ALEA, Lat. Am. J. Probab. Math. Stat. **22**, 627–650 (2025)
DOI: 10.30757/ALEA.v22-24

On the bracketing entropy condition and generalized empirical measures

Davit Varron

Université Marie et Louis Pasteur, CNRS, LmB, F-25000 Besançon, France
E-mail address: davit.varron@univ-fcomte.fr

Abstract. We prove a Donsker and a Glivenko–Cantelli theorem for sequences of random discrete measures generalizing empirical measures. Those two results hold under standard conditions upon bracketing numbers of the indexing class of functions. As a byproduct, we derive a posterior consistency and a Bernstein–von Mises theorem for the Dirichlet process prior, under the topology of total variation, when the observation space is countable. We also obtain new insights upon the Durst–Dudley–Borisov theorem

1. Introduction

1.1. *Frequentist analysis of Bayesian nonparametrics, for the total variation distance.* Nonparametric Bayes theory has been the subject of thorough investigations during the past decades (for a recent monograph on this topic, see e.g. Ghosal and van der Vaart, 2017). This theory focuses on priors taking values in an infinite dimensional space. A typical example is that of a random probability measure Pr on a measurable space $\big(\mathfrak{X}, \mathcal{A}_{\mathfrak{X}}\big)$, for which we assume that a $\mathfrak{X}^n$-valued sample $(X_1, \dots, X_n)$ has law $\mathbf{P}^{\otimes n}$ given $Pr = \mathbf{P}$. The *Bayesian* analysis consists in - given an observed sample $(x_1, \dots, x_n)$ - to update our prior Pr to a posterior distribution $Pr(x_1, \dots, x_n)$, the law of Pr given $(X_1, \dots, X_n) = (x_1, \dots, x_n)$. Now assume that our sampling sequence obeys a *frequentist* model - that is (X_n) is i.i.d with common law $\mathbf{P}_0$ - and that we analyze it with a *Bayes procedure* instead. Can we show that almost every sequence $Pr(x_1, \dots, x_n)$ converges to $\mathbf{P}_0$ in *total variation*, i.e.

$$\sup_{A \in \mathcal{A}_{\mathfrak{X}}} \Big| Pr(x_1, \dots, x_n) - \mathbf{P}_0 \Big| \to_{\mathcal{L}} \ 0? \tag{1.1}$$

Do we have a rate of convergence in the sense that

$$\sqrt{n} \sup_{A \in \mathcal{A}_{\mathfrak{X}}} \Big| Pr(x_1, \dots, x_n) - \mathbf{P}_0 \Big| \to_{\mathcal{L}} \mathcal{U}, \tag{1.2}$$

for a non null distribution $\mathcal{U}$ on $[0, +\infty[$? Those two results belong to the rich theory of posterior consistency and Bernstein-von Mises (BvM) results in an infinite dimensional setting. Several results in this vein have been established (see, e.g., Castillo, 2012; Castillo and Nickl, 2014; James, 2008;

Received by the editors February 7th, 2019; accepted November 16th, 2021.
2010 *Mathematics Subject Classification.* 60F17,60G57.
Key words and phrases. Empirical measures, Donsker theorems, Bayesian nonparametric statistics.

Johnstone, 2010; Kim, 2006; Kim and Lee, 2004; Leahu, 2011; Lo, 1983, 1986 and the references therein) for different choices of distances between probability measures. Total variation induces one of the strongest distances between probability distributions. In this paper, we will prove that both (1.1) and (1.2) may happen when $\mathfrak{X}$ is countable and Pr is a Dirichlet process prior. These results hold under conditions are very similar to those of a result in empirical processes theory called the *Durst-Dudley-Borisov theorem* - see Dudley (2014, Theorem 7.9, p. 279) or Theorem 3 in §4.1. This similarity is not a coincidence: it is a consequence of our investigation upon bracketing entropy and generalized empirical measures (in the vein of Varron, 2014), which is described in the following subsection.

1.2. *Random discrete measures and bracketing entropy.* From now on we will adopt the generic notation (for $r \in [1, \infty]$)

$$\ell^r := \Big\{ \boldsymbol{p} \in \mathbb{R}^{\mathbb{N}},\ ||\, \boldsymbol{p}\, ||_r < \infty \Big\}, \text{ where}$$
$$||\, \boldsymbol{p}\, ||_r^r := \sum_{i \in \mathbb{N}} |\, p_i\, |^r, \text{ for } 1 \le r < \infty, \text{ and where}$$
$$||\, \boldsymbol{p}\, ||_\infty := \sup_{i \in \mathbb{N}} |\, p_i\, |\ , \text{ writing } \boldsymbol{p} = (p_i)_{i \in \mathbb{N}}.$$

We shall also use the notation

$$Q(f) := \int_{\mathfrak{X}} f dQ, \tag{1.3}$$

for a given signed measure Q in $\mathcal{A}_{\mathfrak{X}}$ with finite total variation, and for each $f \in L^1(Q)$. Any $\mathfrak{X}$-valued sequence $\mathbf{y} = (y_i)_{i \in \mathbb{N}}$ combined with an element $\boldsymbol{p} \in \ell^1$ defines a signed discrete measure on $\big(\mathfrak{X}, \mathcal{A}_{\mathfrak{X}}\big)$ - with finite total variation - through the following convex combination:

$$P_{\mathbf{y}, \boldsymbol{p}} := \sum_{i \in \mathbb{N}} p_i \delta_{y_i}. \tag{1.4}$$

Now substitute $\mathbf{y}$ by an $\mathcal{A}_{\mathfrak{X}}^{\otimes \mathbb{N}}$ measurable sequence $\mathbf{Y} = (Y_i)_{i \in \mathbb{N}}$, and $\boldsymbol{p}$ by a ℓ^1-valued Borel random variable $\boldsymbol{\beta} = (\beta_i)_{i \in \mathbb{N}}$ (both of them on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$). Then the composition map $P_{\mathbf{Y}, \boldsymbol{\beta}}$ defines a random signed measure in the following sense: for any specified bounded Borel function f, the map

$$P_{\mathbf{Y}, \boldsymbol{\beta}}(f): \quad \omega \to P_{\mathbf{Y}_n(\omega), \boldsymbol{\beta}_n(\omega)}(f) \tag{1.5}$$

is Borel from $\big(\Omega, \mathcal{A}\big)$ to $\mathbb{R}$. In the sequel we shall continue to adopt the same convention (1.3) for $P(f)$ when P is a random or non random measure, and we shall extend it - when meaningful - to functions f that are not necessarily bounded.
In Varron (2014), the author started the investigation on how well known results in empirical processes theory (see, e.g., Dudley, 2014; van der Vaart and Wellner, 1996 for monographs on the subject) could be carried over sequences of random signed measures of the form $P_{\mathbf{Y}_n, \boldsymbol{\beta}_n}$ where, for each n, the sequence $(Y_{i,n})_{i \in \mathbb{N}}$ is independent and identically distributed *given* $\boldsymbol{\beta}_n$. He showed that the uniform entropy numbers and the Koltchinskii–Pollard uniform entropy integral - two crucial notions in empirical processes theory - both adapt very smoothly to that wider class of random measures, which not only encompasses the empirical measure, but also discrete nonparametric Bayesian priors. The latter notion of uniform entropy integral can be briefly defined as follows for a class $\mathcal{F}$ of real Borel functions on $\big(\mathfrak{X}, \mathcal{A}_{\mathfrak{X}}\big)$:

$$J(\delta, \mathcal{F}) := \int_0^\delta \sqrt{\log \Big(\sup_{Q \ probab.} N\big(\epsilon\, ||\, F\, ||_{Q,2}, \mathcal{F}, ||\cdot||_{Q,2}\, \big)\Big)} d\epsilon, \quad \delta \in (0, \infty]. \tag{1.6}$$

Here $||\cdot||_{Q,2}$ stands for the $L^2(Q)$ norm, $N(\epsilon, \mathcal{F}, ||\cdot||_{Q,2})$ denotes the minimal number of $||\cdot||_{Q,2}$ balls with radius ϵ needed to cover $\mathcal{F}$, F stands for a measurable envelope of the class $\mathcal{F}$ (see, e.g., van der Vaart and Wellner, 1996, p. 85), and the supremum in (1.6) is taken over all probability measures satisfying $\| F \|_{Q,2} > 0$. Note that F can be simply taken as

$$F(y) := \sup\Big\{ \mid f(y) \mid, f \in \mathcal{F}\Big\},\ y \in \mathfrak{X},$$

when $\mathcal{F}$ is countable or pointwise measurable - see §2.2 below. When $J(\infty, \mathcal{F})$ is finite, Varron proved a Donsker theorem under natural asymptotic conditions upon $(\boldsymbol{\beta}_n, \mathbf{Y}_n)_{n\geq 1}$. Those two asymptotic theorems (see Varron, 2014, Theorems 1 and 2) involve processes of the form

$$G_n(f) := \sum_{i\in\mathbb{N}} \beta_{i,n}\Big[f(Y_{i,n}) - \mathbb{E}\Big(f(Y_{i,n}) \mid \beta_{i,n}\Big)\Big],\ f \in \mathcal{F}, \tag{1.7}$$

indexed by a class $\mathcal{F}$ of real Borel functions. A rigorous definition of $G_n(\cdot)$ is not immediate and is therefore voluntarily postponed to §2.2.

A Donsker result for such a sequence involves the approximating "Gaussian analogues" $W_n(\cdot)$ of each $G_n(\cdot)$. Indeed, the $W_n(\cdot)$ are not Gaussian processes but rather mixtures in the following heuristic sense : first randomly generate $\boldsymbol{\beta}_n$, then take $\mathbf{P}_n$ as the law of the $Y_{i,n}$ given $\boldsymbol{\beta}_n$ and generate a $\mathbf{P}_n$-Brownian bridge indexed by $\mathcal{F}$ (for more details and a rigorous construction, see Varron, 2014).

While the uniform entropy has been celebrated as a useful condition to prove that a class $\mathcal{F}$ is Donsker or Glivenko–Cantelli, another condition turned out to be very fruitful as well: bracketing entropy. The bracket $[\![f^-, f^+]\!]$ between two Borel functions f^- and f^+ is defined as the set of Borel functions f fulfilling $f^- \prec f \prec f^+$, the symbol $\prec$ standing for the everywhere pointwise comparison between real functions on $\mathfrak{X}$. Denoting by $N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|_{Q,2})$ the minimal number of brackets with $\|\cdot\|_{Q,2}$ diameter less than ϵ needed to cover $\mathcal{F}$, the Q bracketing entropy of $\mathcal{F}$ is defined as

$$J_{[]}(\delta, \mathcal{F}, Q) := \int_0^\delta \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|_{Q,2})} d\epsilon,\quad \delta \in (0, \infty]. \tag{1.8}$$

A naturally arising question is then: *does bracketing entropy adapt with the same efficiency to sequences of random measures such as in* (1.5)*?* The answer provided in the present article is: *yes, but to a lesser extent.* More restrictions upon the weights are needed. First the $\beta_{i,n}$ have to be non negative, since the idea of bracketing relies on the comparison principle

$$f^- \prec f \prec f^+ \Rightarrow Q(f^-) \leq Q(f) \leq Q(f^+),$$

when Q is a non negative measure. Second, when looking for a Donsker theorem, $\| \boldsymbol{\beta}_n \|_\infty$ has to tend to zero fast enough to counterbalance the possible growth of $\| \boldsymbol{\beta}_n \|_1$. The amount of compensation is directly linked to the moments of $F(Y_{1,n}),\ n \in \mathbb{N}^*$.

Those two conditions were not required under the assumption that $J(\infty, \mathcal{F})$ is finite (see Varron, 2014, Theorems 1 and 2). This difference can be explained by the fact that the use of the Koltchinskii–Pollard entropy is intimately linked to that of symmetrization, namely the study of

$$G_n^0(f) := \sum_{i\in\mathbb{N}} \epsilon_i \beta_{i,n} f(Y_{i,n}),\ f \in \mathcal{F},$$

where the ϵ_i are symmetric Bernoulli (or Rademacher) random variables, independent of $(\mathbf{Y}_n, \boldsymbol{\beta}_n)$. By subgaussianity of Rademacher processes, the $G_n^0(\cdot)$ inherit several properties of infinite dimensional Gaussian analysis. In particular, Hilbert spaces take a predominant role. This explains why the results in Varron (2014) hold under conditions upon $\| \boldsymbol{\beta}_n \|_2$ and $\| \boldsymbol{\beta}_n \|_4$. On the other hand, bracketing methods do not rely on subgaussianity, but on a form of Bernstein's inequality. The latter is a tradeoff between subgaussian and subexponential tails for sums of independent random variables that are uniformly bounded. This roughly explains why $\| \boldsymbol{\beta}_n \|_\infty$ - and its conjugate norm $\| \boldsymbol{\beta}_n \|_1$ - needs to be controlled. Such a difference of extent between bracketing and uniform entropy was not visible on the empirical process for the following simple reason: when taking $\beta_{i,n} \equiv n^{-1/2}$ for $i \leq n$ and $\beta_{i,n} \equiv 0$ otherwise, one has $\| \boldsymbol{\beta}_n \|_\infty \equiv \| \boldsymbol{\beta}_n \|_1^{-1} = n^{-1/2}$. This equality

makes the counterbalance between those two norms hardly visible in the proof of the bracketing Donsker theorem (see, e.g. van der Vaart and Wellner, 1996, p.130, Theorem 2.5.6).
Various interesting classes admit a finite bracketing entropy - see, e.g., van der Vaart and Wellner (1996, Chapter 2.7). In addition, several examples of posterior distributions in (discrete) Bayesian nonparametrics have the form $\mathbf{P}_{\mathbf{Y}_n,\boldsymbol{\beta}_n}$, or at least exhibit a predominant term that can be expressed as such - see Varron (2014, Section 3). Hence our main results may lead to applications of interest, which we will here illustrate through two examples. Our first example is an application in Bayesian nonparametrics : a posterior consistency and a BvM theorem that establish (1.1) and (1.2). Along the proofs we also revisit the Durst–Dudley–Borisov theorem and we obtain new insights about this phenomenon (see Lemma 5.5 in §5.4). To the best of our knowledge, neither does any result in such a vein exists for the topology of discrete total variation, nor do those results fall within the framework of already existing results in functional spaces - see Barrientos and Peña (2020); James (2008); Lo (1983, 1986). Our second example is a Donsker theorem - under a bracketing condition - for a specific form of local empirical measures (see §4.2). The remainder of this article is organized as follows: in §2 we give a careful description of the mathematical framework. Then our two main results are stated in §3. Corollaries and applications follow in §4. The proofs of those results are then written in §5. Finally, the Appendix is dedicated to a minor proof addressing measurability issues and a specific discussion upon the possible interactions of our results and the BvM results of Castillo and Nickl Castillo and Nickl (2014) for continuous priors defined by a random densities that are histograms built upon filtering of Dirichlet processes through dyadic partitions of an interval.

