\mathbb{F}}_n(0, \cdot)$  is a cumulative distribution function on  $[0, 2\pi]^{d-1}$ . Now define

$$\Delta_{1,n} := \sup_{t \geq 0} \left| \frac{1}{k_n} \sum_{i \in \mathcal{I}_1(Y_1, \dots, Y_n)} \mathbb{1}_{]t, +\infty[} \left( \frac{\log(Y_i/Y_{n-k_n:n})}{\hat{\gamma}_n} \right) - \exp(-t) \right|, \quad \text{where} \\ \hat{\gamma}_n := \frac{1}{k_n} \sum_{i \in \mathcal{I}_1(Y_1, \dots, Y_n)} \log(Y_i/Y_{n-k_n:n})$$

is the Hill estimator, known to be  $\sqrt{k_n}$  consistent for  $\gamma := 1/\alpha$  if the first marginal of  $Q$  is  $\text{Pareto}(\alpha)$ . Write  $\mathcal{E}(\alpha)$  for the exponential with expectation  $1/\alpha$ . The idea behind  $\Delta_{1,n}$  is simple: for large  $y$ , if  $Y/y$  is approximately  $\text{Pareto}(\alpha)$  given  $Y > y$ , then  $\log(Y/y)/\gamma$  should be approximately  $\mathcal{E}(1)$ . Hence  $\Delta_{1,n}$  is a good candidate for testing whether the first margin of  $Q$  is Pareto. The independence of the marginals of  $Q$  will be tested with

$$\Delta_{2,n} := \sup_{t \geq 0, \boldsymbol{\theta} \in [0, 2\pi]^{d-1}} \left| \check{\mathbb{F}}_n(t, \boldsymbol{\theta}) - \check{\mathbb{F}}_n^{(1)}(t) \check{\mathbb{F}}_n^{(2)}(\boldsymbol{\theta}) \right|,$$

keeping in mind that  $\log$  is one to one from  $[1, +\infty[$  to  $[0, +\infty[$ . We will prove that, if  $Q$  is of the form  $\text{Pareto}(\alpha) \otimes \nu$

$$\sqrt{k_n}(\Delta_{1,n} \vee \Delta_{2,n}) \xrightarrow{d} \mathcal{Z}, \quad (44)$$

for a limit law  $\mathcal{Z}$ , under the following second order assumption which is the generalization of [23, Assumption 2.1] to dimension  $d$ : we assume that there exists a function  $\varepsilon$  tending to zero at infinity such that, for any choice of  $x_0 > 0$  one has

$$\sup_{\substack{x \geq x_0, \\ \boldsymbol{\theta} \in [0, 2\pi]^{d-1}}} \left| x^\alpha \frac{\mathbb{P}\left(\frac{\|X\|}{y} \in ]x, \infty[, \frac{X}{\|X\|_d} \in O_{\boldsymbol{\theta}}\right)}{\mathbb{P}(\|X\|_d > y)} - v(O_{\boldsymbol{\theta}}) \right| = O(\varepsilon(y)), \text{ as } y \rightarrow \infty. \quad (45)$$

Now write, for a probability measure  $P$

$$\check{F}_P(t, \boldsymbol{\theta}) := P(]t, +\infty[ \otimes O_{\boldsymbol{\theta}}), \quad t \in [0, +\infty[, \boldsymbol{\theta} \in [0, 2\pi]^{d-1}. \quad (46)$$

Proving (44) will be achieved in two steps: We prove a Donsker result (in weighted topology) for  $\sqrt{k_n}(\check{\mathbb{F}}_n - \check{F}_{\mathbf{m}(\alpha)})$ , where

$$\mathbf{m}(\alpha) := \mathcal{E}(\alpha) \otimes \nu, \quad (47)$$

then we express  $\Delta_{1,n} \vee \Delta_{2,n}$  as the norm of  $T(\check{\mathbb{F}}_n) - T(\check{F}_{\mathbf{m}(\alpha)})$ , where  $T$  is a properly differentiable map. In the sequel we shall denote by  $E_\eta$  the space of all real valued functions  $\phi$  on  $[0, +\infty[ \times \mathcal{S}_{d-1}$  for which the following quantity is finite

$$\|\psi\|_{\infty|\eta} := \sup_{t \geq 0, \boldsymbol{\theta} \in [0, 2\pi]^{d-1}} |\exp(\eta t) \psi(t, \boldsymbol{\theta})|,$$

making  $(E_\eta, \|\cdot\|_{\infty|\eta})$  a Banach space.

**Proposition 2.6 (a Donsker result in weighted topology)** *Assume that  $\varepsilon(\cdot)$  is regularly or slowly varying at infinity and that  $\sqrt{k_n} F^{\leftarrow}(1 - \frac{k_n}{n}) \rightarrow 0$ ,  $F$  denoting the c.d.f of  $Y = \|X\|_d$ . Then, for any  $\alpha' \in [0, \alpha[$  one has*

$$\sqrt{k_n}(\check{\mathbb{F}}_n - \check{F}_{\mathbf{m}(\alpha)}) \xrightarrow{d} \mathcal{W}, \text{ in } (E_{\alpha'}, \|\cdot\|_{\infty|\frac{\alpha'}{2}}), \quad (48)$$

where  $\mathcal{W}$  is the centered Gaussian process on  $[0, +\infty[ \times \mathcal{S}_{d-1}$  with covariance function

$$\text{cov}(\mathcal{W}(t_1, \boldsymbol{\theta}_1), \mathcal{W}(t_2, \boldsymbol{\theta}_2)) := \sqrt{\mathbf{P}_1\left(\left([t_1, +\infty[ \times O_{\boldsymbol{\theta}_1}\right) \Delta \left([t_2, +\infty[ \times O_{\boldsymbol{\theta}_2}\right)\right)},$$

the symbol  $\Delta$  standing here for the symmetric difference of sets.

Now the idea is, for fixed  $\alpha' < \alpha$  to consider the difference  $T(\check{\mathbb{F}}_n) - T(\check{F}_{\mathbf{m}(\alpha)})$ , where

$$T(\psi) := \left\{ (t, \boldsymbol{\theta}) \rightarrow \left( \psi \left( t \int_0^{+\infty} x d(-\psi^{(1)})(x) \right), \psi(t, \boldsymbol{\theta}) - \psi^{(1)}(t) \psi^{(2)}(\boldsymbol{\theta}) \right) \right\}, \psi \in D_{\alpha'}, \text{ and where}$$

$$D_{\alpha'} := \left\{ \psi \in E_{\alpha'}, \psi(0, \cdot) - \psi \text{ is a c.d.f on } [0, +\infty[ \times [0, 2\pi]^{d-1} \right\}.$$

Clearly  $\Delta_{1,n} \vee \Delta_{2,n}$  can be written as  $\| T(\check{\mathbb{F}}_n) - T(\check{F}_{\mathbf{m}(\alpha)}) \|_{\infty, \infty}$  with

$$\| (h_1, h_2) \|_{\infty, \infty} := \sup_{t \geq 0} | h_1(t) | \vee \sup_{t \geq 0, \boldsymbol{\theta} \in [0, 2\pi]^{d-1}} | h_2(t, \boldsymbol{\theta}) |.$$

Hence a proper functional Delta method could derive (44) from Proposition 2.6. We will make use of the notion of *quasi Hadamard differentiability* introduced and intensively investigated by Beutner and Zähle (see [1, 2, 3, 4, 5]), in order to take crucial advantage of the fact that  $\sqrt{k_n}(\check{\mathbb{F}}_n - \check{F}_{\mathbf{m}(\alpha)})$  takes its values in

$$E_{0, \frac{\alpha'}{2}} := \left\{ \psi \in E_{\frac{\alpha'}{2}}, \psi^{(1)} \text{ is càdlàg and } \psi^{(1)}(0) = 0 \right\}.$$

It will be checked that  $\mathcal{W}$  admits a version that takes its values in

$$C_{0, \frac{\alpha'}{2}} := \left\{ \psi \in E_{0, \frac{\alpha'}{2}}, \psi^{(1)} \text{ is uniformly continuous with respect to } d \right\}, \text{ where} \quad (49)$$

$$d(t_1, t_2) := | \exp(-\alpha t_1) - \exp(-\alpha t_2) |, (t_1, t_2) \in [0, +\infty]^2.$$

Our next proposition borrows notation from [4].

**Proposition 2.7 (Quasi Hadamard differentiability)** *Fix  $\alpha > 0$  and  $\alpha' \in [0, \alpha[$ . The map  $T$  defined on  $D_{\alpha'}$  is quasi Hadamard differentiable at  $\check{F}_{\mathbf{m}(\alpha)}$  tangentially to  $C_{0, \frac{\alpha'}{2}} \langle E_{0, \frac{\alpha'}{2}} \rangle$ , with derivative*

$$\text{Deriv}_{\check{F}_{\mathbf{m}(\alpha)}}(\psi) := \left\{ (t, \boldsymbol{\theta}) \rightarrow \left( -\alpha^{-1} t \exp(-t) \int \psi^{(1)}(x) dx + \psi^{(1)}(t/\alpha), \right. \right. \\ \left. \left. \psi(t, \boldsymbol{\theta}) - \check{F}_{\mathbf{m}(\alpha)}^{(1)}(t) \psi^{(2)}(\boldsymbol{\theta}) - \psi^{(1)}(t) \check{F}_{\mathbf{m}(\alpha)}^{(2)}(\boldsymbol{\theta}) \right) \right\},$$

for  $\psi \in C_{0, \frac{\alpha'}{2}}$ . In other words, if  $\epsilon_n \rightarrow 0$  is a sequence of positive numbers,  $(\psi_n)$  is a  $E_{0, \frac{\alpha'}{2}}$  valued sequence such that  $\| \psi_n - \psi \|_{\infty, \frac{\alpha'}{2}} \rightarrow 0$  for  $\psi \in C_{0, \frac{\alpha'}{2}}$ , then

$$\left\| \epsilon_n^{-1} \left( T(\check{F}_{\mathbf{m}(\alpha)} + \epsilon_n \psi_n) - T(\check{F}_{\mathbf{m}(\alpha)}) \right) - \text{Deriv}_{\check{F}_{\mathbf{m}(\alpha)}}(\psi) \right\|_{\infty, \infty} \rightarrow 0.$$

**Corollary 1** *Under  $\mathbf{H}_0$ , assertion (44) holds with*

$$\mathcal{Z} := \left\| \text{Deriv}_{\check{F}_{\mathbf{m}(\alpha)}}(\mathcal{W}) \right\|_{\infty, \infty} \\ = \sup_{t \geq 0} \left| - (t/\alpha) \exp(-t) \int \mathcal{W}^{(1)}(x) dx + \mathcal{W}^{(1)}(t/\alpha) \right| \\ + \sup_{t \geq 0, \boldsymbol{\theta} \in [0, 2\pi]^{d-1}} \left| \mathcal{W}(t, \boldsymbol{\theta}) - \check{F}_{\mathbf{m}(\alpha)}^{(1)}(t) \mathcal{W}^{(2)}(\boldsymbol{\theta}) - \mathcal{W}^{(1)}(t) \check{F}_{\mathbf{m}(\alpha)}^{(2)}(\boldsymbol{\theta}) \right|, \quad (50)$$

where  $\alpha$  is the Pareto index of  $\| X \|_d$  and where  $\mathcal{W}$  is as in Proposition 2.6.

The proof is postponed to Section 4.7.

We now address the question of approximating the limit law in (50) by an Efron bootstrap procedure. As it is written in Section 4.5 below, proof of Proposition 2.6 consists in proving that the conditions of Theorem 2 are fulfilled for the choice of  $E_0 := \{u_0\} = \{1\}$ , for  $\mathbf{P}_u$  as in (42) and for the class

$$\mathcal{F} := \left\{ f_{t,\boldsymbol{\theta}}, t \in [0, +\infty[, \boldsymbol{\theta} \in ]0, 2\pi]^{d-1} \right\}, \text{ with} \quad (51)$$

$$f_{t,\boldsymbol{\theta}}(x, s) := \exp\left(\frac{\alpha'}{2}t\right) \mathbf{1}_{]t, +\infty[}(x) \mathbf{1}_{O_{\boldsymbol{\theta}}}(s), (x, s) \in ]0, +\infty[ \times \mathcal{S}_{d-1},$$

and to use the identification  $\sqrt{k_n}(\check{\mathbb{F}}_n - \check{\mathbb{F}}_{m(\alpha)}) = \mathfrak{J}(\mathbb{G}_n^{(n)})$ , where

$$\mathfrak{J}(H) := \left\{ (t, \boldsymbol{\theta}) \rightarrow \exp(-\alpha't)H(f_{t,\boldsymbol{\theta}}) \right\}, H \in \ell^\infty(\mathcal{F}), \quad (52)$$

defines an isometry from  $(\ell^\infty(\mathcal{F}), \|\cdot\|_{\mathcal{F}})$  to  $(E_{\alpha'}, \|\cdot\|_{\infty|\alpha'})$ . Hence a direct byproduct is that Theorem 3 also holds in this setup. We will take advantage of this to prove the consistency of the following bootstrap procedure: for any finitely discrete probability measure  $P$  on  $]0, +\infty[ \times \mathcal{S}_{d-1}$  and  $k \geq 1$  write  $\tilde{\Delta}_k(P)$  for the bounded Lipschitz distance between the law of  $\mathcal{Z}$  and that of the following finitely discrete law

$$\mathcal{Z}_{k,P} := \sup_{t \geq 0} \left| - (t/\alpha(P)) \exp(-t) \int \psi_{k,P}^{(1)}(x) dx + \psi_{k,P}^{(1)}(t/\alpha(P)) \right|$$

$$+ \sup_{t \geq 0, \boldsymbol{\theta} \in ]0, 2\pi]^{d-1}} \left| \psi_{k,P}(t, \boldsymbol{\theta}) - \psi_{k,P}^{(1)}(t) \check{F}_P^{(2)}(\boldsymbol{\theta}) - \check{F}_P^{(1)}(t) \psi_{k,P}^{(2)}(\boldsymbol{\theta}) \right|, \text{ where}$$

$$\alpha(P)^{-1} := \int_{]0, +\infty[ \times \mathcal{S}_{d-1}} x dP(x, s), \quad (53)$$

$$(54)$$

and where - recalling (21) - the random element  $\psi_{k,P} := \mathfrak{J}(\mathbb{G}_{k,P})$  takes its finite number of values in  $E_{\alpha'}$ . Note that  $\tilde{\Delta}_{k_n}(\tilde{\mathbb{P}}_n^{(n)})$  is the quality of approximation of the law of  $\mathcal{Z}$  by that of  $\mathcal{Z}_{k_n,P}$  with  $P$  being the observed value of  $\tilde{\mathbb{P}}_n^{(n)}$ . That law can be computed by Monte-Carlo simulations. Our next proposition validates the consistency of this bootstrap procedure.

**Proposition 2.8 (Consistency of a bootstrap procedure)** *Under the assumptions of Corollary 1 one has*

$$(\tilde{\Delta}_{k_n}(\tilde{\mathbb{P}}_n^{(n)}))^* \xrightarrow{d} 0.$$

Its proof is postponed to Section 4.7.

## 2.4 The nearest neighbor rule for conditional c.d.f.

Take a generic Borel random couple  $(X, Z) \in \mathbb{R}^k \times \mathbb{R}^d$  and denote by  $] -\infty, x]$ , for  $x \in \mathbb{R}^d$ , the lower left orthant defined by  $x$ . For an element  $z \in \mathbb{R}^d$ , we are interested in the

estimation of the c.d.f  $F_z$  of a specific version of the law of  $X$  given  $Z = z$  by a nearest neighbor rule - see (55). Given a norm  $\|\cdot\|$  on  $\mathbb{R}^d$ , this simply consists in building an empirical c.d.f using only the data points  $X_i$  for which the values  $Y_i := (-\|Z_i - z\|)$  are among the  $k_n$ -highest (i.e. the  $\|Z_i - z\|$  are among the  $k_n$  smallest). This leads to the following estimator :

$$\hat{F}_z := \frac{1}{k_n} \sum_{i \in \mathcal{I}_1(Y_1, \dots, Y_n)} \mathbb{1}_{]-\infty, x]}(X_i) = \mathbb{P}_n^{(n)}(]-\infty, x]).$$

We here prove a Donsker theorem for  $\hat{F}_z$  under the following condition:

$$\lim_{h \rightarrow 0} \sup_{x \in \mathbb{R}^d} \left| F^{(h)}(x) - F_z(x) \right| = 0, \text{ where} \quad (55)$$

$$F^{(h)}(x) := \mathbb{P}\left(X \in ]-\infty, x] \mid Z \in B(z, h)\right), \quad x \in \mathbb{R}^d,$$

and where  $B(z, h)$  stands for the closed ball with center  $z$  and radius  $h$ . We also assume that  $z$  is a "true regression" point, namely

$$\mathbb{P}(Z \in B(z, h)) > 0, \text{ for all } h > 0, \quad (56)$$

as well as the following mild assumption (see discussion below)

$$\frac{\mathbb{P}(\|Z - z\| = h)}{\mathbb{P}(\|Z - z\| \leq h)} \rightarrow 0, \text{ as } h \downarrow 0. \quad (57)$$

**Proposition 2.9** *Assume that both (55), (56) and (57) hold. Then if  $k_n \rightarrow \infty$  and  $k_n/n \rightarrow 0$  one has*

$$\sqrt{k_n} \left( \hat{F}_z - F_z^{(1 - Y_{n - k_n : n})} \right) \xrightarrow{d} B, \text{ in } \ell^\infty(\mathbb{R}),$$

with  $B$  the  $F_z$ -Brownian bridge.

