Electronic Journal of Statistics
Vol. 8 (2014) 2296–2320
ISSN: 1935-7524
DOI: 10.1214/14-EJS955

Donsker and Glivenko-Cantelli theorems for a class of processes generalizing the empirical process

Davit Varron

Laboratoire de mathématiques de Besançon
UMR CNRS 6623
Université de Franche-Comté
e-mail: davit.varron@univ-fcomte.fr

Abstract: We establish a Glivenko-Cantelli and a Donsker theorem for a class of random discrete measures which generalize the empirical measure, under conditions on uniform entropy numbers that are common in empirical processes theory. Some illustrative applications in nonparametric Bayesian theory and randomly sized sampling are provided.

MSC 2010 subject classifications: 60F17, 60G57, 62F15.
Keywords and phrases: Empirical processes, random measures, Bayesian nonparametrics.

Received March 2014.

Contents

1. Introduction

The asymptotic properties of empirical processes indexed by functions have been intensively studied during the past decades (see, e.g., Van der Vaart and Wellner (1996) or Dudley (1999) for self-contained, comprehensive books on the topic). Among those results, one of the most used in statistical applications is that a class admiting a finite uniform entropy integral and a square integrable envelope is Donsker (Koltchinskii (1981)). In this note, we will show that this structural condition is strong enough to carry a Donsker and a Glivenko-Cantelli theorem over a wider class of processes which we shall describe at once. For $\boldsymbol{p} = (p_i)_{i\geq 1} \in \mathbb{R}^{\mathbb{N}}$ and $r > 0$, we shall write the (possibly infinite) value

$$|| \boldsymbol{p} ||_r := \Big(\sum_{i\geq 1} | p_i |^r \Big)^{1/r}. \tag{1}$$

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a complete probability space (in the sense that every $\mathbb{P}$-negligible set belongs to $\mathcal{A}$), and let

$$\mathbb{S} := \Big\{ \boldsymbol{p} = (p_i)_{i\geq 1} \in [0, +\infty[^{\mathbb{N}},\ || \boldsymbol{p} ||_1 < \infty \Big\}, \tag{2}$$

be the cone of positive summable sequences, which will be endowed with the product Borel σ-algebra (denoted by Bor). Consider a sequence of $\mathbb{S}$-indexed collections of probability measures $(\mathbf{P}_{n,\boldsymbol{p}})_{n\geq 1, \boldsymbol{p}\in\mathbb{S}}$ on a measurable space $(\mathfrak{X}, \mathcal{X})$. For fixed n, consider a random variable $\boldsymbol{\beta}_n = (\beta_{i,n})_{i\geq 1}$ from $(\Omega, \mathcal{A}, \mathbb{P})$ to $(\mathbb{S}, Bor)$ and a sequence $\mathbf{Y}_n = (Y_{i,n})_{i\geq 1}$ of random variables from $(\Omega, \mathcal{A}, \mathbb{P})$ to $(\mathfrak{X}, \mathcal{X})$, for which the conditional law given $\boldsymbol{\beta}_n = \boldsymbol{p}$ is $\mathbf{P}_{n,\boldsymbol{p}}^{\otimes\mathbb{N}}$ (we also make the assumption that those conditional laws properly define a Markov kernel). We then define the random element

$$Pr_n := \sum_{i\geq 1} \beta_{i,n} \delta_{Y_{i,n}}, \tag{3}$$

which is a random probability measure in the sense that $Pr_n(\mathfrak{X}) \equiv 1$ and $Pr_n(A)$ is Borel measurable for each $A \in \mathcal{X}$. Obtaining asymptotic results (as $n \to \infty$) for Pr_n is of interest for several reasons. The first of them is that Pr_n generalizes the usual empirical measure, and also related objects from the bootstrap theory, such as the empirical bootstrap, or more generally, the exchangeable bootstrap empirical measure (see (Van der Vaart and Wellner, 1996, Section 3.6.2)). Our main motivation to consider such a generalization comes from the second reason: these types of random measures play a role in the modeling of almost surely discrete priors, such as stick breaking priors (including the two parameter Poisson-Dirichlet process) or, more generally, species sampling priors, including the normalized homogenous completely random measures. For an overview of all the previously cited types of random measures, see, e.g., (Hjort et al., 2010, Chapter 3). In addition, there are several situations where objects such as (3) do also appear in the posterior distributions or posterior expected values of some almost surely discrete priors (see §3 in the sequel).

As usual in empirical processes theory, we shall write, for a probability measure $\mathbf{P}$ and an integrable function f:

$$\mathbf{P}(f) := \int_{\mathfrak{X}} f d\mathbf{P}, \tag{4}$$

and we shall adopt the same notation $Pr(f)$ when Pr is a random probability measure, in which case $Pr(f)$ has to be understood as a random variable. As mentioned earlier, we will state a Glivenko-Cantelli and a Donsker result for empirical-like processes indexed by a class of (Borel) functions $\mathcal{F}$. When $\mathcal{F}$ is not uniformly bounded, the definition of such objects requires additional care on the structure of $\mathcal{F}$. Throughout this article, we will assume that $\mathcal{F}$ is pointwise separable relatively to a countable subclass $\mathcal{F}_0$ (see, e.g., Van der Vaart and Wellner (1996, p. 110)). We will denote by $\mathcal{B}(\mathcal{F})$ the space of real bounded functions on $\mathcal{F}$ that are continuous with respect to the topology spanned by the maps $\{f \to f(x),\ x \in \mathcal{X}\}$. We shall also denote the usual sup norm on $\mathcal{B}(\mathcal{F})$

$$|| \psi ||_{\mathcal{F}} := \sup_{f\in\mathcal{F}} | \psi(f) | .$$

We will denote by F the envelope function $x \to \sup\{f(x),\ f \in \mathcal{F}_0\} \vee 1$. For $\epsilon > 0,\ r > 0$ and a probability measure Q, we shall write $N(\epsilon, \mathcal{F}, || \cdot ||_{Q,r})$ for the (possibly infinite) minimal number of balls with radius ϵ needed to cover $\mathcal{F}$, using the usual $L^r(Q)$ norm. We shall also write $\ell^\infty(\mathcal{F}) \supset \mathcal{B}(\mathcal{F})$ for the space of all real bounded functions on $\mathcal{F}$. For $r > 0$, we define the space $\mathcal{E}_{\mathcal{F},r}$ as follows: a map $\Psi : \Omega \to \mathcal{B}(\mathcal{F})$ belongs to $\mathcal{E}_{\mathcal{F},r}$ if and only if $\Psi(f)$ defines a Borel random variable on $(\Omega, \mathcal{A}, \mathbb{P})$ for each $f \in \mathcal{F}$, and if $||| \Psi |||_{\mathcal{F},r}^r := \mathbb{E}(|| \Psi ||_{\mathcal{F}}^r) < \infty$. Under the assumption that $\mathbb{E}(F(Y_{1,n})) < \infty$ for all n, it is possible to define the process

$$G_n(\cdot) :\ f \to \sum_{i\geq 1} \beta_{i,n}\Big(f(Y_{i,n}) - \mathbb{E}\big(f(Y_{i,n}) \mid \boldsymbol{\beta}_n\big)\Big)$$

as the limit, as $k \to \infty$, of the truncated sequence

$$G_n^k(\cdot) :\ f \to \sum_{i=1}^{k} \beta_{i,n}\Big(f(Y_{i,n}) - \mathbb{E}\big(f(Y_{i,n}) \mid \boldsymbol{\beta}_n\big)\Big)$$

in the Banach space $(\mathcal{E}_{\mathcal{F},1}, ||| \cdot |||_{\mathcal{F}})$. Also note that this limit also holds in $\mathcal{E}_{\mathcal{F},r}$ for each $r \geq 1$ such that $\mathbb{E}(F(Y_{1,n})^r) < \infty$. Note that the measurability of the $|| G_n^k ||_{\mathcal{F}}$ is not completely immediate. The corresponding proof can be found at the beginning of §4.

2. Main results

We state here our first result, of Glivenko-Cantelli type.

Theorem 1. *Assume that $|| \boldsymbol{\beta}_n ||_1 = 1$ almost surely for all n, and $|| \boldsymbol{\beta}_n ||_2 \to 0$ in probability. Suppose that*

$$\lim_{M\to\infty} \overline{\lim_{n\to\infty}}\, \mathbb{E}\Big(F(Y_{1,n}) \mathbb{1}_{\{F(Y_{1,n})\geq M\}}\Big) = 0. \tag{5}$$

Also assume that, for each $\epsilon > 0$ and $M > 0$, we have, as $n \to \infty$:

$$\log\Big(N\big(\epsilon, \mathcal{F}_M, || \cdot ||_{\overline{P}(\boldsymbol{\beta}_n, \mathbf{Y}_n),1}\ \big)\Big) = o_{\mathbb{P}}\Big(\ || \boldsymbol{\beta}_n ||_2^{-2}\ \Big), \tag{6}$$

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

$$\overline{P}(\boldsymbol{p},\mathbf{y}) := \sum_{i\ge 1} p_i\delta_{y_i},\ \textit{for}\ \boldsymbol{p}\in\mathbb{S},\ \mathbf{y}\in\mathfrak{X}^{\mathbb{N}}. \tag{7}$$

Then

$$\mathbb{E}\big(\ ||\ G_n\ ||_{\mathcal{F}}\ \big)\to 0. \tag{8}$$

Remark 2.1. Choosing $\beta_{i,n} := n^{-1}$ when $i\le n$ and 0 otherwise, leads to the usual Glivenko-Cantelli theorem under random entropy conditions (see, e.g., Van der Vaart and Wellner (1996, p. 123)), except for the almost sure counterpart of (8). Indeed, that almost sure convergence deeply relies on a reverse submartingale structure, which is not guaranteed under the general conditions of Theorem 1.

Our second result is a Donsker type Theorem. Since, for fixed n, the $Y_{i,n}$ are only conditionally independent, such a result will not involve the Gaussian analogues of the G_n, but mixtures W_n of the $\mathbf{P}_{n,\boldsymbol{\beta}_n}$ Brownian bridges by $\boldsymbol{\beta}_n$, for which a rigorous definition resuires additional care. To properly define them, we proceed as follows: for $p\ge 1$ and $\mathbf{f} := (f_1,\dots,f_p)\in\mathcal{F}^p$ and $\boldsymbol{p}\in\mathbb{S}$, writing $\mathbf{Q}^{\mathbf{f}}_{n,\boldsymbol{p}}$ for the centered Gaussian distribution with variance covariance matrix

$$\Sigma^{\mathbf{f}}_{n,\boldsymbol{p}} := \Big[\mathbf{P}_{n,\boldsymbol{p}}\Big((f_j-\mathbf{P}_{n,\boldsymbol{p}}(f_j))(f_{j'}-\mathbf{P}_{n,\boldsymbol{p}}(f_{j'}))\Big)\Big]_{(j,j')\in\{1,\dots,p\}^2},$$

we set, for each Borel set $A\subset\mathbb{R}^p$:

$$\mathbf{Q}^{\mathbf{f}}_n(A) := \mathbb{E}\Big(\mathbf{Q}^{\mathbf{f}}_{n,\boldsymbol{\beta}_n}(A)\Big).$$

Kolmogorov's extension theorem ensures the existence of a probability measure $\mathbb{P}'_n$ on $\Omega' := \mathbb{R}^{\mathcal{F}}$, endowed with its ($\mathbb{P}'_n$-completed) product Borel σ-algebra $\mathcal{X}'$, which is compatible with the system $\{\mathbf{Q}^{\mathbf{f}}_n,\ \mathbf{f}\in\mathcal{F}^p,\ p\ge 1\}$. We define W_n as the canonical map on $(\Omega',\mathcal{X}',\mathbb{P}'_n)$. Unfortunately, the latter can fail to be measurable with respect to the Borel σ-algebra of $(\ell^\infty(\mathcal{F}),||\cdot||_{\mathcal{F}})$. This lack of measurability will be tackled by introducing outer expectations (see, e.g. Van der Vaart and Wellner (1996, Chapter 1.2)). To simplify the notations, we shall adopt the following convention: each time a map $\mathfrak{h}$ is defined on a probability space, the symbol $\mathbb{E}^*(\mathfrak{h})$ will denote the outer expectation with respect to that probability space. We shall adopt the same convention for outer probabilities $\mathbb{P}^*$.