2. The mathematical framework

In order to properly state our main results, we first need to carefully define their underlying probabilistic framework.

2.1. *The underlying probability space.* Empirical processes carry over some lacks of measurability that are usually tackled by using outer expectations - see, e.g., van der Vaart and Wellner (1996, Chapter 1.2). Mathematical rigor imposes to define the underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$ as a suitable product space. First, for fixed $n \geq 1$, consider a Markov transition kernel from ℓ^1 to $\mathfrak{X}$, i.e., a family $\{\mathbf{P}_{n,\boldsymbol{p}},\ \boldsymbol{p} \in \ell^1\}$ of probability measures for which the maps $\boldsymbol{p} \to \mathbf{P}_{n,\boldsymbol{p}}(A),\ A \in \mathcal{A}_{\mathfrak{X}}$ are measurable from $\big(\ell^1, Bor(\ell^1)\big)$ to $\big([0,1], Bor([0,1])\big)$. Also consider a probability measure Q_n on $\big(\ell^1, Bor(\ell^1)\big)$ and define:

$$\tilde{\Omega} := \ell^1 \times \mathfrak{X}^{\mathbb{N}},\ \text{endowed with its product } \sigma\text{-algebra}$$

$$\tilde{\mathcal{A}} := Bor(\ell^1) \otimes \mathcal{A}_{\mathfrak{X}}^{\otimes \mathbb{N}},\ \text{with probability law defined through the generic formula:}$$

$$\mathbb{P}_n\Big(\big\{(\boldsymbol{p}, \mathbf{y}) \in \tilde{\Omega}, \boldsymbol{p} \in A, \forall j \in \{1, \ldots, k\}, y_i \in B_j\big\}\Big) := \int_{\boldsymbol{p} \in A} \prod_{j=1}^{k} \mathbf{P}_{n,\boldsymbol{p}}(B_j) dQ_n(\boldsymbol{p}).$$

Then define $\Omega := \tilde{\Omega}^{\mathbb{N}^*}, \mathcal{A} := \tilde{\mathcal{A}}^{\mathbb{N}^*}, \mathbb{P} := \bigotimes_{n\geq 1} \mathbb{P}_n$ on $\mathcal{A}$ and define the $\mathbf{Y}_n$ and $\boldsymbol{\beta}_n$ as coordinate maps on Ω :

$$\boldsymbol{\beta}_n(\boldsymbol{p}_1, \mathbf{y}_1, \boldsymbol{p}_2, \mathbf{y}_2, \ldots) := \boldsymbol{p}_n,\ \text{and } \mathbf{Y}_n(\boldsymbol{p}_1, \mathbf{y}_1, \boldsymbol{p}_2, \mathbf{y}_2, \ldots) := \mathbf{y}_n.$$

Note that, for fixed n and $\boldsymbol{p} \in \ell^1$, $\mathbf{P}_{n,\boldsymbol{p}}^{\otimes \mathbb{N}}$ is the law of $\mathbf{Y}_n$ given $\boldsymbol{\beta}_n = \boldsymbol{p}$. We shall denote by $\mathbf{P}_n$ the law of $Y_{1,n}$.

2.2. *Definition of G_n.* From now on, and throughout all this article, we shall make the assumption that $\mathbf{P}_n(F) < \infty$ for all $n \geq 1$. We also assume that $\mathcal{F}$ is *pointwise measurable* with countable separant $\mathcal{F}_0$ in the following sense: for any $f \in \mathcal{F}$, there exists $(f_m)_{m\geq 1} \in \mathcal{F}_0^{\mathbb{N}^*}$ such that $f_m(y) \to f(y)$ for each $y \in \mathfrak{X}$. Such a very standard assumption will be useful to tackle annoying measurability

issues.
Because the symbol $\Sigma_{i\in\mathbb{N}}$ in (1.7) is ambiguous, we need to give a rigorous definition of the processes that will be involved in this article. Our definition differs from that used in Varron (2014) for two reasons. The first (minor) one is to cover the case where the $|| \boldsymbol{\beta}_n ||_1$ are not deterministically equal to 1. The second one is for technical purposes: in our proofs, we shall truncate the $f \in \mathcal{F}$ from above using thresholds that depend upon the weights. First note that, for any bounded function f and any Borel map T from ℓ^1 to $\mathbb{R}^+$, the map

$$\Phi_{f,T}:\ (\mathbf{y},\boldsymbol{p}) \to \sum_{i\in\mathbb{N}} p_i\Big(f\mathbb{1}_{\{F\le T(\boldsymbol{p})\}}(y_i) - \mathbf{P}_{n,\boldsymbol{p}}\big(f\mathbb{1}_{\{F\le T(\boldsymbol{p})\}}(y_i)\big)\Big)$$

is properly defined (through the limits in $\mathbb{R}$ of partial sums) and Borel from $\mathfrak{X}^{\mathbb{N}}\times\ell^1$ to $\mathbb{R}$. One can hence define a random variable $G_n^T(f)$ by composition $G_n^T(f) := \Phi_{f,T}\circ(\mathbf{Y}_n,\boldsymbol{\beta}_n)$. We will say that a map ψ from $\mathcal{F}$ to $\mathbb{R}$ is $\mathcal{F}_0$-*separable* whenever we have $|| \psi ||_{\mathcal{F}} = || \psi ||_{\mathcal{F}_0}$. We shall also denote by $\mathcal{B}(\mathcal{F},\mathcal{F}_0)$ the space of all bounded $\mathcal{F}_0$-separable functions and by $\mathcal{A}_{||\cdot||_{\mathcal{F}}}$ is the σ-algebra spanned by the $||\cdot||_{\mathcal{F}}$-balls.

Lemma 2.1. *For any choice of T as above, the map*

$$\Phi_{\mathcal{F},T}:\ (\mathbf{y},\boldsymbol{p}) \to \big\{f \to \Phi_{f,T}(\mathbf{y},\boldsymbol{p})\big\}$$

is measurable from $\mathfrak{X}^{\mathbb{N}}\times\ell^1$ to $(\mathcal{B}(\mathcal{F},\mathcal{F}_0),\mathcal{A}_{||\cdot||_{\mathcal{F}}})$.

Proof: Fix T. Let us first prove that $\Phi_{\mathcal{F},T}$ takes its values in $\mathcal{B}(\mathcal{F},\mathcal{F}_0)$. Fix $\mathbf{y}\in\mathfrak{X}^{\otimes\mathbb{N}}$ and $\boldsymbol{p}\in\ell^1$. Since, for all $k\in\mathbb{N}$:

$$\sup_{f\in\mathcal{F}}\Big|\sum_{i\ge k+1} p_i\Big[f\mathbb{1}_{\{F\le T(\boldsymbol{p})\}}(y_i) - \mathbf{P}_{n,\boldsymbol{p}}\big(f\mathbb{1}_{\{F\le T(\boldsymbol{p})\}}\big)\Big]\Big| \le 2T(\boldsymbol{p})\sum_{i\ge k+1} |\ p_i\ |,$$

and since $\big(\mathcal{B}(\mathcal{F},\mathcal{F}_0), ||\cdot||_{\mathcal{F}}\ \big)$ is a Banach space, it is sufficient to prove that each trajectory $f\to f\mathbb{1}_{\{F\le T(\boldsymbol{p})\}}(y_i)$, $i\in\mathbb{N}$, and $f\to\mathbf{P}_{n,\boldsymbol{p}}\big(f\mathbb{1}_{\{F\le T(\boldsymbol{p})\}}\big)$ is $\mathcal{F}_0$-separable. To see this, take $f\in\mathcal{F}$ and consider $(f_m)_{m\ge1}\in\mathcal{F}_0^{\mathbb{N}}$ such that $f_m\to f_0$ pointwise. Thus $\mathbf{P}_{n,\boldsymbol{p}}(f_m\mathbb{1}_{\{F\le T(\boldsymbol{p})\}})\to\mathbf{P}_{n,\boldsymbol{p}}\big(f\mathbb{1}_{\{F\le T(\boldsymbol{p})\}}\big)$ by the dominated convergence theorem. Now since $\Phi_{\mathcal{F},T}$ takes its values in $\mathcal{B}(\mathcal{F},\mathcal{F}_0)$, the following equality holds for any $(\mathbf{y},\boldsymbol{p})\in\mathfrak{X}^{\mathbb{N}}\times\ell^1$:

$$\sup_{f\in\mathcal{F}} |\ \Phi_{f,T}(\mathbf{y},\boldsymbol{p})\ | = \sup_{f\in\mathcal{F}_0} |\ \Phi_{f,T}(\mathbf{y},\boldsymbol{p})\ |, \tag{2.1}$$

and then the measurability of each $\Phi_{f,T},\ f\in\mathcal{F}_0$ ensures that of $\Phi_{\mathcal{F}}^T$ with respect to $\mathcal{A}_{||\cdot||_{\mathcal{F}}}$. □
Now denote by $\tilde{\mathcal{E}}_{\mathcal{F},\mathcal{F}_0}$ the space of all measurable maps from $\big(\Omega,\mathcal{A}\big)$ to $\big(\mathcal{B}(\mathcal{F},\mathcal{F}_0),\mathcal{A}_{||\cdot||_{\mathcal{F}}}\big)$. The preceding lemma gives the opportunity to define the processes G_n on any class of functions

$$\mathcal{F}_M := \big\{f\mathbb{1}_{\{F\le M\}},\ f\in\mathcal{F}\big\}$$

since (confounding M with a constant function on ℓ^1) the composition map

$$G_n^M := \Phi_{\mathcal{F},M}\circ(\mathbf{Y}_n,\boldsymbol{\beta}_n)$$

belongs to $\tilde{\mathcal{E}}_{\mathcal{F},\mathcal{F}_0}$. Denote by $\mathcal{E}_{\mathcal{F},\mathcal{F}_0}$ the quotient space of $\tilde{\mathcal{E}}_{\mathcal{F},\mathcal{F}_0}$ with respect to the equivalence class

$$G\sim G' \Leftrightarrow || G-G' ||_{\mathcal{F}} = 0,\ \mathbb{P}\text{-a.s.},$$

and endow $\mathcal{E}_{\mathcal{F},\mathcal{F}_0}$ with the compatible distance

$$d(G,G') := \mathbb{E}\Big(\arctan\big(\ || G-G' ||_{\mathcal{F}}\ \big)\Big), \tag{2.2}$$

which is that of $||\cdot||_{\mathcal{F}}$-convergence in probability. The following lemma defines G_n as a suitable limit of the G_n^M when $M\to\infty$.

Lemma 2.2. *For fixed $n \geq 1$, the sequence $(G_n^M)_{M\geq 1}$ is Cauchy in the complete metric space $\big(\mathcal{E}_{\mathcal{F},\mathcal{F}_0}, d\big)$. It hence converges to a limit which we take as the definition of G_n. Moreover, for any sequence (T_k) of Borel thresholding maps fulfilling $T_k(\boldsymbol{\beta}_n) \to_{\mathbb{P}} \infty$ as $k \to \infty$, we have $d(G_n^{T_k}, G_n) \to 0$ as $k \to \infty$.*

Proof: For integers M, M' we have, writing $f^{M,M'} := f\mathbb{1}_{\{M<F\leq M'\}}$

$$\begin{aligned} & d\Big(G_n^M, G_n^{M'}\Big) \\ =& \mathbb{E}\Bigg(\arctan\Big(\sup_{f\in\mathcal{F}}\Big|\Phi_{f^{M,M'}}(\mathbf{Y}_n, \boldsymbol{\beta}_n)\Big|\Big)\Bigg) \\ \leq& \mathbb{E}\Bigg(\arctan\Bigg(\sum_{i\in\mathbb{N}} \mid \beta_{i,n} \mid \Big[F^{M,M'}(Y_{i,n}) + \mathbb{E}\Big(F^{M,M'}(Y_{i,n})\Big|\boldsymbol{\beta}_n\Big)\Big]\Bigg)\Bigg). \end{aligned} \tag{2.3}$$

Using Fatou's lemma for conditional expectations and the concavity of arctan on $\mathbb{R}^+$ we have, almost surely:

$$\mathbb{E}\Bigg(\arctan\Bigg(\sum_{i\in\mathbb{N}} \mid \beta_{i,n} \mid \Big[F^{M,M'}(Y_{i,n}) + \mathbb{E}\Big(F^{M,M'}(Y_{i,n})\Big|\boldsymbol{\beta}_n\Big)\Big]\Bigg)\Bigg|\boldsymbol{\beta}_n\Bigg)$$

$$\leq \arctan\Bigg(2\sum_{i\in\mathbb{N}} \mid \beta_{i,n} \mid \mathbb{E}\Big(F^{M,M'}(Y_{i,n}) \mid \boldsymbol{\beta}_n\Big)\Bigg)$$

$$= \arctan\Big(2 \mid\mid \boldsymbol{\beta}_n \mid\mid_1 \mathbb{E}\Big(F^{M,M'}(Y_{1,n}) \mid \boldsymbol{\beta}_n\Big)\Big) \tag{2.4}$$

$$\leq \arctan\Big(2 \mid\mid \boldsymbol{\beta}_n \mid\mid_1 \mathbb{E}\Big(F\mathbb{1}_{\{F>M\}}(Y_{1,n}) \mid \boldsymbol{\beta}_n\Big)\Big), \tag{2.5}$$

where (2.4) comes from the fact that the law $\mathbf{Y}_n$ given $\boldsymbol{\beta}_n = \boldsymbol{p}$ is $\mathbf{P}_{n,\boldsymbol{p}}^{\otimes\mathbb{N}}$. It hence suffices to prove that the right hand side (RHS) of (2.5) tends to 0 in probability as $M \to \infty$. This is true since $\mathbf{P}_n(F) < \infty$. Now to prove the last statement of Lemma 2.2, formally replace M by $T_k(\boldsymbol{p})$ in the preceding calculus and let $M' \to \infty$ to obtain, using Fatou's lemma for conditional expectations:

$$d\big(G_n^{T_k}, G_n\big) \leq \mathbb{E}\Bigg(\arctan\Big(2 \mid\mid \boldsymbol{\beta}_n \mid\mid_1 \mathbb{E}\Big(F\mathbb{1}_{\{F>T_k(\boldsymbol{\beta}_n)\}}(Y_{1,n}) \mid \boldsymbol{\beta}_n\Big)\Big)\Bigg), \tag{2.6}$$

which tends to 0 as $k \to \infty$, by assumption upon $(T_k(\boldsymbol{\beta}_n))_{k\geq 1}$ and since $\mathbf{P}_n(F) < \infty$. □

3. Results

Before stating our two main results, let us briefly mention that the maps $\boldsymbol{p} \to N_{[]}\big(\epsilon, \mathcal{F}, ||\cdot||_{\mathbf{P}_{n,\boldsymbol{p}},r}\big)$ and $\boldsymbol{p} \to J_{[]}\big(\delta, \mathcal{F}, \mathbf{P}_{n,\boldsymbol{p}}\big)$ are properly measurable for fixed ϵ and δ. This is proved in §6.1. One of our results is a Donsker theorem (see Theorem 2 below). The limit processes are mixtures of $\mathcal{F}$-indexed Brownian bridges, for which a rigorous definition is not immediate due to the non separability of $\ell^\infty(\mathcal{F})$. We shall use the definition of Varron (2014).