**Short proof:** Applying Proposition 2.2 with the choice of  $\phi(x, u) := x$ , we see that  $\mathbb{P}_n^{(n)}$  satisfies (CEM), with  $\mathbf{P}_u$  equal to the law of  $X$  given  $Y \geq F^{\leftarrow}(u)$ , (note that here  $F$  is the c.d.f of  $-\|Z - z\|$ ). By Assumption (55), we see that, for the choice of  $\mathbf{U} := u_0 = 1$  the collection  $\mathbf{P}_u$ ,  $u \in ]0, 1[$  satisfies (14) for the class  $\mathcal{F} := \{\mathbb{1}_{]-\infty, x]}, x \in \mathbb{R}^d\}$ , and for  $\mathbf{P}_1(]-\infty, x]) := F_z(x)$ ,  $x \in \mathbb{R}$ .  $\square$

**Discussion:** Let us now discuss upon conditions (55) and (56). These are met in several standard situations, for example:

1. When  $z$  is an atom for  $Z$  (note that  $\mathbb{P}(\|Z_z\| = h_n) \rightarrow 0$  for any  $(h_n)$  that is strictly decreasing to 0, as its series is convergent).
2. When  $(X, Z)$  admits a joint density  $f_{X,Z}$  that is continuous on  $\mathbb{R} \times O$ , where  $O$  is a neighborhood of  $z$ , and when the marginal density  $f_Z$  satisfies  $f_Z(z) > 0$ .

3. More interestingly, (55) and (56) also hold in the context of dimension reduction/sparsity for  $Z$ , namely when the law of  $Z$  concentrates on an unknown subspace (of unknown dimension) of  $G \subset \mathbb{R}^d$ , and when  $f_{X,Z}$  as above is defined on  $\mathbb{R} \times G$  and is a density with respect to the restriction (or trace) of the Lebesgue measure to  $\mathbb{R} \times G$ .

In this last case we see that  $\hat{F}_z$  may have advantage over its kernel-smoothing counterpart, namely (see, e.g., [27, Example 3])

$$F_{z,h_n}(x) := \frac{\frac{1}{nh_n^d} \sum_{i=1}^n \mathbb{1}_{]-\infty, x]} K\left(h_n^{-1}(Z_i - z)\right)}{\frac{1}{nh_n^d} \sum_{i=1}^n K\left(h_n^{-1}(Z_i - z)\right)}, \quad x \in \mathbb{R}^d,$$

The function  $K$  denoting a kernel (i.e. a Lebesgue density on  $\mathbb{R}^d$ ). Indeed, denoting by  $p$  the unknown dimension of  $G$  (hence the "true" dimension of the regressor  $Z$ ) one has  $\mathbb{P}(Z \in B(z, h_n)) \simeq f_Z(z)h_n^p$ . Hence the following Donsker theorem holds (see [?]).

$$\sqrt{f_Z(z)nh_n^p} \left( F_{z,h_n} - \mathbb{E}(F_{z,h_n}) \right) \xrightarrow{d} W. \quad (58)$$

However, the rate of convergence (ruling the size of simultaneous confidence intervals) is here completely unknown by the statistician, which makes such a kind of result useless in this context. This caveat does not affect  $\hat{F}_z$  as the rate of convergence is set to  $\sqrt{k_n}$  by the statistician and the construction of  $\hat{F}_z$  adapts to  $k_n$  through the thresholding with  $Y_{n-k_n:n}$ .

Note however that, before being useable in practice, Proposition 2.9 must be supported by

$$\sqrt{k_n} \sup_{x \in \mathbb{R}^d} \left| F_z^{(-Y_{n-k_n:n})}(x) - F_z(x) \right| \xrightarrow{d} 0. \quad (59)$$

This moves the problem of adapting to the dimension of  $Z$  into the bias: a proper choice of  $k_n$  must take  $p$  into account. A possible procedure could be to choose  $k_n = k(x_1, \dots, x_n)$  using a data driven selection method by minimizing in  $k$  a bootstrap estimation of this bias. This opens the door to the investigation of data driven procedures for Theorem 2, possibly through an improvement of this theorem to a uniformity in  $k$  in the same spirit of existing literature on uniform in bandwidth consistency of kernel-type estimators. This problem is beyond the scope of the present paper and will be the subject of future research.

### 3 Proofs related to Section 1

In this section we shall make use of the following notations to unburden formulas. For  $k \geq 1$  we will write  $\vec{x}_k$  for a generic  $k$ -tuple  $(x_1, \dots, x_k) \in \mathfrak{X}^k$ . We shall also write

$$P_{\vec{x}_k} := \frac{1}{k} \sum_{i=1}^k \delta_{x_i},$$

for the empirical measure built from  $\vec{x}_k$ .

### 3.1 Some useful measurability properties

Throughout all the proofs we will take a special care to get rid of outer expectations as soon as possible, in order to make an intensive use of a conditioning upon  $\mathbf{U}_n$  through the CEM assumption. This approach is possible by making an intensive use of the fact that  $\mathcal{F}$  is pointwise measurable - recall (5).

**Lemma 3.1** *Let  $r \geq 1$  and  $P \in \mathcal{M}$  such that  $P(F^r) < \infty$ . Then, for each  $\delta > 0$ , the class*

$$\mathcal{F}_{\delta,P} := \{(f - f'), (f, f') \in \mathcal{F} \times \mathcal{F}, \|f - f'\|_{P,r} < \delta\} \quad (60)$$

*is pointwise measurable with countable separant*

$$\check{\mathcal{F}}_{\delta,P} := \{(f - f'), (f, f') \in \check{\mathcal{F}} \times \check{\mathcal{F}} \mid \|f - f'\|_{P,r} < \delta\}.$$

**Proof :** Take  $(f, f')$  in  $\mathcal{F} \times \mathcal{F}$  such that  $\|f - f'\|_{P,2} < \delta$  and a sequence  $(f_m, f'_m)$  converging pointwise to  $(f, f')$ . Since  $P(F^r) < \infty$ , the dominated convergence theorem yields  $(f_m - f'_m) \in \check{\mathcal{F}}_{\delta,P}$  for all large  $m$ .  $\square$ .

**Lemma 3.2** *Let  $r \geq 1$  and  $P$  and  $P'$  be probability measures such that  $F^r$  is  $P + P'$  integrable. Then we have*

$$\left\| P' - P \right\|_{\mathcal{F}} = \left\| P' - P \right\|_{\check{\mathcal{F}}}, \quad (61)$$

*and for each  $\delta > 0$  we have*

$$\sup_{f \in \mathcal{F}_{\delta,P}} \|f\|_{P',r} = \sup_{f \in \check{\mathcal{F}}_{\delta,P}} \|f\|_{P',r}. \quad (62)$$

**Proof :** We first prove (61). Assume that, for some  $f \in \mathcal{F}$  one has

$$|P(f) - P'(f)| > \sup_{f \in \check{\mathcal{F}}} |P(f) - P'(f)|. \quad (63)$$

Consider a sequence  $(f_m) \in \check{\mathcal{F}}$  such that  $f_m \rightarrow f$  pointwise. Then, since  $F$  is both  $P$  and  $P'$ -integrable, the dominated convergence theorem entails  $|P(f) - P'(f)| = \lim |P(f_m) - P'(f_m)|$ , which contradicts (63). The proof of (62) uses similar arguments, we omit details.  $\square$

**Lemma 3.3** *Fix  $r > 0$ ,  $M > 0$ , and  $\epsilon > 0$ . For each  $P \in \mathfrak{M}$  one has*

$$\begin{aligned} \mathcal{N}(\epsilon, \mathcal{F}_M, \|\cdot\|_{P,r}) &= \mathcal{N}(\epsilon, \check{\mathcal{F}}_M, \|\cdot\|_{P,r}), \text{ where} \\ \check{\mathcal{F}}_M &:= \left\{ f \mathbf{1}_{\{F \leq M\}}, f \in \check{\mathcal{F}} \right\}. \end{aligned} \quad (64)$$

*As a consequence, the map  $P \rightarrow \mathcal{N}(\epsilon, \check{\mathcal{F}}_M, \|\cdot\|_{P,r})$  is measurable from  $(\mathfrak{M}, \mathcal{M})$  to  $\mathbb{N}$ .*

**Proof:** Fix  $r > 0$ ,  $M > 0$ ,  $\epsilon > 0$  and  $P \in \mathfrak{M}$ . Consider an arbitrary finite collection  $\tilde{\mathcal{F}}$  which covers  $\tilde{\mathcal{F}}_M$ , namely

$$\sup_{f \in \tilde{\mathcal{F}}_M} \min_{\tilde{f} \in \tilde{\mathcal{F}}} P(|f - \tilde{f}|^r) \leq \epsilon^r, \quad (65)$$

then this union of balls also covers  $\mathcal{F}_M$ . Indeed, if this was not the case, then there would exist  $f \in \mathcal{F}_M$  such that

$$\min_{\tilde{f} \in \tilde{\mathcal{F}}} P(f - \tilde{f})^r > \epsilon^r. \quad (66)$$

But, taking a sequence  $(f_m)$  of elements of  $\tilde{\mathcal{F}}_M$  converging pointwise to  $f$ , and recalling (65), there would exist  $\tilde{f} \in \tilde{\mathcal{F}}$  such that  $P(f_m - \tilde{f})^r \leq \epsilon^r$  for an infinity of indices  $m$ . On the other hand  $P(|f - f_m|^r) \rightarrow 0$  since all the involved functions are bounded by  $M$ . This would contradict (66). Conclude that (64) holds. Now the last point of Lemma 3.3 is proved by noticing that, for each  $K \in \mathbb{N}$ :

$$\begin{aligned} & \left\{ P \in \mathfrak{M}, \mathcal{N}(\epsilon, \tilde{\mathcal{F}}_M, \|\cdot\|_{P,r}) > K \right\} \\ &= \bigcap_{(f_1, \dots, f_K) \in \tilde{\mathcal{F}}^K} \bigcup_{f \in \tilde{\mathcal{F}}} \left\{ P \in \mathfrak{M}, P(|f - f_k|^r) > \epsilon^r \right\}. \square \end{aligned}$$

## 3.2 Bounds for empirical processes that are explicit in $P$

In this section we make use of a variation of the usual methodology of Koltchinski's Donsker theorem (see [36, Section 2.5.1, p. 127]). With all the measurability properties established in §3.1 at hand, both the finite dimensional convergence and the asymptotic tightness criteria will be expressed through non asymptotic bounds which are explicit in the sample law  $P$ . Several occurrences of a universal constant  $C$  will appear in this subsection. We will assume without loss of generality that we refer to the same constant everywhere in the sequel.

**Lemma 3.4 (Explicit bounds for Glivenko-Cantelli)** *Let  $P$  be a probability measure on  $(\mathfrak{X}, \mathcal{X})$  such that  $P(F) < \infty$ . Then the following quantities are well defined in the sense that their sets are properly measurable.*

$$GC_{\epsilon,k,\mathcal{F}}(P) := P^{\otimes k} \left( \left\{ \vec{x}_k \in \mathfrak{X}^k \mid P_{\vec{x}_k} - P \parallel_{\mathcal{F}} > \epsilon \right\} \right), \quad \epsilon > 0, k \geq 1, \quad (67)$$

$$Entr_{\epsilon,k,\mathcal{F},M}(P) := P^{\otimes k}(B_{\epsilon,k,M}), \quad \text{with}$$

$$B_{\epsilon,k,M} := \left\{ \vec{x}_k \in \mathfrak{X}^k, \sqrt{\frac{2 \log 2N(\epsilon^2, \mathcal{F}_M, \|\cdot\|_{P_{\vec{x}_k},1})}{k}} > \epsilon^2/M \right\}. \quad (68)$$

Moreover, for each  $M > 0$ ,  $\epsilon > 0$  and  $k \geq 1$  one has

$$GC_{\epsilon,k,\mathcal{F}}(P) \leq \frac{2}{\epsilon} P(F \mathbf{1}_{\{F > M\}}) + \frac{2M}{\epsilon} Entr_{\epsilon,k,\mathcal{F},M}(P) + 4\epsilon. \quad (69)$$

In addition, all the involved expressions in  $P$  in this lemma define  $\mathcal{M}$ -measurable maps on the (measurable) set  $M_F := \{P \in \mathfrak{M}, P(F) < \infty\}$ .

**Proof:** First, Lemma 3.2 ensures the claimed measurability of the set in (67). Indeed, for any  $\vec{x}_k$  one has both  $P(F) < \infty$  and  $P_{\vec{x}_k}(F) < \infty$ , from where

$$\{ \| P_{\vec{x}_k} - P \|_{\mathcal{F}} > \epsilon \} = \{ \| P_{\vec{x}_k} - P \|_{\tilde{\mathcal{F}}} > \epsilon \} = \bigcap_{f \in \tilde{\mathcal{F}}} \left\{ \left| \frac{1}{k} \sum_{i=1}^k f(x_i) - P(f) \right| > \epsilon \right\}.$$

Second, Lemma 3.3 combined with the fact that  $\vec{x}_k \rightarrow P_{\vec{x}_k}$  is  $\mathcal{X}^{\otimes k} \rightarrow \mathcal{M}$  measurable shows that  $B_{\epsilon, k, M} \in \mathcal{X}^{\otimes k}$  for each  $\epsilon, k, M$ . In addition, using when needed that, for fixed  $A \in \mathcal{X}^{\otimes k}$  the map  $P \rightarrow P^{\otimes k}(A)$  is Borel measurable, we see that each expression in  $P$  of Lemma 3.4 defines a  $\mathcal{M}$ -measurable map on  $M_F$ . Now fix  $\epsilon > 0$ ,  $M > 0$  and  $k \geq 1$ , then consider any  $(X_1, \dots, X_k) \rightsquigarrow P^{\otimes k}$  on a probability space. Next, note that one can represent  $GC_{\epsilon, k, \mathcal{F}}(P)$  by

$$GC_{\epsilon, k, \mathcal{F}}(P) = \mathbb{P} \left( \| P_{\vec{X}_k} - P \|_{\mathcal{F}} > \epsilon \right)$$

because measurability saves us from using outer probabilities, which themselves strongly depend upon the choice of underlying probability space. Now the proof of (69) is a slight improvement of the arguments of the proof of [36, Theorem 2.4.1, p. 122] to achieve a sufficient sharpness for its later use: A successive combination of symmetrization (see, e.g., [36, Lemma 2.3.1, p. 108]), the triangle inequality and Markov's inequality entail

$$GC_{\epsilon, k, \mathcal{F}}(P) \leq \frac{2}{\epsilon} \mathbb{E} \left( \sup_{f \in \mathcal{F}_M} \left| \frac{1}{k} \sum_{i=1}^k \epsilon_i f(X_i) \right| \right) + \frac{2}{\epsilon} P(F \mathbf{1}_{\{F > M\}}),$$

where  $(\epsilon_1, \dots, \epsilon_k)$  is a Rademacher (or symmetric Bernoulli) sample independent of  $(X_1, \dots, X_k)$ . Now for fixed  $\vec{x}_k \in \mathfrak{X}^k$  one has

$$\mathbb{E} \left( \sup_{f \in \mathcal{F}_M} \left| \frac{1}{k} \sum_{i=1}^k \epsilon_i f(x_i) \right| \right) \leq \epsilon^2 + \mathbb{E} \left( \max_{m=1, \dots, N} |Z_m| \right),$$

where  $N = N(\epsilon^2, \mathcal{F}_M, \|\cdot\|_{P_{\vec{x}_k, 1}})$  and each  $Z_m$  is of the form  $k^{-1} \sum \epsilon_i f_m(x_i)$ ,  $(f_1, \dots, f_m)$  being the centers of balls with radii  $\epsilon^2$  covering  $\mathcal{F}_M$ . Then apply [18, p. 7, Lemma 2.2] to  $Z_1, \dots, Z_N$ , with  $\sigma^2 = M/k$ , to obtain

$$\mathbb{E} \left( \sup_{f \in \mathcal{F}_M} \left| \frac{1}{k} \sum_{i=1}^k \epsilon_i f(x_i) \right| \right) \leq M \wedge \left[ \epsilon^2 + M \sqrt{\frac{2 \log 2N(\epsilon^2, \mathcal{F}_M, \|\cdot\|_{P_{\vec{x}_k, 1}})}{k}} \right],$$

where the first occurrence of  $M$  in this bound comes from the fact that  $\mathcal{F}_M$  is uniformly bounded by  $M$ . Now the RHS is bounded by  $M$  on  $B_{\epsilon, k, M}$  and is bounded by  $\epsilon^2 + M\epsilon^2/M$  on  $B_{\epsilon, k, M}^C$  - recall (68). Hence integrating respectively on those two sets concludes the proof.  $\square$

**Lemma 3.5 (From Ray and Van der Vaart, 2021)** *Let  $P$  be such that  $P(F^2) < \infty$ . For any  $p \geq 1$  and for any  $\mathbf{f} = (f_1, \dots, f_p) \in \mathcal{F}^p$  define*

$$Fidi_{\mathbf{f},k}(P) := \sup_{\phi \in BL_1(\mathbb{R}^p)} \left| \mathbb{E} \left( \phi \left( \sqrt{k} \left( \frac{1}{k} \sum_{i=1}^k Z_i - \mathbb{E}(Z_1) \right) \right) \right) - \mathbb{E} \left( \phi(W) \right) \right|,$$

where  $Z_i := (f_1(X_i), \dots, f_p(X_i))$ , with  $(X_1, \dots, X_k) \rightsquigarrow P^{\otimes k}$ , and where  $W$  is centered Gaussian with the covariances of  $Z_1$ .

For each  $\epsilon > 0$  and  $k \geq 1$  one has, for a universal constant  $C > 0$

$$Fidi_{\mathbf{f},k}(P) \leq 2 \wedge C \left[ \epsilon + \frac{1}{\epsilon^2} H \left( P(F^2 \mathbf{1}_{\{F > \sqrt{k}\epsilon\}}) \right) + H_0(\epsilon P(F^2)) \right], \quad (70)$$

where  $H_0(u) := u + u^{1/3}$  and  $H(u) := u^{1/4}(1 + |\log(u)|^{1/2})$ . In addition, all the terms of the RHS of (70) define  $\mathcal{M}$ -measurable maps on  $M_{F^2} := \{P \in \mathfrak{M}, P(F^2) < \infty\}$ .