Theorem 2. *Assume that, for each* $n\ge 1$, $||\ \boldsymbol{\beta}_n\ ||_2^2 = 1$ *with probability one, and that* $||\ \boldsymbol{\beta}_n\ ||_4\to 0$ *in probability. Also assume that* $\mathbb{E}(F^2(Y_{1,n}))<\infty$ *for all* n*, and that:*

$$\lim_{M\to\infty}\overline{\lim_{n\to\infty}}\ \mathbb{E}\Big(F^2(Y_{1,n})\mathbb{1}_{\{F(Y_{1,n})\ge M\}}\Big) = 0, \tag{9}$$

$$\int_0^\infty\sqrt{\log\Big(\sup_{Q\ probab.} N\big(\epsilon\ ||\ F\ ||_{Q,2},\mathcal{F},||\cdot||_{Q,2}\ \big)\Big)}d\epsilon<\infty. \tag{10}$$

Then, for all $n \geq 1$, W_n is almost surely bounded, and $|| W_n ||_{\mathcal{F}}$ is Borel measurable. Also assume that, for a semimetric ρ_0 which makes $\mathcal{F}$ totally bounded we have,

$$\lim_{\delta \to 0} \overline{\lim_{n\to\infty}}^{\mathbb{P}^*} \sup_{(f_1,f_2)\in\mathcal{F}^2, \rho_0(f_1,f_2)\leq\delta} || f_1 - f_2 ||_{\mathbf{P}_n,\boldsymbol{\beta}_n,2} = 0, \tag{11}$$

where the symbol $\overline{\lim}_{n\to\infty}^{\mathbb{P}^*}$ stands for the lim sup in outer probability. Then

$$d_{BL}\Big(G_n, W_n\Big) := \sup_{B\in BL1} \Big|\mathbb{E}^*\Big(B\big(G_n\big)\Big) - \mathbb{E}^*\Big(B\big(W_n\big)\Big)\Big| \to 0, \tag{12}$$

where $BL1$ is the set of all 1-Lipschitz functions on $(\ell^\infty(\mathcal{F}), ||\cdot||_{\mathcal{F}})$ that are bounded by 1.

Remark 2.2. As pointed out in the introduction, Pr_n encompasses several already well studied random measures, for which Donsker and Glivenko-Cantelli theorems have been established during the past decades. We will hence point out the relevance of our results with respect to the existing literature.

1. When $\mathbf{Y}_n \perp\!\!\!\perp \boldsymbol{\beta}_n$, each W_n is equal in distribution to the $\mathbf{P}_n$-Brownian bridge. In addition, condition (11) is implied by condition (ii) in Sheehy and Wellner (1992, Theorem 3.1), namely:

   For some probability measure $\mathbf{P}_0$, we have:

$$\sup_{(f,g)\in\mathcal{F}^2} \max\Big\{ \mid \mathbf{P}_n\big((f-g)^2\big) - \mathbf{P}_0\big((f-g)^2\big) \mid, \mid \mathbf{P}_n(f) - \mathbf{P}_0(f) \mid \\ \mid \mathbf{P}_n(f^2) - \mathbf{P}_0(f^2) \mid \Big\} \to 0, \tag{13}$$

   where $\mathbf{P}_n$ stands for the law of $Y_{1,n}$. In that case, (12) has the simpler interpretation that $G_n(\cdot)$ converges weakly to the $\mathbf{P}_0$-Brownian bridge $\mathbb{G}_{\mathbf{P}_0}$. Hence, in this setting, Theorem 2 turns out to be a partial generalization of Sheehy and Wellner (1992, Theorem 3.1), where the authors proved, among other results, a Donsker theorem for sequence of $\mathcal{F}$-indexed empirical processes, for which the law of the sample varies with n. The contribution of our result is that it extends to random (possibly infinite) convex combinations of the $\delta_{Y_{i,n}}$.
2. When, again, $\mathbf{Y}_n \perp\!\!\!\perp \boldsymbol{\beta}_n$, when $\mathbf{P}_n := \mathbf{P}_0$ is constant, when $\beta_{i,n} = 0$ for $i \geq n+1$, when $|| \boldsymbol{\beta}_n ||_1 = 1$ almost surely, and when the vector $(\beta_{i,n})_{i=1\ldots,n}$ is exchangeable for all n, then Pr_n coincides with the exchangeable bootstrap empirical measure. Among other results, Praestgaard and Wellner (1993, Theorem 2.2) did establish a Donsker theorem for such objects, under the sole assumption that $\mathcal{F}$ is $\mathbf{P}_0$-Donsker. However, their result holds with sequences of $\mathcal{F}$ indexed processes which differ from G_n, namely:

$$\widetilde{G}_n : f \to \sum_{i=1}^n \Big(\beta_{i,n} - \frac{1}{\sqrt{n}}\Big)\delta_{Y_i}. \tag{14}$$

Not only does the centering parameter differ from G_n to $\widetilde{G}_n$, but their result is a Donsker theorem that holds for the conditional laws of $\widetilde{G}_n$ given $(Y_1, \dots, Y_n)$. Hence, our Theorem 2 has very few similarities with that result.

3. It also has to be noted, when $\mathbf{Y}_n \perp\!\!\!\perp \boldsymbol{\beta}_n$, that the processes G_n are measurelike (see, e.g., (Van der Vaart and Wellner, 1996, Chapter 2.11)). Hence, the result of Alexander (1987) applies, when $\mathcal{F}$ satisfies (10), to sums

$$\mathbf{G}_n = \frac{1}{n} \sum_{i=1}^{n} G_{n,i}, \tag{15}$$

of n independent processes having representation from (3). Though relying on similar ideas, our result is different in the sense that we prove asymptotic theorems for more general sequences, that do not need to be expressed as sums of n independent measurelike processes.

The remainder of this article is organized as follows: in §3, we provide applications of Theorems 1 and 2 to particular types of random probability measures, and provide simple criteria for Glivenko-Cantelli and Donsker theorems to hold. We also discuss the possible applications of those results to Bayesian nonparametrics. Most of the proofs of the results of this section are short, and hence are written in continuation of them. Finally, in §4, the proofs of Theorems 1 and 2 and Corollary 4 are provided.

3. Some applications of Theorems 1 and 2

In this section, we will consider sequences of random measures of different types: modified Kac processes, normalized homogenous completely random measures (NHCRM) and stick breacking random measures. For a given type, we will provide a tractable sufficient criterion for sequences of that type to satisfy a Donsker or a Glivenko Cantelli theorem. Both NHCRM and stick breaking measures are encountered in Bayesian nonparametrics, and also play a role in their posterior distributions given the observed sample. We will hence also discuss of possible uses of Theorems 1 and 2 to frequentist asymptotics in Bayesian nonparametrics. For that aim, let us first provide a concise background of this theory.

Write $\mathfrak{M}$ for the set of all probability measures on $\mathfrak{X}$, endowed with the Borel σ-field $\mathcal{M}$ spanned by the weak topology. The Bayesian analysis with a specified prior Pr exhibits has objects of interest, namely:

1. The posterior distribution:

$$Post_n : \ (x_1, \dots, x_n) \mapsto \tilde{\mathbb{P}}\Big(Pr = \cdot \Big| (X_1, \dots, X_n) = (x_1, \dots, x_n)\Big),$$

where $\tilde{\mathbb{P}}$ is understood as the underlying probability measure of the nonparametric Bayesian model (hence a probability measure on $(\mathfrak{M}, \mathcal{M}) \otimes (\mathfrak{X}, \mathcal{X})^{\otimes \mathbb{N}}$).

2. The posterior expected probability:

$$Expect_n\colon\ (x_1,\dots,x_n)\mapsto\Big\{A\mapsto\mathbb{E}_{\tilde{\mathbb{P}}}\Big(Pr(A)\Big|(X_1,\dots,X_n)=(x_1,\dots,x_n)\Big)\Big\}.$$

Note that those two objects are rigorously classes of equivalences, for which the representant does not have any impact for probability calculus under $\tilde{\mathbb{P}}$. The frequentist analysis for Bayesian nonparametrics, also called "what if" approach, consists in assuming that the X_i are i.i.d. with a distribution $\mathbf{P}_0$, and then obtaining asymptotic results (of first or second order) for $Expect_n(X_1,\dots,X_n)$, and $\mathbf{P}_0^{\otimes\mathbb{N}}$ almost sure limit results for $Post_n(x_1,\dots,x_n)$ (those of the later type are usually called posterior consistency and Bernstein-von Mises theorems). Under the distribution $\mathbf{P}_0^{\otimes\mathbb{N}}$, the choice of representant in $Post_n$ and $Expect_n$ is not harmless anymore. However, the Bayes calculus often exhibits versions that are "natural" from a statistical point of view. For the specific choice of the (strong) topology induced by $||\cdot||_{\mathcal{F}}$, it is our belief that Theorems 1 and 2 can play a role in establishing "what if" asymptotic results. This belief turns out to be true when considering Poisson-Dirichlet priors (see §3.3 below). The case of posterior consistency of NHCRM is also discussed in §3.2. Another brief discussion on Gibbs-type priors is also presented in §3.5. The Bernstein-von Mises phenomenon for Dirichlet processes priors (first proved by Lo (1983), treating the case of the distribution functions of the posterior laws, and later extended by James (2008) to a wider setup) is also revisited in §3.4.

Remark 3.1. Due to the very nature of Theorems 1 and 2, the possible use of those two results in Bayesian nonparametrics is limited to models where the posterior distributions are almost surely discrete, or at least are predominantly built on random discrete probability measures. Such a restriction naturally excludes the vast majority of smooth models (see, e.g., (Hjort et al., 2010, Chapter 3.4)), where the support of Pr (and hence of each $Post_n(x_1,\dots,x_n)$) is the set of continuous measures.

We start this section with an application that is not related to Bayesian nonparametrics: empirical processes with random sample sizes (also called Kac-processes), for which the trust in the observations depends increasingly upon the observed sample size.