3.1. *A Glivenko–Cantelli theorem.* Our first result is a Glivenko–Cantelli theorem. Recall that $\mathbf{P}_n$ is the law of $Y_{1,n}$. We shall denote by $\ell^{1,+} := \ell^1 \cap [0,\infty[^{\mathbb{N}}$ the set of non negative summable sequences.

Theorem 1. *Assume that*

$$\lim_{M\to\infty} \overline{\lim_{n\to\infty}}\, \mathbf{P}_n\Big(F\mathbb{1}_{\{F\geq M\}}\Big) = 0, \tag{3.1}$$

and that, for any $\epsilon > 0$*:*

$$\Big(N_{[]}\big(\epsilon, \mathcal{F}, ||\cdot||_{\mathbf{P}_{n,\boldsymbol{\beta}_n},1}\big) \Big)_{n\geq 1} \text{ is bounded in probability.} \tag{3.2}$$

Also assume that $\boldsymbol{\beta}_n \in \ell^{1,+}$ *almost surely for all* n*, and that*

$$\big(|| \boldsymbol{\beta}_n ||_1 \big)_{n\geq 1} \text{ is bounded in probability.} \tag{3.3}$$

Then, under the condition $|| \boldsymbol{\beta}_n ||_2 \to_{\mathbb{P}} 0$ *we have* $|| G_n ||_{\mathcal{F}} \to_{\mathbb{P}} 0$.

Remark: It is important to compare the assumptions of Theorem 1 to those of Dudley's bracketing Glivenko–Cantelli theorem van der Vaart and Wellner (1996, p. 122, Theorem 2.4.1) for the empirical measure i.e. when $\beta_{i,n}^{Emp} :\equiv n^{-1}$ for $i \leq n$ and is identically null otherwise and when $\mathbf{P}_n = \mathbf{P}_0$ is constant in n. A few easy arguments then show that, in this special case, those two theorems exactly coincide: first, for $\boldsymbol{\beta}_n = \boldsymbol{\beta}_n^{Emp}$ the convergence in probability is equivalent to an almost sure convergence by Pollard's reverse martingale argument (see, e.g., van der Vaart and Wellner, 1996, p. 124, Lemma 2.4.5). Second, (3.1)+(3.2) is here equivalent to the finiteness of $N_{[]}\big(\epsilon, \mathcal{F}, ||\cdot||_{\mathbf{P}_0,1}\big)$ for each $\epsilon > 0$.

3.2. *A Donsker theorem.* For a sequence Z_n of maps from Ω to $\mathbb{R}$ we shall write

$$\overline{\lim_{n\to\infty}}^{\mathbb{P}^*} Z_n := \inf\Big\{ M \in \mathbb{R},\ \lim_{n\to\infty} \mathbb{P}^*\big(Z_n \geq M\big) = 0 \Big\},$$

with the convention $\inf_{\emptyset} = +\infty$, and we shall simply write $\overline{\lim}^{\mathbb{P}}_{n\to\infty} Z_n$ when the maps Z_n are measurable. Our second result is a Donsker theorem. Recall that the processes W_n were informally defined after (1.7) and that their rigorous definition can be found in Varron (2014).

Theorem 2. *Assume that*

$$|| \boldsymbol{\beta}_n ||_2 \to_{\mathbb{P}} 1, \tag{3.4}$$

$$|| \boldsymbol{\beta}_n ||_\infty \to_{\mathbb{P}} 0, \tag{3.5}$$

and that, for some $p \in [2, \infty[$

$$|| \boldsymbol{\beta}_n ||_1 \times || \boldsymbol{\beta}_n ||_\infty^{p-1} \text{ is bounded in probability,} \tag{3.6}$$

$$\lim_{\delta\to 0} \overline{\lim_{n\to\infty}}^{\mathbb{P}} J_{[]}(\delta, \mathcal{F}, \mathbf{P}_{n,\boldsymbol{\beta}_n}) = 0, \tag{3.7}$$

$$\lim_{M\to\infty} \overline{\lim_{n\to\infty}} \mathbf{P}_n\big(F^p \mathbb{1}_{\{F>M\}}\big) = 0. \tag{3.8}$$

Also assume that there exists a semimetric ρ *that makes* $\mathcal{F}$ *totally bounded, and fulfilling*

$$\lim_{\delta\to 0} \overline{\lim_{n\to\infty}}^{\mathbb{P}^*} \sup_{\substack{(f_1-f_2)\in\mathcal{F}^2,\\ \rho(f_1,f_2)<\delta}} \mathbf{P}_{n,\boldsymbol{\beta}_n}\Big((f_1 - f_2)^2\Big) = 0. \tag{3.9}$$

Then

$$d_{BL}\Big(G_n, W_n\Big) := \sup_{B\in BL1} \Big| \mathbb{E}^*\Big(B\big(G_n\big)\Big) - \mathbb{E}^*\Big(B\big(W_n\big)\Big)\Big| \to 0, \tag{3.10}$$

where $BL1$ *is the set of all 1-Lipschitz functions on* $\big(\ell^\infty(\mathcal{F}), ||\cdot||_{\mathcal{F}}\big)$ *that are bounded by 1. Moreover, if* $\mathcal{F}$ *is uniformly bounded, then* (3.10) *holds without assuming* (3.6) *nor* 3.8*.*

Remarks: We chose to state Theorem 2 under the most general assumptions that our methodology can afford. In order to give more substance to those conditions, it seems convenient to discuss on the place of Theorem 2 in the existing literature on Donsker theorems for empirical processes.

(1) When $\boldsymbol{\beta}_n$ is the vector of rescaled empirical weights ($\beta_{i,n} \equiv n^{-1/2}$ for $i \leq n$ and $\beta_{i,n} \equiv 0$ otherwise), and when $\mathbf{P}_n = \mathbf{P}_0$ is constant in n, then $\mathbf{P}_{\mathbf{Y}_n,\boldsymbol{\beta}_n}$ is a sequence of empirical processes. Noting that - for $p = 2$ - the $\boldsymbol{\beta}_n$ obviously satisfy conditions (3.4), (3.5) and (3.6) one can immediately conclude that - in this setup - Theorem 2 exactly coincides with Ossiander's bracketing Donsker theorem (Ossiander, 1987, Theorem 3.1). Andersen et al. (1988) did also prove a Donsker theorem under more general conditions, where the finiteness of $J_{[]}(\infty, \mathcal{F}, \mathbf{P}_0)$ is relaxed a to more abstract assumption, involving majorizing measures on $|| \cdot ||_{\mathbf{P}_0,2}$ balls and "weak $|| \cdot ||_{\mathbf{P}_0,2}$" brackets. This possible extension of Theorem 2 is beyond the scope of the present article and may deserve future investigations.

(2) Let us now relax the assumption that $\mathbf{P}_n$ is constant in n. In that case the G_n fall into the framework of triangular arrays of empirical processes with varying baseline measures, which were studied by Sheehy and Wellner (1992, Section 3). These authors did prove a Donsker result for G_n indexed by classes fulfilling $J(\infty, \mathcal{F}) < \infty$, under the envelope condition (3.8), and assuming that $\mathbf{P}_n$ converges to a limit $\mathbf{P}_0$ in the following sense - see their Corollary 3.1:

$$\sup_{(f_1,f_2)\in\mathcal{F}^2} \max\Big\{ \mid \mathbf{P}_n\big((f_1-f_2)^2\big) - \mathbf{P}_0\big((f_1-f_2)^2\big) \mid, \mid \mathbf{P}_n(f_1) - \mathbf{P}_0(f_1) \mid, \mid \mathbf{P}_n(f_1^2) - \mathbf{P}_0(f_1^2) \mid \Big\} \to 0. \tag{3.11}$$

It is then clear that our Theorem 2 puts forward an analogue of their result, replacing their assumption $J(\infty, \mathcal{F}) < \infty$ by the bracketing condition (3.7). To see this, just note that (3.9) is satisfied under (3.11), by choosing $\rho(f_1, f_2) :=|| f_1 - f_2 ||_{\mathbf{P}_0,2}$.

(3) Let us now discuss on assumption (3.7), which might be the most cumbersome to verify for applications. If $J_{[]}(\infty, \mathcal{F}, \mathbf{P}_0) < \infty$, a simple way to check (3.7) - by direct comparison of bracketing numbers - is to prove that

$$\sup_{(f_1,f_2)\in\mathcal{F}^2} \frac{\mathbf{P}_{n,\boldsymbol{\beta}_n}\big((f_1-f_2)^2\big)}{\mathbf{P}_0\big((f_1-f_2)^2\big)} \text{ is bounded in probability.} \tag{3.12}$$

Such a sufficient condition is quite restrictive and seems far from necessary, but its verification is sometimes very simple to perform. This is for example the case for our application to the local empirical process at fixed point - see §4.2.

(4) We conclude this series of remarks by pointing out that, whereas involving random weights, Theorem 2 has almost no connections with Donsker theorems for bootstrap empirical measures. For more details see Varron (2014, Remark 2.2).

4. Corollaries and applications

4.1. *Application in Bayesian nonparametrics: posterior analysis of the Dirichlet process prior under the discrete total variation.* Assume (in this subsection only) that $\mathfrak{X}$ is infinite countable. The class $\mathcal{F}$ of all indicator functions of subsets of $\mathfrak{X}$:

$$\mathcal{F} := \Big\{ \mathbb{1}_C,\ C \subset \mathfrak{X} \Big\} \tag{4.1}$$

is rich enough to define the discrete total variation between two measures on $\mathfrak{X}$, since

$$|| Q - Q' ||_{\mathcal{F}} = \sup_{C\subset\mathfrak{X}} \mid Q(C) - Q'(C) \mid =:|| Q - Q' ||_{Tot.var.} .$$

Clearly, $\mathcal{F}$ is too large to satisfy $J(\infty, \mathcal{F}) < \infty$. It was however shown in the celebrated Durst–Dudley–Borisov theorem that $\mathcal{F}$ may have a finite bracketing entropy $J_{[]}(\infty, \mathcal{F}, Q)$ under a simple necessary and sufficient criterion upon Q, namely

$$(DDB(Q)) \ : \qquad \sum_{y\in\mathfrak{X}} \sqrt{Q(\{y\})} < \infty.$$

Theorem 3 (Durst–Dudley–Borisov, 1981)**.** *For the class $\mathcal{F}$ defined in (4.1) we have*

$$J_{[]}(\infty, \mathcal{F}, Q) < \infty \quad \Leftrightarrow \quad \big(DDB(Q)\big).$$

We shall combine Theorem 1 with a refinement of Theorem 3 - see Lemma 5.5 - to prove both a posterior consistency and a BvM theorem for the Dirichlet process prior, under the *discrete* total variation. To properly state it, we need to introduce some more notations. From now on we shall denote by $DP(\alpha, M)$ a Dirichlet process with mean probability measure α on $\mathfrak{X}$ and concentration parameter $M > 0$. A possible representation of $DP(\alpha, M)$ is that of Sethuraman (1994):

$$DP(\alpha, M) =_{law} Pr_{\mathbf{Y},\boldsymbol{\beta}}, \tag{4.2}$$

where $\mathbf{Y} \rightsquigarrow \alpha^{\otimes\mathbb{N}}$, and $\beta_i := V_i\, \Pi_{j\leq i-1}(1 - V_j)$ with $(V_i)_{i\in\mathbb{N}} \rightsquigarrow Beta(1, M)^{\otimes\mathbb{N}}$ being independent of $\mathbf{Y}$. Now consider the nonparametric Bayesian model where the prior Pr has distribution $DP(\alpha, M)$ and where the sample $(X_1, \dots, X_n)$ has conditional law $\mathbf{P}^{\otimes n}$ given $Pr = \mathbf{P}$. In this model it is well known (see Ferguson, 1973) that a natural expression of the posterior distribution of Pr given $(X_1, \dots, X_n) = (x_1, \dots, x_n)$ is $Post_n(x_1, \dots, x_n) := DP(\alpha_{(x_1,\dots,x_n)}, M + n)$, where

$$\alpha_{(x_1,\dots,x_n)} :=\theta_n\alpha + (1 - \theta_n)\mathbf{P}_{(x_1,\dots,x_n)}, \text{ with } \theta_n := \frac{M}{M + n}, \text{ and } \mathbf{P}_{(x_1,\dots,x_n)} := \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}.$$

We shall take advantage of that explicit representation to prove the following two results.