**Proof:** The claimed measurability is trivial, while the bound is an immediate consequence of [34, Lemma 1].  $\square$

For our next result recall (60) and notice that, since  $J_{\mathcal{F},F}(\infty) < \infty$ , usual arguments upon covering numbers show that

$$J_{\mathcal{F}_\infty, 2F}(\infty) < \infty, \text{ where } \mathcal{F}_\infty := \left\{ f_1 - f_2, (f_1, f_2) \in \mathcal{F}^2 \right\}. \quad (71)$$

In the sequel we shall use the notation

$$\mathcal{F}_\infty^{\bullet 2} := \left\{ g^2, g \in \mathcal{F}_\infty \right\},$$

and we will keep the generic notation " $\bullet 2$ " for taking the squares of elements of a class of function.

**Lemma 3.6 (Explicit bound for asymptotic tightness)** *For fixed  $P \in \mathfrak{M}$ ,  $\delta > 0$ ,  $\epsilon > 0$  and  $k \geq 1$  the quantity*

$$Osc_{\delta, \epsilon, k}(P) := P^{\otimes k} \left( \left\{ \vec{x}_k \in \mathfrak{X}^k, \sqrt{k} \| P_{\vec{x}_k} - P \|_{\mathcal{F}_{\delta, P}} > \epsilon \right\} \right)$$

is well defined. Moreover one has, for each  $\epsilon > 0$ ,  $M > 0$  and for each  $0 < \delta < \epsilon^2/4$ :

$$Osc_{\delta, \epsilon, k}(P) \leq C \left[ GC_{\delta^2, k, \mathcal{F}_\infty^{\bullet 2}}(P) + \frac{P(F^2)}{M^2} + \frac{1}{\epsilon} M J_{\mathcal{F}_\infty, 2F}(2\delta) \right], \quad (72)$$

for some universal constant  $C$ . In addition each expression in the RHS of (72) defines a  $\mathcal{M}$ -measurable map on  $M_{F^2}$ .

**Proof:** Fix  $P \in \mathcal{M}$ ,  $\epsilon > 0$ ,  $\delta \in (0, \epsilon^2/4)$  and  $k \geq 1$ . By the same measurability arguments as those in the proof of Lemma 3.4 we can represent

$$Osc_{\delta, \epsilon, k}(P) = \mathbb{P}\left(\sqrt{k} \left\| \frac{1}{k} \sum_{i=1}^k \delta_{X_i} - P \right\|_{\mathcal{F}_{\delta, P}} > \epsilon\right),$$

with  $(X_1, \dots, X_k) \rightsquigarrow P^{\otimes k}$ . Now, introducing a Rademacher sample  $(\epsilon_1, \dots, \epsilon_k)$  independent of  $(X_1, \dots, X_k)$  one has, using symmetrization for probabilities

$$\beta_k(\epsilon) \mathbb{P}\left(\sqrt{k} \left\| \frac{1}{k} \sum_{i=1}^k \delta_{X_i} - P \right\|_{\mathcal{F}_{\delta, P}} > \epsilon\right) \leq 2\mathbb{P}\left(\sqrt{k} \left\| \frac{1}{k} \sum_{i=1}^k \epsilon_i \delta_{X_i} \right\|_{\mathcal{F}_{\delta, P}} > \epsilon\right), \quad (73)$$

where  $\beta_k(\epsilon) \geq 1 - (2\delta/\epsilon^2)^2 \geq 1/2$ . To see this, use [36, p. 112, Lemma 2.3.7], borrowing their notation, with  $\sup\{Var(Z_1(f)), f \in \mathcal{F}_{\delta, P}\} \leq \delta^2$ . As a consequence of (73) one has

$$Osc_{\delta, \epsilon, k}(P) \leq 4 \int_{\vec{x}_k \in \mathfrak{X}^k} \mathbb{P}_{\epsilon}\left(\sqrt{k} \sup_{f \in \mathcal{F}_{\delta, P}} \left| \frac{1}{k} \sum_{i=1}^k \epsilon_i f(x_i) \right| > \epsilon\right) dP^{\otimes k}(\vec{x}_k).$$

Now, writing

$$A_1 := \left\{ \vec{x}_k \in \mathfrak{X}^k, \sup_{g \in \mathcal{F}_{\delta, P}^{\bullet 2}} |P_{\vec{x}_k}(g) - P(g)| \leq \delta^2 \right\},$$

$$A_2 := \left\{ \vec{x}_k \in \mathfrak{X}^k, P_{\vec{x}_k}(F^2) \leq M^2 \right\},$$

Bonferroni and Markov's inequalities yield

$$Osc_{\delta, \epsilon, k}(P) \leq 4P^{\otimes k}(A_1^C) + 4P^{\otimes k}(A_2^C)$$

$$+ 4 \int_{A_1 \cap A_2} \frac{1}{\epsilon} \mathbb{E}_{\epsilon}\left(\sqrt{k} \sup_{f \in \mathcal{F}_{\delta, P}} \left| \frac{1}{k} \sum_{i=1}^k \epsilon_i f(x_i) \right|\right) dP^{\otimes k}(\vec{x}_k).$$

The first term is exactly  $4GC(\delta^2, k, \mathcal{F}_{\delta, P}^{\bullet 2}) \leq 4GC(\delta^2, k, \mathcal{F}_{\infty}^{\bullet 2})$ , while the second is obviously bounded by  $4M^{-2}P(F^2)$ . To bound the third term notice that, for  $\vec{x}_k \in A_1$ , the triangle inequality yields

$$\sup_{g \in \mathcal{F}_{\delta, P}^{\bullet 2}} P_{\vec{x}_k}(g) \leq 2\delta^2. \quad (74)$$

Hence, for such  $\vec{x}_k$  one can use the arguments of [36, p. 127-128] to conclude that for a universal constant  $C$

$$\mathbb{E}_{\epsilon}\left(\sqrt{k} \sup_{f \in \mathcal{F}_{\delta, P}} \left| \frac{1}{k} \sum_{i=1}^k \epsilon_i f(x_i) \right|\right) \leq C \mathbb{E}\left(J_{\mathcal{F}_{\infty}, 2F}(\theta_{\vec{x}_k}) \| F \|_{P_{\vec{x}_k}, 2}\right),$$

where  $\theta_{\vec{x}_k}^2 := \frac{\sup_{f \in \mathcal{F}_{\delta, P}} P_{\vec{x}_k}(f^2)}{P_{\vec{x}_k}(F^2)} \leq 2\delta^2$ ,

and where the last inequality comes from (74) together with  $F \geq 1$ . Next, if  $\vec{x}_k \in A_1 \cap A_2$  one has

$$J_{\mathcal{F}_\infty, 2F}(\theta_{\vec{x}_k}) \| F \|_{P_{\vec{x}_k, 2}} \leq M J_{\mathcal{F}_\infty, 2F}(2\delta),$$

which yields (72) by integration in  $\vec{x}_k$ .  $\square$

### 3.3 Proof of Theorem 1

Fix  $\epsilon > 0$ . Recall that  $Q_n$  stands for the law of  $\mathbf{U}_n$ . By the CEM assumption we can condition upon  $\mathbf{U}_n$  to obtain:

$$\begin{aligned} & \mathbb{P}\left(\left\|\mathbb{P}_n^{(n)} - \mathbf{P}_{\mathbf{U}_n}\right\|_{\mathcal{F}} > \epsilon\right) \\ &= \int_E GC_{\epsilon, k_n}(\mathbf{P}_u) dQ_n(u) \\ &\leq \int_E 1 \wedge \left[ \frac{2}{\epsilon} \mathbf{P}_u(F \mathbb{1}_{\{F > M\}}) + \frac{2M}{\epsilon} \text{Entr}_{\epsilon, k_n, \mathcal{F}, M}(\mathbf{P}_u) + 4\epsilon \right] dQ_n(u), \end{aligned} \quad (75)$$

where the last bound uses Lemma 3.4 for fixed  $\mathbf{P}_u \in E$ . Now we have, by (9) together with the bounded convergence theorem (see., e.g. [44, p.130]):

$$\lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E}\left(1 \wedge \frac{2}{\epsilon} \mathbf{P}_{\mathbf{U}_n}(F \mathbb{1}_{\{F > M\}})\right) = 0.$$

Moreover, for fixed  $M > 0$ , condition (10) in conjunction with the bounded convergence theorem entails

$$\lim_{n \rightarrow \infty} \mathbb{E}\left(1 \wedge \text{Entr}_{\epsilon, k_n, \mathcal{F}, M}(\mathbf{P}_{\mathbf{U}_n})\right) = 0.$$

This proves that the limsup of the RHS of (75) is less than  $3\epsilon$ .  $\square$

### 3.4 Proof of Theorem 2

#### Step 1: Finite dimensional convergence

Recall that  $\mathbb{G}_n^{(n)}$  is defined in (11). For any  $p \geq 1$  and  $\mathbf{f} = (f_1, \dots, f_p) \in \mathcal{F}^p$  we shall write

$$\mathbb{G}_n^{(n)}(\mathbf{f}) := \left(\mathbb{G}_n^{(n)}(f_1), \dots, \mathbb{G}_n^{(n)}(f_p)\right).$$

**Proposition 3.1** *For any  $p \geq 1$  and  $\mathbf{f} = (f_1, \dots, f_p) \in \mathcal{F}^p$  we have*

$$\mathbb{G}_n^{(n)}(\mathbf{f}) \xrightarrow{d} \mathbb{G}_{\mathbf{P}_U}(\mathbf{f}),$$

**Proof:** Write  $Q$  for the law of  $\mathbf{U}$ . For fixed  $u \in E$  denote by  $\mathbf{P}_u^{\mathbf{f}}$  the image of  $\mathbf{P}_u$  by  $\mathbf{f}$ , and  $\Sigma_{\mathbf{f}, u}[k, \ell] := \mathbf{P}_u(f_k f_\ell) - \mathbf{P}_u(f_k) \times \mathbf{P}_u(f_\ell)$  for its covariance matrix. Fix  $\mathbf{t} \in \mathbb{R}^d$ . The CEM assumption entails, for each  $n \geq 1$ :

$$\mathbb{E}\left(\exp\left(\mathbf{i}\sqrt{k_n}\mathbb{G}_n^{(n)}(\langle \mathbf{t}, \mathbf{f} \rangle)\right)\right) = \int_E \phi_u\left(\frac{1}{\sqrt{k_n}}\mathbf{t}\right)^{k_n} dQ_n(u),$$

where  $\phi_u$  stands for the characteristic function of  $\mathbf{P}_u^{\mathbf{f}}$ . Since all the integrands are uniformly bounded by 1 the expected convergence

$$\mathbb{E}\left(\exp\left(\mathbf{i}\sqrt{k_n}\mathbb{G}_n^{(n)}(\langle \mathbf{t}, \mathbf{f} \rangle)\right)\right) \rightarrow \int_E \exp\left(-\frac{\mathbf{t}\Sigma_{\mathbf{f},u}\mathbf{t}^\top}{2}\right)dQ(u)$$

will be established if we prove separately that

$$\begin{aligned} \int_E \exp\left(-\frac{\mathbf{t}\Sigma_{\mathbf{f},u}\mathbf{t}^\top}{2}\right)dQ_n(u) &\rightarrow \int_E \exp\left(-\frac{\mathbf{t}\Sigma_{\mathbf{f},u}\mathbf{t}^\top}{2}\right)dQ(u), \text{ and} \\ \int_E \left|\phi_u\left(\frac{1}{\sqrt{k_n}}\mathbf{t}\right)^{k_n} - \exp\left(-\frac{\mathbf{t}\Sigma_{\mathbf{f},u}\mathbf{t}^\top}{2}\right)\right|dQ_n(u) &\rightarrow 0. \end{aligned} \quad (76)$$

Assertion (76) can be seen as  $\mathbb{E}(H'(\mathbf{U}_n)) \rightarrow \mathbb{E}(H'(\mathbf{U}))$  where  $H' : u \rightarrow \exp\left(-\frac{\mathbf{t}\Sigma_{\mathbf{f},u}\mathbf{t}^\top}{2}\right)$  is bounded and continuous at each point of  $E_0$  by a direct use of assumption (14). Therefore (76) holds (apply, e.g. [36, Theorem 1.11.1, p 67], with  $g_n$  constantly equal to  $H'$ ). Now using Lemma 3.5 for fixed  $\mathbf{P}_u$  and integrating we obtain, for any  $\epsilon > 0$  and for each  $n \geq 1$ .

$$\begin{aligned} &\int_E \left|\phi_u\left(\frac{1}{\sqrt{k_n}}\mathbf{t}\right)^{k_n} - \exp\left(-\frac{\mathbf{t}\Sigma_{\mathbf{f},u}\mathbf{t}^\top}{2}\right)\right|dQ_n(u) \\ &\leq \int_E \|\mathbf{t}\|_p \text{Fidi}_{\mathbf{f},k_n}(\mathbf{P}_u)dQ_n(u) \\ &\leq \|\mathbf{t}\|_p \int_E 2 \wedge C \left[\epsilon + \frac{1}{\epsilon^2}H\left(\mathbf{P}_u(F^2\mathbf{1}_{\{F>\sqrt{k_n}\epsilon}\})\right) + H_0(\epsilon\mathbf{P}_u(F^2))\right]dQ_n(u). \end{aligned} \quad (77)$$

Now from (9) and the continuity and boundedness of  $2 \wedge CH(\cdot)$  one has for fixed  $\epsilon > 0$

$$\mathbb{E}\left(H\left(\mathbf{P}_{\mathbf{U}_n}(F^2\mathbf{1}_{\{F>\sqrt{k_n}\epsilon}\})\right)\right) \rightarrow 0,$$

by the bounded convergence theorem. Moreover (9) implies that  $\mathbf{P}_{\mathbf{U}_n}(F^2)$  is a stochastically bounded sequence, from where by continuity and boundedness of  $2 \wedge CH_0(\cdot)$  together with the bounded convergence theorem

$$\lim_{\epsilon \downarrow 0} \limsup_{n \rightarrow \infty} \mathbb{E}\left(2 \wedge H_0(\epsilon\mathbf{P}_{\mathbf{U}_n}(F^2))\right) = 0.$$

this proves that the lim sup of (77) can be made arbitrarily small, and hence concludes the proof.  $\square$

### Step 2: Asymptotic tightness

Note that the tightness of  $\mathbb{G}_{\mathbf{P}_U}$  will be a byproduct of the asymptotic tightness of  $\mathbb{G}_n^{(n)}$  that will be proved in this subsection (see, e.g. [36, Theorem 1.3.9, p. 21]). Fix  $\epsilon > 0$  once for all. We will prove that the sequence  $\mathbb{G}_n^{(n)}$  is asymptotically tight if we exhibit a finite subclass  $\tilde{\mathcal{F}} \subset \mathcal{F}$  such that

$$\limsup_{n \rightarrow \infty} \mathbb{P}\left(\sup_{f \in \mathcal{F}} \min_{\tilde{f} \in \tilde{\mathcal{F}}} \left|\mathbb{G}_n^{(n)}(f - \tilde{f})\right| \geq 2\epsilon\right) \leq 3(1 + C)\epsilon. \quad (78)$$

We first start by noting that (12) entails the existence of  $M$  as well as  $n_0 \geq 1$  such that

$$\forall n \geq n_0, \mathbb{P}\left(\mathbf{U}_n \notin A_0\right) < \epsilon, \text{ where } A_0 := \left\{u \in E, \frac{\mathbf{P}_u(F^2)}{M^2} < \epsilon\right\}. \quad (79)$$

Throughout the rest of our proof of (78),  $M$  will be fixed and chosen accordingly to (79). By (71) one can choose  $0 < \delta < \epsilon^2/4$  such that

$$\frac{M}{\epsilon} J_{\mathcal{F}_\infty, 2F}(2\delta) < \epsilon. \quad (80)$$

Our construction of  $\tilde{\mathcal{F}}$  will be achieved through the following lemma.