3.1. Randomly sized sampling with increasing trust

For fixed n, consider a sequence $\mathbf{Y}_n$ having distribution $\mathbf{P}_n^{\otimes\mathbb{N}}$ and an almost surely finite $\mathbb{N}^*$-valued random variable η_n which is independent of the first sequence. For fixed n, the random probability measure

$$Kac_n:=\frac{1}{\eta_n}\sum_{i=1}^{\eta_n}\delta_{Y_{i,n}},$$

is called a Kac empirical measure. Randomizing the sample size plays a role in probability calculus for the empirical process, where Poissonization techniques are involved (in which case η_n is a Poisson random variable with expectation n). But such a randomization does also reflect a situation sometimes encountered in practice: the sample size is not known in advance but the observations can be considered as i.i.d regardless of the observed value of the sample size. Limit theorems for Kac random measures are well known (see, e.g. (Van der Vaart and Wellner, 1996, Chapter 3.5)). An interesting extension of those objects is to weaken the independence structure to a conditional i.i.d. structure of $(Y_{i,n})_{i\geq 1}$ given η_n. More precisely, given a sequence of probability measures $(\overline{\mathbf{P}}_k)_{k\geq 1}$ one may assume that, for each integers n and k, the law of $(Y_{i,n})_{i\geq 1}$ given $\eta_n = k$ is $\overline{\mathbf{P}}_k^{\otimes \mathbb{N}}$. Moreover, if one makes the assumption that $\overline{\mathbf{P}}_k \to \mathbf{P}_0$, as $k \to \infty$, for some "true" probability measure $\mathbf{P}_0$, then the corresponding model illustrates an *increasing trust* phenomenon: for example, people responding to a survey on a sensible topic may be keener to reveal their true beliefs or habits (modeled by $\mathbf{P}_0$) when they are ensured that a large number of other people are also interrogated. Our application of Theorems 1 and 2 is as follows.

Corollary 1. *Let $\mathcal{F}$ be a pointwise separable class of functions with measurable envelope F. Let $(\overline{\mathbf{P}}_k)_{k\in\mathbb{N}}$ be a sequence of probability measures such that $\overline{\mathbf{P}}_k$ converges to $\mathbf{P}_0$ in the sense of (13). Let η_n be a sequence of $\mathbb{N}^*$ valued random variables satisfying $\eta_n \to_{\mathbb{P}} +\infty$. Assume that, for each integers n and k, the law of $\mathbf{Y}_n$ given $\eta_n = k$ is $\overline{\mathbf{P}}_k^{\otimes\mathbb{N}}$.*

1. *If $\mathcal{F}$ satisfies (6), with $\| \boldsymbol{\beta}_n \|_2^2 = \eta_n$, as well as*

$$\lim_{M\to\infty} \overline{\lim_{k\to\infty}}\, \overline{\mathbf{P}}_k\Big(F\mathbb{1}_{\{F\geq M\}}\Big) = 0, \tag{16}$$

then whe have

$$\sup_{f\in\mathcal{F}} \Big| Kac_n(f) - \mathbf{P}_0(f)\Big| \to_{\mathbb{P}} 0. \tag{17}$$

2. *If $\mathcal{F}$ satisfies (10), and satisfies*

$$\lim_{M\to\infty} \overline{\lim_{k\to\infty}}\, \overline{\mathbf{P}}_k\Big(F^2\mathbb{1}_{\{F\geq M\}}\Big) = 0, \tag{18}$$

then the sequence of $\mathcal{F}$-indexed processes

$$G_n \ : \ \sqrt{\eta_n}\Big(Kac_n(\cdot) - \overline{\mathbf{P}}_{\eta_n}(\cdot)\Big) \tag{19}$$

converges in law to the $\mathbf{P}_0$-Brownian bridge $\mathbb{G}_{\mathbf{P}_0}$. If, in addition, we have

$$\lim_{k\to\infty} \sqrt{k} \sup_{f\in\mathcal{F}} \Big|\overline{\mathbf{P}}_k(f) - \mathbf{P}_0(f)\Big| = 0, \tag{20}$$

then the centering processes $\overline{\mathbf{P}}_{\eta_n}(\cdot)$ can be replaced by $\mathbf{P}_0(\cdot)$ in the above mentioned weak convergence.

Remark 3.2. Condition (20) can be understood as a condition on the rate of trustfullness of $\overline{\mathbf{P}}_k$ as the sample size k grows. It is somehow unavoidable, since, if the (non random) trajectories $\sqrt{k}(\overline{\mathbf{P}}_k(\cdot) - \mathbf{P}_0(\cdot))$ converge in $\ell^\infty(\mathcal{F})$ to a limit trajectory ψ, then the second convergence in point 2 holds toward the drifted Brownian bridge $\mathbb{G}_{\mathbf{P}_0}(\cdot) + \psi(\cdot)$, as it will clearly appear in the proof.

Proof of Corollary 1. We first prove point 1. Looking at (16), we can assume without loss of generality that $\overline{\mathbf{P}}_k(F) < \infty$ for all $k \geq 1$. Since $\eta_n \to_{\mathbb{P}} \infty$, conditioning upon η_n readily gives (5) from (16), by the use of Cesaro's convergence criterion. The same argument combined with (13) shows that

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}} \Big| \mathbb{E}\big(f(Y_{1,n})\big) - \mathbf{P}_0(f) \Big| = 0. \tag{21}$$

Now, since the weights $\beta_{i,n} := \eta_n^{-1} \mathbb{1}_{\{i\leq\eta_n\}}$ satisfy $\|\, \boldsymbol{\beta} \,\|_1 = 1$ and $\|\, \boldsymbol{\beta} \,\|_2^2 = \eta_n^{-1}$, all the conditions of Theorem 1 are met and we apply the latter, justifying the centering terms $\mathbf{P}_0(f)$ by (21). The proof of point 2 is done similarly, by formally replacing F by F^2 and by taking $\beta_{i,n} := \eta_n^{-1/2} \mathbb{1}_{\{i\leq\eta_n\}}$. This entails (19), by Theorem 2 combined with point 1 of Remark 2.2. Now the use of the centering parameter $\mathbf{P}_0(f)$ instead of $\overline{\mathbf{P}}_{\eta_n}(f)$ is guaranteed by (20) and straightforward calculus. ☐

3.2. Application to sequences of normalized homogenous completely random measures

A first general class of random measures admitting representation (3) and playing a role in Bayesian nonparametrics is that of normalized homogenous completely random measures (NHCRM) with infinite activity which can be described as follows (see, e.g., Hjort et al. (2010, p. 84–85)): given any finite measure ν on $(\mathfrak{X}, \mathcal{X})$ and an infinite but σ-finite measure ρ on $\mathbb{R}^+$ fulfilling $\int s d\rho(s) < \infty$, denote by $\mathbf{\Pi}$ a Poisson random measure on $\mathfrak{X} \times \mathbb{R}^+$ with base measure $\nu \otimes \rho$. Then the associated NHCRM Pr is defined as

$$Pr := B \to \frac{\int_{B\times\mathbb{R}^+} s d\mathbf{\Pi}(x,s)}{\int_{\mathfrak{X}\times\mathbb{R}^+} s d\mathbf{\Pi}(x,s)}. \tag{22}$$

Since ρ is σ-finite, we can select a sequence $(A_k)_{k\geq1}$ fulfilling $\rho(A_k) < \infty$, and represent $\mathbf{\Pi}$ as

$$\mathbf{\Pi} =_d \sum_{k\geq1} \sum_{i=1}^{\eta_k} \delta_{(Y_{i,k}, J_{i,k})}, \tag{23}$$

where $(\eta_k)_{k\geq1}$ is a sequence of independent Poisson processes with respective expectations $\rho(A_k) \times \nu(\mathfrak{X})$, and $(Y_{i,n}, J_{i,n})_{n\geq1, i\geq1}$ is an array of independent random variables, independent of $(\eta_k)_{k\geq1}$, and for which any pair $(Y_{i,n}, J_{i,n})$ has distribution $[\nu(\mathfrak{X})^{-1}\nu(\cdot)] \otimes [\rho(A_n)^{-1}\rho(\cdot \cap A_n)]$. Plugging (23) in (22) directly implies that Pr can be represented through (3), with $\mathbf{P}_{n,\boldsymbol{p}} := \nu(\mathfrak{X})^{-1}\nu(\cdot)$ being constant in $\boldsymbol{p}$. As an application of Theorems 1 and 2, we now provide a sufficient and simple criterion for sequences of NHCRM to satisfy a Glivenko-Cantelli or a Donsker theorem.

Corollary 2. *Let $\mathcal{F}$ be a pointwise separable class of functions with measurable envelope F, satisfying (10). Let ν_n (resp ρ_n) be a sequence of finite (resp. σ-finite) measures, and let $\mathbf{\Pi}_n$ be the associated sequence of Poisson random measures. Write $\mathbf{P}_n := \nu_n(\mathfrak{X})^{-1}\nu_n(\cdot)$.*

1. *Assume that $\mathcal{F}$ satisfies (5), with $Y_{1,n} \rightsquigarrow \mathbf{P}_n$ and that, as $n \to \infty$:*

$$K_n^{-2} := \frac{\int_{\mathbb{R}^+} s^2 d\rho_n(s)}{\nu_n(\mathfrak{X}) \times \left(\int_{\mathbb{R}^+} s d\rho_n(s) \right)^2} = o(1). \tag{24}$$

   Then we have

$$\sup_{f\in\mathcal{F}} \Big| \frac{\int_{\mathfrak{X}\times\mathbb{R}^+} sf(x) d\mathbf{\Pi}_n(x,s)}{\int_{\mathfrak{X}\times\mathbb{R}^+} s d\mathbf{\Pi}_n(x,s)} - \int_{\mathfrak{X}} f d\mathbf{P}_n \Big| \to_{\mathbb{P}} 0.$$

2. *Now assume in addition that $\mathcal{F}$ satisfies (9) and that $\mathbf{P}_n$ converges to a probability measure $\mathbf{P}_0$ in the sense of (13). Also assume that*

$$\int_{\mathbb{R}^+} s^4 d\rho_n(s) = o\left(\nu_n(\mathfrak{X}) \times \left(\int_{\mathbb{R}^+} s^2 d\rho_n(s) \right)^2 \right). \tag{25}$$

   Then the sequence of random elements

$$G_n : \ f \mapsto K_n \left(\frac{\int_{\mathfrak{X}\times\mathbb{R}^+} sf(x) d\mathbf{\Pi}_n(x,s)}{\int_{\mathfrak{X}\times\mathbb{R}^+} s d\mathbf{\Pi}_n(x,s)} - \mathbf{P}_n(f) \right) \tag{26}$$

   converges weakly to $\mathbb{G}_{\mathbf{P}_0}$.
   If in addition we have

$$K_n \sup_{f\in\mathcal{F}} \Big| \mathbf{P}_n(f) - \mathbf{P}_0(f) \Big| \to 0, \tag{27}$$

   then the centering trajectories $\mathbf{P}_n(\cdot)$ can be replaced by $\mathbf{P}_0(\cdot)$ in (26).

Remark 3.3. James et al. (2009) did provide the very general form of the posterior probability of priors Pr that are NHCRM. They proved that, given a sample $(x_1, \ldots, x_n)$ with distincts observations $(x_1^*, \ldots, x_k^*)$, and given the value u of a specific random variable U_n, the posterior distribution $Pr^{(u,x_1,\ldots,x_n)}$ of such a prior can be expressed as a convex combination of two independent components: a normalized completely random measure $\overline{Pr}^{(u,x_1,\ldots,x_n)}$, and a (possibly unbalanced) bootstraped empirical measure $\sum_{i=1}^k W_i^{(u,x_1,\ldots,x_n)} \delta_{x_i^*}$.

Hence, the posterior law of Pr given $(x_1, \ldots, x_n)$ is theoretically entirely known, by mixing $Pr^{(u,x_1,\ldots,x_n)}$ along the distribution of U_n given $(x_1, \ldots, x_n)$. Another important feature of this representation is that, if Pr is a NHCRM with location measure ν, then all the $\overline{Pr}^{(u,x_1,\ldots,x_n)}$ are also NHCRM, with location measure ν. Hence, we hope that Corollary 2 could be a contribution toward achieving (at least partially) the aim of first/second order posterior consistency of completely random measures, under norms of the form $|| \cdot ||_{\mathcal{F}}$. However, there is still much work to do, and a very crucial step is to understand the $\mathbf{P}_0^{\otimes \mathbb{N}}$ almost sure asymptotic probabilistic behaviour of U_n given $(x_1, \dots, x_n)$.