Corollary 1 (Posterior consistency)**.** *Take $\mathcal{F}$ as in (4.1). Let $\mathbf{P}_0$ be a probability measure on countable $\mathfrak{X}$, and let $(X_n)_{n\geq1} \rightsquigarrow \mathbf{P}_0^{\mathbb{N}^*}$. Then for almost every sequence $(x_n)_{n\geq1}$ we have*

$$\Big|\Big|Post_n(x_1, \dots, x_n) - \mathbf{P}_{(x_1,\dots,x_n)}\Big|\Big|_{Tot.Var.} \to_{\mathcal{L}} 0.$$

Corollary 2 (Bernstein–von Mises)**.** *Assume in addition $(DDB(\mathbf{P}_0))$ and $(DDB(\alpha))$. Then for almost every sequence $(x_n)_{n\geq1}$ we have*

$$\sqrt{n}\Big(Post_n(x_1, \dots, x_n) - \mathbf{P}_{(x_1,\dots,x_n)}\Big) \to_{\mathcal{L}} \mathbb{G}_{\mathbf{P}_0}, \text{ in } \ell^\infty(\mathcal{F}),$$

where $\mathbb{G}_{\mathbf{P}_0}$ stands for the $\mathbf{P}_0$ Brownian bridge indexed by $\mathcal{F}$ - see, e.g., van der Vaart and Wellner (1996, p. 82). As a consequence, for almost every sequence $(x_n)_{n\geq1}$ we have

$$\sqrt{n}\Big|\Big|Post_n(x_1, \dots, x_n) - \mathbf{P}_{(x_1,\dots,x_n)}\Big|\Big|_{Tot.Var.} \to_{\mathcal{L}} ||\, \mathbb{G}_{\mathbf{P}_0}\, ||_{\mathcal{F}}\, .$$

The corresponding proofs are written in §5.4.

4.2. *Consequence of Theorem 2 to local empirical measures: a bracketing Donsker theorem.* Assume in this subsection that $\mathfrak{X} = \mathbb{R}^d$. The local empirical process indexed by functions, introduced in Einmahl and Mason (1997) has been intensively investigated during the last decades, due to its connections with several smoothing nonparametric methods. One of its particular forms can be written as follows:

$$T_{n,h_n}(f) := \frac{1}{\sqrt{nh_n^d}}\sum_{i=1}^{n}\Big[f\Big(h_n^{-1}(Z_i - z)\Big) - \mathbb{E}\Big(f\Big(h_n^{-1}(Z_i - z)\Big)\Big)\Big],$$

where $(h_n)_{n\geq1}$ is a deterministic non negative sequence tending to 0, and where $(Z_n)_{n\geq1}$ is an i.i.d. sequence. Implicit in the results of Einmahl and Mason (1997, Theorem 1.1) is the following Donsker theorem.

Theorem 4 (Einmahl and Mason, 1997)**.** *Let $(h_n)_{n\geq1}$ be a non random sequence of non negative numbers such that $h_n \to 0$ and $nh_n^d \to \infty$. Assume that $J(\infty, \mathcal{F}) < \infty$. Assume that the support S of F is bounded, and that Z_1 admits a version of Lebesgue density $\mathbf{f}$ on a neighborhood of z that is* *continuous at z and such that $\mathbf{f}(z) > 0$. Also assume that, taking $\mathbf{P}_0$ as the uniform distribution on S, we have $\mathbf{P}_0(F^2) < \infty$. Then we have the following weak convergence*

$$\frac{1}{\sqrt{\lambda(S)\mathbf{f}(z)}} T_{n,h_n}(\cdot) \to_{\mathcal{L}} \mathcal{W}_{\mathbf{P}_0}(\cdot), \text{ in } \ell^\infty(\mathcal{F}),$$

where $\mathcal{W}_{\mathbf{P}_0}(\cdot)$ denotes the $L^2(\mathbf{P}_0)$-isonormal Gaussian process indexed by $\mathcal{F}$ (or $\mathbf{P}_0$-Brownian motion).

Their proof heavily relies on a representation of their own (Einmahl and Mason, 1997, Proposition 3.1):

$$T_{n,h_n}(f) :=_{law} \frac{1}{\sqrt{nh_n^d}} \times \left[\sum_{i=1}^n b_{i,n}\big(f(Y_{i,n}) - \mathbf{P}_n(f)\big) \right] + R_n(f),$$

as processes indexed by $\mathcal{F}$, where:

- The $(b_{i,n})_{i\le n}$ are i.i.d Bernoulli with parameter $a_n := \mathbb{P}\big(h_n^{-1}(Y_1 - z) \in S\big)$;
- The $(Y_{i,n})_{i\le n}$ are i.i.d with law

$$\mathbf{P}_n := \mathbb{P}\big(Y_1 \in \cdot \mid h_n^{-1}(Y_1 - z) \in S\big), \tag{4.3}$$

  with $(b_{i,n})_{i\le n} \perp\!\!\!\perp (Y_{i,n})_{i\le n}$;
- The term

$$R_n(f) := \frac{\sum_{i=1}^n (b_{i,n} - a_n)}{\sqrt{nh_n^d}} \mathbf{P}_n(f) \tag{4.4}$$

  plays the asymptotic role of a correcting drift between the Brownian bridge $\mathbb{G}_{\mathbf{P}_0}$ and the Brownian motion $\mathcal{W}_{\mathbf{P}_0}$.

The following corollary of Theorem 2 is a Donsker theorem for T_{n,h_n} under the condition $J_{[]}(\infty, \mathcal{F}, \mathbf{P}_0) < \infty$.

Corollary 3. *Theorem 4 still holds if assumption $J(\infty, \mathcal{F}) < \infty$ is replaced by $J_{[]}(\infty, \mathcal{F}, \mathbf{P}_0) < \infty$.*

In a sense this result "completes the picture" since it is a counterpart of Einmahl and Mason (1997) for bracketing entropy. Its proof is written in §5.5.

5. Proofs

5.1. *Proof of Theorem 1.* The proof is divided in two lemmas.

Lemma 5.1. *Take $M > 0$. Under the assumptions of Theorem 1 we have $\| G_n^M \|_{\mathcal{F}} \to_{\mathbb{P}} 0$ as $n \to \infty$.*

Proof : Fix $M > 0$, $\epsilon > 0$, and choose - by (3.2) and (3.3) - an integer N for which, $\mathbb{P}(\boldsymbol{\beta}_n \in \mathbb{S}_n) > 1 - \epsilon$ for all $n \ge 1$, with

$$\mathbb{S}_n := \Big\{ \boldsymbol{p} \in \ell^{1,+},\ \| \boldsymbol{p} \|_1 \le N \text{ and } N_{[]}\big(\epsilon, \mathcal{F}, \|\cdot\|_{\mathbf{P}_{n,\boldsymbol{p}},1}\big) \le N \Big\}.$$

Now fix $n \ge 1$, and $\boldsymbol{p} \in \mathbb{S}_n$ and denote by $Br(n, \boldsymbol{p}) = \big\{ (f_j^-, f_j^+),\ j = 1, \dots, N \big\}$ a covering bracket of $\mathcal{F}$ with $\max_{j\le N} \| f_j^+ - f_j^- \|_{\mathbf{P}_{n,\boldsymbol{p}},1} \le \epsilon/N$. Using the same comparison argument as in van der Vaart and Wellner (1996, p. 122) we have, for $(f^-, f^+) \in Br(n, \boldsymbol{p})$, $f \in [\![f^-, f^+]\!]$ and $\mathbf{y} \in \mathfrak{X}^{\mathbb{N}}$ (recall

(1.4)):

$$\begin{aligned}&\Big[P_{\mathbf{y},\boldsymbol{p}}\Big(f^-\mathbb{1}_{\{F\le M\}}-\mathbf{P}_{n,\boldsymbol{p}}\big(f^-\mathbb{1}_{\{F\le M\}}\big)\Big)\Big]-N\mathbf{P}_{n,\boldsymbol{p}}(f^+-f^-)\\ \le& P_{\mathbf{y},\boldsymbol{p}}\Big(f\mathbb{1}_{\{F\le M\}}-\mathbf{P}_{n,\boldsymbol{p}}\big(f\mathbb{1}_{\{F\le M\}}\big)\Big)\\ \le&\Big[P_{\mathbf{y},\boldsymbol{p}}\Big(f^+\mathbb{1}_{\{F\le M\}}-\mathbf{P}_{n,\boldsymbol{p}}\big(f^+\mathbb{1}_{\{F\le M\}}\big)\Big)\Big]+N\mathbf{P}_{n,\boldsymbol{p}}(f^+-f^-),\end{aligned} \tag{5.1}$$

from where, writing $B_{n,\boldsymbol{p}}$ for the set of functions f that are a side of a bracket (hence $\sharp B_{n,\boldsymbol{p}}\le 2N$, where "$\sharp$" stands for "cardinal"):

$$\begin{aligned}&\sup_{f\in\mathcal{F}}\Big|P_{\mathbf{y},\boldsymbol{p}}\Big(f\mathbb{1}_{\{F\le M\}}-\mathbf{P}_{n,\boldsymbol{p}}\big(f\mathbb{1}_{\{F\le M\}}\big)\Big)\Big|\\ \le&\max_{f\in B_{n,\boldsymbol{p}}}\Big|P_{\mathbf{y},\boldsymbol{p}}\Big(f\mathbb{1}_{\{F\le M\}}-\mathbf{P}_{n,\boldsymbol{p}}\big(f\mathbb{1}_{\{F\le M\}}\big)\Big)\Big|+\epsilon.\end{aligned}$$

Note that the condition $\boldsymbol{p}\in\ell^{1,+}$ is crucial to obtain (5.1). Now formally replacing $\mathbf{y}$ by an i.i.d sequence $(Y_i)_{i\in\mathbb{N}}$ having distribution $\mathbf{P}_{n,\boldsymbol{p}}$ we obtain, for $\boldsymbol{p}\in\mathbb{S}_n$:

$$\mathbb{E}\left(\sup_{f\in\mathcal{F}}\Big|\sum_{i\in\mathbb{N}}p_i\big(f(Y_i)-\mathbf{P}_{n,\boldsymbol{p}}(f)\big)\Big|\right)\le\Delta_n(\boldsymbol{p})+\epsilon,\ \text{where} \tag{5.2}$$

$$\begin{aligned}\Delta_n^2(\boldsymbol{p}):=&\mathbb{E}^2\left(\sum_{f\in B_{n,\boldsymbol{p}}}\Big|\sum_{i\in\mathbb{N}}p_i\Big(f(Y_i)-\mathbf{P}_{n,\boldsymbol{p}}(f)\Big)\Big|\right)\\ \le&\mathbb{E}\left(\left(\sum_{f\in B_{n,\boldsymbol{p}}}\Big|\sum_{i\in\mathbb{N}}p_i\Big(f(Y_i)-\mathbf{P}_{n,\boldsymbol{p}}(f)\Big)\Big|\right)^2\right)\\ \le&(2N)^2\max_{f\in B_{n,\boldsymbol{p}}}\mathbb{E}\left(\Big|\sum_{i\in\mathbb{N}}p_i\Big(f(Y_i)-\mathbf{P}_{n,\boldsymbol{p}}(f)\Big)\Big|^2\right)\\ \le&(2N)^2\max_{f\in B_{n,\boldsymbol{p}}}\sum_{i\in\mathbb{N}}p_i^2\mathrm{Var}\Big(f(Y_1)\Big)\\ \le&(2NM\ ||\ \boldsymbol{p}\ ||_2)^2.\end{aligned} \tag{5.3}$$

Combining (5.2) and (5.3) yields, almost surely

$$\mathbb{E}\Big(\ ||\ G_n^M\ ||_{\mathcal{F}}\ \Big|\boldsymbol{\beta}_n\Big)\mathbb{1}_{\mathbb{S}_n}(\boldsymbol{\beta}_n)\le 2NM\ ||\ \boldsymbol{\beta}_n\ ||_2+\epsilon.$$

This concludes the proof, since $||\ \boldsymbol{\beta}_n\ ||_2\to_{\mathbb{P}}\ 0$ by assumption and since $\mathbb{P}\big(\boldsymbol{\beta}_n\notin\mathbb{S}_n\big)<\epsilon$. □
With Lemma 5.1 at hand, the proof of Theorem 1 will be concluded as follows:

Lemma 5.2. *We have*

$$\lim_{M\to\infty}\overline{\lim_{n\to\infty}}\ d(G_n^M,G_n)=0.$$

Proof: In view of (3.3) it is sufficient to show that

$$\forall\epsilon>0,\ \overline{\lim_{M\to\infty}}\ \overline{\lim_{n\to\infty}}\ \mathbb{P}\Big(\mathbb{E}\Big(F\mathbb{1}_{\{F>M\}}\big(Y_{1,n}\big)\Big|\boldsymbol{\beta}_n\Big)>\epsilon\Big)\le\epsilon.$$

This is immediate by (3.1) combined with Markov's inequality. □

5.2. *Proof of Theorem 2*. By (3.4) we can assume without loss of generality that $\|\boldsymbol{\beta}_n\|_2 \equiv 1$ for all n. First note that (3.7) immediately implies

$$\forall \delta > 0, \Big(N_{[]}\big(\delta, \mathcal{F}, \|\cdot\|_{\mathbf{P}_{n,\boldsymbol{\beta}_n},2}\big) \Big)_{n\geq 1} \text{ is bounded in probability.} \tag{5.4}$$

The proof of Theorem 2 follows the same directions as in that of Theorem 2 in Varron (2014). The only crucial point that changes is that of proving the following asymptotic equicontinuity condition

$$\lim_{\delta\to 0} \overline{\lim_{n\to\infty}}^{\mathbb{P}^*} \sup_{(f_1,f_2)\in\mathcal{F}^2,\ \rho(f_1,f_2)<\delta} | G_n(f_1) - G_n(f_2) | = 0, \tag{5.5}$$

which would be the only missing ingredient to complete the proof of Theorem 2. Proving (5.5) will be achieved by conditioning upon $\boldsymbol{\beta}_n$ and using the following chaining argument. It is an extension of usual chaining arguments for the bracketing entropy van der Vaart (1998, p. 286, Lemma 19.34) to unbalanced empirical measures. Due to the fact that infinitely many weights are involved, only uniformly bounded classes of functions are treated here for simplicity. This will be largely sufficient for our purposes.