**Lemma 3.7** *There exists  $n_1 \geq 1$  and  $\tilde{\mathcal{F}} \subset \mathcal{F}$  finite such that*

$$\begin{aligned} \forall n \geq n_1, \mathbb{P}\left(\mathbf{U}_n \notin A_1\right) < \epsilon, \text{ where} \\ A_1 := \left\{u \in E, \sup_{f \in \mathcal{F}} \min_{\tilde{f} \in \tilde{\mathcal{F}}} \|f - \tilde{f}\|_{\mathbf{P}_{u,2}} \leq 2\delta\right\}. \end{aligned}$$

**Proof:** Since  $\mathbf{U}_n \xrightarrow{d} \mathbf{U}$  and since  $E_0$  is a (separable) support of  $\mathbf{U}$  we can exhibit a compact set  $K \subset E_0$  such that,

$$\limsup_{n \rightarrow \infty} \mathbb{P}\left(\mathbf{U}_n \notin K\right) < \epsilon. \quad (81)$$

The compactness of  $K$  together with the continuity condition (14) entails the existence of a finite  $\tilde{K} \subset K$  such that

$$u \in K \Rightarrow \exists u' \in \tilde{K}, \left\| \mathbf{P}_u - \mathbf{P}_{u'} \right\|_{\mathcal{G}} \leq \delta.$$

Now, for each  $u' \in \tilde{K}$ , since  $\mathcal{F}$  is totally bounded for  $\|\cdot\|_{\mathbf{P}_{u',2}}$  (a consequence of  $J_{\mathcal{F},F}(\infty) < \infty$ ), we can exhibit a finite subset of  $\tilde{\mathcal{F}}_{u'} \subset \mathcal{F}$  for which the union of their  $\|\cdot\|_{\mathbf{P}_{u',2}}$  open balls with radius  $\delta$  covers  $\mathcal{F}$ . Take  $\tilde{\mathcal{F}}$  as the union of all  $\tilde{\mathcal{F}}_{u'}$  for  $u' \in \tilde{K}$ . By this construction, if one takes  $f \in \mathcal{F}$  and  $u \in \tilde{K}$  then there exists  $\tilde{f} \in \tilde{\mathcal{F}}$  such that  $\|f - \tilde{f}\|_{\mathbf{P}_{u,2}} \leq \delta$ . Note that  $K \subset A_1$  and use (81) to conclude the proof.  $\square$

We now choose  $\tilde{\mathcal{F}}$  and  $A_1$  as enabled by the preceding lemma. Since  $\mathcal{G} \supset \mathcal{F}_\infty^{\bullet 2}$  satisfies the conditions of Theorem 1, there exists  $n_2$  such that

$$\begin{aligned} \forall n \geq n_2, \mathbb{P}\left(\mathbf{U}_n \notin A_{2,n}\right) < \epsilon, \text{ where} \\ A_{2,n} := \left\{u \in E, GC_{\delta^2, k_n, \mathcal{F}_\infty^{\bullet 2}}(\mathbf{P}_u) < \epsilon\right\}. \end{aligned} \quad (82)$$

We now have all the required tools to prove that our choice of  $\tilde{\mathcal{F}}$  satisfies (78). For fixed  $n \geq \max\{n_0, n_1, n_2\}$  we have, in view of (79),(82) and Lemma 3.7:

$$\begin{aligned} & \mathbb{P}\left(\sup_{f \in \mathcal{F}} \min_{\tilde{f} \in \tilde{\mathcal{F}}} \left| \sqrt{k_n} \left( \mathbb{P}_n^{(n)} - \mathbf{P}_{\mathbf{U}_n} \right) (f - \tilde{f}) \right| > 2\epsilon\right) \\ & \leq 3\epsilon + \int_{A_0 \cap A_1 \cap A_{2,n}} O_{Sc_{\epsilon, 2\delta, k_n}}(\mathbf{P}_u) dQ_n(u), \end{aligned}$$

where the majoration by  $Osc_{\epsilon, 2\delta, k_n}(\mathbf{P}_u)$ , for fixed  $u \in A_1$ , is immediate from the construction of  $\tilde{\mathcal{F}}$ . Now, using Lemma 3.6 for fixed  $\mathbf{P}_u$  yields

$$\begin{aligned} & \int_{A_0 \cap A_1 \cap A_{2,n}} Osc_{\epsilon, 2\delta, k_n}(\mathbf{P}_u) dQ_n(u) \\ & \leq \int_{A_0 \cap A_1 \cap A_{2,n}} C \left[ GC_{\delta^2, \mathcal{F}_\infty, 4F^2}(\mathbf{P}_u) + \frac{\mathbf{P}_u(F^2)}{M^2} + \frac{1}{\epsilon} MJ_{\mathcal{F}_\infty, 2F}(2\delta) \right] dQ_n(u) \\ & \leq 3C\epsilon, \end{aligned}$$

by definition of  $A_0$  and  $A_{2,n}$  and since  $\delta$  satisfies (80). This concludes the proof of (78), and hence finishes the proof of Theorem 2.

### 3.5 Proof of Theorem 3

Our proof is based on the following split for fixed  $P \in \mathfrak{M}_0$  and  $k \geq 1$

$$\begin{aligned} Donsker_k(P) & \leq \sup_{\phi \in BL_1(\ell^\infty(\mathcal{F}))} \left| \mathbb{E} \left( \phi \left( \sqrt{k} \left( \frac{1}{k} \sum_{i=1}^k \delta_{X_i} - P \right) \right) \right) - \mathbb{E} \left( \phi(\mathbb{G}_P) \right) \right| \\ & \quad + \sup_{\phi \in BL_1(\ell^\infty(\mathcal{F}))} \left| \mathbb{E} \left( \phi(\mathbb{G}_P) \right) - \mathbb{E} \left( \phi(\mathbb{G}_{\mathbf{P}_{u_0}}) \right) \right| \\ & =: A_k(P) + B(P). \end{aligned}$$

Now the idea of the proof is to take  $k := k_n$  and substitute  $P$  by  $\mathbb{P}_n^{(n)}$  in the preceding expression, then bound the random elements  $A_{k_n}(\mathbb{P}_n^{(n)})$  and  $B(\mathbb{P}_n^{(n)})$  *everywhere* on  $\Omega$  by measurable random variables converging to 0 in probability.

**Step 1: bounds for  $A_{k_n}(\mathbb{P}_n^{(n)})$ .**

We now make use - for fixed  $P$  and  $k$  - of the arguments of, e.g., [36, Proof of Theorem 2.9.6, p. 182], recalling that  $(\mathcal{F}, \|\cdot\|_{P,2})$  is totally bounded since  $J_{\mathcal{F}, F}(\infty) < \infty$ . We hence have, for fixed  $\epsilon > 0$  and  $\delta > 0$

$$\begin{aligned} A_k(P) & \leq Osc_{\delta, \epsilon, k}(P) + \mathbb{P} \left( \|\mathbb{G}_P\|_{\mathcal{F}_{\delta, P}} > \epsilon \right) + 2\epsilon + \sup_{p \geq 1, \mathbf{f} \in \mathcal{F}^p} Fidi_{k, \mathbf{f}}(P) \\ & \leq Osc_{\delta, \epsilon, k}(P) + \frac{\mathbb{E} \left( \|\mathbb{G}_P\|_{\mathcal{F}_{\delta, P}} \right)}{\epsilon} + 2\epsilon \\ & \quad + C \left[ \epsilon + \frac{1}{\epsilon^2} H \left( P(F^2 \mathbb{1}_{\{F > \sqrt{k}\epsilon\}}) \right) + H_0(\epsilon P(F^2)) \right], \end{aligned}$$

where the last bound is a consequence of Lemma 3.5. Now using, e.g., [36, Corollary 2.2.5, p. 98] we have

$$\mathbb{E} \left( \|\mathbb{G}_P\|_{\mathcal{F}_{\delta, P}} \right) \leq C \int_0^{diam(\mathcal{F}, \|\cdot\|_{P,2})} \sqrt{\log(2N(\epsilon, \mathcal{F}, \|\cdot\|_{P,2}))} d\epsilon,$$

$C$  denoting a universal constant for which we can assume without loss of generality that it has the same value as each of the universal constants defined up to now. By the change of variable  $\epsilon' = \epsilon \|F\|_{P,2}$  and since  $F \geq 1$  implies

$$\text{diam}(\mathcal{F}_{\delta,P}, \|\cdot\|_{P,2}) / \|F\|_{P,2} \leq \delta,$$

we conclude that

$$\begin{aligned} A_k(P) \leq & \text{Osc}_{\delta,\epsilon,k}(P) + \frac{C}{\epsilon} \|F\|_{P,2} J_{\mathcal{F},F}(\delta) + 2\epsilon \\ & + C \left[ \epsilon + \frac{1}{\epsilon^2} H\left(P(F^2 \mathbf{1}_{\{F > \sqrt{k}\epsilon}\})\right) + H_0(\epsilon P(F^2)) \right], \end{aligned}$$

where the RHS defines a Borel measurable map on  $(M_{F^2}, \mathcal{M})$ . As a consequence,  $A_{k_n}(\mathbb{P}_n^{(n)})$  will be proved to converge to zero in outer probability if we prove the following lemma (recall that  $A_k(P)$  is always less than 2).

**Lemma 3.8** *The following sequences of Borel random variables satisfy*

$$\forall \epsilon > 0, \lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty}^{\mathbb{P}} \text{Osc}_{\delta,\epsilon,k_n}(\mathbb{P}_n^{(n)}) = 0, \quad (83)$$

$$\forall \epsilon > 0, \mathbb{P}_n^{(n)}(F^2 \mathbf{1}_{\{F > \sqrt{k}\epsilon}\}) \xrightarrow{d} 0, \quad (84)$$

$$\limsup_{n \rightarrow \infty}^{\mathbb{P}} \mathbb{P}_n^{(n)}(F^2) < \infty. \quad (85)$$

**Proof:** Notice that, for fixed  $u \in E$ , and  $M > 0$ , if  $(X_1, \dots, X_{k_n}) \rightsquigarrow \mathbf{P}_u^{\otimes k_n}$  then

$$\mathbb{E} \left( \frac{1}{k_n} \sum_{i=1}^{k_n} F^2(X_i) \mathbf{1}_{\{F^2(X_i) > M\}} \right) = \mathbf{P}_u(F^2 \mathbf{1}_{\{F^2 > M\}}).$$

As a consequence we have almost surely

$$\mathbb{E} \left( \mathbb{P}_n^{(n)}(F^2 \mathbf{1}_{\{F^2(X_i) > M\}}) \middle| \mathbf{U}_n \right) = \mathbf{P}_{\mathbf{U}_n}(F^2 \mathbf{1}_{\{F^2(X_i) > M\}}),$$

which proves both (84) and (85) by (12). Now using (72) with sufficiently large  $M$  for fixed  $\epsilon > 0$  we see that (83) will be proved if we establish, for fixed  $\epsilon > 0$ :

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty}^{\mathbb{P}} GC_{\delta^2, k_n, \mathcal{F}_{\infty}^{\bullet 2}}(\mathbb{P}_n^{(n)}) = 0. \quad (86)$$

In view of (69) it will be sufficient to prove that, for fixed  $\delta > 0$  and  $M > 0$  one has

$$\text{Entr}_{\delta^2, k_n, \mathcal{F}_{\infty}^{\bullet 2}, M}(\mathbb{P}_n^{(n)}) \rightarrow 0. \quad (87)$$

Now fix  $k \geq 1$  and  $\vec{x}_k \in \mathfrak{X}^k$ . Since, for fixed  $f$  and  $g$  of  $\mathcal{F}_{\infty}$  one has

$$\|f^2 - g^2\|_{P_{\vec{x}_k}, 1} = \|(f-g)(f+g)\|_{P_{\vec{x}_k}, 1} \leq \|f-g\|_{P_{\vec{x}_k}, 2} \times \|4f\|_{P_{\vec{x}_k}, 2},$$

comparison of covering numbers leads to the following bound

$$N(\delta^4, \mathcal{F}_\infty^{\bullet 2}, \|\cdot\|_{P_{\vec{x}_k, 1}}) \leq N(4\delta^4 \|F\|_{P_{\vec{x}_k, 2}}, \mathcal{F}_\infty, \|\cdot\|_{P_{\vec{x}_k, 2}}),$$

from where, by monotonicity of the integrand in the definition of  $J_{\mathcal{F}_\infty, 2F}$

$$\sqrt{\log 2N(\delta^4, \mathcal{F}_\infty^{\bullet 2}, \|\cdot\|_{P_{\vec{x}_k, 1}})} \leq \frac{1}{2\delta^4} J_{\mathcal{F}_\infty, 2F}(2\delta^4),$$

which does not depend upon  $\vec{x}_k$ . As a consequence, the set  $B_{\delta^2, M, k_n}$  is empty for all  $n$  large enough - recall (68) - implying  $\text{Entr}_{\delta^2, k_n, \mathcal{F}_\infty^{\bullet 2}, M}(P) = 0$  for *any*  $P \in \mathcal{M}$ . This is more than sufficient to prove (86).  $\square$

**Step 2: bounds for  $B(\mathbf{P}_{\mathbf{U}_n})$**

Fix  $P \in M_{F^2}$ , take  $\delta > 0$  and cover  $\mathcal{F}$  by a finite number  $p$  of  $\|\cdot\|_{\mathbf{P}_{u_0, 2}}$ -balls with radius  $\delta$ . Denote the  $p$ -tuple of their centers by  $\mathbf{f} = (f_1, \dots, f_p)$ . By the same arguments as those invoked at the beginning of Step 1 we have, for any  $\epsilon > 0$ :

$$\begin{aligned} B(P) &\leq \mathbb{P}\left(\|\mathbb{G}_P\|_{\mathcal{F}_{\delta, \mathbf{P}_{u_0}}} > \epsilon\right) + \mathbb{P}\left(\|\mathbb{G}_{\mathbf{P}_{u_0}}\|_{\mathcal{F}_{\delta, \mathbf{P}_{u_0}}} > \epsilon\right) + 2\epsilon + B_{\mathbf{f}}(P), \text{ where} \\ B_{\mathbf{f}}(P) &:= \sup_{\phi \in BL_1(\mathbb{R}^p)} \left| \mathbb{E}\left(\phi(\mathbb{G}_P(\mathbf{f}))\right) - \mathbb{E}\left(\phi(\mathbb{G}_{\mathbf{P}_{u_0}}(\mathbf{f}))\right) \right|. \end{aligned} \quad (88)$$

The first two terms will be bounded by using Dudley's bound for suprema of Gaussian processes. Apply, e.g., [36, Corollary 2.2.8] with (using their notation)  $T := \mathcal{F}$ ,  $d(f, f') := \|f - f'\|_{P, 2}$  and  $\delta := \sup\{\|f - f'\|_{P, 2}, \|f - f'\|_{\mathbf{P}_{u_0, 2}} \leq \delta\}$ , to obtain the following bound,  $C$  denoting a universal constant.

$$\begin{aligned} &\mathbb{P}\left(\|\mathbb{G}_P\|_{\mathcal{F}_{\delta, \mathbf{P}_{u_0}}} > \epsilon\right) \\ &\leq \frac{C}{\epsilon} \int_0^{\text{diam}(\mathcal{F}_{\delta, P}, \|\cdot\|_{\mathbf{P}_{u_0, 2}})} \sqrt{\log(N(\epsilon, \mathcal{F}, \|\cdot\|_{P, 2}))} d\epsilon \\ &= \frac{C}{\epsilon} \|F\|_{P, 2} \int_0^{\frac{\text{diam}(\mathcal{F}_{\delta, P}, \|\cdot\|_{\mathbf{P}_{u_0, 2}})}{\|F\|_{P, 2}}} \sqrt{\log(N(\epsilon \|F\|_{P, 2}, \mathcal{F}, \|\cdot\|_{P, 2}))} d\epsilon \\ &\leq \frac{C}{\epsilon} \|F\|_{P, 2} J_{\mathcal{F}, F}\left(\text{diam}(\mathcal{F}_{\delta, P}, \|\cdot\|_{\mathbf{P}_{u_0, 2}})\right) =: \frac{C}{\epsilon} B_{\epsilon, \delta}(P). \end{aligned} \quad (89)$$

By similar argument we have

$$\mathbb{P}\left(\|\mathbb{G}_{\mathbf{P}_{u_0}}\|_{\mathcal{F}_{\delta, \mathbf{P}_{u_0}}} > \epsilon\right) \leq \frac{C}{\epsilon} \|F\|_{\mathbf{P}_{u_0, 2}} J_{\mathcal{F}, F}(\delta),$$

which does not depend upon  $P$  and which, for any  $\epsilon > 0$  can be made arbitrarily small by a right choice of  $\delta > 0$ . As a consequence, proving that  $B(\mathbf{P}_{\mathbf{U}_n})^* \xrightarrow{d} 0$  will be a consequence of the following lemma.

**Lemma 3.9** *For any  $p \geq 1$ ,  $\mathbf{f} \in \mathcal{F}^p$  (resp.  $\epsilon > 0, \delta > 0$ ) the map  $B_{\mathbf{f}}(\cdot)$  (resp.  $B_{\epsilon, \delta}(\cdot)$ ) is Borel measurable from  $(M_{F^2}, \mathcal{M})$  to  $[0, \infty[$ . Moreover we have*

$$\forall \epsilon > 0, \lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty}^{\mathbb{P}} B_{\epsilon, \delta}(\mathbf{P}_{\mathbf{U}_n}) = 0, \quad (90)$$

$$\forall p \geq 1, \forall \mathbf{f} \in \mathcal{F}^p, B_{\mathbf{f}}(\mathbf{P}_{\mathbf{U}_n}) \xrightarrow{d} 0. \quad (91)$$

**Proof:** For fixed  $\epsilon, \delta$  the measurability of  $B_{\epsilon, \delta}(\cdot)$  is a consequence of Lemma 3.2. For fixed  $p \geq 1$ ,  $\mathbf{f} \in \mathcal{F}^p$ , the measurability of  $B_{\mathbf{f}}(P)$  is a consequence of the fact that it is a continuous function of the covariance matrix

$$\Sigma_{\mathbf{f}}(P) := \left[ P(f_k, f_\ell) - P(f_k)P(f_\ell) \right]_{k, \ell \in \llbracket 1, p \rrbracket^2}.$$

This also proves (91) since condition (14) entails

$$\Sigma_{\mathbf{f}}(\mathbf{P}_{\mathbf{U}_n}) \rightarrow_{\mathbb{P}} \Sigma_{\mathbf{f}}(\mathbf{P}_{u_0}).$$

To prove (90) note first that  $\|F\|_{\mathbf{P}_{\mathbf{U}_n, 2}}$  is bounded in probability. In addition we have, by the triangle inequality:

$$\text{diam}(\mathcal{F}_{\delta, \mathbf{P}_{\mathbf{U}_n}}, \|\cdot\|_{\mathbf{P}_{u_0, 2}}) \leq \delta + \|\mathbf{P}_{\mathbf{U}_n} - \mathbf{P}_{u_0}\|_{\mathcal{F}_\infty}, \quad (92)$$

everywhere on  $\Omega$ . But the second term converges to zero in outer probability by (14). This concludes the proof.  $\square$

## 4 Proofs related to Section 2

### 4.1 Proof of Proposition 2.1

Fix  $(y_1, \dots, y_n)$ . In this proof we shall omit the dependency of  $\mathcal{I}_>, \mathcal{I}_=, \mathcal{I}_<, \mathcal{I}_1, \mathcal{I}_2$  upon  $(y_1, \dots, y_n)$  in order to unburden notation. By construction the (random) set  $\mathcal{I}_2$  almost surely satisfies the following inclusions between deterministic sets

$$\mathcal{I}_> \subset \mathcal{I}_2 \subset \mathcal{I}_> \cup \mathcal{I}_=.$$

Recalling the definition of  $\mathcal{I}_1$ , we see that it is sufficient to prove that the random set  $\mathcal{I}_2 \setminus \mathcal{I}_>$  is equally distributed among all subsets of  $\mathcal{I}_=$  with cardinality  $k_n - \#\mathcal{I}_>$ . To prove this, notice that, with probability one

$$\begin{aligned} & \mathcal{I}_2 \setminus \mathcal{I}_> \\ &= \left\{ i \in \mathcal{I}_=, \#\{j \in \llbracket 1, n \rrbracket, F(y_j) + \Delta F(y_j)\boldsymbol{\eta}_j \leq F(y_i) + \Delta F(y_i)\boldsymbol{\eta}_i\} \geq n - k_n + 1 \right\} \\ &= \left\{ i \in \mathcal{I}_=, \#\mathcal{I}_< + \#\{j \in \mathcal{I}_=, F(y_j) + \Delta F(y_j)\boldsymbol{\eta}_j \leq F(y_i) + \Delta F(y_i)\boldsymbol{\eta}_i\} \geq n - k_n + 1 \right\} \\ &= \left\{ i \in \mathcal{I}_=, \#\mathcal{I}_< + \#\{j \in \mathcal{I}_=, \boldsymbol{\eta}_j \leq \boldsymbol{\eta}_i\} \geq n - k_n + 1 \right\} \end{aligned}$$

Since  $n = \#\mathcal{I}_> + \#\mathcal{I}_= + \#\mathcal{I}_<$  we have

$$n - k_n + 1 = n - \#\mathcal{I}_> - (k_n - \#\mathcal{I}_>) + 1 = \#\mathcal{I}_< + \#\mathcal{I}_= - (k_n - \#\mathcal{I}_>) + 1,$$

whence we conclude that

$$\mathcal{I}_2 \setminus \mathcal{I}_> = \left\{ i \in \mathcal{I}_=, \#\{j \in \mathcal{I}_=, \boldsymbol{\eta}_j \leq \boldsymbol{\eta}_i\} \geq \#\mathcal{I}_= - (k_n - \#\mathcal{I}_>) + 1 \right\}.$$

By generating i.i.d uniform random variables  $(\boldsymbol{\eta}_i)_{i \in \mathcal{I}_=}$  and taking the subset of indices corresponding to the  $(k_n - \#\mathcal{I}_>)$  largest values among  $(\boldsymbol{\eta}_i)_{i \in \mathcal{I}_=}$ , we select a subset of  $\mathcal{I}_=$  with cardinality  $(k_n - \#\mathcal{I}_>)$  uniformly at random. This concludes the proof  $\square$ .

