Bayesian Analysis (2012) **7**, Number 2, pp. 277–310

On the Support of MacEachern's Dependent Dirichlet Processes and Extensions

Andrés F. Barrientos*, Alejandro Jara† and Fernando A. Quintana‡

Abstract. We study the support properties of Dirichlet process–based models for sets of predictor–dependent probability distributions. Exploiting the connection between copulas and stochastic processes, we provide an alternative definition of MacEachern's dependent Dirichlet processes. Based on this definition, we provide sufficient conditions for the full weak support of different versions of the process. In particular, we show that under mild conditions on the copula functions, the version where only the support points or the weights are dependent on predictors have full weak support. In addition, we also characterize the Hellinger and Kullback–Leibler support of mixtures induced by the different versions of the dependent Dirichlet process. A generalization of the results for the general class of dependent stick–breaking processes is also provided.

Keywords: Related probability distributions, Bayesian nonparametrics, Copulas, Weak support, Hellinger support, Kullback–Leibler support, Stick–breaking processes

1 Introduction

This paper focuses on the support properties of probability models for sets of predictor–dependent probability measures, $\{G_x : x \in \mathscr{X}\}$, where the G_x's are probability measures defined on a common measurable space $(S, \mathscr{S})$ and indexed by a p–dimensional vector of predictors $x \in \mathscr{X}$. The problem of defining probability models of this kind has received increasing recent attention in the Bayesian literature, motivated by the construction of nonparametric priors for the conditional densities estimation problem. To date, much effort has focused on constructions that generalize the widely used class of Dirichlet process (DP) priors (Ferguson 1973, 1974), and, consequently, the class of DP mixture models (Ferguson 1983; Lo 1984; Escobar and West 1995) for single density estimation. A random probability measure G is said to be a DP with parameters (α, G_0), where $\alpha \in \mathbb{R}_0^+ = [0, +\infty)$ and G_0 is a probability measure on $(S, \mathscr{S})$, written as $G \mid \alpha, G_0 \sim DP(\alpha G_0)$, if for any measurable nontrivial partition $\{B_l : 1 \le l \le k\}$ of S, the vector $\{G(B_l) : 1 \le l \le k\}$ has a Dirichlet distribution with parameters $(\alpha G_0(B_1), \dots, \alpha G_0(B_k))$. It follows that $G(B) \mid \alpha, G_0 \sim \text{Beta}(\alpha G_0(B), \alpha G_0(B^c))$, and therefore, $E[G(B) \mid \alpha, G_0] = G_0(B)$ and $Var[G(B) \mid \alpha, G_0] = G_0(B)G_0(B^c)/(\alpha + 1)$. These results show the role of G_0 and α, namely, that G is centered around G_0 and that α is a precision parameter.

*Department of Statistics, Pontificia Universidad Católica de Chile, Santiago, Chile, afbarrie@uc.cl
†Department of Statistics, Pontificia Universidad Católica de Chile, Santiago, Chile, atjara@uc.cl
‡Department of Statistics, Pontificia Universidad Católica de Chile, Santiago, Chile, quintana@mat.puc.cl

 DOI:10.1214/12-BA709

An early reference on predictor–dependent DP models is Cifarelli and Regazzini (1978), who defined a model for related probability measures by introducing a regression model in the centering measure of a collection of independent DP random measures. This approach is used, for example, by Muliere and Petrone (1993), who considered a linear regression model for the centering distribution $G_x^0 \equiv N(x'\beta, \sigma^2)$, where $x \in \mathbb{R}^p$, $\beta \in \mathbb{R}^p$ is a vector of regression coefficients, and $N(\mu, \sigma^2)$ stands for a normal distribution with mean μ and variance σ^2, respectively. Similar models were discussed by Mira and Petrone (1996) and Giudici et al. (2003). Linking nonparametric models through the centering distribution, however, limits the nature of the dependence of the process. A more flexible construction, the dependent Dirichlet process (DDP), was proposed by MacEachern (1999, 2000). The key idea behind the DDP is the construction of a set of random measures that are marginally (i.e. for every possible predictor value) DP–distributed random measures. In this framework, dependence is introduced through a modification of the stick–breaking representation of each element in the set. If $G \mid \alpha, G_0 \sim DP(\alpha G_0)$, then the trajectories of the process can be almost surely represented by the following stick–breaking representation provided by Sethuraman (1994):

$$G(B) = \sum_{i=1}^{\infty} W_i \delta_{\theta_i}(B), \quad B \in \mathscr{S}, \tag{1}$$

where $\delta_\theta(\cdot)$ is the Dirac measure at θ, $W_i = V_i \prod_{j<i}(1 - V_j)$ for all $i \geq 1$, with $V_i \mid \alpha \overset{iid}{\sim} \text{Beta}(1, \alpha)$, and $\theta_i \mid G_0 \overset{iid}{\sim} G_0$. MacEachern (1999, 2000) generalized expression (1) by considering

$$G_x(B) = \sum_{i=1}^{\infty} W_i(x) \delta_{\theta_i(x)}(B), \quad B \in \mathscr{S},$$

where the support points $\theta_i(x)$, $i = 1, \ldots$, are independent stochastic processes with index set $\mathscr{X}$ and G_x^0 marginal distributions, and the weights take the form $W_i(x) = V_i(x) \prod_{j<i}[1 - V_j(x)]$, where $\{V_i(x) : i \geq 1\}$ are independent stochastic processes with index set $\mathscr{X}$ and Beta $(1, \alpha_x)$ marginal distributions.

MacEachern (2000) showed that the DDP exists and can have full weak support, provided a flexible specification for the point mass processes $\{\theta_i(x) : x \in \mathscr{X}\}$ and simple conditions for the weight processes $\{V_i(x) : x \in \mathscr{X}\}$ are assumed. Based on the latter result, he also proposed a version of the process with predictor–independent weights, $G_x(B) = \sum_{i=1}^{\infty} W_i \delta_{\theta_i(x)}(B)$, called the single weights DDP model. Versions of the single weights DDP have been applied to ANOVA (De Iorio et al. 2004), survival (De Iorio et al. 2009; Jara et al. 2010), spatial modeling (Gelfand et al. 2005), functional data (Dunson and Herring 2006), time series (Caron et al. 2008), discriminant analysis (De la Cruz et al. 2007), and longitudinal data analysis (Müller et al. 2005). We refer the reader to Müller et al. (1996), Dunson et al. (2007), Dunson and Park (2008), and Chung and Dunson (2009), and references therein, for other DP–based models for related probability distributions.

Although there exists a wide variety of probability models for related probability distributions, there is a scarcity of results characterizing the support of the proposed processes. The large support is a minimum requirement and almost a "necessary condition" for a nonparametric model to be considered "nonparametric", because it ensures that a nonparametric prior does not assign too much mass on small sets of probability measures. This property is also important because it is a typically required condition for frequentist consistency of the posterior distribution. Some recent results have been provided by Pati et al. (2011) and Norets and Pelenis (2011), in the context of dependent mixtures of Gaussians induced by probit stick–breaking processes (Chung and Dunson 2009), and dependent mixtures of location–scale distributions induced by finite mixing distributions (Norets 2010) and kernel stick–breaking processes (Dunson and Park 2008), respectively.