Lemma 5.3. *Let $\boldsymbol{p} \in \ell^{1,+}$ such that $\|\boldsymbol{p}\|_2 = 1$ let Q be a probability measure and let $\mathcal{G}$ be a uniformly bounded pointwise measurable class of functions with countable separant $\mathcal{G}_0$. Let $\delta \in (0, \infty]$ be such that*

$$\sup_{g\in\mathcal{G}} \| g \|_{Q,2} \leq \delta \ \textit{and} \tag{5.6}$$

$$\sup_{g\in\mathcal{G},\ y\in\mathfrak{X}} | g(y) | \leq \|\boldsymbol{p}\|_\infty^{-1} \mathfrak{a}(\delta, Q),\ \textit{where} \tag{5.7}$$

$$\mathfrak{a}(\delta, Q) := \delta \Big/ \sqrt{\log N_{[]}\big(\delta, \mathcal{G}, \|\cdot\|_{Q,2}\big)}.$$

Then for any i.i.d sequence $(Y_i)_{i\in\mathbb{N}}$ with distribution Q, we have

$$\mathbb{E}\Big(\sup_{g\in\mathcal{G}} \Big| \sum_{i\in\mathbb{N}} p_i \Big(g(Y_i) - Q(g) \Big) \Big| \Big) \leq \mathfrak{C}_1 J_{[]}\big(\delta, \mathcal{G}, Q\big), \tag{5.8}$$

where $\mathfrak{C}_1$ is a universal constant.

Proof: We shall use the notations

$$\Delta^2(\mathcal{G}, Q) := \sup_{g\in\mathcal{G}} Q(g^2), \text{ and } \Gamma(\mathcal{G}) := \sup_{g\in\mathcal{G},\ y\in\mathfrak{X}} | g(y) | .$$

Given a finite class of functions $\tilde{\mathcal{G}}$ and given $m \geq 1$ we have, by combining Lemmas 2.2.9 and 2.2.10 in van der Vaart and Wellner (1996, p. 102):

$$\begin{aligned}
&\mathbb{E}\Big(\max_{g\in\tilde{\mathcal{G}}} \Big| \sum_{i=0}^m p_i \Big(g(Y_i) - Q(g) \Big) \Big| \Big) \\
\leq & 24 \Big[\sqrt{\sum_{i=0}^m p_i^2 \Delta^2(\tilde{\mathcal{G}}, Q) \log(1 + \sharp\tilde{\mathcal{G}})} + \max_{i\leq m} | p_i | \Gamma(\tilde{\mathcal{G}}) \log(1 + \sharp\tilde{\mathcal{G}}) \Big], \\
\leq & 24 \Big[\|\boldsymbol{p}\|_2 \Delta(\tilde{\mathcal{G}}, Q) \sqrt{\log(1 + \sharp\tilde{\mathcal{G}})} + \|\boldsymbol{p}\|_\infty \Gamma(\tilde{\mathcal{G}}) \log(1 + \sharp\tilde{\mathcal{G}}) \Big],
\end{aligned} \tag{5.9}$$

where the possible choice of factor 24 was actually shown in van der Vaart (1998, p. 285, Lemma 19.33). Since (5.9) does not depend upon m we then have, as soon as $\Gamma(\tilde{\mathcal{G}}) < \infty$ (and recalling that

$\| \boldsymbol{p} \|_2 = 1$):

$$\mathbb{E}\Big(\max_{g\in\tilde{\mathcal{G}}}\Big|\sum_{i\in\mathbb{N}} p_i\Big(g(Y_i)-Q(g)\Big)\Big|\Big)$$

$$=\lim_{m\to\infty}\mathbb{E}\Big(\max_{g\in\tilde{\mathcal{G}}}\Big|\sum_{i=0}^{m} p_i\Big(g(Y_i)-Q(g)\Big)\Big|\Big) \tag{5.10}$$

$$\leq 24\Big[\Delta(\tilde{\mathcal{G}},Q)\sqrt{\log(1+\sharp\tilde{\mathcal{G}})}+\| \boldsymbol{p} \|_\infty \Gamma(\tilde{\mathcal{G}})\log(1+\sharp\tilde{\mathcal{G}})\Big], \tag{5.11}$$

where (5.10) is an application of the monotone convergence theorem, since all the involved random variables are bounded by $2\Gamma(\tilde{\mathcal{G}})$. Now with (5.11) at hand, the remainder of the proof is as follows: a careful look at all the arguments of the proof of van der Vaart (1998, p. 286, Lemma 19.34) - noting that their truncating argument is not needed here - shows that the latter are still true with the systematic formal change of $\sqrt{n}$ by $\| \boldsymbol{p} \|_\infty^{-1}$. □

We can now start our proof of (5.5). First fix $\epsilon > 0$. Using (3.7) and (3.9) there exist $\delta_1, \delta_2 > 0$ and n_0 such that for all $n \geq n_0$ we have $1-\epsilon \leq \mathbb{P}_*\big(\boldsymbol{\beta}_n \in \mathbb{S}'_n\big)$, where $\mathbb{S}'_n$ is the set of all $\boldsymbol{p} \in \ell^{1,+}$ satisfying the following conditions:

$$2\sqrt{2}\mathfrak{C}_1 J_{[]}\Big(\frac{\delta_1}{2},\mathcal{F},\mathbf{P}_{n,\boldsymbol{p}}\Big)\leq\epsilon \tag{5.12}$$

$$\sup_{f\in\mathcal{F}_{\delta_2}} \| f \|_{\mathbf{P}_{n,\boldsymbol{p}},2} < \delta_1, \text{ where}$$

$$\mathcal{F}_{\delta_2} := \Big\{ f_1 - f_2,\ (f_1,f_2)\in\mathcal{F}^2,\ \rho(f_1,f_2)<\delta_2\Big\},$$

and where $\mathfrak{C}_1$ denotes the universal constant in (5.8). Now fix $\boldsymbol{p}$, write

$$T(\boldsymbol{p}) := \| \boldsymbol{p} \|_\infty^{-1} \mathfrak{a}(\delta_1, \mathbf{P}_{n,\boldsymbol{p}})\mathbb{1}_{\{\|\boldsymbol{p}\|_\infty>0\}}, \text{ and define}$$

$$\mathcal{F}_{\boldsymbol{p},\delta_1} := \Big\{ (f_1-f_2)\mathbb{1}_{\{F\leq T(\boldsymbol{p})\}},\ (f_1,f_2)\in\mathcal{F}^2,\ \| f_1-f_2 \|_{\mathbf{P}_{n,\boldsymbol{p}},2} < \delta_1 \Big\}.$$

Next, apply Lemma 5.3 for fixed $\boldsymbol{p} \in \mathbb{S}'_n$ to obtain (noticing that $\mathcal{F}_{\boldsymbol{p},\delta_1}$ satisfies (5.6) and (5.7) for the choice of $Q := \mathbf{P}_{n,\boldsymbol{p}}$ and $\delta := \delta_1$)

$$\mathbb{E}\Big(\sup_{f\in\mathcal{F}_{\boldsymbol{p},\delta_1}}\Big|\sum_{i\in\mathbb{N}} p_i\Big(f(Y_i)-\mathbf{P}_{n,\boldsymbol{p}}(f)\Big)\Big|\Big) \leq \mathfrak{C}_1 J_{[]}\Big(\delta_1,\mathcal{F}_{\boldsymbol{p},\delta_1},\mathbf{P}_{n,\boldsymbol{p}}\Big)$$

$$\leq 2\sqrt{2}\mathfrak{C}_1 J_{[]}\Big(\frac{\delta_1}{2},\mathcal{F},\mathbf{P}_{n,\boldsymbol{p}}\Big), \tag{5.13}$$

where the Y_i are i.i.d with law $\mathbf{P}_{n,\boldsymbol{p}}$ and where (5.13) is a consequence of (5.12) and standard comparisons of entropy numbers. Now since the latter inequality is valid for all $\boldsymbol{p} \in \mathbb{S}'_n$ we have

$$\mathbb{E}\Big(\| G_n^T \|_{\mathcal{F}_{\boldsymbol{\beta}_n,\delta_1}} \Big| \boldsymbol{\beta}_n\Big)\mathbb{1}_{\mathbb{S}'_n}\big(\boldsymbol{\beta}_n\big) \leq \epsilon\mathbb{1}_{\mathbb{S}'_n}\big(\boldsymbol{\beta}_n\big), \text{ almost surely.} \tag{5.14}$$

Note that the measurability $\| G_n^T \|_{\mathcal{F}_{\boldsymbol{\beta}_n,\delta_1}}$ is not immediate at all, but can be proved using the same arguments as in Varron (2014, proof of Proposition 4.2). In view of (5.14), and since $\mathcal{F}_{\delta_2} \subset \mathcal{F}_{\boldsymbol{p},\delta_1}$ for $\boldsymbol{p} \in \mathbb{S}'_n$, the proof of (5.5) will be completed if we prove the following lemma.

Lemma 5.4. *We have $d(G_n^T, G_n) \to 0$ as $n \to \infty$.*

Proof: From (2.6) we have (noting that $\| \beta_n \|_2 \equiv 1$ implies $\| \beta_n \|_\infty > 0$ a.s.)

$$\begin{aligned} & d\big(G_n^T, G_n\big) \\ \leq & \mathbb{E}\Bigg(\arctan\bigg(2 \, \| \boldsymbol{\beta}_n \|_1 \, \mathbb{E}\Big(F \mathbb{1}_{\{F > T(\boldsymbol{\beta}_n)\}}(Y_{1,n}) \mid \boldsymbol{\beta}_n\Big)\bigg)\Bigg) \qquad (5.15) \\ = & \mathbb{E}\Bigg(\arctan\bigg(2 \, \| \boldsymbol{\beta}_n \|_1 \, \mathbb{E}\Big(F \mathbb{1}_{\{F > T(\boldsymbol{\beta}_n)\}}(Y_{1,n}) \mid \boldsymbol{\beta}_n\Big)\bigg)\Bigg) \\ \leq & \mathbb{E}\Bigg(\arctan\bigg(2 \frac{\| \boldsymbol{\beta}_n \|_1}{T(\boldsymbol{\beta}_n)^{p-1}} \mathbb{E}\Big(F^p \mathbb{1}_{\{F > T(\boldsymbol{\beta}_n)\}}(Y_{1,n}) \mid \boldsymbol{\beta}_n\Big)\bigg)\Bigg) \\ = & \mathbb{E}\Bigg(\arctan\bigg(2 \frac{\| \boldsymbol{\beta}_n \|_1 \times \| \boldsymbol{\beta}_n \|_\infty^{p-1}}{\mathfrak{a}(\delta_1, \mathbf{P}_{n,\boldsymbol{\beta}_n})^{p-1}} \mathbb{E}\Big(F^p \mathbb{1}_{\{F > T(\boldsymbol{\beta}_n)\}}(Y_{1,n}) \mid \boldsymbol{\beta}_n\Big)\bigg)\Bigg). \end{aligned}$$

Since, by (3.6) and (5.4), the sequence $\| \boldsymbol{\beta}_n \|_1 \times \big(\| \boldsymbol{\beta}_n \|_\infty / \mathfrak{a}(\delta_1, \mathbf{P}_{n,\boldsymbol{\beta}_n})\big)^{p-1}$ is bounded in probability, it only remains to prove that

$$E_n := \mathbb{E}\Big(F^p \mathbb{1}_{\{F > T(\boldsymbol{\beta}_n)\}}(Y_{1,n}) \mid \boldsymbol{\beta}_n\Big) \to_{\mathbb{P}} 0. \qquad (5.16)$$

To prove this, fix $\epsilon > 0$ and choose M large enough so that

$$\mathbb{E}\Big(F^p \mathbb{1}_{\{F > M\}}(Y_{1,n})\Big) \leq \epsilon^2,$$

for all $n \geq 1$, which is possible by (3.8). Next apply Markov's inequality to E_n on the set $\{T(\boldsymbol{\beta}_n) > M\}$ and then note that $\mathbb{P}\big(T(\boldsymbol{\beta}_n) \leq M\big) \to 0$ by (3.5) and (5.4). To conclude the proof, let us now consider the isolated case where (3.6) and (3.8) are removed from the set of assumptions of Theorem 2, but $\mathcal{F}$ is uniformly bounded, i.e., $F \leq M$ for some constant $M > 0$. Then a look at (5.15) immediately yields the claim, noticing that $T(\boldsymbol{\beta}_n) \to_{\mathbb{P}} \infty$. □

5.3. *Proof of Corollary 1.* With (4.2) in mind, let us define

$$A := \Big\{ (x_n)_{n \geq 1}, \ \Big\| \alpha_{(x_1,\ldots,x_n)} - \mathbf{P}_0 \Big\|_{\mathcal{F}} \to 0 \Big\}. \qquad (5.17)$$

The class of indicators of subsets of a countable set is universally Glivenko–Cantelli - see, e.g., Dudley (2014, p. 217, Remark 6.4.3). Therefore, since $\theta_n \to 0$, the triangle inequality entails $\mathbf{P}_0^{\mathbb{N}^*}(A) = 1$. Now take an arbitrary sequence $(x_n)_{n \geq 1} \in A$. We shall apply Theorem 1 to the sequence $Post_n(x_1, \ldots, x_n)$. In this setup we have $\mathbf{P}_{n,\boldsymbol{p}} = \mathbf{P}_n = \alpha_{(x_1,\ldots,x_n)}$ for all $\boldsymbol{p} \in \ell^1$, and

$$\beta_{i,n} := V_{i,n} \prod_{j=0}^{i-1} (1 - V_{j,n}), \ i \in \mathbb{N}, \ n \geq 1,$$

with $(V_{i,n})_{i \in \mathbb{N}} \rightsquigarrow Beta(1, M+n)^{\otimes \mathbb{N}}$. To prove (3.2) let us first remark that if $[\![f^-, f^+]\!]$ is a bracket between two indicator functions fulfilling $\| f^+ - f^- \|_{\mathbf{P}_{(x_1,\ldots,x_n)},1} \leq \epsilon$ then $\| f^+ - f^- \|_{\alpha_{(x_1,\ldots,x_n)},1} \leq \theta_n + (1-\theta_n)\epsilon$. Moreover, since the pointwise supremum/infimum of a set of indicator functions is itself an indicator function, any covering of $\mathcal{F}$ by brackets can be converted into another covering with the same number of brackets, each of one between two indicator functions. Hence, since $\theta_n \to 0$, we conclude that it is sufficient to prove that $N_{[]}(\epsilon, \mathcal{F}, \mathbf{P}_{(x_1,\ldots,x_n)})$ is a bounded sequence for fixed $\epsilon > 0$. This is done as follows: let us first choose a finite set $C_0 \subset \mathfrak{X}$ such that $\mathbf{P}_0(C_0) > 1 - \epsilon$. Then by definition of A one has $\mathbf{P}_{(x_1,\ldots,x_n)}(C_0) > 1 - \epsilon$ for all large enough n.

$$\forall C \subset \mathfrak{X}, C \cap C_0 \subset C \subset \big(C \cap C_0\big) \cup C_0^c. \qquad (5.18)$$

Hence the finite collection

$$\Big\{ [\![\mathbb{1}_C, \mathbb{1}_{C\cup C_0^c}]\!],\ C \subset C_0 \Big\}$$

defines a covering of $2^{\sharp C_0}$ brackets having $||\cdot||_{\mathbf{P}_{(x_1,\ldots,x_n)},1}$ diameters less than ϵ. This proves that $N_{[]}(\epsilon, \mathcal{F}, \mathbf{P}_{(x_1,\ldots,x_n)}) \leq 2^{\sharp C_0}$ for all large n, and hence proves (3.2). Now conditions (3.1) and (3.3) are immediate since $\mathcal{F}$ is uniformly bounded and $||\,\boldsymbol{\beta}_n\,||_1 \equiv 1$ - see, e.g. Hjort et al. (2010, p. 112). Finally, standard calculus on beta distributions shows that $\mathbb{E}\big(\ ||\,\boldsymbol{\beta}_n\,||_2^2\ \big) \sim n^{-1}$, from where one can apply Theorem 1 and conclude the proof.