## 4.2 Proof of Proposition 2.2

We first prove the claimed equality in laws. Recall that the  $R_i$  are defined in *tie breaking algorithm 2*. Fix arbitrary  $A \in \mathcal{M}$  and  $B$  Borel. Since  $(Y_1, \dots, Y_n)$  has no ties with probability one, one can write

$$\begin{aligned} & \mathbb{P}\left(\left(\mathbb{P}_n^{(n)}, \mathbf{U}_n\right) \in A \times B\right) \\ &= \sum_{\sigma} \mathbb{P}\left(\left(\frac{1}{k_n} \sum_{\substack{i \in \llbracket 1, n \rrbracket \\ R_i \geq n - k_n + 1}} \delta_{\phi(X_i, U_{n-k_n:n}), U_{n-k_n:n}}\right) \in A \times B, U_{\sigma(1)} < U_{\sigma(2)} < \dots < U_{\sigma(n)}\right), \end{aligned}$$

where the first sum holds over all permutations  $\sigma$  of  $\llbracket 1, n \rrbracket$ . Now using the constraints  $U_{\sigma(1)} < \dots < U_{\sigma(n)}$  then noting that  $(X_{\sigma(i)}, U_{\sigma(i)})_{i \in \llbracket 1, n \rrbracket} =_d (X_i, U_i)_{i \in \llbracket 1, n \rrbracket}$  for each fixed permutation  $\sigma$ , the last sum equals

$$\begin{aligned} & \sum_{\sigma} \mathbb{P}\left(\left(\frac{1}{k_n} \sum_{i=n-k_n+1}^n \delta_{\phi(X_{\sigma(i)}, U_{\sigma(n-k_n)}), U_{\sigma(n-k_n)}}\right) \in A \times B, U_{\sigma(1)} < U_{\sigma(2)} < \dots < U_{\sigma(n)}\right) \\ &= n! \mathbb{P}\left(\left(\frac{1}{k_n} \sum_{i=n-k_n+1}^n \delta_{\phi(X_i, U_{n-k_n}), U_{n-k_n}}\right) \in A \times B, U_1 < U_2 < \dots < U_n\right). \end{aligned} \quad (93)$$

But also note that

$$\begin{aligned} & \mathbb{P}\left(\left(\frac{1}{k_n} \sum_{i=n-k_n+1}^n \delta_{\phi(X_i, U_{n-k_n}), U_{n-k_n}}\right) \in A \times B \mid U_1 < U_2 < \dots < U_n\right) \\ &= n! \mathbb{P}\left(\left(\frac{1}{k_n} \sum_{i=n-k_n+1}^n \delta_{\phi(X_i, U_{n-k_n}), U_{n-k_n}}\right) \in A \times B, U_1 < U_2 < \dots < U_n\right), \end{aligned}$$

which equals (93). This proves the claimed equality in laws. Let us now prove the claimed CEM property. Fix  $A \in \mathcal{M}$  and  $B$  Borel. Write  $E := [0, 1[$ . We have to prove that

$$\mathbb{P}\left(\left(\mathbb{P}_n^{(n)}, \mathbf{U}_n\right) \in A \times B\right) = \int_B \overline{\mathbb{P}}\left(\frac{1}{k_n} \sum_{i=1}^{k_n} \delta_{X_i^{(u)}} \in A\right) d\mathbb{P}^{\mathbf{U}_n}(u), \quad (94)$$

where, for given  $u \in E$ , the family  $(X_i^{(u)})_{i \in \llbracket 1, k_n \rrbracket}$  is i.i.d. with law  $\mathbf{P}_u$ , from some probability space  $(\overline{\Omega}, \overline{\mathcal{A}}, \overline{\mathbb{P}})$  to  $\mathfrak{X}$ .

To prove (94), we use the just proved equality in laws, namely

$$\begin{aligned}
& \mathbb{P}\left(\left(\mathbb{P}_n^{(n)}, \mathbf{U}_n\right) \in A \times B\right) \\
&= n! \mathbb{P}\left(\left(\frac{1}{k_n} \sum_{i=n-k_n+1}^n \delta_{\phi(X_i, U_{n-k_n})}, U_{n-k_n}\right) \in A \times B \text{ and } U_1 < U_2 < \dots < U_n\right) \\
&= n! \int_{(\mathfrak{X}' \times E)^n} \mathbb{1}_A\left(\frac{1}{k_n} \sum_{i=n-k_n+1}^n \delta_{\phi(x_i, u_{n-k_n})}\right) \mathbb{1}_B(u_{n-k_n}) \mathbb{1}_{\{u_1 < \dots < u_n\}} \prod_{i=1}^n d\mathbb{P}^{X, U}(x_i, u_i) \\
&= n! \int_{u \in B} du \\
&\quad \left[ \int_{u < u_{n-k_n+1} < \dots < u_n} du_{n-k_n+1} \dots du_n \right. \\
&\quad \left[ \int_{u_1 < \dots < u_{n-k_n-1} < u} du_1 \dots du_{n-k_n-1} \right. \\
&\quad \left. \left. \left[ \int_{(x_1, \dots, x_n)} \mathbb{1}_A\left(\frac{1}{k_n} \sum_{i=n-k_n+1}^n \delta_{\phi(x_i, u)}\right) \prod_{i=1}^n d\mathbb{P}^{X|U=u_i}(x_i) \right] \right] \right], \tag{95}
\end{aligned}$$

where, in the first integral,  $u$  could have been written  $u_{n-k_n}$  (we here make the choice of unburdening further notations and formulas).

We will handle each of those four embedded integrals successively from last to first. In the fourth integral, the term involving  $\mathbb{1}_A$  does only depend upon  $(x_{n-k_n+1}, \dots, x_n, u)$  from where the integral equals

$$\int_{(x_{n-k_n+1}, \dots, x_n) \in \mathfrak{X}'^{k_n}} \mathbb{1}_A\left(\frac{1}{k_n} \sum_{i=n-k_n+1}^n \delta_{\phi(x_i, u)}\right) \prod_{i=n-k_n+1}^n d\mathbb{P}^{X|U=u_i}(x_i),$$

and hence does not depend upon  $(u_1, \dots, u_{n-k_n-1})$ . Integrating out in  $(u_1, \dots, u_{n-k_n-1})$  we see that the third integral in (95) equals

$$\begin{aligned}
& \left( \int_{u_1 < \dots < u_{n-k_n-1} < u} du_{n-k_n+1} \dots du_n \right) \\
& \quad \times \left( \int_{(x_{n-k_n+1}, \dots, x_n) \in \mathfrak{X}'^{k_n}} \mathbb{1}_A\left(\frac{1}{k_n} \sum_{i=n-k_n+1}^n \delta_{\phi(x_i, u)}\right) \prod_{i=n-k_n+1}^n d\mathbb{P}^{X|U=u_i}(x_i) \right) \\
&= \frac{u^{n-k_n-1}}{(n-k_n-1)!} \int_{(x_{n-k_n+1}, \dots, x_n) \in \mathfrak{X}'^{k_n}} \mathbb{1}_A\left(\frac{1}{k_n} \sum_{i=n-k_n+1}^n \delta_{\phi(x_i, u)}\right) \prod_{i=n-k_n+1}^n d\mathbb{P}^{X|U=u_i}(x_i).
\end{aligned}$$

Now integrating out in  $(u_{n-k_n+1}, \dots, u_n)$  and artificially inserting  $(1-u)^{k_n}$  we see that the second integral in (95) equals

$$\begin{aligned} & \frac{u^{n-k_n-1}(1-u)^{k_n}}{(n-k_n-1)!} \int_{(u_{n-k_n+1}, \dots, u_n) \in [u, 1]^{k_n}} \mathbb{1}_{\{u_{n-k_n+1} < \dots < u_n\}} \frac{du_{n-k_n+1} \dots du_n}{(1-u)^{k_n}} \\ & \left[ \int_{(x_{n-k_n+1}, \dots, x_n) \in \mathcal{X}^{k_n}} \mathbb{1}_A \left( \frac{1}{k_n} \sum_{i=n-k_n+1}^n \delta_{\phi(x_i, u)} \right) \prod_{i=n-k_n+1}^n d\mathbb{P}^{X|U=u_i}(x_i) \right] \\ & = \frac{u^{n-k_n-1}(1-u)^{k_n}}{(n-k_n-1)!} \int_{[u, 1]^{k_n}} \mathbb{1}_{\{u_{n-k_n+1} < \dots < u_n\}} H_u(u_{n-k_n+1}, \dots, u_n) \frac{du_{n-k_n+1} \dots du_n}{(1-u)^{k_n}}, \quad (96) \end{aligned}$$

where we wrote

$$H_u(u_{n-k_n+1}, \dots, u_n) = \mathbb{P} \left( \frac{1}{k_n} \sum_{i=n-k_n+1}^n \delta_{\phi(X_i, u)} \in A \mid U_{n-k_n+1} = u_{n-k_n+1}, \dots, U_n = u_n \right).$$

But since  $H$  is symmetric in its arguments, (96) equals

$$\frac{u^{n-k_n-1}(1-u)^{k_n}}{k_n!(n-k_n-1)!} \int_{[u, 1]^{k_n}} H_u(u_{n-k_n+1}, \dots, u_n) \frac{du_{n-k_n+1} \dots du_n}{(1-u)^{k_n}}. \quad (97)$$

The key is to interpret the integral in (97) as an expectation with respect to the following probabilistic mechanism:

- Generate  $(U_{n-k_n+1}^{(u)}, \dots, U_n^{(u)})$  according to the law of  $(U_{n-k_n+1}, \dots, U_n)$  conditional to  $U_{n-k_n+1} > u, \dots, U_n > u$ , i.e. generate each  $\tilde{U}_i$  independently with common distribution  $\mathbb{P}^U \mid U > u$ . Therefore that sample is i.i.d and uniformly distributed on  $[u, 1]$ .

Write  $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbb{P}})$  for a probability space holding  $(U_{n-k_n+1}^{(u)}, \dots, U_n^{(u)})$ , and write  $(u_{n-k_n+1}, \dots, u_n)$  for its realization.

- One the same  $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbb{P}})$  generate each  $\tilde{X}_i(u_i)$ ,  $i \in \llbracket n - k_n + 1, n \rrbracket$  independently from the rest, and respectively distributed as  $\mathbb{P}^{X|U=u_i}$ .
- The resulting random variables  $\tilde{X}_i(\tilde{U}_i^{(u)})$ ,  $i \in \llbracket n - k_n + 1, n \rrbracket$  are hence mutually independent realizations of  $\mathbb{P}^{X|U>u}$ . As a consequence,  $X_i^{(u)} := \phi(\tilde{X}_i(\tilde{U}_i^{(u)}), u)$ ,  $i \in \llbracket n - k_n + 1, n \rrbracket$  is a  $\mathbf{P}_u$ -i.i.d family.

With this interpretation at hand one can express (97) as

$$\begin{aligned}
&= \frac{u^{n-k_n-1}(1-u)^{k_n}}{k_n!(n-k_n-1)!} \mathbb{P} \left( \frac{1}{k_n} \sum_{i=n-k_n+1}^n \delta_{\phi(X_i, u) \in A} \middle| U_{n-k_n+1} > u, \dots, U_n > u \right) \\
&= \frac{u^{n-k_n-1}(1-u)^{k_n}}{k_n!(n-k_n-1)!} \overline{\mathbb{P}} \left( \frac{1}{k_n} \sum_{i=n-k_n+1}^n \delta_{\phi(\tilde{X}_i(\tilde{U}_i^{(u)}), u) \in A} \right) \\
&= \frac{u^{n-k_n-1}(1-u)^{k_n}}{k_n!(n-k_n-1)!} \overline{\mathbb{P}} \left( \frac{1}{k_n} \sum_{i=n-k_n+1}^n \delta_{X_i^{(u)} \in A} \right) \\
&= \frac{u^{n-k_n-1}(1-u)^{k_n}}{k_n!(n-k_n-1)!} \overline{\mathbb{P}} \left( \frac{1}{k_n} \sum_{i=1}^{k_n} \delta_{X_i^{(u)} \in A} \right).
\end{aligned}$$

Now integrating out in  $u$ , we finally conclude that (95) equals

$$\int_{u \in [0,1]} \overline{\mathbb{P}} \left( \frac{1}{k_n} \sum_{i=1}^{k_n} \delta_{X_i^{(u)} \in A} \right) \frac{n!}{k_n!(n-k_n-1)!} u^{n-k_n-1} (1-u)^{k_n} du.$$

This proves (94) by identifying the density of  $\mathbf{U}_n = U_{n-k_n:n}$ .  $\square$

### 4.3 Proof of Proposition 2.3

Fix  $g \in \mathcal{G}$  and  $u \in [0, 1[$ , and write

$$\begin{aligned}
&\mathbb{E} \left( g \circ \phi(X, u) \mathbf{1}_{\{U \geq u\}} \right) \\
&= \mathbb{E} \left( g \circ \tilde{\phi}(X, F^{\leftarrow}(u)) \left( \int_{[0,1]} \mathbf{1}_{\{F_-(Y) + \Delta F(Y) \eta \geq u\}} d\eta \right) \right). \tag{98}
\end{aligned}$$

But, for fixed  $y$  one has

$$\begin{aligned}
\int_{[0,1]} \mathbf{1}_{\{F_-(y) + \Delta F(y) \eta \geq u\}} d\eta &= 1, \text{ if } y > F^{\leftarrow}(u), \\
&= \frac{F(F^{\leftarrow}(u)) - u}{\Delta F(y)} \mathbf{1}_{\{\Delta F(y) > 0\}}, \text{ if } y = F^{\leftarrow}(u), \tag{99} \\
&= 0, \text{ otherwise,}
\end{aligned}$$

or more compactly

$$\int_{[0,1]} \mathbf{1}_{\{F_-(y) + \Delta F(y) \eta \geq u\}} d\eta = \mathbf{1}_{\{y > F^{\leftarrow}(u)\}} + \theta_u \mathbf{1}_{\{y = F^{\leftarrow}(u)\}}, \tag{100}$$

with  $\theta_u \in [0, 1]$  defined in (99). Using (100) in (98) gives

$$\begin{aligned}
&\mathbb{E} \left( g \circ \tilde{\phi}(X, F^{\leftarrow}(u)) \mathbf{1}_{\{U \geq u\}} \right) \\
&= \mathbb{E} \left( g \circ \tilde{\phi}(X, F^{\leftarrow}(u)) \mathbf{1}_{\{Y > F^{\leftarrow}(u)\}} \right) + \theta_u \mathbb{E} \left( g \circ \tilde{\phi}(X, F^{\leftarrow}(u)) \mathbf{1}_{\{Y = F^{\leftarrow}(u)\}} \right).
\end{aligned}$$