In this paper we provide an alternative characterization of the weak support of the two versions of MacEachern's DDP discussed above, namely, a version where both weights and support points are functions of the predictors, and a version where only the support points are functions of the predictors. We also characterize the weak support of a version of the DDP model where only the weights depend on predictors. Finally, we provide sufficient conditions for the full Hellinger support of mixture models induced by DDP priors, and characterize their Kulback–Leibler support. Our results are based on an alternative definition of MacEachern's DDP, which exploits the connection between stochastic processes and copulas. Specifically, families of copulas are used to define the finite dimensional distributions of stochastic processes with given marginal distributions. The alternative formulation of the DDP makes explicit the parameters of the process, and their role on the support properties. The rest of this paper is organized as follows. Section 2 provides the alternative definition of MacEachern's DDP. Section 3 contains the main results about the support of the various DDP versions, as well as extensions to more general stick–breaking constructions. A general discussion and possible future research lines are given in Section 4.

2 MacEachern's dependent Dirichlet processes

MacEachern (1999, 2000) defined the DDP by using transformations of independent stochastic processes. Let $\alpha_{\mathscr{X}} = \{\alpha_x : x \in \mathscr{X}\}$ be a set such that, for every $x \in \mathscr{X}$, $\alpha_x \in \mathbb{R}_0^+$, and let $G^0_{\mathscr{X}} = \left\{G^0_x : x \in \mathscr{X}\right\}$ be a set of probability distributions with support on $(S, \mathscr{S})$. Let $Z^{\theta_i}_{\mathscr{X}} = \left\{Z^{\theta}_i(x) : x \in \mathscr{X}\right\}$, $i \in \mathbb{N}$, be independent and identically distributed real–valued processes with marginal distributions $\left\{F^{\theta}_x : x \in \mathscr{X}\right\}$. Similarly, let $Z^{V_i}_{\mathscr{X}} = \left\{Z^{V}_i(x) : x \in \mathscr{X}\right\}$, $i \in \mathbb{N}$, be independent and identically distributed real–valued processes with marginal distributions $\left\{F^{V}_x : x \in \mathscr{X}\right\}$. For every $x \in \mathscr{X}$, let $T^V_x : \mathbb{R} \longrightarrow [0,1]$ and $T^{\theta}_x : \mathbb{R} \longrightarrow S$ be transformations that specify a mapping of $Z^V_i(x)$ into $V_i(x)$, and $Z^{\theta}_i(x)$ into $\theta_i(x)$, respectively. Furthermore, set $T^V_{\mathscr{X}} = \left\{T^V_x : x \in \mathscr{X}\right\}$ and $T^{\theta}_{\mathscr{X}} = \left\{T^{\theta}_x : x \in \mathscr{X}\right\}$. In MacEachern's definition, the DDP is parameterized by

$$\left(\alpha_{\mathscr{X}}, \left\{Z^{V_i}_{\mathscr{X}}\right\}_{i=1}^{\infty}, \left\{Z^{\theta_i}_{\mathscr{X}}\right\}_{i=1}^{\infty}, T^V_{\mathscr{X}}, T^{\theta}_{\mathscr{X}}\right).$$

To induce the desired marginal distributions of the weights and support point processes, MacEachern defined the transformations as a composition of appropriate measurable mappings. Specifically, for every $x \in \mathscr{X}$, he wrote $T_x^V = B_x^{-1} \circ F_x^V$ and $T_x^\theta = G_x^{0^{-1}} \circ F_x^\theta$, where B_x^{-1} and $G_x^{0^{-1}}$ are the inverse cumulative density function (CDF) of the Beta$(1, \alpha_x)$ distribution and G_x^0, respectively.

We provide an alternative definition of MacEachern's DDP that explicitly exploits the connection between copulas and stochastic processes. The basic idea is that many properties of stochastic processes can be characterized by their finite–dimensional distributions. Therefore, copulas can be used for their analysis. Note however, that many concepts associated with stochastic processes are stronger than the finite–dimensional distribution approach. In order to make this paper self–contained, we provide below a brief discussion about the definition of stochastic processes through the specification of finite dimensional copula functions.

2.1 Copulas and stochastic processes

Copulas are functions that are useful for describing and understanding the dependence structure between random variables. The basic idea is the ability to express a multivariate distribution as a function of its marginal distributions. If H is a d–variate CDF with marginal CDF's given by $F_1, \ldots, F_d$, then by Sklar's theorem (Sklar 1959), there exists a copula function $C : [0,1]^d \longrightarrow [0,1]$ such that $H(t_1, \ldots, t_d) = C(F_1(t_1), \ldots, F_d(t_d))$, for all $t_1, \ldots, t_d \in \mathbb{R}$, and this representation is unique if the marginal distributions are absolutely continuous w.r.t. Lebesgue measure. Thus by the probability integral transform, a copula function is a d–variate CDF on $[0,1]^d$ with uniform marginals on $[0,1]$, which fully captures the dependence among the associated random variables, irrespective of the marginal distributions. Examples and properties of copulas can be found, for example, in Joe (1997).

Under certain regularity conditions a stochastic process is completely characterized by its finite–dimensional distributions. Therefore, it is possible –and useful– to use copulas to define stochastic processes with given marginal distributions. The basic idea is to specify the collection of finite dimensional distributions of a process through a collection of copulas and marginal distributions. The following result is a straightforward consequence of Kolmogorov's consistency theorem (Kolmogorov 1933, page 29) and of Sklar's theorem (Sklar 1959).

Corollary 1. *Let $\mathscr{C}_{\mathscr{X}} = \{C_{x_1,\ldots,x_d} : x_1, \ldots, x_d \in \mathscr{X}, d > 1\}$ be a collection of copula functions and $\mathscr{D}_{\mathscr{X}} = \{F_x : x \in \mathscr{X}\}$ a collection of one–dimensional probability distributions defined on a common measurable space $(\mathscr{D}, \mathscr{B}(\mathscr{D}))$, where $\mathscr{D} \subseteq \mathbb{R}$. Assume that for every integer $d > 1$, $x_1, \ldots, x_d \in \mathscr{X}$, $u_i \in [0,1]$, $i = 1, \ldots, d$, $k \in \{1, \ldots, d\}$, and permutation $\pi = (\pi_1, \ldots, \pi_d)$ of $\{1, \ldots, d\}$, the elements in $\mathscr{C}_{\mathscr{X}}$ satisfy the following consistency conditions:*

(i) $C_{x_1,\ldots,x_d}(u_1, \ldots, u_d) = C_{x_{\pi_1},\ldots,x_{\pi_d}}(u_{\pi_1}, \ldots, u_{\pi_d})$, *and*

(ii) $C_{x_1,\ldots,x_d}(u_1,\ldots,u_{k-1},1,u_{k+1},\ldots,u_d) =$

$$C_{x_1,\ldots,x_{k-1},x_{k+1},\ldots,x_d}(u_1,\ldots,u_{k-1},u_{k+1},\ldots,u_d).$$

Then there exists a probability space $(\Omega,\mathscr{A},P)$ *and a stochastic process*

$$Y:\mathscr{X}\times\Omega\to\Delta,$$

such that

$$P\{\omega\in\Omega:Y(x_1,\omega)\le t_1,\ldots,Y(x_d,\omega)\le t_d\} = C_{x_1,\ldots,x_d}(F_{x_1}(t_1),\ldots,F_{x_d}(t_d)),$$

for any $t_1,\ldots,t_d\in\mathbb{R}$.

Notice that conditions (i) and (ii) above correspond to the definition of a consistent system of probability measures, applied to probability measures defined on appropriate unitary hyper–cubes. Notice also that finite–dimensional distributions of $[0,1]$–valued stochastic processes necessarily satisfy conditions (i) and (ii), i.e., they form a consistent system of probability measures. Kolmogorov's consistency theorem states that conversely, if the sample space is a subset of the real line, every consistent family of measures is in fact the family of finite–dimensional distributions of some stochastic process. Since the unitary hyper–cube is a subset of a Euclidean space, Kolmogorov's consistency theorem implies that every family of distributions satisfying conditions (i) and (ii), is the family of finite–dimensional distributions of an $[0,1]$–valued stochastic process.