Proof of Corollary 2. We start the proof with the following straightforward lemma.

Lemma 3.1. *For a non negative Borel function ψ on $[0, \infty)$, if we have*

$$\frac{\int_{\mathbb{R}^+} \psi^2(s) d\rho_n(s)}{v_n(\mathfrak{X}) \times \Big(\int_{\mathbb{R}^+} \psi(s) d\rho_n(s) \Big)^2} = o(1),$$

as $n \to \infty$, then, since $\mathbf{\Pi}_n$ is a Poisson measure with mean measure $\rho_n \otimes v_n$, we have

$$\mathrm{Var}\Big(\int_{\mathbb{R}^+} \psi(s) d\mathbf{\Pi}_n(x, s) \Big) = o\bigg(\mathbb{E}\Big(\int_{\mathbb{R}^+} \psi(s) d\mathbf{\Pi}_n(x, s) \Big)^2 \bigg),$$

and hence the involved random variable is an equivalent of its expectation, in probability, as $n \to \infty$.

We first prove point 1. With Theorem 1 at hand, it is sufficient to prove that

$$|| \boldsymbol{\beta}_n ||_2^2 = \frac{\int_{\mathfrak{X} \times \mathbb{R}^+} s^2 d\mathbf{\Pi}_n(x, s)}{\Big(\int_{\mathfrak{X} \times \mathbb{R}^+} s d\mathbf{\Pi}_n(x, s) \Big)^2} \to_{\mathbb{P}} 0. \tag{28}$$

By Markov's inequality, the numerator is $O_{\mathbb{P}}(v_n(\mathfrak{X}) \int s^2 d\rho_n(s))$. By Lemma 3.1 (taking $\psi(s) := s$), we have

$$\int_{\mathfrak{X} \times \mathbb{R}^+} s d\mathbf{\Pi}_n(x, s) \sim_{\mathbb{P}} v_n(\mathfrak{X}) \times \int_{\mathbb{R}^+} s d\rho_n(s). \tag{29}$$

Hence $|| \boldsymbol{\beta}_n ||_2^2 = O_{\mathbb{P}}(K_n^{-2})$, which tends to 0 by (24). This proves point 1 of Corollary 2. Proving (26) of point 2 is done in a similar way, formally replacing s by s^2 and using (25) instead of (24). This entails

$$\int_{\mathfrak{X} \times \mathbb{R}^+} s^2 d\mathbf{\Pi}_n(x, s) \sim_{\mathbb{P}} v_n(\mathfrak{X}) \times \int_{\mathbb{R}^+} s^2 d\rho_n(s), \tag{30}$$

$$\frac{\int_{\mathfrak{X} \times \mathbb{R}^+} s^4 d\mathbf{\Pi}_n(x, s)}{\Big(\int_{\mathfrak{X} \times \mathbb{R}^+} s^2 d\mathbf{\Pi}_n(x, s) \Big)^2} \to_{\mathbb{P}} 0,$$

and hence Theorem 2 applies. The choice of normalization K_n in (26) is justified because (29) and (30) imply

$$|| \boldsymbol{\beta}_n ||_2^{-2} = \frac{\Big(\int_{\mathbb{R}^+} s d\mathbf{\Pi}_n(x, s) \Big)^2}{\int_{\mathbb{R}^+} s^2 d\mathbf{\Pi}_n(x, s)} \sim_{\mathbb{P}} K_n^2. \tag{31}$$

This concludes the proof of Corollary 2. $\square$

3.3. Stick breaking random measures

Another general class of discrete priors (see, e.g., (Hjort et al., 2010, Chapter 3)) is that of stick breaking priors, which can be represented as in (3), with $\mathbf{Y}_n$ being independent of $\boldsymbol{\beta}_n$, and with $\boldsymbol{\beta}_n$ constructed as follows:

$$\beta_{i,n} := V_{i,n} \prod_{j=1}^{i-1} \big(1 - V_{j,n}\big). \tag{32}$$

Here the $V_{i,n}$ are mutually independent (but not necessarily identically distributed), satisfying

$$\sum_{i=1}^{\infty} - \log \big(1 - \mathbb{E}(V_{i,n})\big) = \infty,$$

to ensure that $\|\, \boldsymbol{\beta}_n \,\|_1 = 1$ with probability 1. The class of stick breaking priors is not included in that of NHCRM, and encompasses discrete priors of interest. A crucial example is the two parameter Poisson-Dirichlet process. More precisley, a Poisson-Dirichlet process $PD(\alpha_n, M_n, \mathbf{P}_n)$ with base probability measure $\mathbf{P}_n$, concentration parameter $M_n > 0$ and shape parameter $\alpha_n \in [0, 1)$ can be represented as a stick breaking random probability measure with $Y_{1,n} \rightsquigarrow \mathbf{P}_n$ and $V_{i,n} \rightsquigarrow Beta(1 - \alpha_n, M_n + i\alpha_n)$ (see, e.g., James (2008)). Except for the case $\alpha_n = 0$, which corresponds to the Dirichlet process, a Poisson-Dirichlet process does not belong to the NHCRM class. Our next result provides a sufficient criterion for sequences of stick breaking random measures to satisfy a Donsker or a Glivenko-Cantelli theorem. As in Corollary 2, the criterion can be divided into a convergence criterion of the underlying measures $\mathbf{P}_n$ and a sufficient condition for the weights to behave as demanded in Theorems 1 and 2. Motivated by the Poisson-Dirichlet process, our proposed criterion can be roughly described as a sufficent set of conditions to ensure that, for some non random sequence $K_n \to \infty$:

1. The number of predominant weights $\beta_{i,n}$ is of order K_n.
2. Those weights behave as if they where i.i.d with common order of magnitude K_n^{-1}.

Due to the particular structure of the stick breaking construction, we propose three conditions (see (34), (35) and (36) below) to ensure that 1 and 2 are satisfied. To heuristically explain why we propose these conditions, let us first assume that, for fixed n, the weight constructing random variables $(V_{i,n})_{i\geq 1}$ are i.i.d. with expectation K_n^{-1}. The stick breaking construction shows a picture of first weights $\beta_{i,n}$ that are not too far from the respective $V_{i,n}$, as long as the correcting factor $\prod_{j\leq i-1}(1 - V_{j,n})$ remains close enough to 1. Roughly speaking, such a closeness is guaranteed as long as no $V_{i,n}$ departs too far away from its expectation. Indeed, when such a departure happens, the unexpectedly small factor $(1 - V_{i_0,n})$ introduces a significant gap between the $\beta_{i,n},\ i \leq i_0 - 1$ and the $\beta_{i,n},\ i \geq i_0$. We make the assumption that $i_0(n, \omega)$ is at least of order K_n

(see condition (34)), ensuring that such a gap does not happen too soon. Conditions (35) and (36) are introduced to handle the case where the independent $V_{i,n}$ fail to be identically distributed, but the consecutive laws of $V_{i+1,n}$ and $V_{i,n}$ are nevertheless similar enough. This is the case for the Poisson-Dirichlet process, as it will be shown in Corollary 4.

Corollary 3. *Let $\mathcal{F}$ be a pointwise separable class of functions satisfying (10), with measurable envelope F. Let Pr_n be a sequence of stick breaking measures, with respective underlying distributions $\mathbf{P}_n$, and with respective sequences of weight-constructing random variables $(V_{i,n})_{i\geq 1}$, $n \geq 1$.*

1. *Assume that F satisfies (5), with $Y_{1,n} \rightsquigarrow \mathbf{P}_n$, and that*
$$\sum_{i\geq 1} \mathbb{E}\big(V_{i,n}^2\big) \prod_{j=1}^{i-1} \mathbb{E}\big((1-V_{j,n})^2\big) \to 0. \tag{33}$$
Then we have
$$\sup_{f\in\mathcal{F}} \Big| Pr_n(f) - \mathbf{P}_n(f) \Big| \to_{\mathbb{P}} 0.$$
2. *Assume that F satisfies (9) and that $\mathbf{P}_n$ converges to $\mathbf{P}_0$ in the sense of (13). Assume that, for some determistic sequence $K_n \to \infty$, writing $Z_{i,n} := K_n V_{i,n}$, the following conditions are satisfied:*
$$\forall \epsilon > 0,\ \exists \delta > 0, \overline{\lim}_{n\to\infty} \mathbb{P}\Big(\max_{i\leq [\delta K_n]} Z_{i,n} \geq \log(K_n) \Big) \leq \epsilon, \tag{34}$$
$$\overline{\lim}_{n\to\infty} \tfrac{1}{K_n} \sum_{i\geq 1} \mathbb{E}\Big(Z_{i,n}^4 \Big) \prod_{j=1}^{i-1} \mathbb{E}\bigg(\Big(1 - \tfrac{Z_{i,n}}{K_n}\Big)^4 \bigg) < \infty, \tag{35}$$
$$\forall \delta > 0, \underline{\lim}_{n\to\infty} \mathbb{P}\, K_n^{-1} \sum_{i=1}^{[\delta K_n]} Z_{i,n}^2 > 0. \tag{36}$$
Then the sequence of processes
$$f \to \frac{1}{\|\boldsymbol{\beta}_n\|_2} \Big(Pr_n(f) - \mathbf{P}_n(f) \Big), \tag{37}$$
with $\boldsymbol{\beta}_n$ defined as in (32), converges weakly to $\mathbb{G}_{\mathbf{P}_0}$ in $\ell^\infty(\mathcal{F})$. If, in addition there exists $\delta > 0$ such that
$$K_n^{1+\delta} \sup_{f\in\mathcal{F}} \Big| \mathbf{P}_n(f) - \mathbf{P}_0(f) \Big| \to 0,$$
as a deterministic sequence, then the centering process $\mathbf{P}_n(\cdot)$ can be replaced by $\mathbf{P}_0(\cdot)$ in (37).