Similarly one has

$$\mathbb{P}(U \geq u) = \mathbb{P}(Y > F^{\leftarrow}(u)) + \theta_u \mathbb{P}(Y = F^{\leftarrow}(u)),$$

from where

$$\begin{aligned} & \mathbb{E}\left(g \circ \tilde{\phi}(X, F^{\leftarrow}(u)) \mid U \geq u\right) \\ &= \frac{\mathbb{E}\left(g \circ \tilde{\phi}(X, F^{\leftarrow}(u)) \mathbf{1}_{\{Y > F^{\leftarrow}(u)\}}\right) + \theta_u \mathbb{E}\left(g \circ \tilde{\phi}(X, F^{\leftarrow}(u)) \mathbf{1}_{\{Y = F^{\leftarrow}(u)\}}\right)}{\mathbb{P}(Y > F^{\leftarrow}(u)) + \theta_u \mathbb{P}(Y = F^{\leftarrow}(u))} \\ &= \mathbb{E}\left(g \circ \tilde{\phi}(X, F^{\leftarrow}(u)) \mid Y > F^{\leftarrow}(u)\right) (1 + a_g(u))(1 + b(u))^{-1}, \text{ where} \\ & a_g(u) := \theta_u \frac{\mathbb{E}\left(g \circ \tilde{\phi}(X, F^{\leftarrow}(u)) \mathbf{1}_{\{Y = F^{\leftarrow}(u)\}}\right)}{\mathbb{P}(Y > F^{\leftarrow}(u))}, \text{ and} \\ & b(u) := \theta_u \frac{\mathbb{P}(Y = F^{\leftarrow}(u))}{\mathbb{P}(Y > F^{\leftarrow}(u))}. \end{aligned}$$

By (32) and since  $\theta_u$  is bounded by 1 we have  $b(u) = O(\varepsilon(F^{\leftarrow}(u)))$ . Similarly (33) entails  $a_g(u) = O(\varepsilon(F^{\leftarrow}(u)))$  uniformly in  $g \in \mathcal{G}$ . Combine this with (30) to conclude that  $\|\mathbf{P}_u - \mathbf{P}_1\|_{\mathcal{G}} = O(\varepsilon(F^{\leftarrow}(u)))$ , which proves (34) as  $\varepsilon'(\cdot)$  was arbitrarily chosen. Finally (35) is proved in a very similar way. We omit details.  $\square$

## 4.4 Proof of Proposition 2.5

Fix  $\delta > 0$  and write

$$f_x := V(x)^{-1/2+\delta} \mathbf{1}_{]x, \infty[}, \text{ for } x \in [0, x^+].$$

Note that we can assume  $\delta \leq 1/4$  without loss of generality. Taking  $\tilde{\phi}(x, y) := [x - b(1/\bar{F}(y))]/a(1/\bar{F}(y))$  we here need to prove that all the conditions of Theorem 2 are fulfilled for  $\mathbb{P}_n^{(n)}$  defined as in (29), with  $\mathcal{F} := \{f_x, x \in [0, x^+]\}$ ,  $E := [0, 1]$ ,  $E_0 := \{1\}$ ,  $\mathbf{P}_1(\cdot, \infty) := V(\cdot)$ , and

$$\mathbf{P}_u(\cdot) := \mathbb{P}\left(\frac{Y - b(1/\bar{F}(F^{\leftarrow}(u)))}{a(1/\bar{F}(F^{\leftarrow}(u)))} \in \cdot \mid U \geq u\right),$$

with  $U := F_-(Y) + \Delta F(Y)\eta$ , and  $\eta \perp Y$  uniform on  $[0, 1]$ . Clearly  $\mathcal{F}$  has envelope

$$F : x \rightarrow V(x)^{-1/2+\delta} \mathbf{1}_{[0, x^+]}(x). \quad (101)$$

Moreover  $J_{\mathcal{F}, F}(\infty) < \infty$  as  $\mathcal{F}$  is a VC subgraph class (see, e.g., [36, p. 141]), and the derived class  $\mathcal{G}$  involved in condition (14) has envelope  $G \leq 4F^2$ . We will now prove that the conditions of Proposition 2.3 are fulfilled with the choice of  $\epsilon(y) := A(1/\bar{F}(y))$ ,

which will prove (18) and (19) and hence prove (14) and (12).  
Let us first prove (30). Write

$$\Delta_y(x) := \mathbb{P}\left(\frac{Y - b(1/\overline{F}(y))}{a(1/\overline{F}(y))} > x \mid Y > y\right) - \mathbf{P}_1(]x, \infty[),$$

and fix  $\eta \in ]0, \delta[$  so that  $\rho - 1 + \eta \leq -1/2 + \delta$ . With this choice of  $\eta$ , Proposition 2.4 with  $t := 1/\overline{F}(y)$  reads

$$\limsup_{y \uparrow F^{\leftarrow}(1)} \sup_{x \in [0, x^+]} V(x)^{\rho-1+\eta} \left| \frac{\Delta_y(x)}{A(1/\overline{F}(y))} - \Psi_{\gamma, \rho}(V(x)) \right| = 0,$$

with  $\Psi_{\gamma, \rho}(v) = o(v^{1-\rho-\delta/2})$  as  $v \rightarrow 0$ , from where  $\Psi_{\gamma, \rho}(V)$  is bounded on  $[0, x^+]$ . A use of the triangle inequality yields (since  $\rho - 1 + \eta \leq -1/2 + \delta$ )

$$\begin{aligned} & \limsup_{y \uparrow F^{\leftarrow}(1)} \frac{1}{A(1/\overline{F}(y))} \sup_{x \in [0, x^+]} V(x)^{-1/2+\delta} |\Delta_y(x)| \\ & \leq \sup_{x \in [0, x^+]} |V(x)^{-1/2+\delta} \Psi_{\gamma, \rho}(V(x))| < \infty, \end{aligned} \quad (102)$$

which proves that condition (30) is satisfied for the class  $\mathcal{F}$ . Using  $(a-b)^2 = a^2 - 2ab + b^2$ , we see that it will be also the case for the class  $\mathcal{G}$  if we show that

$$\sup_{(x, x') \in [0, x^+]^2} \frac{1}{A(1/\overline{F}(y))} \left| \mathbb{E}\left((f_x f_{x'}) (\tilde{\phi}(X, y)) \mid Y > y\right) - \mathbf{P}_1(f_x f_{x'}) \right| < \infty,$$

which reads

$$\lim_{y \uparrow F^{\leftarrow}(1)} \sup_{(x, x') \in [0, x^+]^2} \frac{(V(x)V(x'))^{-1/2+\delta}}{A(1/\overline{F}(y))} |\Delta_y(x \vee x')| < \infty. \quad (103)$$

But, by arguments that are very similar to those used to obtain (102) we have

$$\limsup_{y \uparrow F^{\leftarrow}(1)} \frac{1}{A(1/\overline{F}(y))} \sup_{x \in [0, x^+]} V(x)^{-1+2\delta} |\Delta_y(x)| < \infty,$$

which proves (103) by noting that  $(V(x)V(x'))^{-1/2+\delta} \leq V(x \vee x')^{-1+2\delta}$ . This concludes the proof of (30).

We will now prove that that (31) holds. Write

$$H(x) := V(x)^{(-1+2\delta)(1+\delta)}, \quad x \in [0, x^+],$$

so that  $G^{1+\delta} \leq H \mathbf{1}_{[0, x^+]}$ . Now fix  $\epsilon > 0$  and notice that, since  $H$  is continuous on  $[0, x^+]$  and continuously differentiable on  $]0, x^+[$ :

$$\begin{aligned} & \mathbb{E}\left((H \mathbf{1}_{[0, x^+]}) \circ \tilde{\phi}(X, y) \mid Y > y\right) \\ & = \int_{\omega \in \Omega} \mathbf{1}_{[0, x^+]}(\tilde{\phi}(X(\omega), y)) \left[ H(\epsilon) + \int_{\epsilon}^{\tilde{\phi}(X(\omega), y)} H'(x) dx \right] d\mathbb{P}_{|Y > y}(\omega). \end{aligned}$$

Now fix  $\epsilon > 0$ . By Fubini's theorem the preceding expectation equals

$$\begin{aligned} & H(\epsilon)\mathbb{P}\left(\tilde{\phi}(X, y) \in [0, x^+]\right) + \int_{x \in [\epsilon, x^+]} H'(x)\mathbb{P}\left(\tilde{\phi}(X, y) \in ]x, x^+]\middle| Y > y\right) dx \\ & \leq H(\epsilon) + \int_{x \in [\epsilon, x^+]} H'(x)\mathbb{P}\left(\tilde{\phi}(X, y) \in ]x, x^+]\middle| Y > y\right) dx. \end{aligned}$$

Write  $t := 1/\bar{F}(y)$ . We will now prove that one can substitute  $H'(x)(V_t(x) - V_t(x^+)) = H'(x)V_t(x)$  to the last integrand. This is done by noticing that for all  $y$  such that  $|cA(t)| < \epsilon$ , and for all  $x \geq \epsilon$

$$\begin{aligned} & \mathbb{P}\left(\tilde{\phi}(X, y) \in ]x, x^+], Y \leq y\right) \\ & \leq \mathbb{P}\left(y \geq Y \geq b(t) + \epsilon a(t)\right) \\ & = \mathbb{P}\left(y \geq Y \geq U(t) + a(t)(\epsilon + cA(t))\right) \\ & \leq \mathbb{P}\left(y \geq Y > U(1/\bar{F}(y))\right) \\ & = 0, \end{aligned} \tag{104}$$

which is a consequence of the fact that  $F$  is the c.d.f of  $Y$  (use the representation  $Y = F^{\leftarrow}(V)$  with uniform for example).

Now (104) enables the sought substitution: for all large  $y$  one has

$$\begin{aligned} & \int_{x \in [\epsilon, x^+]} H'(x)\mathbb{P}\left(\tilde{\phi}(X, y) \in ]x, x^+]\middle| Y > y\right) dx \\ & = \int_{x \in [\epsilon, x^+]} H'(x) \frac{\mathbb{P}\left(\tilde{\phi}(X, y) \in ]x, x^+]\right)}{\mathbb{P}(Y > y)} dx \\ & = \int_{x \in [\epsilon, x^+]} H'(x)V_t(x) dx. \end{aligned}$$

Now writing

$$\eta_t(x) := V(x)^{\rho-1+\eta} \left( \frac{\Delta_y(x)}{A(1/\bar{F}(y))} - \Psi_{\gamma, \rho}(V(x)) \right),$$

one has

$$\begin{aligned} \int_{x \in [\epsilon, x^+]} H'(x)V_t(x) dx & = \int_{x \in [\epsilon, x^+]} H'(x)V(x) dx \\ & \quad + A(t) \int_{x \in [\epsilon, x^+]} H'(x)\Psi_{\gamma, \rho}(V(x)) dx \\ & \quad + A(t) \int_{x \in [\epsilon, x^+]} H'(x)V(x)^{1-\rho-\eta} \eta_t(x) dx. \end{aligned}$$

Now because  $H' = V'V^{-(1+\delta)(1-2\delta)-1}$ ,  $|\eta_t|$  is uniformly bounded by 1 for all large  $y$ ,  $\Psi_{\gamma,\rho}(V) = o(V^{1-\rho-\delta/2})$  around  $x^+$  and  $V^{-(1+\delta)(1-2\delta)}V^{1-\rho-\eta}$  has a finite limit at  $x^+$ , straightforward analysis shows that all these three integrals are finite for all large  $y$ . This proves that (31) holds.

We will now prove (32). Fix  $y$ , write again  $t := 1/\bar{F}(y)$  and note that,

$$\begin{aligned} \frac{\mathbb{P}(Y = y)}{\mathbb{P}(Y > y)} &\leq \frac{\mathbb{P}\left(\frac{Y-b(t)}{a(t)} \in \left] \frac{y-b(t)-a(t)\epsilon(t)}{a(t)}, \frac{y-b(t)}{a(t)} \right] \right)}{\mathbb{P}(Y > y)} \\ &= V_t(x_t - \epsilon(t) + cA(t)) - V_t(x_t + cA(t)), \end{aligned}$$

where  $x_t := (y - U(t))/a(t) \rightarrow 0$  as  $y \rightarrow F^{\leftarrow}(1)$  (see, e.g., [8, p. 11]). By Proposition 2.4 one has (ignoring the weights which play no role here)

$$\begin{aligned} &V_t(x_t - \epsilon(t) + cA(t)) - V_t(x_t + cA(t)) \\ &= V(x_t - \epsilon(t) + cA(t)) - V(x_t + cA(t)) \\ &\quad + A(t) \left[ \Psi_{\gamma,\rho}(x_t - \epsilon(t) + cA(t)) - \Psi_{\gamma,\rho}(x_t + cA(t)) + o(1) \right]. \end{aligned}$$

Because  $x_t \rightarrow 0$  and the involved functions are continuously differentiable, a Taylor expansion shows that this expression is  $O(A(t))$  as  $t \rightarrow \infty$ . This proves (32). Now (33) is proved by noticing that

$$\begin{aligned} \frac{\mathbb{E}\left(G(\tilde{\phi}(X, y))\mathbf{1}_{\{Y=y\}}\right)}{\mathbb{P}(Y > y)} &= G\left(\frac{y - b(1/\bar{F}(y))}{a(t)}\right) \frac{\mathbb{P}(Y = y)}{\mathbb{P}(Y > y)} \\ &= G\left(x_t + cA(t)\right) \frac{\mathbb{P}(Y = y)}{\mathbb{P}(Y > y)}, \end{aligned}$$

and noticing that  $G \leq 4F^2$ , which has limit zero at zero. We proved that all the conditions of Proposition 2.3 are fulfilled. In particular

$$\|\mathbf{P}_u - \mathbf{P}_1\|_{\mathcal{G}} = O(A(1/\bar{F}(F^{\leftarrow}(u)))) \text{, as } u \rightarrow 1,$$

which will be used to check the very last condition of Theorem 2, namely:

$$\sqrt{k_n} \sup_{x \in [0, x^+]} \left| \mathbf{P}_{U_{n-k_n:n}}(]x, \infty[) - \mathbf{P}_1(]x, \infty[) \right| \xrightarrow{d} 0. \quad (105)$$

Recall that  $A(\cdot) \simeq \tilde{A}(\cdot)$  and  $\tilde{A}$  is  $\rho$ -varying. As a consequence, since

$$\bar{F}(F^{\leftarrow}(U_{n-k_n:n})) = \bar{F}(Y_{n-k_n:n}) \simeq \frac{k_n}{n},$$

we conclude that

$$\sqrt{k_n} A(1/\bar{F}(F^{\leftarrow}(U_{n-k_n:n}))) \simeq \sqrt{k_n} A\left(\frac{n}{k_n}\right),$$

which tends to 0 by assumption. This proves (105) and hence concludes the proof of Proposition 2.5□.

## 4.5 Proof of Proposition 2.6

The methodology of the proof is very similar to that of Proposition 2.5, hence our proof will be voluntarily more elusive for sake of brevity. Fix  $\alpha' \in [0, \alpha[$ . We need to show that all the conditions of Theorem 2 are satisfied for the choice of  $\mathbf{P}_u$  as in (42) and  $\mathbf{P}_1 := \mathbf{m}(\alpha)$  as defined in (47), for the class  $\mathcal{F}$  defined in (51) and with  $E := [0, 1]$ ,  $E_0 := \{1\}$ . Clearly  $\mathcal{F}$  has envelope function  $F : (x, s) \rightarrow \exp(\frac{\alpha'}{2}x)$ . We will now prove that  $J_{\mathcal{F}, F}(+\infty) < \infty$  by proving that  $\mathcal{F}$  is VC-subgraph (see, .e.g. [36, p.141]). A close look at (43) shows that, for a one-to-one map  $\phi$  from  $]0, 2\pi]^{d-1}$  to  $\mathcal{S}_{d-1}$ , each  $O_\theta$  can be written  $\phi([0, \theta_1] \times \dots \times [0, \theta_{d-1}])$ . Apply [36, p. 147, Lemma 2.6.17] to conclude that the class of all these  $O_\theta$  is VC and then

$$\mathcal{C} := \left\{ ]t + \infty[ \times O_\theta, t \geq 0, \theta \in ]0, 2\pi]^{d-1} \right\} \text{ is VC,} \quad (106)$$

which then proves that the subgraphs of  $\mathcal{F}$  form a VC class of sets. Let  $\mathcal{G}$  be as in (13) and let  $\varepsilon$  be as in (45). We will first prove that all the conditions of Proposition 2.3 are fulfilled, which will prove (18) and (19) and hence prove (14) and (12). First note that we have for any  $(t_1, \theta_1, t_2, \theta_2)$ , by definition of  $O_\theta$ :

$$f_{t_1, \theta_1} f_{t_2, \theta_2} = \exp\left(\frac{\alpha'}{2}(t_1 + t_2 - t_1 \vee t_2)\right) f_{t_1 \vee t_2, \theta_1 \wedge \theta_2} = \exp\left(\frac{\alpha'}{2}(t_1 \wedge t_2)\right) f_{t_1 \vee t_2, \theta_1 \wedge \theta_2}, \quad (107)$$

where  $\theta_1 \wedge \theta_2$  is understood as componentwise. As a consequence, the envelope  $G$  of  $\mathcal{G}$  is less than  $4F^2$ . Let us first prove (30). Taking  $t := \log x$  we see that (45) implies

$$\Delta_y := \sup_{t \geq 0, \theta \in ]0, 2\pi]^{d-1}} \exp\left(\left(\alpha - \frac{\alpha'}{2}\right)t\right) \left| Q_y(f_{t, \theta}) - \mathbf{P}_1(f_{t, \theta}) \right| = O(\varepsilon(y)), \quad (108)$$

where  $Q_y$  is the law of  $(\log(\|X\|_d / y), \|X\|_d^{-1} X)$  given  $Y > y$ . Now for any  $(t_1, \theta_1, t_2, \theta_2)$  one has

$$\begin{aligned} & \left| Q_y(f_{t_1, \theta_1} f_{t_2, \theta_2}) - \mathbf{P}_1(f_{t_1, \theta_1} f_{t_2, \theta_2}) \right| \\ &= \exp\left(\frac{\alpha'}{2}(t_1 \wedge t_2)\right) \left| Q_y(f_{t_1 \vee t_2, \theta_1 \wedge \theta_2}) - \mathbf{P}_1(f_{t_1 \vee t_2, \theta_1 \wedge \theta_2}) \right| \\ &\leq \exp\left(\frac{\alpha'}{2}(t_1 \wedge t_2 + t_1 \vee t_2) - \alpha(t_1 \vee t_2)\right) \left[ \exp\left(\left(\alpha - \frac{\alpha'}{2}\right)t_1 \vee t_2\right) \left| Q_y(f_{t_1 \vee t_2, \theta_1 \wedge \theta_2}) - \mathbf{P}_1(f_{t_1 \vee t_2, \theta_1 \wedge \theta_2}) \right| \right] \\ &\leq \exp\left(\left(\alpha - \frac{\alpha'}{2}\right)t_1 \vee t_2\right) \left| Q_y(f_{t_1 \vee t_2, \theta_1 \wedge \theta_2}) - \mathbf{P}_1(f_{t_1 \vee t_2, \theta_1 \wedge \theta_2}) \right| \leq \Delta_y, \end{aligned}$$

where the second inequality is a consequence of

$$\frac{\alpha'}{2}(t_1 \wedge t_2 + t_1 \vee t_2) \leq \alpha'(t_1 \vee t_2) \leq \alpha(t_1 \vee t_2).$$