A consequence of the previous result is that it is possible to interpret a stochastic process in terms of a simpler process of uniform variables transformed by the marginal distributions via a quantile mapping. The use of copulas to define stochastic processes was first considered by Darsow et al. (1992), who studied the connection between Markov processes and copulas, and provided necessary and sufficient conditions for a process to be Markovian, based on the copula family. Although in an approach completely different to the one considered here, copulas have been used to define dependent Bayesian nonparametric models by Epifani and Lijoi (2010) and Leisen and Lijoi (2011). These authors consider a Lévy copula to define dependent versions of neutral to the right and two–parameter Poisson–Dirichlet processes (Pitman and Yor 1997), respectively.

From a practical point of view, it is easy to specify a family of copulas satisfying conditions (i) and (ii) in Corollary 1. An obvious approach is to consider the family of copula functions arising from the finite–dimensional distributions of known and tractable stochastic processes. The family of copula functions associated with Gaussian or t–Student processes could be considered as natural candidates in many applications for which $\mathscr{X}\subseteq\mathbb{R}^p$. The finite–dimensional copula functions of Gaussian processes are given by

$$C_{x_1,\ldots,x_d}(u_1,\ldots,u_d) = \Phi_{\mathbf{R}(x_1,\ldots,x_d)}(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_d)),$$

where $\Phi_{\mathbf{R}(x_1,\ldots,x_d)}$ is the CDF of a d–variate normal distribution with mean zero, variance one and correlation matrix $\mathbf{R}(x_1,\ldots,x_d)$, arising from the corresponding correlation function, and Φ is the CDF of a standard normal distribution.

In the context of longitudinal or spatial modeling, natural choices for correlation functions are the Matérn, powered exponential and spherical. The elements of the correlation matrix induced by the Matérn covariance function are given by

$$\mathbf{R}(x_1,\ldots,x_d)_{(i,j)} = \left\{2^{\kappa-1}\Gamma(\kappa)\right\}^{-1}\left(\frac{||x_i - x_j||_2}{\tau}\right)^{\kappa}\mathscr{K}_{\kappa}\left(\frac{||x_i - x_j||_2}{\tau}\right),$$

where $\kappa \in \mathbb{R}^+$, $\tau \in \mathbb{R}^+$ and $\mathscr{K}_\kappa(\cdot)$ is the modified Bessel function of order κ (Abramowitz and Stegun 1964). The elements of the correlation matrix under the powered exponential covariance function are given by

$$\mathbf{R}(x_1,\ldots,x_d)_{(i,j)} = \exp\left\{-\left(\frac{||x_i - x_j||_2}{\tau}\right)^{\kappa}\right\},$$

where $\kappa \in (0,2]$ and $\tau \in \mathbb{R}^+$. Finally, the elements of the correlation matrix induced by the spherical covariance function are given by

$$\mathbf{R}(x_1,\ldots,x_d)_{(i,j)} = \begin{cases} 1 - \frac{3}{2}\left(\frac{||x_i-x_j||_2}{\tau}\right) + \frac{1}{2}\left(\frac{||x_i-x_j||_2}{\tau}\right)^3, & \text{if } ||x_i - x_j||_2 \le \tau, \\ 0, & \text{if } ||x_i - x_j||_2 > \tau, \end{cases}$$

where $\tau \in \mathbb{R}^+$.

2.2 The alternative definition

Let $\mathscr{C}^V_{\mathscr{X}}$ and $\mathscr{C}^\theta_{\mathscr{X}}$ be two sets of copulas satisfying the consistency conditions of Corollary 1. As earlier, let $\alpha_{\mathscr{X}} = \{\alpha_x : x \in \mathscr{X}\}$ be a set such that, for every $x \in \mathscr{X}$, $\alpha_x \in \mathbb{R}^+_0$, and let $G^0_{\mathscr{X}} = \left\{G^0_x : x \in \mathscr{X}\right\}$ be a set of probability measures defined on a common measurable space $(S,\mathscr{S})$, where $S \subseteq \mathbb{R}^q$, $q \in \mathbb{N}$, and $\mathscr{S} = \mathscr{B}(S)$ is the Borel σ–field of S. Finally, let $\mathscr{P}(S)$ be the set of all Borel probability measures defined on $(S,\mathscr{S})$.

Definition 1. *Let $\{G_x : x \in \mathscr{X}\}$ be a $\mathscr{P}(S)$–valued stochastic process on an appropriate probability space $(\Omega,\mathscr{A},P)$ such that:*

(i) $V_1, V_2, \ldots$ are independent stochastic processes of the form $V_i : \mathscr{X} \times \Omega \to [0,1]$, $i \ge 1$, with common finite dimensional distributions determined by the set of copulas $\mathscr{C}^V_{\mathscr{X}}$ and the set of Beta marginal distributions with parameters $(1,\alpha_x)$, $\{\text{Beta}(1,\alpha_x) : x \in \mathscr{X}\}$.

(ii) $\theta_1, \theta_2, \ldots$ are independent stochastic processes of the form $\theta_i : \mathscr{X} \times \Omega \to S$, $i \ge 1$, with common finite dimensional distributions determined by the set of copulas $\mathscr{C}^\theta_{\mathscr{X}}$ and the set of marginal distributions $G^0_{\mathscr{X}}$.

(iii) For every $x \in \mathscr{X}$, $B \in \mathscr{S}$ and almost every $\omega \in \Omega$,

$$G(x,\omega)(B) = \sum_{i=1}^{\infty} \left\{ V_i(x,\omega) \prod_{j<i} [1 - V_j(x,\omega)] \right\} \delta_{\theta_i(x,\omega)}(B).$$

Such a process $\mathscr{H} = \{G_x \doteq G(x,\cdot) : x \in \mathscr{X}\}$ will be referred to as a dependent Dirichlet process with parameters $\left(\alpha_{\mathscr{X}}, \mathscr{C}^{\theta}_{\mathscr{X}}, \mathscr{C}^{V}_{\mathscr{X}}, G^0_{\mathscr{X}}\right)$, and denoted by $\mathrm{DDP}\left(\alpha_{\mathscr{X}}, \mathscr{C}^{\theta}_{\mathscr{X}}, \mathscr{C}^{V}_{\mathscr{X}}, G^0_{\mathscr{X}}\right)$.

In what follows, two simplifications of the general definition of the process will be of interest. If the stochastic processes in (i) of Definition 1 are replaced by independent and identically distributed Beta$(1, \alpha)$ random variables, with $\alpha > 0$, the process will be referred to as "single weights" DDP. This is to emphasize the fact that the weights in the stick–breaking representation (iii) of Definition 1, are not indexed by predictors x.

Definition 2. *Let $\{G_x : x \in \mathscr{X}\}$ be a $\mathscr{P}(S)$–valued stochastic process on an appropriate probability space $(\Omega, \mathscr{A}, P)$ such that:*

(i) $V_1, V_2, \ldots$ are independent random variables of the form $V_i : \Omega \to [0,1]$, $i \geq 1$, with common Beta distribution with parameters $(1, \alpha)$.

(ii) $\theta_1, \theta_2, \ldots$ are independent stochastic processes of the form $\theta_i : \mathscr{X} \times \Omega \to S$, $i \geq 1$, with common finite dimensional distributions determined by the set of copulas $\mathscr{C}^{\theta}_{\mathscr{X}}$ and the set of marginal distributions $G^0_{\mathscr{X}}$.