Proof of Corollary 3. Point 1 is proved by noticing that (33) implies $\|\boldsymbol{\beta}_n\|_2 \to_{\mathbb{P}} 0$ by Markov's inequality. To prove point 2, we need to show that $\|\boldsymbol{\beta}_n\|_4 / \|\boldsymbol{\beta}_n\|_2 \to_{\mathbb{P}} 0$. By (35), we already have $\|\boldsymbol{\beta}_n\|_4^4 = O_{\mathbb{P}}(K_n^{-3})$. Hence it is largely sufficient to prove that
$$\exists \delta_0,\ \forall \delta \leq \delta_0,\ K_n^{-1-3\delta} = o_{\mathbb{P}}\Big(\|\boldsymbol{\beta}_n\|_2^2 \Big). \tag{38}$$

For fixed $\delta > 0$, $\delta' > 0$ and $n \geq 1$, write

$$A_{n,\delta,\delta'} := \Big\{ \max_{i \leq [\delta n]} Z_{i,n} \leq \log(K_n) \Big\} \cap \Big\{ K_n^{-1} \sum_{i=1}^{[\delta K_n]} Z_{i,n}^2 \geq \delta' \Big\}.$$

Note that, by (34) and (36), we have

$$\lim_{\delta \to 0} \lim_{\delta' \to 0} \varliminf_{n \to \infty} \mathbb{P}\Big(A_{n,\delta,\delta'} \Big) = 1. \tag{39}$$

Introducing the equalities $Z_{i,n} = K_n V_{i,n}$ in the expression of $|| \boldsymbol{\beta}_n ||_2$ entails:

$$\begin{aligned} K_n^2 \, || \boldsymbol{\beta}_n ||_2^2 \, \mathbb{1}_{A_{n,\delta,\delta'}} &= \sum_{i=1}^{+\infty} Z_{i,n}^2 \prod_{j=1}^{i-1} \Big(1 - \frac{Z_{j,n}}{K_n} \Big)^2 \mathbb{1}_{A_{n,\delta,\delta'}} \\ &\geq \Big(1 - \frac{\log(K_n)}{K_n} \Big)^{2\delta K_n} \sum_{i=1}^{[\delta K_n]} Z_{i,n}^2 \mathbb{1}_{A_{n,\delta,\delta'}} \\ &\geq \exp\Bigg(-2\delta \log(K_n) - \frac{\delta \log^2(K_n)}{K_n} \Bigg) \delta' K_n \mathbb{1}_{A_{n,\delta,\delta'}} \\ &\geq \delta' K_n^{1-3\delta} \mathbb{1}_{A_{n,\delta,\delta'}}. \end{aligned}$$

Now choosing $\delta > 0$ and $\delta' > 0$ small enough and using (39) proves (38). The last statement of Corollary 3 is also a consequence of (38). $\square$

Remark 3.4. Given a prior $DP(\alpha, M, \mathbf{P})$ a version of $Post_n$ is the map for which, given $(x_1, \ldots, x_n)$:

$$Post_n(x_1, \ldots, x_n) =_d \; R_k \times DP(\alpha, M + k\alpha, \mathbf{P}) + (1 - R_k) \sum_{i=1}^{k} \Delta_j x_i^*, \tag{40}$$

with k being the number of distinct obervations $(x_1^*, \ldots, x_k^*)$ with respective numbers of ties $(n_1, \ldots, n_k)$, with $(\Delta_1, \ldots, \Delta_k) \rightsquigarrow Dirichlet(n_1 - \alpha, \ldots, n_k - \alpha)$, with $R_k \rightsquigarrow Beta(M + k\alpha, n - k\alpha)$, and with all the three random variables/processes in (40) being mutually independent. Hence, Glivenko-Cantelli and Donsker theorems for sequences Poisson-Dirichlet processes are definitely a valuable tool to derive posterior first/second order consistency under $|| \cdot ||_{\mathcal{F}}$.

By a use of Corollary 3, we can deduce asymptotics for sequences of Poisson-Dirichlet processes in strong topology (see Corollary 4 below). It has to be mentioned that, among other results, (James, 2008, Section 4) did also prove such kind of results, when $\alpha_n = \alpha$ and $\mathbf{P}_n = \mathbf{P}$ are constant in n, and $K_n + 1 = M + \alpha n$, and hence extending the pioneering result of Lo (1983). James took another route, by expressing $PD(\alpha, M + \alpha n, \mathbf{P})$ as a randomly weighted convex combination of $n + 1$ independent Poisson-Dirichlet processes. Then, he used standard methods in empirical processes theory: convergence of sums of independent measurelike processes (see, e.g., (Van der Vaart and Wellner, 1996, Chapter 2.11) for a definition), and a multiplier central limit theorem. Finally he used those asymptotic results together with the theory of exchangeable bootstrap, to prove, under the $|| \cdot ||_{\mathcal{F}}$ topology:

- A first and a second order posterior consistency result for the Dirichlet process prior.
- A first and a second order posterior consistency result for the Poisson-Dirichlet process prior, when $\mathbf{P}_0$ is discrete.
- A first and second order posterior convergence result, for the Poisson-Dirichlet process prior, to $\alpha\mathbf{P} + (1-\alpha)\mathbf{P}_0$, when $\alpha \in (0,1)$ and $\mathbf{P}_0$ is continuous (hence an unconsistency result).

All his results were restricted to the case where $\mathcal{F}$ is a V-C subgraph class of functions (see, e.g., (Van der Vaart and Wellner, 1996, p. 141)), which is an assumption stronger than (10). However, a close look at his arguments leads to the conclusion that all his results still hold when the V-C subgraph assumption is relaxed to (10). Our next result, which generalizes Theorems 4.1 and 4.2 of James (2008), can also be seen as an alternate route to prove all the three above-mentioned posterior convergence results.

Corollary 4. *Let K_n and α_n be two sequences of non negative real numbers such that $K_n \to \infty$ and $\alpha_n \in [0, 1-\epsilon)$ for some $\epsilon > 0$ not depending upon n. Let $\mathbf{P}_n$ be a sequence of probability measures on $\mathfrak{X}$, and let $\mathcal{F}$ be a pointwise separable class of functions with measurable envelope F satisfying (10). Let Pr_n be a sequence of random probability measures having distributions $PD(\alpha_n, K_n + 1, \mathbf{P}_n)$.*

1. *Assume that F satisfies (5), with $Y_{1,n} \rightsquigarrow \mathbf{P}_n$. Then we have*

$$\sup_{f \in \mathcal{F}} \Big| Pr_n(f) - \mathbf{P}_n(f) \Big| \to_{\mathbb{P}} 0.$$

2. *Assume that F satisfies (9) and that $\mathbf{P}_n$ converges to $\mathbf{P}_0$ in the sense of (13). Then the sequence of processes*

$$f \to \frac{1}{|| \boldsymbol{\beta}_n ||_2} \Big(Pr_n(f) - \mathbf{P}_n(f) \Big), \tag{41}$$

with $\boldsymbol{\beta}_n$ defined as in (32), converges weakly to $\mathbb{G}_{\mathbf{P}_0}$ in $\ell^\infty(\mathcal{F})$. If, in addition, there exists $\delta > 0$ such that

$$K_n^{1+\delta} \sup_{f \in \mathcal{F}} \Big| \mathbf{P}_n(f) - \mathbf{P}_0(f) \Big| \to 0,$$

as a deterministic sequence, then the centering process $\mathbf{P}_n(\cdot)$ can be replaced by $\mathbf{P}_0(\cdot)$ in (41).

The corresponding proof relies on cumbersome calculations based on the usual properties of the Beta distribution. It is postponed to §4.

3.4. A Bernstein-von Mises phenomenon under $||\cdot||_{\mathcal{F}}$ for sequences of Dirichlet processes priors

Using the results of the preceding subsections, it is now possible to extend James' results to a Berstein-von Mises theorem for posterior distributions of *sequences* of Dirichlet processes priors $PD(0, M_n, \mathbf{P}_n)$ with varying baseline probability measures $\mathbf{P}_n$ and concentration parameters M_n, as long as the latters are negligible in front of the sample information n. Such an approach of using priors distributions that change with the sample size is already present in Castillo and Nickl (2014). Note that, for a Dirichlet process $PD(0, M_n, \mathbf{P}_n)$, the (natural version of) the posterior distribution given a sample has the following simpler expression (Ferguson (1973)):

$$Post_n(x_1, \dots, x_n) := PD\Big(0, M_n + n, \theta_n \mathbf{P}_n + (1-\theta_n)\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\Big), \text{ where} \quad (42)$$

$$\theta_n := \frac{M_n}{M_n + n}.$$

Corollary 5. *Let M_n be a sequence of positive numbers which is $o(n^{1/2})$, let $\mathbf{P}_n$ be a sequence of probability measures on $\mathfrak{X}$, and let $PD(0, M_n, \mathbf{P}_n)$ be the associated sequence of Dirichlet processes. For fixed n, define $Post_n(x_1, \dots, x_n)$ as in (42). Let $\mathcal{F}$ be a pointwise separable class of functions with measurable envelope F satisfying (10) and*

$$\overline{\lim_{n\to\infty}}\, \mathbf{P}_n(F^2) < +\infty. \quad (43)$$

Let $\mathbf{P}_0$ be a probability measure for which F is square integrable. Then for $\mathbf{P}_0^{\otimes \mathbb{N}}$ almost every sequence $(x_i)_{i\geq 1}$ we have

$$\sqrt{n}\bigg(Post_n(x_1, \dots, x_n) - \frac{1}{n}\sum_{i=1}^n \delta_{x_i}\bigg) \to_{\mathcal{L}} \mathbb{G}_{\mathbf{P}_0}, \ \textit{in } \ell^\infty(\mathcal{F}). \quad (44)$$

Proof. Write

$$\mathcal{G} := \Big\{(f-g)^2,\ (f,g) \in \mathcal{F}^2, \Big\} \cup \Big\{f^2,\ f \in \mathcal{F}\Big\} \cup \mathcal{F},$$

$$A := \Big\{(x_i)_{i\geq 1},\ \sqrt{n}\sup_{f\in\mathcal{G}}\Big|\theta_n\Big(\mathbf{P}_n(f) - \frac{1}{n}\sum_{i=1}^n f(x_i)\Big)\Big| \to 0\Big\},$$

$$B := \Big\{(x_i)_{i\geq 1},\ \sup_{f\in\mathcal{G}}\Big|\frac{1}{n}\sum_{i=1}^n f(x_i) - \mathbf{P}_0(f)\Big| \to 0\Big\},$$

$$C := \Big\{(x_i)_{i\geq 1},\ \forall M \in \mathbb{N},\ \frac{1}{n}\sum_{i=1}^n F^2(x_i)\mathbb{1}_{\{F(x_i)\geq M\}} \to \mathbf{P}_0\big(F^2\mathbb{1}_{\{F\geq M\}}\big)\Big\}.$$

Since $\theta_n = o(n^{-1/2})$, since $\mathbf{P}_0(F^2) < \infty$ and by (43), we have $\mathbf{P}_0^{\otimes\mathbb{N}}(A) = 1$. Moreover, by (10) combined with standard covering numbers arguments, the class $\mathcal{G}$ is $\mathbf{P}_0$-Glivenko-Cantelli, and hence $\mathbf{P}_0^{\otimes\mathbb{N}}(B) = 1$. Finally $\mathbf{P}_0^{\otimes\mathbb{N}}(C) = 1$ by the strong law of large numbers. Now, for every $(x_i)_{i\geq 1} \in A \cap B \cap C$ the sequence $(Post_n(x_1, \ldots, x_n))_{n\geq 1}$ is a sequence of NHCRM (see, e.g. Hjort et al. (2010)), with base measures

$$\begin{aligned} d\rho_n(s) &:= \frac{1}{s}\exp(-s)\mathbb{1}_{(0,+\infty)}(s)ds \\ v_n &:= (M_n + n) \times \alpha_n, \text{ with} \\ \alpha_n &:= \theta_n \mathbf{P}_n + (1-\theta_n)\frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}. \end{aligned}$$

Note that, by elementary analysis, we have (using the notations of Corollary 2) $K_n^2 \sim (M_n + n) \sim n$. Moreover, by the triangle inequality we have

$$\lim_{M\to\infty} \overline{\lim_{n\to\infty}}\, \alpha_n\Big(F^2\mathbb{1}_{\{F\geq M\}}\Big) = 0. \tag{45}$$

Hence, the sequence $(Post_n(x_1, \ldots, x_n))_{n\geq 1}$ satisfies all the required conditions to use (26) in Corollary 2, with the formal change of $\mathbf{P}_n$ into α_n (note that condition (13) is guaranteed by the definition of A and B together with the triangle inequality for $||\cdot||_{\mathcal{F}}$). We hence have

$$\sqrt{n}\Big(Post_n(x_1, \ldots, x_n) - \alpha_n\Big) \to_{\mathcal{L}} \mathbb{G}_{\mathbf{P}_0}, \text{ in } \ell^\infty(\mathcal{F}). \tag{46}$$

Now the substitution of α_n to $n^{-1}\sum_{i=1}^{n}\delta_{x_i}$ is possible by definition of A. □

3.5. A discussion on Gibbs-type priors

A very important class of discrete priors Pr is that of Gibbs-type priors with parameter $\sigma \in (-\infty, 1)$, for which

$$Expect_n(x_1, \ldots, x_n) := \frac{V_{n+1,k+1}}{V_{n,k}}P^* + \frac{V_{n+1,k}}{V_{n,k}}\sum_{i=1}^{k}(n_i - \sigma)\delta_{x_i^*},$$

with P^* being the mean value of Pr, and with $V_{n,k}$ satisfying the recursion

$$V_{n,k} := (n - \sigma k)V_{n+1,k} + V_{n+1,k+1}.$$

De Blasi et al. (2013) did prove a first order posterior consistency result for $Post_n(x_1, \ldots, x_n)$ under the weak topology, and hence it would be of interest to strenghten their result under $||\cdot||_{\mathcal{F}}$. Their proofs, however, do not relies on any explicit representation of $Post_n(x_1, \ldots, x_n)$, which are not known in generality. Instead, they successfully take advantage of the exchangeability of $(X_1, \ldots, X_n)$ (in the Bayesian model) to obtain suitable bounds for the posterior variances of $Pr(A)$, given $(x_1, \ldots, x_n)$, for fixed $A \in \mathcal{X}$. Since our results are limited to posterior distributions that are explicit, we doubt that Theorems 1 and 2 could be useful toward that aim of generalization.