This is enough to prove (30). We will now prove (31). Write the survival functions on  $]0, +\infty]$

$$\begin{aligned} S &:= \mathbf{P}_1(\cdot, +\infty[ \times \mathcal{S}_{d-1}) = \exp(-\alpha \cdot), \text{ and} \\ S_y &:= Q_y(\cdot, +\infty[ \times \mathcal{S}_{d-1}). \end{aligned}$$

Taking  $\boldsymbol{\theta} := (2\pi, \dots, 2\pi)$  in (108) entails

$$S_y(t) - S(t) = \exp(-\alpha t)r_y(t), \text{ with } r_y(t) \rightarrow 0 \text{ uniformly in } t \text{ as } y \rightarrow \infty. \quad (109)$$

As a consequence taking  $\delta > 0$  such that  $\alpha'(1 + \delta) < \alpha$ ,

$$\begin{aligned} \mathbb{E}\left(F^{2(1+\delta)}(\log Y - \log y) \mid Y > y\right) &= \int_0^{+\infty} \alpha'(1 + \delta) \exp(\alpha'(1 + \delta)t) S_y(t) dt \\ &\int_0^{+\infty} \alpha'(1 + \delta) \exp(\alpha'(1 + \delta)x) S(x) + \int_0^{+\infty} \alpha'(1 + \delta) \exp((\alpha'(1 + \delta) - \alpha)t) r_y(t) dt, \end{aligned}$$

which converges to a finite value as  $y \rightarrow \infty$ : the first term is finite and constant in  $n$ , while the second term converges to zero by (109). This proves (31) by Markov's inequality. We now prove (32) by using (45): take for example  $x_0 := 1/2$  and write

$$\frac{\mathbb{P}(Y = y)}{\mathbb{P}(Y > y)} = \frac{\mathbb{P}\left(\frac{\|X\|}{y} = 1\right)}{\mathbb{P}(Y > y)} \leq \frac{\mathbb{P}\left(\frac{\|X\|}{y} \in ]x(y), 1]\right)}{\mathbb{P}(Y > y)},$$

where we chose  $x(y) \uparrow 1$  such that

$$\int_{]x(y), 1]} \alpha t^{-\alpha} dt = o(\epsilon(y)), \text{ as } y \rightarrow \infty.$$

Finally (31) is proved by noticing that  $F^2(\log Y - \log y) \mathbb{1}_{\{Y=y\}} \equiv 1$  and using the preceding result. We have now proved that all the conditions of Proposition 2.3 are fulfilled. In particular, we have  $\|\mathbf{P}_u - \mathbf{P}_1\|_{\mathcal{G}} = O(\epsilon(y))$ , which permits to verify the bias condition (16) exactly the same way as for (105), formally replacing  $A(\cdot)$  by  $\epsilon(\cdot)$ , and hence proves Proposition 2.6 by using Theorem 2. We omit details.  $\square$

## 4.6 Proof of Proposition 2.7

We first prove (49). The fact that  $\mathbb{P}(\mathcal{W} \in E_{0, \frac{\alpha'}{2}}) = 1$  is a consequence of Proposition 2.6 combined with the fact that  $\sqrt{k_n}(\check{F}_n - \check{F}_{m(\alpha)})$  takes its values in  $E_{0, \frac{\alpha'}{2}}$ , which is closed for  $\|\cdot\|_{\infty, \frac{\alpha'}{2}}$ . Now by (106) combined with [36, p. 136, Theorem 2.6.4] and Dudley's chaining method [36, p. 101, Corollary 2.2.8] we deduce that  $\mathcal{W}$  admits a version that is uniformly continuous on  $[0, +\infty[ \times \mathcal{S}_{d-1}$  with respect to its intrinsic semimetric. In particular  $\mathcal{W}^{(1)}$  is almost surely uniformly continuous on  $[0, +\infty[$  with respect to  $\sqrt{d}(t_1, t_2) = \sqrt{|\exp(-\alpha t_1) - \exp(-\alpha t_2)|}$ . We now start the proof of Proposition 2.7. Let  $t_n \rightarrow 0$  and  $(\psi_n)$  be a sequence of  $E_{0, \frac{\alpha'}{2}}$  such that  $\|\psi_n - h\|_{\infty, \frac{\alpha'}{2}} \rightarrow 0$  for some  $h \in C_{0, \alpha}$  and such that  $\Psi_n := \check{F}_{m(\alpha)} + \epsilon_n \psi_n \in D_{\alpha'}$  for each  $n$ . Now consider the increment  $\epsilon_n^{-1}(T(\Psi_n) - T(\check{F}_{m(\alpha)}))$ , which has two coordinates. Elementary algebra combined with the non weighted convergence  $\psi_n \rightarrow \psi$  in  $\ell^\infty([0, +\infty[ \times \mathcal{S}_{d-1})$  shows that its second coordinate converges to  $\psi - \check{F}_{m(\alpha)}^{(1)} \psi^{(2)} + \check{F}_{m(\alpha)}^{(2)} \psi^{(1)}$  uniformly in  $(t, \boldsymbol{\theta})$ . The first coordinate

needs more care. It equals

$$\begin{aligned}
& \epsilon_n^{-1} \left[ \Psi_n^{(1)} \left( \cdot \int x d(-\Psi_n^{(1)}(x)) \right) - \check{F}_{\mathfrak{m}(\alpha)}^{(1)} \left( \cdot \int x d(-\check{F}_{\mathfrak{m}(\alpha)}^{(1)}(x)) \right) \right] \\
&= \epsilon_n^{-1} \left[ \check{F}_{\mathfrak{m}(\alpha)}^{(1)} \left( \cdot \int x d(-\Psi_n^{(1)}(x)) \right) - \check{F}_{\mathfrak{m}(\alpha)}^{(1)} \left( \cdot \int x d(-\check{F}_{\mathfrak{m}(\alpha)}^{(1)}(x)) \right) \right] + \psi_n^{(1)} \left( \cdot \int x d(-\Psi_n^{(1)}(x)) \right) \\
&=: a_n(\cdot) + b_n(\cdot).
\end{aligned}$$

Now notice that

$$\begin{aligned}
\alpha_n^{-1} &:= \int_0^\infty x d(-\Psi_n^{(1)}(x)) \\
&= \int_0^\infty x d(-\check{F}_{\mathfrak{m}(\alpha)}^{(1)}) + \epsilon_n \int_0^\infty x d(-\psi_n^{(1)}(x)) \\
&= \alpha^{-1} + \epsilon_n \int_0^\infty \psi_n^{(1)}(x) dx,
\end{aligned}$$

where the integration by parts used for the last term is possible because  $\psi_n \in E_{0, \frac{\alpha'}{2}}$ . Moreover since  $\|\psi_n - h\|_{\infty | \frac{\alpha'}{2}} \rightarrow 0$  we easily conclude that

$$\alpha_n^{-1} = \alpha^{-1} + \epsilon_n \int_0^\infty \psi^{(1)}(x) dx (1 + o(1)). \quad (110)$$

As a consequence one has, for fixed  $t \geq 0$

$$\begin{aligned}
|b_n(t) - \psi^{(1)}(t)| &\leq |\psi_n^{(1)}(t/\alpha_n) - \psi^{(1)}(t/\alpha_n)| + |\psi^{(1)}(t/\alpha_n) - \psi^{(1)}(t/\alpha)| \\
&\leq \sup_{t \geq 0} |\psi_n^{(1)}(t) - \psi^{(1)}(t)| + \eta \left( \exp\left(-t \frac{\alpha}{\alpha_n}\right) - \exp(-t) \right),
\end{aligned}$$

for a function  $\eta$  that has limit zero at zero (recall that  $\psi \in C_{0, \frac{\alpha'}{2}}$ ). This proves that  $b_n(t) \rightarrow \psi^{(1)}(t)$  uniformly in  $t \geq 0$ . Let us now deal with  $a_n(\cdot)$ . For fixed  $t$  one has, by definition of  $\check{F}_{\mathfrak{m}(\alpha)}$  and by (110)

$$\begin{aligned}
a_n(t) &= \epsilon_n^{-1} \left[ \exp(-\alpha t \alpha_n^{-1}) - \exp(-t) \right] \\
&= \epsilon_n^{-1} \left[ \exp\left(-\alpha t (\alpha^{-1} + \epsilon_n \int \psi^{(1)}(x) dx (1 + o(1)))\right) - \exp(-t) \right] \\
&= \exp(-t) \epsilon_n^{-1} \left[ \exp\left(-\alpha \epsilon_n t \int \psi^{(1)}(x) dx (1 + o(1))\right) - 1 \right],
\end{aligned}$$

where the symbols " $o(1)$ " are uniform in  $t$  and  $n \rightarrow \infty$ . Now a second order Taylor expansion combined with boundedness arguments show that  $a_n(t) \rightarrow -\alpha^{-1} t \exp(-t) \int \psi^{(1)}(x) dx$  uniformly in  $t \geq 0$ , which concludes the proof of Proposition 2.7.  $\square$

## 4.7 Proof of Corollary 1 and Proposition 2.8

Let  $\mathcal{T}_0$  be any countable dense subset of  $[0, +\infty[ \times \mathcal{S}_{d-1}$ . A function  $\psi \in \ell^\infty([0, +\infty[ \times \ell^\infty([0, +\infty[ \times \mathcal{S}_{d-1})$  is said to be  $\mathcal{T}_0$ -separable if, for any  $(t, s) \in [0, +\infty[ \times \mathcal{S}_{d-1}$ , there exists a  $\mathcal{T}_0$  valued sequence  $(t_m, s_m) \rightarrow (t, s)$  such that  $\psi(t_m, s_m) \rightarrow \psi(t, s)$  in  $\mathbb{R}^2$ . Write  $\mathbf{V}'$  for the closed subspace of  $\ell^\infty([0, +\infty[ \times \ell^\infty([0, +\infty[ \times \mathcal{S}_{d-1})$  of all these functions. By "right continuity" properties of distribution functions it is easy to see that  $T(D_{\alpha'}) \subset \mathbf{V}'$ . We will now apply [1, Theorem 4.1] to  $T_n := \check{\mathbb{F}}_n$ ,  $\theta_0 := \check{F}_{\mathfrak{m}(\alpha)}$ ,  $V := \mathcal{W}$ ,  $f := T$ ,  $\mathbb{C}_0 := C_{0, \frac{\alpha'}{2}}$ ,  $\mathbf{V} := E_{\alpha'}$ ,  $\mathbf{V}_f := D_{\alpha'}$ ,  $(\mathbf{V}_0, \|\mathbf{v}_0\|) := (E_{0, \frac{\alpha'}{2}}, \|\cdot\|_{\infty | \frac{\alpha'}{2}})$ , with  $(\mathbf{V}', \|\cdot\|_{\infty, \infty})$  as above, endowed with  $\mathcal{V}'$ , the  $\sigma$ -algebra spanned by its open balls (we define  $\mathcal{V}_0$  likewise for  $\mathbf{V}_0$ ). The measurability condition in (ii) is met because the class  $\mathcal{F}$  involved in the proof of Proposition 2.6 is pointwise measurable. It remains to verify (iii) before being able to apply [1, Theorem 4.1]. That condition is quite cumbersome to verify in its full generality. However, as pointed out at the beginning of the proof of [1, Theorem 4.1], we only need to prove that  $T(\check{\mathbb{F}}_n)$  is  $\mathcal{V}'$  measurable. This is verified by noticing that, for fixed  $\psi \in \mathbf{V}'$  and  $\epsilon > 0$ , recalling the definition of  $\mathbf{V}'$ :

$$\left\{ \left\| T(\check{\mathbb{F}}_n) - \psi \right\|_{\infty, \infty} \leq \epsilon \right\} = \bigcap_{(t,s) \in \mathcal{T}_0} \left\{ T(\check{\mathbb{F}}_n)(t, s) \in B(\psi(t, s), \epsilon) \right\},$$

where each  $B(\psi(t, s), \epsilon)$  is a closed set of  $\mathbb{R}^2$ . We can now apply [1, Theorem 4.1] to conclude that

$$\sqrt{k_n} \left( T(\check{\mathbb{F}}_n) - T(\check{F}_{\mathfrak{m}(\alpha)}) \right) \xrightarrow{d} \text{Deriv}_{\check{F}_{\mathfrak{m}(\alpha)}}(\mathcal{W}), \text{ in } \left( \ell^\infty([0, +\infty[ \times \ell^\infty([0, +\infty[ \times \mathcal{S}_{d-1}), \|\cdot\|_{\infty, \infty} \right),$$

and apply the continuous mapping theorem with the map  $\|\cdot\|_{\infty, \infty}$  to conclude the proof of Corollary 1.  $\square$

We will now prove Proposition 2.8. To that end, we will make use of an extended continuous mapping (see [43, Corollary 1]) that is suitably adapted to sequence of mapping  $g_n$  that depend upon the observed sample. Borrowing the notation of the just cited article, we take  $(\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$  as any probability space supporting the sequence  $(X_n)$ , and we take  $(\Omega_2, \mathcal{A}_2, \mathbb{P}_2)$  as any probability space supporting an independent sequence  $(\vec{M}_n)$  such that each  $\vec{M}_n$  is  $(n_{k_n}, n_{k_n}^{-1}, \dots, n_{k_n}^{-1})$  multinomial. Note that, for fixed  $n \geq$  and  $\omega_1 \in \Omega_1$  one has (recall that  $Y_i := \|X_i\|_d$ )

$$\tilde{\Delta}_{k_n}(\tilde{\mathbb{P}}_n^{(n)}(\omega_1)) = \sup_{\phi \in BL_1} \left| \mathbb{E}_{\vec{M}_n} \left( \phi \left( \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} (M_{i,n} - 1) \delta_{Z_{r(i)}(\omega_1)} \right) - \mathbb{E}(\phi(\mathbb{G})) \right) \right|,$$

where  $Z_i := (\log(Y_i / \log Y_{n-k_n \cdot n}), Y_i^{-1} X_i)$ , where  $r(i)$  is the  $i$ -th first indice among  $\mathcal{I}_1(Y_1(\omega_1), \dots, Y_n(\omega_1))$ . By the isometry property of the map  $\mathfrak{J}$  defined in (52) we see that Theorem 3 applied in the setup of Proposition 2.6 yields

$$\left( \sup_{\phi \in BL_1(E_{\alpha'})} \left| \mathbb{E}_2(\phi(\mathcal{Z}_n)) - \mathbb{E}(\phi(\mathcal{W})) \right| \right)^* \rightarrow_{\mathbb{P}_1} 0, \text{ where}$$

$$\mathcal{Z}_n(\omega_1, \omega_2) := \mathfrak{J} \left( \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} (M_{i,n}(\omega_2) - 1) \delta_{Z_{r_{\omega_1}(i)}(\omega_1)} \right), (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2.$$

Now define the maps

$$\begin{aligned}
g_n(\omega_1, \psi) &:= h(\tilde{\mathbb{P}}_n^{(n)}(\omega_1), \psi), \\
g(\psi) &:= h(\mathbf{P}_1, \psi), \quad \psi \in D_{\alpha'}, \text{ with} \\
h(P, \psi) &:= (t, \boldsymbol{\theta}) \rightarrow \left( - (t/\alpha(P)) \exp(-t) \int \psi^{(1)}(x) dx + \psi^{(1)}(t/\alpha(P)), \right. \\
&\quad \left. \psi(t, \boldsymbol{\theta}) - \psi^{(1)}(t) \check{F}_P^{(2)}(\boldsymbol{\theta}) - \check{F}_P^{(1)}(t) \psi^{(2)}(\boldsymbol{\theta}) \right),
\end{aligned}$$

and with  $\alpha(P)$  defined as in (53). By separability properties of cumulative distribution functions, each  $g_n(\omega_1, \cdot)$  maps  $D_{\alpha'}$  to the space  $\mathbf{V}'$  defined at the beginning of this section. This point is crucial to verify that each  $g_n$  satisfies the measurability assumption (5) in [43], with

$$(D, d_D) := (D_{\alpha'}, \|\cdot\|_{\infty, \frac{\alpha'}{2}}), \quad (E, d_E) := (\mathbf{V}', \|\cdot\|_{\infty, \infty}), \quad D_n := D_{\alpha'}, \quad D_0 := C_{0, \frac{\alpha'}{2}}. \quad (111)$$