(iii) For every $x \in \mathscr{X}$, $B \in \mathscr{S}$ and almost every $\omega \in \Omega$,

$$G(x,\omega)(B) = \sum_{i=1}^{\infty} \left\{ V_i(\omega) \prod_{j<i} [1 - V_j(\omega)] \right\} \delta_{\theta_i(x,\omega)}(B).$$

Such a process $\mathscr{H} = \{G_x \doteq G(x,\cdot) : x \in \mathscr{X}\}$ will be referred to as a single weights dependent Dirichlet process with parameters $\left(\alpha, \mathscr{C}^{\theta}_{\mathscr{X}}, G^0_{\mathscr{X}}\right)$, and denoted by $\mathrm{wDDP}\left(\alpha, \mathscr{C}^{\theta}_{\mathscr{X}}, G^0_{\mathscr{X}}\right)$.

The second simplification is when the stochastic processes in (ii) of Definition 1 are replaced by independent random vectors with common distribution G^0, where G^0 is supported on the measurable space $(S, \mathscr{S})$. In this case the process will be referred to as "single atoms" DDP, to emphasize the fact that the support points in the stick–breaking representation are not indexed by predictors x.

Definition 3. *Let $\{G_x : x \in \mathscr{X}\}$ be a $\mathscr{P}(S)$–valued stochastic process on an appropriate probability space $(\Omega, \mathscr{A}, P)$ such that:*

(i) $V_1, V_2, \ldots$ *are independent stochastic processes of the form* $V_i : \mathscr{X} \times \Omega \to [0,1]$, $i \geq 1$, *with common finite dimensional distributions determined by the set of copulas* $\mathscr{C}_{\mathscr{X}}^{V}$ *and the set of Beta marginal distributions with parameters* $(1, \alpha_x)$, $\{\text{Beta}(1, \alpha_x) : x \in \mathscr{X}\}$.

(ii) $\theta_1, \theta_2, \ldots$ *are independent* S*–valued random variables/vectors,* $i \geq 1$, *with common distribution* G^0.

(iii) For every $x \in \mathscr{X}$, $B \in \mathscr{S}$ *and almost every* $\omega \in \Omega$,

$$G(x, \omega)(B) = \sum_{i=1}^{\infty} \left\{ V_i(x, \omega) \prod_{j<i} [1 - V_j(x, \omega)] \right\} \delta_{\theta_i(\omega)}(B).$$

Such a process $\mathscr{H} = \{G_x \doteq G(x, \cdot) : x \in \mathscr{X}\}$ *will be referred to as a single atoms dependent Dirichlet process with parameters* $\big(\alpha_{\mathscr{X}}, \mathscr{C}_{\mathscr{X}}^{V}, G^0\big)$, *and denoted by* $\theta\text{DDP}\big(\alpha_{\mathscr{X}}, \mathscr{C}_{\mathscr{X}}^{V}, G^0\big)$.

3 The main results

3.1 Preliminaries

As is widely known, the definition of the support of probability models on functional spaces depends on the choice of a "distance" defining the basic neighborhoods. The results presented here are based on three different notions of distance between probability measures. Theorems 1 through 3 below are based on neighborhoods created using any distance that metrizes the weak star topology, namely, any distance d_W such that, for two probability measures F and G_n defined on a common measurable space, $d_W(G_n, F) \longrightarrow 0$ if and only if G_n converges weakly to F as n goes to infinity. If F and G are probability measures absolutely continuous with respect to a common dominating measure, stronger notions of distance can be considered. The results summarized in Theorems 4 and 5 are based on neighborhoods created using the Hellinger distance, which is topologically equivalent to the L_1 distance, and the Kullback–Leibler divergence, respectively. If f and g are versions of the densities of F and G with respect to a dominating measure λ, respectively, the L_1 distance, Hellinger distance and the Kullback–Leibler divergence are defined by

$$d_{L_1}(f, g) = \int |f(y) - g(y)| \lambda(dy),$$

$$d_H(f, g) = \int \left(\sqrt{f(y)} - \sqrt{g(y)} \right)^2 \lambda(dy),$$

and

$$d_{KL}(f, g) = \int f(y) \log \left(\frac{f(y)}{g(y)} \right) \lambda(dy),$$

respectively.

The support of a probability measure μ defined on a space of probability measures is the smallest closed set of μ–measure one, say $C(\mu)$, which can be expressed as

$$C(\mu) = \{F: \ \mu(N_\epsilon(F)) > 0, \forall \epsilon > 0\},$$

where $N_\epsilon(F) = \{G: \ d(F,G) < \epsilon\}$, with d being any notion of "distance". The different types of "metrics" discussed above, therefore, induce different types of supports. Let $C_W(\mu)$, $C_{L_1}(\mu)$, $C_H(\mu)$ and $C_{KL}(\mu)$ be the support induced by d_W, d_{L_1}, d_H and d_{KL}, respectively. The relationships among these different supports are completely defined by the relationships between the different "metrics". Since L_1 convergence implies weak convergence, the topology generated by any distance metrizing the weak convergence (e.g., the Prokhorov or Lévy metric) is coarser than the one generated by the L_1 distance, i.e., $C_W(\mu) \supset C_{L_1}(\mu)$. Hellinger distance is equivalent to the L_1 distance since $d_{L_1}(f,g) \leq d_H^2(f,g) \leq 4d_{L_1}(f,g)$, which implies that $C_H(\mu) = C_{L_1}(\mu)$. Finally, the relation between the L_1 distance and Kullback–Leibler divergence is given by the inequality $d_{KL}(f,g) \geq \frac{1}{4}d_{L_1}(f,g)$, implying that $C_{L_1}(\mu) = C_H(\mu) \supset C_{KL}(\mu)$.

3.2 Weak support of dependent Dirichlet processes

Let $\mathscr{P}(S)^{\mathscr{X}}$ be the set of all $\mathscr{P}(S)$–valued functions defined on $\mathscr{X}$. Let $\mathscr{B}\left\{\mathscr{P}(S)^{\mathscr{X}}\right\}$ be the Borel σ-field generated by the product topology of weak convergence. The support of the DDP is the smallest closed set in $\mathscr{B}\left\{\mathscr{P}(S)^{\mathscr{X}}\right\}$ with $P \circ \mathscr{H}^{-1}$–measure one.

Assume that $\Theta \subseteq S$ is the support of G_x^0, for every $x \in \mathscr{X}$. The following theorem provides sufficient conditions under which $\mathscr{P}(\Theta)^{\mathscr{X}}$ is the weak support of the DDP, where $\mathscr{P}(\Theta)^{\mathscr{X}}$ is the set of all $\mathscr{P}(\Theta)$–valued functions defined on Θ, with $\mathscr{P}(\Theta)$ being the set of all probability measures defined on $(\Theta, \mathscr{B}(\Theta))$.

Theorem 1. *Let $\{G_x : x \in \mathscr{X}\}$ be a* DDP $\left(\alpha_{\mathscr{X}}, \mathscr{C}_{\mathscr{X}}^{\theta}, \mathscr{C}_{\mathscr{X}}^{V}, G_{\mathscr{X}}^0\right)$. *If $\mathscr{C}_{\mathscr{X}}^{\theta}$ and $\mathscr{C}_{\mathscr{X}}^{V}$ are collections of copulas with positive density w.r.t. Lebesgue measure, on the appropriate unitary hyper–cubes, then $\mathscr{P}(\Theta)^{\mathscr{X}}$ is the weak support of the process, i.e., the DDP has full weak support.*

Proof: The proof has two parts. The first part shows that a sufficient condition for the full weak support result is that the process assigns positive probability mass to a product space of particular simplices. The second part of the proof shows that the DDP assigns positive probability mass to that product space of simplices.