4. Proofs

Proof of Theorem 1. The proof is an adaptation of Van der Vaart and Wellner (1996, p. 123). We first prove the Borel measurability of the $\|\ G_n^k\ \|_{\mathcal{F}}$. Since $\mathbb{E}(F(Y_{1,n})) < \infty$, for all n, there exists Ω_0 fulfilling $\mathbb{P}(\Omega_0) = 1$ and such that $\mathbb{E}(F(Y_{1,n}) \mid \boldsymbol{\beta}_n)(\omega) < \infty$ for all $n \geq 1$ and $\omega \in \Omega_0$. Applying, for fixed $\omega \in \Omega_0$, the dominated convergence theorem with underlying measure $\mathbf{P}_{n,\boldsymbol{\beta}_n}(\omega)$, we conclude that the pointwise dense class $\mathcal{F}_0$ is also dense in $(\mathcal{F}, \|\cdot\|_{\mathbf{P}_{n,\boldsymbol{\beta}_n(\omega)}})$. This ensures the Borel measurability of $\|\ G_n^k\ \|_{\mathcal{F}}$ for all $n \geq 1$ and $k \geq 1$. We now fix $M > 0$, and we introduce an independent, identically distributed Rademacher sequence $(\epsilon_i)_{i\geq 1}$ (namely $\mathbb{P}(\epsilon_i = 1) = \mathbb{P}(\epsilon_i = -1) = 1/2$), which is independent of both $(\boldsymbol{\beta}_n)_{n\geq 1}$ and $(\mathbf{Y}_n)_{n\geq 1}$. For fixed $n \geq 1$ and $k \geq 1$, we apply Lemma 2.3.1 in Van der Vaart and Wellner (1996, p. 108) conditionally to $\boldsymbol{\beta}_n$, for $\sum_{i=1}^k \beta_{i,n}\delta_{Y_{i,n}}$. Because $\mathcal{F}_M$ is uniformly bounded by M, we can take the limit in $(\mathcal{E}_{\mathcal{F}_M,1}, |||\cdot|||_{\mathcal{F}_M,1})$ of the latter inequality, as $k \to \infty$, to obtain

$$\mathbb{E}\Big(\ \|\ G_n\ \|_{\mathcal{F}_M}\ \Big|\boldsymbol{\beta}_n\Big) \leq 2\mathbb{E}\Big(\sup_{f\in\mathcal{F}_M}\ |\sum_{i\geq 1}\beta_{i,n}\epsilon_i f(Y_{i,n})\ |\ \ \Big|\boldsymbol{\beta}_n\Big),\ \text{with probability 1.}$$

Now fix $\epsilon > 0$, $\boldsymbol{p} = (p_i)_{i\geq 1} \in \mathbb{S}$ and $\mathbf{y} = (y_i)_{i\geq 1} \in \mathfrak{X}^{\mathbb{N}}$. By (6) we have,

$$\mathbb{E}\bigg(\sup_{f\in\mathcal{F}_M}\ |\sum_{i\geq 1} p_i\epsilon_i f(y_i)\ |\ \bigg) \leq \epsilon + \mathbb{E}\bigg(\max_{f\in\mathcal{F}_{\boldsymbol{p},\mathbf{y}}}\ |\sum_{i\geq 1} p_i\epsilon_i f(y_i)\ |\ \bigg), \tag{47}$$

where $\mathcal{F}_{\boldsymbol{p},\mathbf{y}} \subset \mathcal{F}_M$ has cardinality less than $N(\epsilon, \mathcal{F}_M, \overline{P}(\boldsymbol{p},\mathbf{y}))$ (recall (7)). The maximal inequality for sub-gaussian random variables (see, e.g., Van der Vaart and Wellner (1996, Lemma 2.2.2, p. 96)) yields, almost surely:

$$\begin{aligned}
&\mathbb{E}\Big(\max_{f\in\mathcal{F}_{\boldsymbol{p},\mathbf{y}}}\ |\sum_{i\geq 1} p_i\epsilon_i f(y_i)\ |\ \Big)\\
&\leq \sqrt{1+\log\Big(N\big(\epsilon,\mathcal{F}_M,\overline{P}(\boldsymbol{p},\mathbf{y})\big)\Big)}\sqrt{\sum_{i\geq 1} p_i^2 F^2(y_i)\mathbb{1}_{\{F(y_i)\leq M\}}}\\
&\leq \sqrt{1+\log\Big(N\big(\epsilon,\mathcal{F}_M,\overline{P}(\boldsymbol{p},\mathbf{y})\big)\Big)}\times \|\ \boldsymbol{p}\ \|_2 \times M,\ \text{from where, integrating (47):}\\
&\mathbb{E}\bigg(\sup_{f\in\mathcal{F}_M}\ |\sum_{i\geq 1}\beta_{i,n}\epsilon_i f(Y_{i,n})\ |\ \bigg|\boldsymbol{\beta}_n\bigg) \to_{\mathbb{P}} 0,\ \text{because}\ \ \|\ \boldsymbol{\beta}_n\ \|_2 \to_{\mathbb{P}} 0,
\end{aligned}$$

and that convergence in probability also holds in expectation, since the involved random variables take values in $[-M, M]$ with probability one (see, e.g., Williams (1991, p. 130)). The proof is then concluded by noticing that

$$\mathbb{E}\bigg(\sup_{f\in\mathcal{F}}\Big|G_n(f) - G_n\big(f\mathbb{1}_{\{F\leq M\}}\big)\Big|\bigg) \leq 2\mathbb{E}\Big(F(Y_{1,n})\mathbb{1}_{\{F(Y_{1,n})\geq M\}}\Big), \tag{48}$$

which can be rendered arbitrarily small by (5). $\square$

Proof of Theorem 2. Invoking similar arguments as in the beginning of the proof of Theorem 1, there exist Ω_0 having probability one, such that, for each $\omega \in \Omega_0$, $\mathcal{F}_0$ is dense in $(\mathcal{F}, ||\cdot||_{\mathbf{P}_{n,\boldsymbol{\beta}_n(\omega)},2})$. Moreover, given $\boldsymbol{\beta}_n$, W_n is Gaussian with intrinsic semimetric bounded by $||\cdot||_{\mathbf{P}_{n,\boldsymbol{\beta}_n},2}$. Hence (see, e.g., Van der Vaart and Wellner (1996, p. 100)) we have, almost surely (K denoting a universal constant):

$$\begin{aligned}
&\mathbb{E}\Big(\sup_{(f_1,f_2)\in\mathcal{F}_0^2} \mid W_n(f_1)-W_n(f_2)\mid \mathbb{1}_{\{||f_1-f_2||_{\mathbf{P}_{n,\boldsymbol{\beta}_n},2}\leq\delta\}}\Big)\\
&=\mathbb{E}\Big(\sup_{(f_1,f_2)\in\mathcal{F}_0^2,\ ||f_1-f_2||_{\mathbf{P}_{n,\boldsymbol{\beta}_n},2}\leq\delta} \mid W_n(f_1)-W_n(f_2)\mid\Big)\\
&=\mathbb{E}\Big(\mathbb{E}\Big(\sup_{(f_1,f_2)\in\mathcal{F}_0^2,\ ||f_1-f_2||_{\mathbf{P}_{n,\boldsymbol{\beta}_n},2}\leq\delta} \mid W_n(f_1)-W_n(f_2)\mid\Big)\Big|\boldsymbol{\beta}_n\Big)\\
&\leq K\mathbb{E}^*\Big(\int_0^\delta\sqrt{\log\big(N(\epsilon,\mathcal{F}_0,||\cdot||_{\mathbf{P}_{n,\boldsymbol{\beta}_n},2})\big)}d\epsilon\Big)\\
&\leq K\int_0^\delta\sqrt{\log\sup_{Q\ prob.}\big(N(\epsilon\,||\,F\,||_{Q,2},\mathcal{F}_0,||\cdot||_{Q,2})\big)}d\epsilon,\ \text{since } F\geq 1.
\end{aligned}$$

Hence, for any $k\geq 1$, that bound can be rendered less than 2^{-2k} for a suitable choice of $\delta_k>0$. Markov's inequality and the Borel-Cantelli lemma yield

$$\mathbb{P}\Big(\sup_{(f_1,f_2)\in\mathcal{F}_0,\ ||f_1-f_2||_{\mathbf{P}_{n,\boldsymbol{\beta}_n},2}\leq\delta_k} \mid W_n(f_1)-W_n(f_2)\mid\geq 2^{-k} \text{ infinitely often}\Big)=0.$$

This entails the existence of $\Omega_1\subset\Omega_0$ having probability one, on which the process W_n is $||\cdot||_{\mathbf{P}_{n,\boldsymbol{\beta}_n},2}$ uniformly continuous on $\mathcal{F}_0$. Combine this with the density (for fixed ω) of $\mathcal{F}_0$ in $(\mathcal{F},||\cdot||_{\mathbf{P}_{n,\boldsymbol{\beta}_n},2})$ to conclude that, for any (possibly infinite) $\delta>0$, we have, on Ω_1:

$$\sup_{\substack{(f_1,f_2)\in\mathcal{F}^2,\\ ||f_1-f_2||_{\mathbf{P}_{n,\boldsymbol{\beta}_n},2}<\delta}} \mid W_n(f_1)-W_n(f_2)\mid=\sup_{\substack{(f_1,f_2)\in\mathcal{F}_0^2,\\ ||f_1-f_2||_{\mathbf{P}_{n,\boldsymbol{\beta}_n},2}<\delta}} \mid W_n(f_1)-W_n(f_2)\mid, \tag{49}$$

which is almost surely finite. The proof of Theorem 2 is divided in two steps.