5.4. *Proof of Corollary 2.* We shall now assume without loss of generality that the support of $\mathbf{P}_0$ is infinite.

5.4.1. *Two preliminary results.* Theorem 3 states that the finiteness of $\Sigma_{y\in\mathfrak{X}}\sqrt{\mathbf{P}_0(\{y\})}$ is equivalent to that of $J_{[]}(\infty, \mathcal{F}, \mathbf{P}_0)$. Our next lemma goes one step further: it shows that it is possible to control the magnitude of $J_{[]}(\delta, \mathcal{F}, \mathbf{P}_0)$, for small $\delta > 0$, by "tail" sums of the $\sqrt{\mathbf{P}_0(\{y\})}$.

Lemma 5.5. *Define, for* $k \in \mathbb{N}$:

$$\mathbf{j}_{\mathbf{P}_0}(k) := \min\Big\{ J \in \mathbb{N},\ \sum_{y\in\mathfrak{X}\,:\,\mathbf{P}_0(\{y\})\leq 16^{-J}} \mathbf{P}_0(\{y\}) \leq 4^{-k} \Big\}. \tag{5.19}$$

Then, for all $p \geq 1$ *we have, for a universal constant* $\mathfrak{C}_2$

$$J_{[]}\big(2^{-(p-1)}, \mathcal{F}, \mathbf{P}_0\big) \leq \mathfrak{C}_2 \sqrt{\sum_{y\in\mathfrak{X}} \sqrt{\mathbf{P}_0(\{y\})}} \times \sqrt{\sum_{y:\ \mathbf{P}_0(\{y\})\leq 16^{-\mathbf{j}_{\mathbf{P}_0}(p)+1}} \sqrt{\mathbf{P}_0(\{y\})}}.$$

Moreover if the support of $\mathbf{P}_0$ *is infinite we have* $\mathbf{j}_{\mathbf{P}_0}(p) \to \infty$ *as* $p \to \infty$.

Proof : The very last statement is obvious. We shall now write $\mathbf{j}(\cdot)$ instead of $\mathbf{j}_{\mathbf{P}_0}(\cdot)$ for concision. The proof consists in enriching the arguments of Dudley (2014, p. 245-246) with additional analytical precisions. We shall hence borrow his notations. First, for $j \in \mathbb{N}$ write

$$A_j := \Big\{ y \in \mathfrak{X},\ 16^{-j-1} < \mathbf{P}_0(\{y\}) \leq 16^{-j} \Big\},\ \text{and}\ r_j := \sharp A_j.$$

Now define the following maps on $\mathbb{N}$

$$m(\cdot):\ k \to \sum_{j=0}^{\mathbf{j}(k)} r_j = \sharp \bigcup_{j=0}^{\mathbf{j}(k)} A_j,$$

$$k(\cdot):\ J \to \min\Big\{ p \geq 1,\ 4^{-p} < \sum_{y:\ \mathbf{P}_0(\{y\})\leq 16^{-J}} \mathbf{P}_0(\{y\}) \Big\},$$

$$\kappa(\cdot):\ k \to \min\Big\{ \kappa \in \mathbb{N},\ \mathbf{j}(\kappa) = \mathbf{j}(k) \Big\}.$$

For consistency of notations in the following calculus, we shall also define $k(-1) := 0$. Note that, writing $\mathcal{K}$ for the range of $\kappa(\cdot)$, the map $\mathbf{j}(\cdot)$ is one to one on $\mathcal{K}$.

Fix $k \geq 1$. Similarly as in Dudley (2014, p. 245-246) we see that one can use the same arguments as for (5.18), with the formal replacement of ϵ by 4^{-k} and C_0 by

$$C_k := \bigcup_{j=0}^{\mathbf{j}(k)} A_j,$$

which satisfies $\mathbf{P}_0(C_k) \geq 1 - 4^{-k}$ by (5.19). This implies

$$\forall k \geq 1,\ N_{[]}\big(2^{-k}, \mathcal{F}, ||\cdot||_{\mathbf{P}_0,2}\ \big) \leq 2^{m(k)}. \tag{5.20}$$

Now, for any $p \geq 1$, by monotonicity of the involved functions:

$$J_{[]}\big(2^{-(p-1)}, \mathcal{F}, \mathbf{P}_0\big) \leq \sum_{k \geq p} \sqrt{\log N_{[]}\big(2^{-k}, \mathcal{F}, ||\cdot||_{\mathbf{P}_0,2}\big)}\big(2^{-(k-1)} - 2^{-k}\big)$$
$$\leq \sqrt{\log(2)} \sum_{k \geq p} \frac{\sqrt{m(k)}}{2^k} \text{ by (5.20).}$$

Next, fix $p \geq 1$ and write

$$\sum_{k \geq p} \frac{\sqrt{m(k)}}{2^k} \leq \sum_{k \geq p} \sum_{j=0}^{\mathbf{j}(k)} \sqrt{r_j}\, 2^{-k}$$
$$\leq \sum_{k \geq p} \sum_{j=0}^{\mathbf{j}(k)} 2 \sqrt{\sum_{y \in A_j} \sqrt{\mathbf{P}_0(\{y\})}}\, 2^{j-k}, \text{ since } r_j \leq \sum_{y \in A_j} 4^{j+1} \sqrt{\mathbf{P}_0(\{y\})}$$
$$= 2 \sum_{j \geq 0} \sqrt{\sum_{y \in A_j} \sqrt{\mathbf{P}_0(\{y\})}} \sum_{\substack{k:\, k \geq p,\\ \mathbf{j}(k) \geq j}} 2^{j-k}$$
$$\leq 2 \sqrt{\sum_{j \geq 0} \sum_{y \in A_j} \sqrt{\mathbf{P}_0(\{y\})}} \times \sqrt{\sum_{j \geq 0} \Big(\sum_{\substack{k:\, k \geq p,\\ \mathbf{j}(k) \geq j}} 2^{j-k} \Big)^2}, \text{ using Cauchy–Schwartz}$$
$$= 2 \sqrt{\sum_{y \in \mathfrak{X}} \sqrt{\mathbf{P}_0(\{y\})}} \times \sqrt{\sum_{j \geq 0} \Big(\sum_{\substack{k \geq p,\\ \mathbf{j}(k) \geq j}} 2^{j-k} \Big)^2}.$$

Now we have

$$\sum_{j \geq 0} \Big(\sum_{\substack{k:\, k \geq p,\\ \mathbf{j}(k) \geq j}} 2^{j-k} \Big)^2$$
$$\leq 4 \sum_{j \geq 0} 4^{j-k(j-1)} \wedge 4^{j-p}, \text{ since } \mathbf{j}(k) \geq j \text{ implies } k \geq k(j-1)$$
$$= 4 \sum_{k \geq 0} \sum_{j:\, \mathbf{j}(k-1) \leq j-1 < \mathbf{j}(k)} 4^{j-k} \wedge 4^{j-p}, \text{ since } k(j) = k \text{ for } \mathbf{j}(k-1) \leq j < \mathbf{j}(k)$$
$$= 4 \sum_{k \leq p} \sum_{j:\, \mathbf{j}(k-1) \leq j-1 < \mathbf{j}(k)} 4^{j-p} + 4 \sum_{k \geq p+1} \sum_{j:\, \mathbf{j}(k-1) \leq j-1 < \mathbf{j}(k)} 4^{j-k}$$
$$\leq 8 \Big[4^{\mathbf{j}(p)-p} + \sum_{k \geq p+1} 4^{\mathbf{j}(k)-k} \Big]$$
$$= 8 \sum_{k \geq p} 4^{\mathbf{j}(k)-k}$$
$$\leq 8 \sum_{k \geq p} 4^{\mathbf{j}(k)-k+\kappa(k)} \times \sum_{\ell\, :\, \mathbf{P}_0(\{y\}) \leq 16^{-\mathbf{j}(k)+1}} \mathbf{P}_0(\{y\}), \text{ by (5.19) and since } \mathbf{j}(k) = \mathbf{j}(\kappa(k))$$
$$= 8 \sum_{k \geq p} 4^{\mathbf{j}(k)-k+\kappa(k)} \times \sum_{j \geq \mathbf{j}(k)-1} \sum_{\ell \in A_j} \mathbf{P}_0(\{y\})$$
$$\leq 8 \sum_{j \geq 0} \Big(\sum_{y \in A_j} \mathbf{P}_0(\{y\}) \Big) \Big(\sum_{k:\, k \geq p,\, \mathbf{j}(k)-1 \leq j} 4^{\mathbf{j}(k)-k+\kappa(k)} \Big).$$

Now notice that, when $\mathbf{j}(p) > j+1$ the set of indices $\{k \geq p,\ \mathbf{j}(k)-1 \leq j\}$ is empty, from where

$$
\begin{aligned}
&\sum_{j\geq 0}\Big(\sum_{y\in A_j}\mathbf{P}_0(\{y\})\Big)\Big(\sum_{k:\,k\geq p,\,\mathbf{j}(k)-1\leq j} 4^{\mathbf{j}(k)-k+\kappa(k)}\Big)\\
\leq &\sum_{j\geq \mathbf{j}(p)-1}\Big(\sum_{y\in A_j}\mathbf{P}_0(\{y\})\Big)\Big(\sum_{k:\,\mathbf{j}(k)-1\leq j} 4^{\mathbf{j}(k)+\kappa(k)-k}\Big)\\
\leq &\sum_{j\geq \mathbf{j}(p)-1}\Big(\sum_{y\in A_j}\mathbf{P}_0(\{y\})\Big)\Big(\sum_{k'\in\mathcal{K},\,\mathbf{j}(k')\leq j+1} 4^{\mathbf{j}(k')+k'}\sum_{k:\,\kappa(k)=k'}4^{-k}\Big)\\
\leq &4\sum_{j\geq \mathbf{j}(p)-1}\Big(\sum_{y\in A_j}\mathbf{P}_0(\{y\})\Big)\Big(\sum_{k'\in\mathcal{K},\,\mathbf{j}(k')\leq j+1} 4^{\mathbf{j}(k')}\Big),\text{ since } \kappa(k)=k' \text{ implies } k\geq k'\\
\leq &32\sum_{j\geq \mathbf{j}(p)-1}\sum_{y\in A_j}\mathbf{P}_0(\{y\})4^j,\ \text{ since } \kappa(\cdot) \text{ is one to one on } \mathcal{K}\\
\leq &32\sum_{j\geq \mathbf{j}(p)-1}\sum_{y\in A_j}\sqrt{\mathbf{P}_0(\{y\})},\ \text{ since } y\in A_j \text{ implies } \mathbf{P}_0(\{y\})4^j\leq\sqrt{\mathbf{P}_0(\{y\})}\\
= &32\sum_{y:\,\mathbf{P}_0(\{y\})\leq 16^{-\mathbf{j}(p)+1}}\sqrt{\mathbf{P}_0(\{y\})}.
\end{aligned}
$$

This concludes the proof.□

Our second preliminary result is as follows.

Lemma 5.6. *Write*

$$I_\epsilon := \big\{y\in\mathfrak{X},\ \mathbf{P}_0(\{y\})\leq\epsilon\big\}, \epsilon\in\mathbb{Q}^+.$$

Then for $\mathbf{P}_0^{\otimes\mathbb{N}^*}$*-almost any sequence* $(x_n)_{n\geq 1}$ *we have:*

$$\forall\epsilon\in\mathbb{Q}^+,\ \lim_{n\to\infty}\sum_{y\in I_\epsilon}\sqrt{\mathbf{P}_{(x_1,\ldots,x_n)}(\{y\})}=\sum_{y\in I_\epsilon}\sqrt{\mathbf{P}_0(\{y\})}. \tag{5.21}$$

Proof: Since the class $\mathcal{F}$ is $\mathbf{P}_0$-Donsker and admits a square integrable envelope ($F\equiv 1$), the conditional multiplier Donsker theorem applies for a suitable i.i.d. standard normal sequence $(\xi_n)_{n\geq 1}$ - see, e.g., van der Vaart and Wellner (1996, p. 183, Theorem 2.9.7). Hence for $\mathbf{P}_0^{\mathbb{N}^*}$-almost every sequence $(x_n)_{n\geq 1}$ we have - recalling that $\mathcal{W}_{\mathbf{P}_0}$ stands for the $L^2(\mathbf{P}_0)$-isonormal Gaussian process indexed by $\mathcal{F}$:

$$\Big(\mathcal{W}_{(x_1,\ldots,x_n)}(f)\Big)_{f\in\mathcal{F}}\to_{\mathcal{L}}\ \Big(\mathcal{W}_{\mathbf{P}_0}(f)\Big)_{f\in\mathcal{F}},\ \text{where} \tag{5.22}$$

$$\mathcal{W}_{(x_1,\ldots,x_n)}(f):=\frac{1}{\sqrt{n}}\sum_{i=1}^n\xi_i f(x_i),\ f\in\mathcal{F}.$$