Let $\mathscr{P}_n = \{P_x^n : x \in \mathscr{X}\} \in \mathscr{P}(\Theta)^{\mathscr{X}}$ be a collection of probability measures with support contained in Θ. Let $\{\mathscr{P}_n\}_{n\geq 1} \subset \mathscr{P}(\Theta)^{\mathscr{X}}$ be a sequence of such collections, satisfying the condition that for all $x \in \mathscr{X}$, $P_x^n \overset{weakly}{\longrightarrow} P_x$, when $n \longrightarrow \infty$, where P_x

is a probability measure. Since S is closed and $P_x^n \overset{weakly}{\longrightarrow} P_x$, Portmanteau's theorem implies that $P_x(\Theta) \geq \limsup_n P_x^n(\Theta)$, for every $x \in \mathscr{X}$. It follows that $\mathscr{P}(\Theta)^{\mathscr{X}}$ is a closed set. Now, let $\Theta^{\mathscr{X}} = \prod_{x \in \mathscr{X}} \Theta$. Since Θ is the support of G_x^0, for every $x \in \mathscr{X}$, it follows that

$$P\left\{\omega \in \Omega : \theta_i(\cdot, \omega) \in \Theta^{\mathscr{X}},\ i = 1, 2, \ldots\right\} = 1,$$

i.e.,

$$P\left\{\omega \in \Omega : G(\cdot, \omega) \in \mathscr{P}(\Theta)^{\mathscr{X}}\right\} = 1.$$

To show that $\mathscr{P}(\Theta)^{\mathscr{X}}$ is the smallest closed set with $P \circ \mathscr{H}^{-1}$–measure one, it suffices to prove that any basic open set in $\mathscr{P}(\Theta)^{\mathscr{X}}$ has positive $P \circ \mathscr{H}^{-1}$–measure. Now, it is easy to see that the measure of a basic open set for $\left\{P_x^0 : x \in \mathscr{X}\right\} \in \mathscr{P}(\Theta)^{\mathscr{X}}$ is equal to the measure of a set of the form

$$\prod_{i=1}^{T} \left\{P_{x_i} \in \mathscr{P}(\Theta) : \left|\int f_{ij} dP_{x_i} - \int f_{ij} dP_{x_i}^0\right| < \epsilon_i,\ j = 1, \ldots, K_i\right\}, \tag{2}$$

where $x_1, \ldots, x_T \in \mathscr{X}$, T and K_i, $i = 1, \ldots, T$, are positive integers, f_{ij}, $i = 1, \ldots, T$, and $j = 1, \ldots, K_i$, are bounded continuous functions and ϵ_i, $i = 1, \ldots, T$, are positive constants. To show that neighborhoods of the form (2) have positive probability mass, it suffices to show they contain certain subset–neighborhoods with positive probability mass. In particular, we consider subset–neighborhoods of probability measures which are absolutely continuous w.r.t. the corresponding centering distributions and that adopt the form

$$U\left(Q_{x_1}, \ldots, Q_{x_T}, \{A_{ij}\}, \epsilon^*\right) = \prod_{i=1}^{T} \left\{P_{x_i} \in \mathscr{P}(\Theta) : \left|P_{x_i}(A_{ij}) - Q_{x_i}(A_{ij})\right| < \epsilon^*,\ j = 1 \ldots m_i\right\},$$

where Q_{x_i} is a probability measure absolutely continuous w.r.t. $G_{x_i}^0$, $i = 1, \ldots, T$, $A_{i1}, \ldots, A_{im_i} \subseteq \Theta$ are measurable sets with Q_{x_i}–null boundary, and $\epsilon^* > 0$. For discrete centering distributions $G_{x_1}^0, \ldots, G_{x_T}^0$, the existence of a subset–neighborhood $U\left(Q_{x_1}, \ldots, Q_{x_T}, \{A_{ij}\}, \epsilon^*\right)$ of the set (2) is immediately ensured. The case of centering distributions that are absolutely continuous w.r.t. Lebesgue measure follows after Lemma 1 in Appendix A.

Next, borrowing the trick in Ferguson (1973), for each $\nu_{ij} \in \{0, 1\}$, we define sets $B_{\nu_{11} \ldots \nu_{m_T T}}$ as

$$B_{\nu_{11} \ldots \nu_{m_T T}} = \bigcap_{i=1}^{T} \bigcap_{j=1}^{m_i} A_{ij}^{\nu_{ij}},$$

where A_{ij}^1 is interpreted as A_{ij} and A_{ij}^0 is interpreted as A_{ij}^c. Note that

$$\left\{B_{\nu_{11} \ldots \nu_{m_T T}}\right\}_{\nu_{ij} \in \{0,1\}},$$

is a measurable partition of Θ such that

$$A_{ij} = \bigcup_{\{\nu_{11},\ldots,\nu_{m_T T}:\, \nu_{ij}=1\}} B_{\nu_{11}\ldots\nu_{m_T T}}.$$

It follows that sets of the form

$$\prod_{i=1}^{T} \Big\{ P_{x_i} \in \mathscr{P}(\Theta) : \Big| P_{x_i}\left(B_{\nu_{11}\ldots\nu_{m_T T}}\right) - Q_{x_i}\left(B_{\nu_{11}\ldots\nu_{m_T T}}\right)\Big| < 2^{-\sum_{l=1}^{T} m_l}\epsilon^*, \forall\ (\nu_{11},\ldots,\nu_{m_T T})\Big\},$$

are contained in $U\left(Q_{x_1},\ldots,Q_{x_T},\{A_{ij}\},\epsilon^*\right)$. To simplify the notation, set

$$J_\nu = \Big\{\nu_{11}\ldots\nu_{m_T T} : G_x^0\left(B_{\nu_{11}\ldots\nu_{m_T T}}\right) > 0\Big\},$$

and let M be a bijective mapping from J_ν to $\{0,\ldots,k\}$, where k is the cardinality of J_ν minus 1. Therefore, $A_{M(\nu)} = B_\nu$, for all $\nu \in J_\nu$. Now, set

$$\mathbf{s}_{x_i} = \left(w_{(x_i,0)},\ldots,w_{(x_i,k)}\right) = \left(Q_{x_i}(A_0),\ldots,Q_{x_i}(A_k)\right) \in \Delta_k,\ i=1,\ldots,T,$$

where $\Delta_k = \Big\{(w_0,\ldots,w_k) : w_i \geq 0,\ i=0,\ldots,k,\ \sum_{i=0}^{k} w_i = 1\Big\}$ is the k-simplex, and, for $i=1,\ldots,T$, set

$$B\left(\mathbf{s}_{x_i},\epsilon\right) = \left\{(w_0,\ldots,w_k) \in \Delta_k : w_{(x_i,j)} - \epsilon < w_j < w_{(x_i,j)} + \epsilon,\ j=0,\ldots,k\right\},$$

where $\epsilon = 2^{-\sum_{l=1}^{T} m_l}\epsilon^*$. Note that

$$\begin{aligned}\{\omega \in \Omega : [G(x_1,\omega),\ldots,G(x_T,\omega)] \in U\left(Q_{x_1},\ldots,Q_{x_T},\{A_{ij}\},\epsilon\right)\} \supseteq \\ \{\omega \in \Omega : [G(x_i,\omega)(A_0),\ldots,G(x_i,\omega)(A_k)] \in B\left(\mathbf{s}_{x_i},\epsilon\right),\ i=1,\ldots,T\}.\end{aligned}$$

Thus, to show that (2) has positive P–measure, it suffices to show that

$$P\{\omega \in \Omega : [G(x_i,\omega)(A_0),\ldots,G(x_i,\omega)(A_k)] \in B\left(\mathbf{s}_{x_i},\epsilon\right),\ i=1,\ldots,T\} > 0. \quad (3)$$

Now, consider a subset $\Omega_0 \subseteq \Omega$, such that for every $\omega \in \Omega_0$ the following conditions are met:

(A.1) For $i=1,\ldots,T$,

$$w_{(x_i,0)} - \frac{\epsilon}{2} < V_1(x_i,\omega) < w_{(x_i,0)} + \frac{\epsilon}{2}.$$

(A.2) For $i=1,\ldots,T$ and $j=1,\ldots,k-1$,

$$\frac{w_{(x_i,j)} - \frac{\epsilon}{2}}{\prod_{l<j+1}\left(1 - V_l(x_i,\omega)\right)} < V_{j+1}(x_i,\omega) < \frac{w_{(x_i,j)} + \frac{\epsilon}{2}}{\prod_{l<j+1}\left(1 - V_l(x_i,\omega)\right)}.$$

(A.3) For $i = 1, \ldots, T$,

$$\frac{1 - \sum_{j=0}^{k-1} W_j(x_i, \omega) - \frac{\epsilon}{2}}{\prod_{l<k+1}(1 - V_l(x_i, \omega))} < V_{k+1}(x_i, \omega) < \frac{1 - \sum_{j=0}^{k-1} W_j(x_i, \omega)}{\prod_{l<k+1}(1 - V_l(x_i, \omega))},$$

where for $j = 1, \ldots, k-1$,

$$W_{j-1}(x_i, \omega) = V_j(x_i, \omega) \prod_{l<j} (1 - V_l(x_i, \omega)).$$

(A.4) For $j = 0, \ldots, k$,

$$[\theta_{j+1}(x_1, \omega), \ldots, \theta_{j+1}(x_T, \omega)] \in A_j^T.$$

Now, to prove the theorem, it suffices to show that $P(\{\omega : \omega \in \Omega_0\}) > 0$. It is easy to see that if assumptions (A.1) – (A.4) hold, then for $i = 1, \ldots, T$,

$$[G(x_i, \omega)(A_0), \ldots, G(x_i, \omega)(A_k)] \in B(\mathbf{s}_{x_i}, \epsilon).$$

It then follows from the DDP definition that

$$\begin{aligned} P\{\omega \in \Omega \quad : \quad & [G(x_i, \omega)(A_0), \ldots, G(x_i, \omega)(A_k)] \in B(\mathbf{s}_{x_i}, \epsilon), i = 1, \ldots, T\} \geq \\ & P\left\{\omega \in \Omega : [V_j(x_1, \omega), \ldots, V_j(x_T, \omega)] \in Q_j^\omega, j = 1, \ldots, k+1\right\} \times \\ & \prod_{j=1}^{k+1} P\left\{\omega \in \Omega : [\theta_j(x_1, \omega), \ldots, \theta_j(x_T, \omega)] \in A_{j-1}^T\right\} \times \\ & \prod_{j=k+2}^{\infty} P\left\{\omega \in \Omega : [V_j(x_1, \omega), \ldots, V_j(x_T, \omega)] \in [0, 1]^T\right\} \times \\ & \prod_{j=k+2}^{\infty} P\left\{\omega \in \Omega : [\theta_j(x_1, \omega), \ldots, \theta_j(x_T, \omega)] \in \Theta^T\right\}, \end{aligned}$$

where,

$$Q_1^\omega = \prod_{i=1}^{T} \left[w_{(x_i,0)} - \frac{\epsilon}{2}, w_{(x_i,0)} + \frac{\epsilon}{2}\right],$$

$$\begin{aligned} Q_{j+1}^\omega &= Q_{j+1}^\omega(V_1(x_1, \omega), \ldots, V_j(x_T, \omega)) \\ &= \prod_{i=1}^{T} \left[\frac{w_{(x_i,j)} - \frac{\epsilon}{2}}{\prod_{l<j+1}(1 - V_l(x_i, \omega))}, \frac{w_{(x_i,j)} + \frac{\epsilon}{2}}{\prod_{l<j+1}(1 - V_l(x_i, \omega))}\right], \end{aligned}$$

for $j = 1, \ldots, k-1$, and

$$\begin{aligned} Q_{k+1}^\omega &= Q_{k+1}^\omega(V_1(x_1, \omega), \ldots, V_k(x_T, \omega)) \\ &= \prod_{i=1}^{T} \left[\frac{1 - \sum_{j=0}^{k-1} W_j(x_i, \omega) - \frac{\epsilon}{2}}{\prod_{l<k+1}(1 - V_l(x_i, \omega))}, \frac{1 - \sum_{j=0}^{k-1} W_j(x_i, \omega)}{\prod_{l<k+1}(1 - V_l(x_i, \omega))}\right]. \end{aligned}$$

By the definition of the process,

$$P\left\{\omega \in \Omega : \left[V_j\left(x_1,\omega\right),\ldots,V_j\left(x_T,\omega\right)\right] \in \left[0,1\right]^T\right\} = 1,$$

and

$$P\left\{\omega \in \Omega : \left[\theta_j\left(x_1,\omega\right),\ldots,\theta_j\left(x_T,\omega\right)\right] \in \Theta^T\right\} = 1.$$

It follows that

$$\begin{aligned} P\left\{\omega \in \Omega\right. &: \left.\left[G\left(x_i,\omega\right)\left(A_0\right),\ldots,G\left(x_i,\omega\right)\left(A_k\right)\right] \in B\left(\mathbf{s}_{x_i},\epsilon\right), i=1,\ldots,T\right\} \geq \\ & P\left\{\omega \in \Omega : \left[V_j\left(x_1,\omega\right),\ldots,V_j\left(x_T,\omega\right)\right] \in Q_j^{\omega}, j=1,\ldots,k+1\right\} \times \\ & \prod_{j=1}^{k+1} P\left\{\omega \in \Omega : \left[\theta_j\left(x_1,\omega\right),\ldots,\theta_j\left(x_T,\omega\right)\right] \in A_{j-1}^T\right\}. \end{aligned}$$

Since by assumption $\mathscr{C}_{\mathscr{X}}^V$ is a collection of copulas with positive density w.r.t. Lebesgue measure, the non–singularity of the Beta distribution implies that

$$\begin{aligned} & P\left\{\omega \in \Omega : \left[V_j\left(x_1,\omega\right),\ldots,V_j\left(x_T,\omega\right)\right] \in Q_j^{\omega}, j=1,\ldots,k+1\right\} = \\ & \int_{Q_1}\int_{Q_2(\mathbf{v}_1)}\cdots\int_{Q_{k+1}(\mathbf{v}_1,\ldots,\mathbf{v}_k)} f_{x_1,\ldots,x_T}^V\left(\mathbf{v}_1\right)\cdots f_{x_1,\ldots,x_T}^V\left(\mathbf{v}_{k+1}\right) d\mathbf{v}_{k+1}\cdots d\mathbf{v}_2 d\mathbf{v}_1 > 0, \end{aligned} \tag{4}$$

where $f_{x_1,\ldots,x_T}^V\left(\mathbf{v}_j\right)$, $j=1,\ldots,k+1$, is the density function of

$$C_{x_1,\ldots,x_T}^V\left(\mathrm{B}\left(v_1 \mid 1,\alpha_{x_1}\right),\ldots,\mathrm{B}\left(v_T \mid 1,\alpha_{x_T}\right)\right),$$

with $\mathrm{B}(\cdot \mid a,b)$ denoting the CDF of a Beta distribution with parameters (a,b). Finally, since by assumption $\mathscr{C}_{\mathscr{X}}^{\theta}$ is a collection of copulas with positive density w.r.t. Lebesgue measure and, for all $x \in \mathscr{X}$, Θ is the topological support of G_x^0, it follows that

$$\begin{aligned} & P\left\{\omega \in \Omega : \left[\theta_j\left(x_1,\omega\right),\ldots,\theta_j\left(x_T,\omega\right)\right] \in A_{j-1}^T\right\} = \\ & \qquad\qquad \int I_{A_{j-1}^T}\left(\theta\right) dC_{x_1,\ldots,x_T}^{\theta}\left(G_{x_1}^0\left(\theta_1\right),\ldots,G_{x_T}^0\left(\theta_T\right)\right) > 0, \end{aligned}$$

where $I_A(\cdot)$ is the indicator function for the set A. This completes the proof of the theorem. □

The successful results obtained in applications of the single weights DDP in a variety of applications (see, e.g. De Iorio et al. 2004; Müller et al. 2005; De Iorio et al. 2009; Gelfand et al. 2005; De la Cruz et al. 2007; Jara et al. 2010), suggest that simplified versions of the DDP can be specified to have large support. The following theorem provides sufficient conditions under which $\mathscr{P}\left(\Theta\right)^{\mathscr{X}}$ is the weak support of the single–weights DDP.