Step 1: Convergence of the marginals

For $p\geq 1$, write $|\cdot|_p$ for the usual Euclidean norm on $\mathbb{R}^p$. Also write d_{LP} for the Levy-Prokhorov distance between (Borel) probability measure on $\mathbb{R}^p$, generated by $|\cdot|_p$, namely, for two probability measures P and Q

$$d_{LP}(P,Q):=\inf\big\{\lambda>0,\ \Pi(P,Q,\lambda)\leq\lambda\big\},\ \text{where we write} \tag{50}$$

$$\Pi(P,Q,\lambda):=\sup_{A\ Borel}\max\big\{P(A)-Q(A^\lambda),Q(A)-P(A^\lambda)\big\}, \text{for } \lambda>0, \tag{51}$$

and where, for a set A, we defined

$$A^\lambda:=\Big\{x\in\mathbb{R}^p,\ \inf_{y\in A}\mid x-y\mid_p<\lambda\Big\}. \tag{52}$$

Our first proposition controls the distance of the marginals of G_n and those of W_n. Given $\psi \in \ell^\infty(\mathcal{F})$, $\psi(\mathbf{f})$ stands for $(\psi(f_1), \ldots, \psi(f_p))$.

Proposition 4.1. *For each $p \geq 1$ and $\mathbf{f} \in \mathbb{R}^p$, we have*

$$d_{LP}\big(G_n(\mathbf{f}), W_n(\mathbf{f})\big) \to 0.$$

Proof. For $\delta > 0$, write

$$Z_{1,n}(\delta) := \sum_{i \geq 1} \beta_{i,n} \bigg[\mathbb{1}_{\left\{\beta_{i,n} F(Y_{i,n}) > \delta\right\}} \mathbf{f}(Y_{i,n}) - \mathbb{E}\Big(\mathbb{1}_{\left\{\beta_{i,n} F(Y_{i,n}) > \delta\right\}} \mathbf{f}(Y_{i,n}) \mid \boldsymbol{\beta}_n \Big) \bigg],$$

$$Z_{2,n}(\delta) := \sum_{i \geq 1} \beta_{i,n} \bigg[\mathbb{1}_{\left\{\beta_{i,n} F(Y_{i,n}) \leq \delta\right\}} \mathbf{f}(Y_{i,n}) - \mathbb{E}\Big(\mathbb{1}_{\left\{\beta_{i,n} F(Y_{i,n}) \leq \delta\right\}} \mathbf{f}(Y_{i,n}) \mid \boldsymbol{\beta}_n \Big) \bigg].$$

Now, for $\boldsymbol{p} \in \mathbb{S}$, denote by $\mathbf{P}^{(1)}_{n,\boldsymbol{p}}(\delta)$ (resp. $\mathbf{P}^{(2)}_{n,\boldsymbol{p}}(\delta)$) the law of $Z_{1,n}(\delta)$ (resp. $Z_{2,n}(\delta)$) conditionally to $\boldsymbol{\beta}_n = \boldsymbol{p}$, and $\mathbf{Q}^{(1)}_{n,\boldsymbol{p}}(\delta)$ (resp. $\mathbf{Q}^{(2)}_{n,\boldsymbol{p}}(\delta)$) the Gaussian analogue of $\mathbf{P}^{(1)}_{n,\boldsymbol{p}}(\delta)$ (resp. $\mathbf{P}^{(2)}_{n,\boldsymbol{p}}(\delta)$). We will also write $V_{1,n}(\delta)$ (resp. $V_{2,n}(\delta)$) for a (generic) random vector, for which the law is the mixture of $(\mathbf{P}^{(1)}_{n,\boldsymbol{p}}(\delta))_{\boldsymbol{p}\in\mathbb{S}}$ (resp. $(\mathbf{P}^{(2)}_{n,\boldsymbol{p}}(\delta))_{\boldsymbol{p}\in\mathbb{S}}$ by $\boldsymbol{\beta}_n$. The proof is then divided into two separate lemmas.

Lemma 4.1. *For any $\delta > 0$ we have $Z_{1,n}(\delta) \to_{\mathcal{L}} 0$ and $V_{1,n}(\delta) \to_{\mathcal{L}} 0$.*

Proof. We can assume without loss of generality that $p = 1$. Fix δ and $\epsilon > 0$, and choose (by (9)) $\eta > 0$ small enough to have

$$\overline{\lim_{n\to\infty}}\, \mathbb{E}\Big(F(Y_{1,n}) \mathbb{1}_{\left\{F(Y_{1,n}) \geq \delta/\eta\right\}} \Big) \leq \epsilon^3. \tag{53}$$

The union bound entails, for fixed n:

$$\mathbb{P}\Big(\mid Z_{1,n}(\delta) \mid_p > \epsilon \Big) \leq \mathbb{P}\Big(\mid\mid \boldsymbol{\beta}_n \mid\mid_4 \geq \eta \Big) + \mathbb{P}\Big(\mid\mid \boldsymbol{\beta}_n \mid\mid_4 \leq \eta,\ \mid Z_{1,n}(\delta) \mid_p > \epsilon \Big).$$

The first term of the RHS of the preceding inequality tends to 0 by assumption. Applying Tchebychev's inequality conditionally to $\boldsymbol{\beta}_n$, yields:

$$\begin{aligned}
&\mathbb{P}\Big(\mid\mid \boldsymbol{\beta}_n \mid\mid_4 \leq \eta,\ \mid Z_{1,n}(\delta) \mid_p > \epsilon \Big) \\
\leq\ & \frac{1}{\epsilon^2} \mathbb{E}\bigg(\mathbb{1}_{\left\{\mid\mid \boldsymbol{\beta}_n \mid\mid_4 \leq \eta\right\}} \sum_{i \geq 1} \beta_{i,n}^2 \mathrm{Var}\Big(\mathbf{f}(Y_{i,n}) \mathbb{1}_{\{\beta_{i,n} F(Y_{i,n}) > \delta\}} \mid \boldsymbol{\beta}_n \Big) \bigg) \\
\leq\ & \frac{1}{\epsilon^2} \mathbb{E}\bigg(\mathbb{1}_{\left\{\mid\mid \boldsymbol{\beta}_n \mid\mid_4 \leq \eta\right\}} \sum_{i \geq 1} \beta_{i,n}^2 \mathbb{E}\Big(F^2(Y_{1,n}) \mathbb{1}_{\{F(Y_{1,n}) \geq \delta/\eta\}} \Big) \bigg) \\
\leq\ & \epsilon,
\end{aligned}$$

for all n large enough, by (53) and because $\mid\mid \boldsymbol{\beta}_n \mid\mid_2^2 = 1$ with probability one. This proves the first assertion of Lemma 4.1. The second assertion is treated in a very similar way. We omit details. ∎

Lemma 4.2. *There exists $c > 0$ such that, for any $p \geq 1$, $\mathbf{f} \in \mathcal{F}^p$, $\delta > 0$ and $n \geq 1$, we have (as a comparison of real functions on $[0, \infty[$):*

$$\sup_{\boldsymbol{p} \in \mathbb{S}} \Pi\Big(\mathbf{P}^{(2)}_{n,\boldsymbol{p}}(\delta), \mathbf{Q}^{(2)}_{n,\boldsymbol{p}}(\delta), \bullet\Big) \leq cp^{5/2} \exp\Big(-\frac{\bullet}{2cp^{5/2}\delta}\Big), \textit{which implies} \tag{54}$$

$$\Pi\Big(Z_{2,n}(\delta), V_{2,n}(\delta), \bullet\Big) \leq cp^{5/2} \exp\Big(-\frac{\bullet}{2cp^{5/2}\delta}\Big), \textit{which entails} \tag{55}$$

$$d_{LP}\big(Z_{2,n}(\delta), V_{2,n}(\delta)\big) \leq cp^{5/2} e^{-1} 2\delta(\mid \ln(2\delta) \mid +1). \tag{56}$$

Proof. Fix $p \geq 1$. Since Theorem 1.1 in Zaitsev (1987) provides bounds that do not depend on the power of convolutions, as soon as the latter are properly defined as limits under weak convergence. Hence, for fixed $\boldsymbol{p} \in \mathbb{S}$ we apply the second part of Theorem 1.1 in Zaitsev (1987), with $\tau = 2\delta$, for which the upper bound does not depend on $\boldsymbol{p}$. This proves (54), then (55) by integration, then (56) by definition of d_{LP}. □

Now a combination of Lemmas 4.1 and 4.2 concludes the proof of Proposition 4.1. □

Step 2: Asymptotic equicontinuity

We now turn on to proving that both $(W_n)_{n \geq 1}$ and $(G_n)_{n \geq 1}$ are asymptotically equicontinuous.

Proposition 4.2. *We have*

$$\forall \epsilon > 0, \lim_{\delta \to 0} \overline{\lim_{n \to \infty}}^{\mathbb{P}^*} \sup_{(f_1, f_2) \in \mathcal{F}^2,\ \rho_0(f_1, f_2) < \delta} \mid G_n(f_1) - G_n(f_2) \mid = 0,$$

$$\forall \epsilon > 0, \lim_{\delta \to 0} \overline{\lim_{n \to \infty}}^{\mathbb{P}^*} \sup_{(f_1, f_2) \in \mathcal{F}^2,\ \rho_0(f_1, f_2) < \delta} \mid W_n(f_1) - W_n(f_2) \mid = 0.$$

Proof. First, from assumption (11), the proof boils down to showing that

$$\forall \epsilon > 0, \lim_{\delta \to 0} \overline{\lim_{n \to \infty}}^{\mathbb{P}} \sup_{(f_1, f_2) \in \mathcal{F}^2,\ ||f_1 - f_2||_{\mathbf{P}_{n,\boldsymbol{\beta}_n},2} < \delta} \mid G_n(f_1) - G_n(f_2) \mid = 0, \tag{57}$$

$$\forall \epsilon > 0, \lim_{\delta \to 0} \overline{\lim_{n \to \infty}}^{\mathbb{P}} \sup_{(f_1, f_2) \in \mathcal{F}^2,\ ||f_1 - f_2||_{\mathbf{P}_{n,\boldsymbol{\beta}_n},2} < \delta} \mid W_n(f_1) - W_n(f_2) \mid = 0, \tag{58}$$

and that the involved maps are measurable, which will justify the use of $\mathbb{P}$ instead of $\mathbb{P}^*$. The required measurability of (58) has already been proved through (49). Moreover, since $\sum_{i \geq 1} \beta_{i,n} F(Y_{i,n}) < \infty$ almost surely for all n, and by a simple truncation argument, we see that there exists Ω_2 having probability 1, such that, for each $\omega \in \Omega_2$, the sets

$$\mathcal{F}_{n,\boldsymbol{\beta}_n,\delta}(\omega) := \big\{f_1 - f_2,\ (f_1, f_2) \in \mathcal{F}^2,\ \mid\mid f_1 - f_2 \mid\mid_{\mathbf{P}_{n,\boldsymbol{\beta}_n(\omega)},2} < \delta\big\},\ \delta > 0$$

are open for the topology $\mathfrak{O}$ generated by the maps $\{f \to f(x),\ x \in \mathfrak{X}\}$. This proves the measurability requirement of (57), since G_n is $\mathfrak{O}$- continuous for $\omega \in \Omega_0$ (here, we take the set Ω_0 which was exhibited in the beginning of the proof of Theorem 2). We will now complete the proof of (57). Fix $\epsilon > 0$. Conditioning by $\boldsymbol{\beta}_n$ and using symmetrization as in the proof of Theorem 1, we have, almost surely:

$$\begin{aligned}
&\mathbb{E}\Big(\mid\mid G_n \mid\mid_{\mathcal{F}_{n,\boldsymbol{\beta}_n,\delta}} \ \Big| \boldsymbol{\beta}_n \Big) \\
&\leq 2\mathbb{E}\Big(\sup_{f \in \mathcal{F}_{n,\boldsymbol{\beta}_n,\delta}} \mid \sum_{i \geq 1} \beta_{i,n}\epsilon_i f\big(Y_{i,n}\big) \mid \ \Big| \boldsymbol{\beta}_n \Big) \\
&\leq 2C \int_0^{\theta_n} \sqrt{\log\big(\epsilon \mid\mid F \mid\mid_{\mathbf{P}_{n,\boldsymbol{\beta}_n},2}, \mathcal{F}_{n,\boldsymbol{\beta}_n,\delta}, \mid\mid \cdot \mid\mid_{\mathbf{P}_{n,\boldsymbol{\beta}_n},2}\big)}\, d\epsilon, \ \text{where} \qquad (59) \\
\theta_n^2 &:= \frac{\sup_{f \in \mathcal{F}_{n,\boldsymbol{\beta}_n,\delta}} \sum_{i \geq 1} \beta_{i,n}^2 f^2(Y_{i,n})}{\mathbb{E}\Big(F^2(Y_{1,n}) \mid \boldsymbol{\beta}_n\Big)}.
\end{aligned}$$