Here the weak convergence holds in the sense of Hoffman-Jőrgensen holds taking the underlying probability space as the canonical product space for $(\xi_n)_{n\geq 1}$. Moreover the involved processes are Gaussian, hence weak convergence implies convergence of first moments of absolute suprema. As a consequence, for such a sequence $(x_n)_{n\geq 1}$ fulfilling (5.22) we have, for all $\epsilon\in\mathbb{Q}^+$

$$\mathbb{E}\left(2\sup_{A\subset I_\epsilon}\frac{1}{\sqrt{n}}\Big|\sum_{i=1}^n\xi_i\mathbb{1}_A(x_i)\Big|-\frac{1}{\sqrt{n}}\Big|\sum_{i=1}^n\xi_i\mathbb{1}_{I_\epsilon}(x_i)\Big|\right)\to\mathbb{E}\left(2\sup_{A\subset I_\epsilon}\Big|\mathcal{W}_{\mathbf{P}_0}(\mathbb{1}_A)\Big|-\Big|\mathcal{W}_{\mathbf{P}_0}(\mathbb{1}_{I_\epsilon})\Big|\right).$$

Finally, by the standard equality

$$\sup_{A\subset I_\epsilon}\Big|\sum_{y\in A}g(y)\Big|=\frac{1}{2}\Big(\sum_{y\in I_\epsilon}\mid g(y)\mid+\Big|\sum_{y\in I_\epsilon}g(y)\Big|\Big),$$

we have (with $g(y) := n^{-1/2} \sum_{i=1}^{n} \xi_i \mathbb{1}_{\{y\}}(x_i)$)

$$\begin{aligned}
&\mathbb{E}\left(2 \sup_{A\subset I_\epsilon} \frac{1}{\sqrt{n}}\Big|\sum_{i=1}^{n} \xi_i \mathbb{1}_A(x_i)\Big| - \frac{1}{\sqrt{n}}\Big|\sum_{i=1}^{n} \xi_i \mathbb{1}_{I_\epsilon}(x_i)\Big|\right)\\
=&\mathbb{E}\left(\sum_{y\in I_\epsilon}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i \mathbb{1}_{\{y\}}(x_i)\Big|\right)\\
=&\sum_{y\in I_\epsilon}\sqrt{\mathbf{P}_{(x_1,\dots,x_n)}(\{y\})},
\end{aligned}$$

and similarly (with now $g(y) := \mathcal{W}_{\mathbf{P}_0}(\mathbb{1}_{\{y\}})$)

$$\mathbb{E}\left(2 \sup_{A\subset I_\epsilon}\Big|\mathcal{W}_{\mathbf{P}_0}(\mathbb{1}_A)\Big| - \Big|\mathcal{W}_{\mathbf{P}_0}(\mathbb{1}_{I_\epsilon})\Big|\right) = \sum_{y\in I_\epsilon}\sqrt{\mathbf{P}_0(\{y\})}$$

which concludes the proof.□

5.4.2. *Use of Theorem 2*. Recall that A was defined in (5.17) and has probability one. Let us consider the set

$$B := \left\{(x_n)_{n\geq 1},\ \forall\epsilon\in\mathbb{Q}^+,\ \overline{\lim_{n\to\infty}} \sum_{y:\ \alpha_{(x_1,\dots,x_n)}(\{y\})\leq\epsilon}\sqrt{\alpha_{(x_1,\dots,x_n)}(\{y\})} \leq \sum_{y\in I_{2\epsilon}}\sqrt{\mathbf{P}_0(\{y\})}\right\}.$$

We have, for any $n\geq 1$ and $\epsilon\in\mathbb{Q}^+$ (using $\sqrt{a+b}\leq\sqrt{a}+\sqrt{b}$)

$$\begin{aligned}
&\sum_{y:\ \alpha_{(x_1,\dots,x_n)}(\{y\})\leq\epsilon}\sqrt{\alpha_{(x_1,\dots,x_n)}(\{y\})}\\
\leq&\sqrt{\theta_n}\sum_{y:\ \alpha_{(x_1,\dots,x_n)}(\{y\})\leq\epsilon}\sqrt{\alpha(\{y\})} + \sqrt{1-\theta_n}\sum_{y:\ \alpha_{(x_1,\dots,x_n)}(\{y\})\leq\epsilon}\sqrt{\mathbf{P}_{(x_1,\dots,x_n)}(\{y\})}.
\end{aligned}$$

Now if $(x_n)_{n\geq 1}$ belongs to A and since $\mathcal{F}$ induces the total variation distance we have, for all n large enough

$$\big\{y\in\mathfrak{X},\ \alpha_{(x_1,\dots,x_n)}(\{y\})\leq\epsilon\big\}\subset\big\{y\in\mathfrak{X},\ \mathbf{P}_0(\{y\})\leq 2\epsilon\big\} = I_{2\epsilon}.$$

We hence conclude that $\mathbf{P}_0^{\mathbb{N}^*}(A\cap B) = 1$ - recalling that $\theta_n\to 0$ and $(DDB(\alpha))$ holds. Let us now consider a sequence $(x_n)_{n\geq 1}\in A\cap B$. Similarly as in §5.3, we shall prove Corollary 2 by verifying all the assumptions of Theorem 2, for the choice of $\mathbf{P}_{n,\boldsymbol{p}} = \mathbf{P}_n := \mathbf{P}_{(x_1,\dots,x_n)}$. Because the class $\mathcal{F}$ is uniformly bounded by 1, the conditions upon

$$\beta_{i,n} := \sqrt{n}V_{i,n}\prod_{j=0}^{i-1}(1-V_{j,n}),\ i\in\mathbb{N},$$

that we need to check are (3.4) and (3.5), or equivalently

$$||\ \boldsymbol{\beta}_n\ ||_2\to_{\mathbb{P}}\ 1,\ \text{and}\ \ ||\ \boldsymbol{\beta}_n\ ||_4\to_{\mathbb{P}}\ 0.$$

These are respectively proved by direct computations of expectations and variances. It now remains to verify (3.7) and (3.9). By definition of A and since $\theta_n\to 0$, the sequence $\mathbf{P}_n$ obviously fulfills (3.11) and therefore satisfies (3.9). Now in view of Lemma 5.5, assertion (3.7) will be proved if we show that

$$\lim_{p\to\infty}\ \overline{\lim_{n\to\infty}} \sum_{y:\ \alpha_{(x_1,\dots,x_n)}(\{y\})\leq 16^{-\mathbf{j}_{\mathbf{P}_n}(p)+1}}\sqrt{\mathbf{P}_n(\{y\})} = 0. \tag{5.23}$$

Lemma 5.7. *Take $(x_n)_{n\geq 1} \in A$. For any $p \geq 1$ we have $\mathbf{j}_{\mathbf{P}_n}(p) \geq \mathbf{j}_{\mathbf{P}_0}(p)$ for all large enough n.*

Proof: Fix $p \geq 1$. By definition of $\mathbf{j}_{\mathbf{P}_0}$ we have

$$\sum_{y:\, \mathbf{P}_0(\{y\}) \leq 16^{-\mathbf{j}_{\mathbf{P}_0}(p)+1}} \mathbf{P}_0(\{y\}) > 4^{-p}.$$

Now since $(x_n)_{n\geq 1} \in A$, we have, for all n large enough:

$$\sum_{y:\, \mathbf{P}_0(\{y\}) \leq 16^{-\mathbf{j}_{\mathbf{P}_0}(p)+1}} \mathbf{P}_n(\{y\}) > 4^{-p},$$

whence $\mathbf{j}_{\mathbf{P}_0}(p) - 1 \leq \mathbf{j}_{\mathbf{P}_n}(p) - 1$ by definition of $\mathbf{j}_{\mathbf{P}_n}$. □
Now applying Lemma 5.7 we have, writing $\epsilon(p) := 16^{-\mathbf{j}_{\mathbf{P}_0}(p)+1}$:

$$\begin{aligned}
&\lim_{p\to\infty} \overline{\lim_{n\to\infty}} \sum_{y:\ \alpha_{(x_1,\dots,x_n)}(\{y\}) \leq 16^{-\mathbf{j}_{\mathbf{P}_n}(p)+1}} \sqrt{\mathbf{P}_n(\{y\})} \\
\leq &\lim_{p\to\infty} \overline{\lim_{n\to\infty}} \sum_{y:\ \alpha_{(x_1,\dots,x_n)}(\{y\}) \leq \epsilon(p)} \sqrt{\mathbf{P}_n(\{y\})} \\
\leq &\lim_{p\to\infty} \overline{\lim_{n\to\infty}} \sum_{y \in I_{2\epsilon(p)}} \sqrt{\mathbf{P}_0(\{y\})}, \text{ since } (x_n)_{n\geq 1} \in B \\
= &0,
\end{aligned}$$

by $(DDB(\mathbf{P}_0))$ together with $\lim \mathbf{j}_{\mathbf{P}_0}(p) \to \infty$. This proves (5.23) and we can now apply Theorem 2 to obtain

$$d_{BL}\left(\sqrt{n}\Big(Post_n(x_1,\dots,x_n) - \alpha_{(x_1,\dots,x_n)}\Big), \mathbb{G}_{\alpha_{(x_1,\dots,x_n)}}\right) \to 0.$$

But since $(x_n)_{n\geq 1} \in A$ the sequence $\mathbf{P}_n := \alpha_{(x_1,\dots,x_n)}$ satisfies (3.11) from where (see Varron, 2014, Remark 2.2):

$$\mathbb{G}_{\alpha_{(x_1,\dots,x_n)}} \to_{\mathcal{L}} \mathbb{G}_{\mathbf{P}_0}, \text{ in } \ell^\infty(\mathcal{F}),$$

which concludes the proof of Corollary 2. □

5.5. *Proof of Corollary 3.* Recall that $\mathbf{P}_0$ here denotes the uniform distribution on S and that $\mathbf{P}_n$ has been defined in (4.3). Let $\mathcal{V}$ be a neighborhood of z on which Z_1 admits the density $\mathbf{f}$. Since S is bounded and $h_n \to 0$ we have $z + h_n S \subset \mathcal{V}$ for n large enough. We may assume without loss of generality that this is the case for each $n \geq 1$.

Lemma 5.8. *We have (taking here the convention $0/0 = 0$)*

$$\left(\sum_{i=1}^n \frac{b_{i,n}}{\sqrt{\sum_{i=1}^n b_{i,n}^2}} \Big(f(Y_{i,n}) - \mathbf{P}_n(f)\Big)\right)_{f\in\mathcal{F}} \to_{\mathcal{L}} \mathbb{G}_{\mathbf{P}_0}.$$

Proof: Write

$$\beta_{i,n} := \frac{b_{i,n}}{\sqrt{\sum_{i=1}^n b_{i,n}^2}}, \text{ for } i = 1, \dots, n.$$

Since $\mathbf{f}$ is continuous at z we have

$$a_n \sim \lambda(S)\mathbf{f}(z)h_n^d, \text{ from where } na_n \to \infty \text{ and } a_n \to 0. \tag{5.24}$$

This property ensures that the sequence $\boldsymbol{\beta}_n$ satisfies (3.4), (3.5) and (3.6) of Theorem 2 - taking $p := 2$ and recalling that $b_{i,n} \equiv b_{i,n}^2$. In order to verify (3.7) and (5.4) we will now prove (3.12), noting here that $\mathbf{P}_{n,\boldsymbol{p}} := \mathbf{P}_n$ for all $\boldsymbol{p} \in \ell^{1,+}$. The usual change of variable $u = h_n^{-1}(v-z)$ in the next integrals gives, for an arbitrary non negative function g with support included in S

$$\begin{aligned}\mathbf{P}_n(g) =& \frac{1}{a_n}\int_{z+h_nS} g\Big(h_n^{-1}(v-z)\Big)\mathbf{f}(v)dv\\ =& \frac{h_n^d}{a_n}\int_S g(u)\mathbf{f}(z+h_nu)du\\ \leq& \sup_{u\in S}\mathbf{f}(z+h_nu)\ \frac{h_n^d}{a_n}\int_S g(u)du\\ =& \sup_{u\in S}\mathbf{f}(z+h_nu)\ \frac{h_n^d}{a_n}\lambda(S)\mathbf{P}_0(g).\end{aligned}$$

This proves (3.12) by applying that inequality to elements of the form $(f_1-f_2)^2$, $(f_1,f_2)\in\mathcal{F}^2$ and recalling (5.24) together with the continuity of $\mathbf{f}$ at z. This also proves (3.8), taking $g := F^2\mathbb{1}_{\{F>M\}}$. Let us now verify (3.9) by proving (3.11). Using a calculus similar as above we have, for an arbitrary function $g \prec (2F)^2 \vee (2F)$

$$\begin{aligned}&\mathbf{f}(z)\lambda(S)\Big|\mathbf{P}_n(g)-\mathbf{P}_0(g)\Big|\\ =&\Big|\frac{\mathbf{f}(z)\lambda(S)h_n^d}{a_n}\int_S g(u)\mathbf{f}(z+h_nu)du - \int_S g(u)\mathbf{f}(z)du\Big|\\ \leq&\frac{\mathbf{f}(z)\lambda(S)h_n^d}{a_n}\times\Big|\int_S g(u)\mathbf{f}(z+h_nu)du - \mathbf{f}(z)\int_S g(u)du\Big|\\ &+\Big|\frac{\mathbf{f}(z)\lambda(S)h_n^d}{a_n}-1\Big|\times\mathbf{f}(z)\int_S \mid g(u)\mid du\\ \leq&\frac{\mathbf{f}(z)\lambda(S)h_n^d}{a_n}\times\sup_{v\in S}\mid\mathbf{f}(z+h_nv)-\mathbf{f}(z)\mid\times\int_S(2F)^2\vee(2F)du\\ &+\Big|\frac{\mathbf{f}(z)\lambda(S)h_n^d}{a_n}-1\Big|\times\mathbf{f}(z)\int_S(2F)^2\vee(2F)du,\end{aligned}\tag{5.25}$$

which tends to zero independently of $g \prec (2F)^2\vee(2F)$. This proves (3.11) and concludes the proof of Lemma 5.8.□

Let us now continue the proof of Corollary 3. First, note that we have

$$\sum_{i=1}^n b_{i,n}^2 \sim \lambda(S)\mathbf{f}(z)nh_n^d \text{ in probability, from where}\tag{5.26}$$

$$\left(\frac{1}{\sqrt{\mathbf{f}(z)\lambda(S)nh_n^d}}\sum_{i=1}^n\beta_{i,n}\Big(f(Y_{i,n})-\mathbb{E}\big(f(Y_{i,n})\big)\Big)\right)_{f\in\mathcal{F}} \to_{\mathcal{L}}\ \mathbb{G}_{\mathbf{P}_0},\tag{5.27}$$

and hence that sequence of processes is asymptotically tight (see, e.g., van der Vaart and Wellner, 1996, p. 20, Definition 1.3.7). Now elementary probability calculus shows that

$$\frac{\sum_{i=1}^n(b_{i,n}-a_n)}{\sqrt{\mathbf{f}(z)\lambda(S)nh_n^d}} \to_{\mathcal{L}}\ Z,\tag{5.28}$$

where Z is standard normal. Moreover, since $\mathbf{P}_n$ satisfies (3.11) and since $f \to \mathbf{P}_0(f)$ is continuous with respect to $||\cdot||_{\mathbf{P}_0,2}$, which makes $\mathcal{F}$ totally bounded, the (deterministic) sequence $\mathbf{P}_n(\cdot)$ is relatively compact in $\ell^\infty(\mathcal{F})$. This, combined with (5.28), implies that the sequence $R_n(\cdot)$ - defined in (4.4) - is asymptotically tight, and hence so is $T_{n,h_n}(\cdot)$ by summation. It will hence be proved to converge to $\mathcal{W}_{\mathbf{P}_0}$ if we prove finite marginal convergences. This is done by elementary analysis of characteristic functions, using the change of variable $u = h_n^{-1}(v - z)$ in the integrals. We omit details. □

6. Appendix

6.1. *A measurability lemma.* In this section we prove the measurability properties claimed in §3.