Theorem 2. *Let $\{G_x : x \in \mathscr{X}\}$ be a* wDDP $\left(\alpha, \mathscr{C}^{\theta}_{\mathscr{X}}, G^0_{\mathscr{X}}\right)$. *If $\mathscr{C}^{\theta}_{\mathscr{X}}$ is a collection of copulas with positive density w.r.t. Lebesgue measure, on the appropriate unitary hyper–cubes, then $\mathscr{P}(\Theta)^{\mathscr{X}}$ is the weak support of the process.*

Proof: Using a similar reasoning as in the proof of Theorem 1, it suffices to prove (3), that is

$$P\left\{\omega \in \Omega : \left[G\left(x_i, \omega\right)\left(A_0\right), \ldots, G\left(x_i, \omega\right)\left(A_k\right)\right] \in B\left(\mathbf{s}_{x_i}, \epsilon\right),\, i = 1, \ldots, T\right\} > 0.$$

As in the proof of Theorem 1, we consider constraints for the elements of the wDDP that imply the previous relation. Since the rational numbers are dense in $\mathbb{R}$, there exist $M_i, m_{ij} \in \mathbb{N}$ such that for $i = 1, \ldots, T$, and $j = 0, \ldots, k-1$,

$$w_{(x_i,j)} - \frac{\epsilon}{4} < \frac{m_{ij}}{M_i} < w_{(x_i,j)} + \frac{\epsilon}{4}.$$

Now, let $N = M_1 \times \ldots \times M_T$ and $n_{ij} = m_{ij} \prod_{l \neq i} M_l$. It follows that, for $i = 1, \ldots, T$, and $j = 0, \ldots, k-1$,

$$w_{(x_i,j)} - \frac{\epsilon}{4} < \frac{n_{ij}}{N} < w_{(x_i,j)} + \frac{\epsilon}{4}.$$

Therefore, for any

$$(p_1, \ldots, p_N) \in \Delta_{N-1} = \left\{(w_1, \ldots, w_N) : w_i \geq 0,\, 1 \leq i \leq N,\, \sum_{i=1}^{N} w_i = 1\right\},$$

that verifies

$$\frac{1}{N} - \frac{\epsilon}{4N} < p_l < \frac{1}{N} + \frac{\epsilon}{4N}, \qquad \text{for } l = 1, \ldots, N,$$

we have

$$w_{(x_i,0)} - \frac{\epsilon}{2} < \sum_{l=1}^{n_{i0}} p_l < w_{(x_i,0)} + \frac{\epsilon}{2},\; i = 1, \ldots, T,$$

and

$$w_{(x_i,j)} - \frac{\epsilon}{2} < \sum_{l=n_{i(j-1)}+1}^{n_{ij}} p_l < w_{(x_i,j)} + \frac{\epsilon}{2},$$

for $i = 1, \ldots, T$ and $j = 1, \ldots, k-1$.

On the other hand, let $a(i,l)$ be a mapping such that

$$a(i,l) = \begin{cases} 0 & \text{if} \quad l \leq n_{i0} \\ 1 & \text{if} \quad n_{i0} < l \leq n_{i0} + n_{i1} \\ \vdots & \vdots \quad \vdots \\ k-1 & \text{if} \quad \sum_{k'=0}^{k-2} n_{ik'} < l \leq \sum_{k'=0}^{k-1} n_{ik'} \\ k & \text{if} \quad \sum_{k'=0}^{k-1} n_{ik'} < l \leq N \end{cases},$$

$i = 1, \ldots, T$, and $l = 1, \ldots, N$. Note that the previous function defines a possible path for the functions $\theta_1(\cdot, \omega), \theta_2(\cdot, \omega), \ldots$ through the measurable sets $A_0, \ldots, A_k$.

The required constraints are defined next. Consider a subset $\Omega_0 \subseteq \Omega$, such that for every $\omega \in \Omega_0$ the following conditions are met:

(B.1) For $l = 1$,

$$\frac{1}{N} - \frac{\epsilon}{4N} < V_l(\omega) < \frac{1}{N} + \frac{\epsilon}{4N}.$$

(B.2) For $l = 2, \ldots, N-1$,

$$\frac{\frac{1}{N} - \frac{\epsilon}{4N}}{\prod_{l'<l}(1 - V_{l'}(\omega))} < V_l(\omega) < \frac{\frac{1}{N} + \frac{\epsilon}{4N}}{\prod_{l'<l}(1 - V_{l'}(\omega))}.$$

(B.3) For $l = N$,

$$\frac{1 - \sum_{l'=1}^{N-1} W_{l'}(\omega) - \frac{\epsilon}{2}}{\prod_{l'<N}(1 - V_{l'}(\omega))} < V_l(\omega) < \frac{1 - \sum_{l'=1}^{N-1} W_{l'}(\omega)}{\prod_{l'<N}(1 - V_{l'}(\omega))},$$

where for $l = 1, 2, \ldots$

$$W_{l-1}(\omega) = V_l(\omega) \prod_{l'<l} \left[1 - V_{l'}(\omega)\right].$$

(B.4) For $i = 1, \ldots, T$ and $l = 1, \ldots, N$,

$$\left(\theta_l\left(x_1, \omega\right), \ldots, \theta_l\left(x_T, \omega\right)\right) \in A_{a(1,l)} \times \ldots \times A_{a(T,l)}.$$

Now, to prove the theorem, it suffices to show that $P(\{\omega : \omega \in \Omega_0\}) > 0$. It is easy to see that if assumptions (B.1) – (B.4) hold, then, for $i = 1, \ldots, T$,

$$\left[G\left(x_i, \omega\right)\left(A_0\right), \ldots, G\left(x_i, \omega\right)\left(A_k\right)\right] \in B\left(\mathbf{s}_{x_i}, \epsilon\right).$$