Since the integrand in (59) is bounded by the integrand in (10), the proof of (57) will be achieved if we establish that $\theta_n \to_{\mathbb{P}} 0$. Since $F \geq 1$, it suffices to prove that the class $\mathcal{G} := \{(f_1 - f_2)^2,\ (f_1, f_2) \in \mathcal{F}^2\}$ satisfies the conditions of Theorem 1, with the formal changes of $\beta_{i,n}$ to $\beta_{i,n}^2$. All these conditions are straightforwardly satisfied, except (6) for which we invoke the almost sure comparison, for fixed $\epsilon > 0$:

$$N\big(\epsilon, \mathcal{G}, \mid\mid \cdot \mid\mid_{\overline{P}(\boldsymbol{\beta}_n, \mathbf{Y}_n),1}\big) \leq N\Big(\mid\mid F \mid\mid^{-1}_{\overline{P}(\boldsymbol{\beta}_n,\mathbf{Y}_n),2} \epsilon/8, \mathcal{F}, \mid\mid \cdot \mid\mid_{\overline{P}(\boldsymbol{\beta}_n,\mathbf{Y}_n),2}\Big)^2,$$

for which the RHS is bounded in probability as $n \to \infty$ by (9) and (10). This proves (57). The proof of (58) is very similar. We omit details. □

The proof of Theorem 2 is then achieved by combining Step 1 and Step 2 with a straightforward adaptation of Van der Vaart and Wellner (1996, p. 72, Theorem 1.12.2). □

Proof of Corollary 4. Assume without loss of generality that $K_n \geq 3$ for all $n \geq 1$. We will prove assertions (34), (35) and (36) with K_n as written in the statement of Corollary 4. Note that $Z_{i,n} \rightsquigarrow K_n \times Beta(1 - \alpha_n, K_n + i\alpha_n)$. For fixed $\delta > 0$ and $n \geq 1$ we have

$$\begin{aligned}
&\mathbb{P}\Big(\max_{i \leq [\delta K_n]} Z_{i,n} \geq \log(K_n)\Big) \\
&\leq [\delta K_n] \max_{i \leq [\delta K_n]} \mathbb{P}\Big(Z_{i,n} \geq \log(K_n)\Big) \\
&= [\delta K_n]\mathbb{P}\Big(Z_{1,n} \geq \log(K_n)\Big), \ \text{by usual properties of the Beta distribution} \\
&= [\delta K_n] \frac{\Gamma(1 - \alpha_n + K_n + 1)}{\Gamma(1 - \alpha_n)\Gamma(K_n + 1)} K_n^{-1} \int_{\log(K_n)}^{K_n} \big(\frac{x}{K_n}\big)^{-\alpha_n} \big(1 - \frac{x}{K_n}\big)^{K_n} dx \\
&\leq [\delta K_n] \frac{\Gamma(1 - \alpha_n + K_n + 1)}{\Gamma(1 - \alpha_n)\Gamma(K_n + 1)} K_n^{-1+\alpha_n} \int_{\log(K_n)}^{+\infty} \exp\big(-x\big) dx,
\end{aligned}$$

$\Gamma(\cdot)$ denoting the Gamma function. Now, by a use of Stirling's formula for $\Gamma(\cdot)$, and because $1-\alpha_n \geq \epsilon$ for all n, we have

$$\frac{\Gamma(1-\alpha_n+K_n+1)}{\Gamma(1-\alpha_n)\Gamma(K_n+1)} K_n^{-1+\alpha_n} \leq 2$$

for all n large enough. This proves (34). To prove (35), fix $n \geq 1$ and first notice that, by the moment formulas of the Beta distribution, we have

$$\begin{aligned}
&\frac{1}{K_n}\sum_{i\geq 1}\mathbb{E}\Big(Z_{i,n}^4\Big)\prod_{j=1}^{i-1}\mathbb{E}\Big(\Big(1-\frac{Z_{i,n}}{K_n}\Big)^4\Big)\\
\leq\;& \frac{1}{K_n}\sum_{i\geq 1}\mathbb{E}\Big(Z_{i,n}^4\Big)\prod_{j=1}^{i-1}\mathbb{E}\Big(1-\frac{Z_{i,n}}{K_n}\Big)\\
\leq\;& \frac{1}{K_n}\sum_{i\geq 1}\frac{(1-\alpha_n)(2-\alpha_n)(3-\alpha_n)(4-\alpha_n)}{\big(1+\frac{i\alpha_n}{K_n}\big)^4}\prod_{j=1}^{i-1}\Big(1-\frac{1-\alpha_n}{1-\alpha_n+K_n+1+j\alpha_n}\Big).
\end{aligned} \tag{60}$$

Now consider the two following complementary cases.

Case 1: $\alpha_n = 0$: In that case the RHS of (60) is bounded by

$$\frac{24}{K_n}\sum_{i\geq 1}\prod_{j=1}^{i-1}\Big(1-\frac{1}{K_n+2}\Big) \leq \frac{24(K_n+2)}{K_n} \leq 48,$$

because $K_n \geq 3$.

Case 2: $\alpha_n > 0$: In that case the RHS of (60) can be bounded by 64 by the following chain of inequalities (we use the classical comparisons of sums/integral of monotonic functions in (61) and (62)):

$$\begin{aligned}
=\;& \frac{24}{K_n}\sum_{i\geq 1}\frac{1}{\big(1+\frac{i\alpha_n}{K_n}\big)^4}\prod_{j=1}^{i-1}\Big(1-\frac{1-\alpha_n}{1-\alpha_n+K_n+1+j\alpha_n}\Big)\\
\leq\;& \frac{24}{K_n}\Bigg[2+\sum_{i\geq 3}\frac{1}{\big(1+\frac{i\alpha_n}{K_n}\big)^4}\exp\Bigg(-\sum_{j=1}^{i-1}\frac{1-\alpha_n}{1-\alpha_n+K_n+1+\alpha_n j}\Bigg)\Bigg]\\
\leq\;& \frac{24}{K_n}\Bigg[2+\sum_{i\geq 3}\Big(1+\frac{i\alpha_n}{K_n}\Big)^{-4}\exp\Bigg(-\frac{1-\alpha_n}{\alpha_n}\Big(\log\Big(\frac{1-\alpha_n+K_n+1+i\alpha_n}{1-\alpha_n+K_n+1+2\alpha_n}\Big)\Big)\Bigg)\Bigg]
\end{aligned} \tag{61}$$

$$\begin{aligned}
=\;& \frac{24}{K_n}\Bigg[2+\sum_{i\geq 3}\Big(1+\frac{i\alpha_n}{K_n}\Big)^{-4}\Big(1+\frac{(i-2)\alpha_n}{1-\alpha_n+K_n+1+2\alpha_n}\Big)^{-\frac{1-\alpha_n}{\alpha_n}}\Bigg]\\
\leq\;& \frac{48}{K_n}+\frac{24}{\alpha_n}\frac{K_n+3}{K_n}\times\frac{\alpha_n}{K_n+3}\sum_{i\geq 3}\Big(1+\frac{(i-2)\alpha_n}{K_n+3}\Big)^{-4}\Big(1+\frac{(i-2)\alpha_n}{K_n+3}\Big)^{-\frac{1-\alpha_n}{\alpha_n}}
\end{aligned}$$

$$\leq 16 + \frac{48}{\alpha_n} \times \frac{1}{4 + \frac{1-\alpha_n}{\alpha_n}} \tag{62}$$
$$= 16 + \frac{48}{1 + 3\alpha_n}$$
$$\leq 64.$$

Now, in order to prove (36), in virtue of the Bienaymé-Tchebytchev inequality, it is sufficient to establish that

$$\underline{\lim}_{n\to\infty} \min_{i\leq n} \mathbb{E}\left(Z_{i,n}^2\right) > 0, \tag{63}$$

$$\overline{\lim}_{n\to\infty} \max_{i\leq n} \mathrm{Var}\left(Z_{i,n}^2\right) < \infty, \tag{64}$$

which is straightforward using the moments expressions of the Beta distribution. We omit details. □

Acknowledgement

The author would like to thank two anonymous referees for their very helpful suggestions and comments, which contributed to several improvements of the present article.

References

Alexander, K. S., 1987. Central limit theorems for stochastic processes under random entropy conditions. Prob. Theo. Related Fields 75 (3), 351–378. MR0890284

Castillo, I., Nickl, R., 2014. On the bernstein-von mises phenomenon for nonparametric bayes procedures. Ann. Probab. 42 (5), 1941–1969. MR3262473

De Blasi, P., Lijoi, A., Prünster, I., 2013. An asymptotic analysis of a class of discrete nonparametric priors. Stat. Sinica 23, 1299–1322. MR3114715

Dudley, R., 1999. Uniform Central Limit Theorems. Cambridge Univ. Press. MR1720712

Ferguson, T. S., 1973. A bayesian analysis of some nonparametric problems. Ann. Statist., 209–230. MR0350949

Hjort, N., Holmes, C., Müller, P., Walker, S., 2010. Bayesian Nonparametrics. Vol. 28. Cambridge University Press. MR2722987

James, L. F., 2008. Large sample asymptotics for the two-parameter poisson-dirichlet process. In: Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh. Institute of Mathematical Statistics, pp. 187–199. MR2459225

James, L. F., Lijoi, A., Prünster, I., 2009. Posterior analysis for normalized random measures with independent increments. Scand. J. Statist. 36 (1), 76–97. MR2508332

Koltchinskii, V. I., 1981. On the central limit theorem for empirical measures. Theory Probab. Math. Statist. 24, 71–82. MR0628431

Lo, A. Y., 1983. Weak convergence for dirichlet processes. Sankhyā Ser. A, 105–111. MR0749358

Praestgaard, J., Wellner, J. A., 1993. Exchangeably weighted bootstraps of the general empirical process. Ann. Probab., 2053–2086. MR1245301

Sheehy, A., Wellner, J. A., 1992. Uniform donsker classes of functions. Ann. Probab., 1983–2030. MR1188051

Van der Vaart, A., Wellner, J., 1996. Weak Convergence and Empirical Processes. Springer. MR1385671

Williams, D., 1991. Probability with Martingales. Cambridge University Press. MR1155402

Zaitsev, A., 1987. On the Gaussian approximation of convolutions under multi-dimensional analogues of S. N. Bernstein's inequality. Probab. Theory Related Fields 74 (4), 535–566. MR0876255