Lemma 6.1. *For fixed $r \geq 1$ and $n \geq 1$, the map $(\epsilon, \boldsymbol{p}) \to N_{[]}\Big(\epsilon, \mathcal{F}, ||\cdot||_{\mathbf{P}_{n,\boldsymbol{p}},r}\Big)$ is Borel from $]0,\infty[\times\ell^1$ to $\mathbb{R}^+$. As a consequence, the maps*

$$\boldsymbol{p} \to J_{[]}\big(\delta, \mathcal{F}, \mathbf{P}_{n,\boldsymbol{p}}\big),\ \delta > 0,$$

are Borel.

Proof: Fix $r \geq 1$ and $n \geq 1$. Any bracket is closed for the the pointwise topology, i.e., the topology spanned by the evaluation maps $\big\{\{f \to f(y)\},\ y \in \mathfrak{X}\big\}$. Hence so is any finite union of brackets that covers $\mathcal{F}_0$. Since $\mathcal{F}$ is included in the closure of $\mathcal{F}_0$ for the pointwise topology, we deduce that

$$\forall(\epsilon, \boldsymbol{p}) \in]0,\infty[\times\ell^1,\ N_{[]}\Big(\epsilon, \mathcal{F}, ||\cdot||_{\mathbf{P}_{n,\boldsymbol{p}},r}\Big) = N_{[]}\Big(\epsilon, \mathcal{F}_0, ||\cdot||_{\mathbf{P}_{n,\boldsymbol{p}},r}\Big). \tag{6.1}$$

Now the proof of Lemma 6.1 boils down to proving the measurability of

$$H:\ (\epsilon, \boldsymbol{p}) \to N_{[]}\Big(\epsilon, \mathcal{F}_0, ||\cdot||_{\mathbf{P}_{n,\boldsymbol{p}},r}\Big).$$

This is done by noting that, for any $K \in \mathbb{N}$, the set

$$B_K := \Big\{(f_j^-, f_j^+)_{j=1,\dots,K} \in \big(\mathcal{F}_0^2\big)^K,\ \mathcal{F}_0 \subset \bigcup_{j=1}^K [\![f_j^-, f_j^+]\!]\Big\}$$

is countable, and that

$$H(\epsilon, \boldsymbol{p}) > K \quad \Leftrightarrow \quad \forall(f_j^-, f_j^+)_{j=1,\dots,K} \in B_K,\ \exists j \in \{1,\dots,K\},\ ||\ f_j^+ - f_j^-\ ||_{\mathbf{P}_{n,\boldsymbol{p}},r} > \epsilon,$$

which yields the claimed result, since for fixed Borel non negative g, the map $\boldsymbol{p} \to ||\ g\ ||_{\mathbf{P}_{n,\boldsymbol{p}},r}$ is Borel (recall that $\{\mathbf{P}_{n,\boldsymbol{p}},\ \boldsymbol{p} \in \ell^1\}$ is regular).□

6.2. *A discussion upon the possible links of our results to those of Castillo and Nickl (2014).* As requested by a referee we will now discuss whether the results of the present paper (as well as those of Varron, 2014) have any connections with the results of Castillo and Nickl (2014), where the authors investigated the Dirichlet process filtered through successive histograms of dyadic depths L_n. Their article states BvM phenomena for continuous priors, assuming that the "true" density generating $(X_1,\dots,X_n)$ satisfies regularity conditions. Their methods are directed towards handling continuous models, and the authors prove BvM theorems for posterior distributions of *sequences* of priors, having complexities growing with the sample size. This is why they consider sequences of random histograms of the following form (we here correct what seems to be a misprint of their definition (**H**)).

$$\mathbf{f}_n := \sum_{i=0}^{2^{L_n}-1} h_{k,n} 2^{L_n} \mathbb{1}_{I_{k,n}},\ \text{where}$$

$$(h_{k,n})_{k=0,\dots,2^{L_n}-1} \rightsquigarrow \mathit{Dirichlet}\big(\alpha_{0,n},\dots,\alpha_{2^{L_n}-1,n}\big),\ \text{with} \tag{6.2}$$

$$c_1 2^{-aL_n} \leq \alpha_{k,n} \leq c_2,\ \text{with } c_1, c_2 \text{ and } a \text{ not depending upon } k \text{ nor } n. \tag{6.3}$$

Of course it is naturally possible to represent such a random histogram prior as the image of a Dirichlet process by the map

$$\Psi_n:\ \mathbf{P}\rightarrow\sum_{k=0}^{2^{L_n}-1}\mathbf{P}(I_{k,n})\lambda\big(\cdot\mid I_{k,n}\big).$$

Taking $\mathcal{F}_0 := \Big\{\mathbb{1}_{]s,t]},\ (s,t)\in\mathbb{R}^2\Big\}$ and $\mathcal{F}_1 := \Big\{\mathbb{1}_{]-\infty,t]},\ t\in\mathbb{R}\Big\}$, each map Ψ_n is 1-Lipschitz under the norm $||\cdot||_{\mathcal{F}_0}+||\cdot||_{\mathcal{F}_1}$. Hence, in the particular case where $\boldsymbol{\beta}_n \perp\!\!\!\perp \mathbf{Y}_n$, and under the conditions of Varron (2014, Theorem 2), we can have (by linearity)

$$d_{BL1}\Big(\sqrt{n}\Big(\Psi_n\big(Pr_n\big)-\Psi_n\big(\mathbf{P}_n\big)\Big),\Psi_n\Big(\mathbb{G}_{\mathbf{P}_n}\Big)\Big)\rightarrow 0. \tag{6.4}$$

While undeniably true, (6.4) is not a BvM result in the Bayesian model involved in Castillo and Nickl (2014). Indeed their model assumes that, at stage n, the sample is i.i.d. with density f, given the random histogram prior $\mathbf{f}_n$ takes value f. On the other hand, (6.4) is a result of BvM-type for the sequence posterior distributions in another model, where the law of the sample is discrete given the value $\mathbf{P}$ of the Dirichlet prior. Call this second model *model 2*. Even if $\mathbf{f}_n$ can be expressed as $\Psi_n(Pr)$ (Pr denoting a Dirichlet process), one cannot link the posterior distributions in models 1 and 2 by a simple, general argument. However, due to conjugacy of the Dirichlet distribution and the multinomial distribution, it can be readily proved that, if the random histogram prior has law defined through (6.2), then its posterior law given $(x_1,\ldots,x_n)$ in *model 1* is that of the random densities

$$\sum_{k=0}^{2^L_n-1} h_{k,n}^{(n)}2^{L_n}\mathbb{1}_{I_{k,n}},\ \text{where}$$

$$\begin{aligned}(h_{k,n}^{(n)})_{k=0,\ldots,2^{L_n}-1} \rightsquigarrow & \mathit{Dirichlet}\big(\alpha_{0,n}+n\overline{\alpha}_{(x_1,\ldots,x_n)}(I_{0,n}),\ldots,\\ & \alpha_{2^{L_n}-1,n}+n\overline{\alpha}_{(x_1,\ldots,x_n)}(I_{2^{L_n}-1,n})\big),\ \text{with}\\ \overline{\alpha}_{(x_1,\ldots,x_n)} := & \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}.\end{aligned} \tag{6.5}$$

Hence, for the particular case of priors involving the Dirichlet distribution, we can write the posterior law of densities as $\Psi_n(Pr_n)$ where, writing $M_n := \sum_k \alpha_{k,n}$ and $\theta_n := M_n/(M_n+n)$, Pr_n is a Dirichlet process with concentration parameter M_n+n and with mean measure $\theta_n\alpha_n+(1-\theta_n)\overline{\alpha}_{(x_1,\ldots,x_n)}$, where α_n can be chosen as *any* probability measure fulfilling $\alpha_n(I_{k,n})=\alpha_{n,k}/M_n,\ k\leq 2^{L_n}-1$. We will call such a probability measure *compatible*. Now, taking an arbitrary sequence (α_n) of probability measure that is compatible with the sequence $(\alpha_{k,n},\ k\leq 2^{L_n}-1)_{n\geq 1}$, an application of Varron (2014, Corollary 5) shows that the sequence Pr_n satisfies Donsker theorem (indexing by the class $\mathcal{F}_0\cup\mathcal{F}_1$). We can hence recover Corollary 1 of Castillo and Nickl, but with centering $\mathbb{T}_n=P(L_n)$, borrowing their notations. Then the substitution of $P_n(L_n)$ by F_n can be done only through regularity assumptions upon the regularity of the true f_0 (see their Theorem 3). However, we have the feeling that this is more a coincidence than the open door to a wide range of applications of Theorem 2 (of the present paper, or that of Varron, 2014) to BvM theorems for *continuous* prior. By "coincidence", we mean that the conjugacy properties of the Dirichlet distribution and the multinomial distribution are heavily involved in the representation of the posterior law in *model 1*.

Acknowledgement

The author warmly thanks the anonymous referees for their remarks and suggestions that significantly improved the present manuscript.

References

Andersen, N. T., Giné, E., Ossiander, M., and Zinn, J. The central limit theorem and the law of iterated logarithm for empirical processes under local conditions. *Probab. Theory Related Fields*, **77** (2), 271–305 (1988). MR927241.

Barrientos, A. F. and Peña, V. Bayesian bootstraps for massive data. *Bayesian Anal.*, **15** (2), 363–388 (2020). MR4078718.

Castillo, I. A semiparametric Bernstein–von Mises theorem for Gaussian process priors. *Probab. Theory Related Fields*, **152** (1-2), 53–99 (2012). MR2875753.

Castillo, I. and Nickl, R. On the Bernstein-von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist.*, **42** (5), 1941–1969 (2014). MR3262473.

Dudley, R. M. *Uniform central limit theorems*, volume 142 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, New York, second edition (2014). ISBN 978-0-521-73841-5; 978-0-521-49884-5. MR3445285.

Einmahl, U. and Mason, D. M. Gaussian approximation of local empirical processes indexed by functions. *Probab. Theory Related Fields*, **107** (3), 283–311 (1997). MR1440134.

Ferguson, T. S. A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, **1**, 209–230 (1973). MR350949.

Ghosal, S. and van der Vaart, A. *Fundamentals of nonparametric Bayesian inference*, volume 44 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge (2017). ISBN 978-0-521-87826-5. MR3587782.

Hjort, N. L., Holmes, C., Müller, P., and Walker, S. G., editors. *Bayesian nonparametrics*, volume 28 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge (2010). ISBN 978-0-521-51346-3. MR2722987.

James, L. F. Large sample asymptotics for the two-parameter Poisson-Dirichlet process. In *Pushing the limits of contemporary statistics: contributions in honor of Jayanta K. Ghosh*, volume 3 of *Inst. Math. Stat. (IMS) Collect.*, pp. 187–199. Inst. Math. Statist., Beachwood, OH (2008). MR2459225.

Johnstone, I. M. High dimensional Bernstein–von Mises: simple examples. In *Borrowing strength: theory powering applications—a Festschrift for Lawrence D. Brown*, volume 6 of *Inst. Math. Stat. (IMS) Collect.*, pp. 87–98. Inst. Math. Statist., Beachwood, OH (2010). MR2798513.

Kim, Y. The Bernstein-von Mises theorem for the proportional hazard model. *Ann. Statist.*, **34** (4), 1678–1700 (2006). MR2283713.

Kim, Y. and Lee, J. A Bernstein-von Mises theorem in the nonparametric right-censoring model. *Ann. Statist.*, **32** (4), 1492–1512 (2004). MR2089131.

Leahu, H. On the Bernstein-von Mises phenomenon in the Gaussian white noise model. *Electron. J. Stat.*, **5**, 373–404 (2011). MR2802048.

Lo, A. Y. Weak convergence for Dirichlet processes. *Sankhyā Ser. A*, **45** (1), 105–111 (1983). MR749358.

Lo, A. Y. A remark on the limiting posterior distribution of the multiparameter Dirichlet process. *Sankhyā Ser. A*, **48** (2), 247–249 (1986). MR905464.

Ossiander, M. A central limit theorem under metric entropy with L_2 bracketing. *Ann. Probab.*, **15** (3), 897–919 (1987). MR893905.

Sethuraman, J. A constructive definition of Dirichlet priors. *Statist. Sinica*, **4** (2), 639–650 (1994). MR1309433.

Sheehy, A. and Wellner, J. A. Uniform Donsker classes of functions. *Ann. Probab.*, **20** (4), 1983–2030 (1992). MR1188051.

van der Vaart, A. W. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge (1998). ISBN 0-521-49603-9; 0-521-78450-6. MR1652247.

van der Vaart, A. W. and Wellner, J. A. *Weak convergence and empirical processes. With applications to statistics.* Springer Series in Statistics. Springer-Verlag, New York (1996). ISBN 0-387-94640-3. MR1385671.

Varron, D. Donsker and Glivenko-Cantelli theorems for a class of processes generalizing the empirical process. *Electron. J. Stat.*, **8** (2), 2296–2320 (2014). MR3275745.