That assumption is the Borel measurability, for fixed  $n \geq 1$ ,  $\delta > 0$  and  $\psi \in D_0$ , of

$$\begin{aligned}
&\omega_1 \rightarrow \sup_{\psi' \in D_{\alpha'}, \|\psi' - \psi\|_{\infty, \frac{\alpha'}{2}} < \delta} \|h(\tilde{\mathbb{P}}_n^{(n)}(\omega_1), \psi') - h(\mathbf{P}_1, \psi)\|_{\infty, \infty} \\
&\equiv \sup_{\psi' \in D_{\alpha'}, \|\psi' - \psi\|_{\infty, \frac{\alpha'}{2}} < \delta} \sup_{(t, \boldsymbol{\theta}) \in \mathcal{T}_0} \left| - (t/\alpha(\tilde{\mathbb{P}}_n^{(n)})) \exp(-t) \int \psi'^{(1)}(x) dx + \psi'^{(1)}(t/\alpha(\tilde{\mathbb{P}}_n^{(n)})) - a(\psi, t, \boldsymbol{\theta}) \right| \\
&\quad \vee \left| \psi'(t, \boldsymbol{\theta}) - \check{\mathbb{F}}_n^{(1)}(t) \psi'^{(2)}(\boldsymbol{\theta}) - \psi'^{(1)}(t) \check{\mathbb{F}}_n^{(2)}(\boldsymbol{\theta}) - b(\psi, t, \boldsymbol{\theta}) \right|,
\end{aligned}$$

where  $h(\mathbf{P}_1, \psi) =: \{(t, \boldsymbol{\theta}) \rightarrow (a(\psi, t, \boldsymbol{\theta}), b(\psi, t, \boldsymbol{\theta}))\}$ . Now, inverting the order of suprema, the preceding expression equals

$$\begin{aligned}
&\sup_{(t, \boldsymbol{\theta}) \in \mathcal{T}_0} \sup_{\|\psi' - \psi\|_{\infty, \frac{\alpha'}{2}} < \delta} \left| \left( \frac{-te^{-t}}{\alpha(\tilde{\mathbb{P}}_n^{(n)})} \int \psi'^{(1)} + \psi'^{(1)} \left( \frac{t}{\alpha(\tilde{\mathbb{P}}_n^{(n)})} \right), \psi'(t, \boldsymbol{\theta}) - \check{\mathbb{F}}_n^{(1)}(t) \psi'^{(2)}(\boldsymbol{\theta}) - \psi'^{(1)}(t) \check{\mathbb{F}}_n^{(2)}(\boldsymbol{\theta}) \right) \right. \\
&\quad \left. - (a(\psi, t, \boldsymbol{\theta}), b(\psi, t, \boldsymbol{\theta})) \right|_2,
\end{aligned}$$

where  $\|(x, y)\|_2 := |x| \vee |y|$ . Now, for fixed  $(t, \boldsymbol{\theta})$  the map  $(\alpha(\tilde{\mathbb{P}}_n^{(n)}), \check{\mathbb{F}}_n^{(1)}(t), \check{\mathbb{F}}_n^{(2)}(\boldsymbol{\theta}))$  is Borel measurable (recall that  $\tilde{\mathbb{P}}_n^{(n)}$  is  $\mathcal{M}$ -measurable). In addition, for fixed  $(\boldsymbol{\theta}, t)$  the map

$$\begin{aligned}
S_{t, \boldsymbol{\theta}}(x_1, x_2, x_3) &:= \sup_{\|\psi' - \psi\|_{\infty, \frac{\alpha'}{2}} < \delta} \left| \left( \frac{-te^{-t}}{x_1} \int \psi'^{(1)} + \psi'^{(1)} \left( \frac{t}{x_1} \right), \psi'(t, \boldsymbol{\theta}) - x_2 \psi'^{(2)}(\boldsymbol{\theta}) - \psi'^{(1)}(t) x_3 \right) \right. \\
&\quad \left. - (a(\psi, t, \boldsymbol{\theta}), b(\psi, t, \boldsymbol{\theta})) \right|_2, \quad (x_1, x_2, x_3) \in ]0, +\infty[ \times [0, 1]^2
\end{aligned}$$

can be seen as the supremum of uncountably many lower semicontinuous functions indexed by  $\psi'$ . Indeed for fixed  $\psi' \in D_{\alpha'}$  the map  $x_3 \rightarrow \psi'^{(1)}(t/x_3)$  is lower semicontinuous

because  $\psi^{(1)}$  is a survival function on  $[0, +\infty[$ . Hence  $S_{t,\boldsymbol{\theta}}$  is Borel measurable, which proves assumption (5) in [43] by the composition  $S_{t,\boldsymbol{\theta}} \circ (\alpha(\tilde{\mathbb{P}}_n^{(n)}), \check{\mathbb{F}}_n^{(1)}(t), \check{\mathbb{F}}_n^{(2)}(\boldsymbol{\theta}))$  followed with taking the countable supremum in  $(t, \boldsymbol{\theta}) \in \mathcal{T}_0$ .

The final step is to verify assumption  $(Hg')$  in [43] in order to apply the Corollary 1 therein. A careful look of the arguments of the proofs in [43] show that it is sufficient to verify the following weaker form of  $(Hg')$ : for any compact set  $K$  of  $D_0$  and for any  $\delta > 0$  the following convergence holds:

$$\mathbb{P}_1^* \left( \left\{ \omega_1 \in \Omega_1, \exists z \in K, \forall \tau > 0, \sup_{z' \in D_n, d_D(z', z) < \tau} d_E(g_n(\omega_1, z'), g(z)) > \delta \right\} \right) \rightarrow 0.$$

Note that the occurrences of  $z'$  instead of  $y$  are here to correct a misprint in [43]. Recalling (111), we choose We have to prove that

$$\mathbb{P}^* \left( \exists \psi \in K, \forall \tau > 0, \sup_{\substack{\psi' \in D_{\alpha'} \\ \|\psi' - \psi\|_{\infty|\frac{\alpha'}{2}} < \tau}} \left\| h(\tilde{\mathbb{P}}_n^{(n)}, \psi') - h(\mathbf{P}_1) \right\|_{\infty, \infty} > \delta \right) \rightarrow 0. \quad (112)$$

Elementary algebra shows that:

$$\begin{aligned} & \left\| h(\tilde{\mathbb{P}}_n^{(n)}, \psi') - h(\mathbf{P}_1, \psi) \right\|_{\infty, \infty} \\ \leq & \left( \sup_{t \geq 0} t e^{-t} \right) \left[ \frac{\int |\psi'^{(1)} - \psi^{(1)}|}{\alpha(\mathbf{P}_1)} + \left( \frac{1}{\alpha(\tilde{\mathbb{P}}_n^{(n)})} - \frac{1}{\alpha(\mathbf{P}_1)} \right) \int |\psi'^{(1)}| \right] \\ & + 4 \|\psi' - \psi\|_{\infty|\frac{\alpha'}{2}} + \sup_{t \geq 0} \left| \psi^{(1)} \left( \frac{t}{\alpha(\tilde{\mathbb{P}}_n^{(n)})} \right) - \psi^{(1)} \left( \frac{t}{\alpha(\mathbf{P}_1)} \right) \right| + 2 \|\psi'\|_{\infty|\frac{\alpha'}{2}} \|\tilde{\mathbb{P}}_n^{(n)} - \mathbf{P}_1\|_{\mathcal{F}}, \end{aligned}$$

which, if  $\|\psi' - \psi\|_{\infty|\frac{\alpha'}{2}} < \tau$  is less than

$$\begin{aligned} \leq & \left( \sup_{t \geq 0} t e^{-t} \right) \left[ \frac{2\tau}{\alpha' \alpha(\mathbf{P}_1)} + \left( \frac{1}{\alpha(\tilde{\mathbb{P}}_n^{(n)})} - \frac{1}{\alpha(\mathbf{P}_1)} \right) \left( \int |\psi^{(1)}| + \frac{2\tau}{\alpha'} \right) \right] \\ & + 4\tau + \sup_{t \geq 0} \left| \psi^{(1)} \left( \frac{t}{\alpha(\tilde{\mathbb{P}}_n^{(n)})} \right) - \psi^{(1)} \left( \frac{t}{\alpha(\mathbf{P}_1)} \right) \right| + 2(\|\psi\|_{\infty|\frac{\alpha'}{2}} + \tau) \|\tilde{\mathbb{P}}_n^{(n)} - \mathbf{P}_1\|_{\mathcal{F}}, \end{aligned} \quad (113)$$

where the last inequality is valid as soon as  $\|\psi' - \psi\|_{\infty|\frac{\alpha'}{2}} < \tau$ . Now as  $K$  is a  $\|\cdot\|_{\infty|\frac{\alpha'}{2}}$ -totally bounded subset of  $C_{0, \frac{\alpha'}{2}}$  the quantity  $\|\psi\|_{\infty|\frac{\alpha'}{2}}$  is bounded in  $\psi \in K$ . Moreover,  $K$  is also totally bounded for the (weaker) uniform norm over  $[0, +\infty[ \times \mathcal{S}_{d-1}$ . Hence it satisfies (by Ascoli's theorem)

$$\lim_{\delta \rightarrow 0} \sup_{\psi \in K} \sup_{\substack{t_1, t_2 \geq 0 \\ |e^{-\alpha t_1} - e^{-\alpha t_2}| \leq \delta}} \left| \psi^{(1)}(t_1) - \psi^{(1)}(t_2) \right| = 0. \quad (114)$$

This is enough to conclude that  $\int |\psi|$  is also bounded in  $\psi \in K$ . As a consequence there exists a constant  $C$  depending only upon  $K$  such that (113) is less than  $C$  times

$$\frac{2\tau}{\alpha' \alpha(\mathbf{P}_1)} + \left( \frac{1}{\alpha(\tilde{\mathbb{P}}_n^{(n)})} - \frac{1}{\alpha(\mathbf{P}_1)} \right) + 4\tau + \sup_{t \geq 0} \left| \psi^{(1)}\left(\frac{t}{\alpha(\tilde{\mathbb{P}}_n^{(n)})}\right) - \psi^{(1)}\left(\frac{t}{\alpha(\mathbf{P}_1)}\right) \right| + \|\tilde{\mathbb{P}}_n^{(n)} - \mathbf{P}_1\|_{\mathcal{F}}.$$

For fixed  $\tau > 0$ , we see that the lim sup in probability of the preceding display is less than  $(2/\alpha'\alpha + 6)\tau$  as soon as we prove that

$$\limsup \sup_{\psi \in K} \sup_{t \geq 0} \left| \psi^{(1)}\left(\frac{t}{\alpha(\tilde{\mathbb{P}}_n^{(n)})}\right) - \psi^{(1)}\left(\frac{t}{\alpha(\mathbf{P}_1)}\right) \right| \leq \tau, \text{ in probability.} \quad (115)$$

But elementary analysis shows that (recall that  $\alpha = \alpha(\mathbf{P}_1)$ )

$$\sup_{t \geq 0} \left| \exp\left(-\frac{\alpha t}{\alpha(\tilde{\mathbb{P}}_n^{(n)})}\right) - \exp\left(-\frac{\alpha t}{\alpha(\mathbf{P}_1)}\right) \right| \leq \left( \sup_{t \geq 0} t e^{-t} \right) \left| \frac{\alpha(\mathbf{P}_1)}{\alpha(\tilde{\mathbb{P}}_n^{(n)})} - 1 \right|, \quad (116)$$

which converges to zero as  $n \rightarrow \infty$ , by  $\mathcal{F}$ -continuity of  $P \rightarrow \alpha(P)$ . This concludes the proof using (115).  $\square$

## Acknowledgement

The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-20-CE40-0025-01 (T-REX project).

## References

- [1] E. Beutner and H. Zähle. A modified functional delta method and its application to the estimation of risk functionals. *Journ. Multiv. Anal.*, 101(10):2452–2463, 2010.
- [2] E. Beutner and H. Zähle. Deriving the asymptotic distribution of U- and V-statistics of dependent data using weighted empirical processes. *Bernoulli*, 18(3):803–822, 2012.
- [3] E. Beutner and H. Zähle. Continuous mapping approach to the asymptotics of  $u$ - and  $v$ -statistics. *Bernoulli*, 20(2):846–877, 2014.
- [4] E. Beutner and H. Zähle. Functional delta-method for the bootstrap of quasi-Hadamard differentiable functionals. *Electron. Journ. Statist.*, 10(1):1181–1222, 2016.
- [5] E. Beutner and H. Zähle. Bootstrapping average value at risk of single and collective risks. *Risks*, 6(3):96–125, 2018.
- [6] S. Bouzebda and I. Elhattab. Uniform-in-bandwidth consistency for kernel-type estimators of Shannon’s entropy. *Electron. J. Statist.*, 5:440–459, 2011.

- [7] S. Bouzebda and B. Nemouchi. Uniform consistency and uniform in bandwidth consistency for nonparametric regression estimates and conditional u-statistics involving functional data. *Journal of Nonparametric Statistics*, 32(2):452–509, 2020.
- [8] L. De Haan and A. Ferreira. *Extreme value theory: an introduction*. Springer Science & Business Media, 2007.
- [9] P. Deheuvels, U. Einmahl, and D. Mason. Asymptotic independence of the local empirical process indexed by functions. In *High dimensional probability, II*, pages 183–205, 2000.
- [10] P. Deheuvels and D.M. Mason. Nonstandard functional laws of the iterated logarithm for tail empirical and quantile processes. *Ann. Probab.*, 18(4):1693–1722, 1990.
- [11] P. Deheuvels and D.M. Mason. A tail empirical process approach to some nonstandard laws of the iterated logarithm. *J. Theoret. Probab.*, 4:53–85, 1991.
- [12] P. Deheuvels and D.M. Mason. Functional laws of the iterated logarithm for the increments of empirical and quantile processes. *Ann. Probab.*, 20(3):1248–1287, 1992.
- [13] P. Deheuvels and D.M. Mason. Functional laws of the iterated logarithm for local empirical processes indexed by sets. *Ann. Probab.*, 22(3):1619–1661, 1994.
- [14] P. Deheuvels and D.M. Mason. Nonstandard local empirical processes indexed by sets. *J. Statist. Plann. Inference*, 23(1):355–387, 1995.
- [15] P. Deheuvels and D.M. Mason. On the fractal nature of empirical increments. *Ann. Probab.*, 45(1-2):91–112, 1995.
- [16] P. Deheuvels and D.M. Mason. General asymptotic confidence bands based on kernel-type function estimators. *Stat. Inference Stoch. Process.*, 7(3):225–277, 2004.
- [17] P. Deheuvels and S. Ouadah. Uniform-in-Bandwidth Functional Limit Laws. *J. Theoret. Probab.*, pages 1–25, 2011.
- [18] L. Devroye and G. Lugosi. *Combinatorial methods in density estimation*. Springer, New York, 2001.
- [19] J. Dony and U. Einmahl. Uniform in bandwidth consistency of kernel regression estimators at a fixed point. In *High Dimensional Probability V*, pages 308–325, 2009.
- [20] J. Dony, U. Einmahl, and D. Mason. Uniform in bandwidth consistency of local polynomial regression function estimators. *Austr. J. Stat.*, 35:105–120, 2006.
- [21] J. Dony and D. M. Mason. Uniform in bandwidth consistency of conditional  $U$ -statistics. *Bernoulli*, 14(4):1108–1133, 2008.

- [22] H. Drees, L. de Haan, and D. Li. Approximations to the tail empirical distribution function with application to testing extreme value conditions. *Statist. Plann. Inference*, 136(10):3498–3538, 2006.
- [23] J. H. J. Einmahl, F. Yang, and C. Zhou. Testing the multivariate regular variation model. *J. Bus. Econom. Statist.*, 39:1–13, 2020.
- [24] J.H.J. Einmahl. The as behavior of the weighted empirical process and the lil for the weighted tail empirical process. *Ann. Probab.*, 20(2):681 – 695, 1992.
- [25] J.H.J. Einmahl. Poisson and gaussian approximation of weighted local empirical processes. *Stochastic Process. and their Appl.*, 70(1):31 – 58, 1997.
- [26] U. Einmahl and D.M. Mason. Gaussian approximation of local empirical processes indexed by functions. *Probab. Theory Related Fields*, 107(3):283–311, 1997.
- [27] U. Einmahl and D.M. Mason. An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoret. Probab.*, 13:1–37, 2000.
- [28] U. Einmahl and D.M. Mason. Uniform in bandwidth consistency of variable bandwidth kernel estimators. *Ann. Statist.*, 33(3):1380–1403, 2005.
- [29] E. Giné and D. Mason. Uniform in bandwidth estimation of integral functionals of the density function. *Scand. J. of Statist.*, 35(4):739–761, 2008.
- [30] J. Hoffmann-Jorgensen. *Stochastic processes on Polish spaces*. Number 39 in Various Publication Series. Aarhus Universitet, Aarhus, Denmark, 1991.
- [31] D. M. Mason. A uniform functional law of the logarithm for the local empirical process. *Ann. Probab.*, 32(2):1391–1418, 2004.
- [32] David M Mason. A uniform functional law of the logarithm for a local gaussian process. In *High Dimensional Probability III*, pages 135–151. Springer, 2003.
- [33] S. Y. Novak. *Extreme value methods with applications to finance*. CRC Press, 2011.
- [34] K. Ray and A. van der Vaart. On the bernstein-von mises theorem for the dirichlet process. *Electronic Journal of Statistics*, 15(1):2224–2246, 2021.
- [35] A. Sheehy and Jon A Wellner. Uniform Donsker classes of functions. *Ann. Probab.*, 20(4):1983–2030, 1992.
- [36] A.W. Van der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes*. Springer, New-York, 1996.
- [37] I. Van Keilegom and D. Varron. Uniform in bandwidth exact rates for a class of kernel estimators. *Ann. Inst. Statist. Math.*, 63(6):1077–1102, 2011.

- [38] D. Varron. A limited in bandwidth uniformity for the functional limit law for the increments of the empirical process. *Electron. J. Statist.*, 2:1043–1064, 2008.
- [39] D. Varron. Donsker and Glivenko-Cantelli theorems for a class of processes generalizing the empirical process. *Electron. J. Stat.*, 8:2301–2325, 2014.
- [40] D. Varron. Functional limit laws for the local empirical process in a spatial setting. *Stochastics*, 2015.
- [41] D. Varron. Limit theorems for occupation rates of local empirical processes. *Sankhyā Ser. A*, In press, 2015.
- [42] D. Varron. The almost sure topological limits of collections of local empirical processes at many different scales. *Stoch. Proc. Appl.*, 126(4):997–1018, 2016.
- [43] D. Varron. An extended continuous mapping theorem for outer almost sure weak convergence. *Theory of Probability & Its Applications*, 64(2):304–323, 2019.
- [44] David Williams. *Probability with martingales*. Cambridge university press, 1991.