Thus, from the definition of the wDDP, it follows that

$$\begin{aligned} P\left\{\omega \in \Omega \right. &: \left[G\left(x_i, \omega\right)\left(A_0\right), \ldots, G\left(x_i, \omega\right)\left(A_k\right)\right] \in B\left(\mathbf{s}_{x_i}, \epsilon\right), i = 1, \ldots, T\} \geq \\ & P\left\{\omega \in \Omega : V_l\left(\omega\right) \in Q_l^\omega, l = 1, \ldots, N\right\} \times \\ & \prod_{l=1}^{N} P\left\{\omega \in \Omega : \left[\theta_l\left(x_1, \omega\right), \ldots, \theta_l\left(x_T, \omega\right)\right] \in A_{a(1,l)} \times \ldots \times A_{a(T,l)}\right\} \times \\ & \prod_{l=N+1}^{\infty} P\left\{\omega \in \Omega : V_l\left(\omega\right) \in [0,1]\right\} \times \\ & \prod_{l=N+1}^{\infty} P\left\{\omega \in \Omega : \left[\theta_l\left(x_1, \omega\right), \ldots, \theta_l\left(x_T, \omega\right)\right] \in \Theta^T\right\}, \end{aligned}$$

where,

$$Q_1^{\omega} = \left[\frac{1}{N} - \frac{\epsilon}{4N}, \frac{1}{N} + \frac{\epsilon}{4N}\right],$$

$$\begin{aligned} Q_{l+1}^{\omega} &= Q_{l+1}^{\omega}\left\{V_1(\omega), \ldots, V_l(\omega)\right\} \\ &= \left[\frac{\frac{1}{N} - \frac{\epsilon}{4N}}{\prod_{l'<l+1}\left(1 - V_{l'}(\omega)\right)}, \frac{\frac{1}{N} + \frac{\epsilon}{4N}}{\prod_{l'<l+1}\left(1 - V_{l'}(\omega)\right)}\right], \end{aligned}$$

$l = 1, \ldots, N-2$, and

$$\begin{aligned} Q_N^{\omega} &= Q_N^{\omega}\left\{V_1(\omega), \ldots, V_{N-1}(\omega)\right) \\ &= \left[\frac{1 - \sum_{l'=1}^{N-1} W_{l'}(\omega) - \frac{\epsilon}{2}}{\prod_{l'<N}\left(1 - V_{l'}(\omega)\right)}, \frac{1 - \sum_{l'=1}^{N-1} W_{'l}(\omega)}{\prod_{l'<N}\left(1 - V_{l'}(\omega)\right)}\right]. \end{aligned}$$

From the definition of the process, $P\left\{\omega \in \Omega : V_l\left(\omega\right) \in [0,1], l \in \mathbb{N}\right\} = 1$, and

$$P\left\{\omega \in \Omega : \left[\theta_l\left(x_1, \omega\right), \ldots, \theta_l\left(x_T, \omega\right)\right] \in \Theta^T, l \in \mathbb{N}\right\} = 1.$$

It follows that

$$\begin{aligned} P\left\{\omega \in \Omega\right. &: \left[G\left(x_i, \omega\right)\left(A_0\right), \ldots, G\left(x_i, \omega\right)\left(A_k\right)\right] \in B\left(\mathbf{s}_{x_i}, \epsilon\right), i = 1, \ldots, T\right\} \geq \\ & P\left\{\omega \in \Omega : V_l\left(\omega\right) \in Q_l^{\omega}, l = 1, \ldots, N\right\} \times \\ & \prod_{l=1}^{N} P\left\{\omega \in \Omega : \left[\theta_l\left(x_1, \omega\right), \ldots, \theta_l\left(x_T, \omega\right)\right] \in A_{a(1,l)} \times \ldots \times A_{a(T,l)}\right\}. \end{aligned}$$

The non–singularity of the Beta distribution implies that

$$P\left\{\omega \in \Omega : V_l\left(\omega\right) \in Q_l^{\omega}, l = 1, \ldots, N\right\} > 0. \tag{5}$$

Finally, since by assumption $\mathscr{C}_{\mathscr{X}}^{\theta}$ is a collection of copulas with positive density w.r.t. Lebesgue measure and, for all $x \in \mathscr{X}$, Θ is the topological support of G_x^0, it follows that

$$\begin{aligned} P\left\{\omega \in \Omega : \left[\theta_l\left(x_1, \omega\right), \ldots, \theta_l\left(x_T, \omega\right)\right] \in A_{l-1}^T\right\} = \\ \int I_{A_{l-1}^T}\left(\theta\right) dC_{x_1,\ldots,x_T}^{\theta}\left(G_{x_1}^0\left(\theta_1\right), \ldots, G_{x_T}^0\left(\theta_T\right)\right) > 0, \end{aligned}$$

which completes the proof. □

In the search of a parsimonious model, the previous result shows that full weak support holds for the single–weights DDP for which only the atoms are subject to a flexible specification. The following theorem provides sufficient conditions under which $\mathscr{P}\left(\Theta\right)^{\mathscr{X}}$ is the weak support of the single–atoms DDP.

Theorem 3. *Let $\{G_x : x \in \mathscr{X}\}$ be a θDDP $\left(\alpha_{\mathscr{X}}, \mathscr{C}^V_{\mathscr{X}}, G^0\right)$, where the support of G_0 is Θ. If $\mathscr{C}^V_{\mathscr{X}}$ is a collection of copulas with positive density w.r.t. to Lebesgue measure, on the appropriate unitary hyper–cubes, then the support of the process is $\mathscr{P}(\Theta)^{\mathscr{X}}$.*

Proof: In analogy with the proofs of Theorems 1 and 2, it suffices to prove (3), that is

$$P\left\{\omega \in \Omega : \left[G\left(x_i,\omega\right)\left(A_0\right),\ldots,G\left(x_i,\omega\right)\left(A_k\right)\right] \in B\left(\mathbf{s}_{x_i},\epsilon\right),\, i=1,\ldots,T\right\} > 0.$$

Consider a subset $\Omega_0 \subseteq \Omega$, such that for every $\omega \in \Omega_0$ the following conditions are met:

(C.1) For $i = 1,\ldots,T$,

$$w_{(x_i,0)} - \frac{\epsilon}{2} < V_1\left(x_i,\omega\right) < w_{(x_i,0)} + \frac{\epsilon}{2}.$$

(C.2) For $i = 1,\ldots,T$ and $j = 1,\ldots,k-1$,

$$\frac{w_{(x_i,j)} - \frac{\epsilon}{2}}{\prod_{l<j+1}\left(1 - V_l\left(x_i,\omega\right)\right)} < V_{j+1}\left(x_i,\omega\right) < \frac{w_{(x_i,j)} + \frac{\epsilon}{2}}{\prod_{l<j+1}\left(1 - V_l\left(x_i,\omega\right)\right)}.$$

(C.3) For $i = 1,\ldots,T$,

$$\frac{1 - \sum_{j=0}^{k-1} W_j\left(x_i,\omega\right) - \frac{\epsilon}{2}}{\prod_{l<k+1}\left(1 - V_l\left(x_i,\omega\right)\right)} < V_{k+1}\left(x_i,\omega\right) < \frac{1 - \sum_{j=0}^{k-1} W_j\left(x_i,\omega\right)}{\prod_{l<k+1}\left(1 - V_l\left(x_i,\omega\right)\right)},$$

where,

$$W_{j-1}\left(x_i,\omega\right) = V_j\left(x_i,\omega\right)\prod_{l<j}\left(1 - V_l\left(x_i,\omega\right)\right),$$

for $j = 1,\ldots,k-1$.

(C.4) For $j = 0,\ldots,k$,

$$\theta_{j+1}(\omega) \in A_j.$$

Now, to prove the theorem, it suffices to show that $P(\{\omega : \omega \in \Omega_0\}) > 0$. It is easy to see that if assumptions (C.1) – (C.4) hold, then, for $i = 1,\ldots,T$,

$$\left[G\left(x_i,\omega\right)\left(A_0\right),\ldots,G\left(x_i,\omega\right)\left(A_k\right)\right] \in B\left(\mathbf{s}_{x_i},\epsilon\right).$$

Thus, from the definition of the θDDP, it follows that

$$\begin{aligned} P\{\omega \in \Omega \quad : \quad & \left[G\left(x_i,\omega\right)\left(A_0\right),\ldots,G\left(x_i,\omega\right)\left(A_k\right)\right] \in B\left(\mathbf{s}_{x_i},\epsilon\right), i=1,\ldots,T\} \geq \\ & P\left\{\omega \in \Omega : \left[V_j\left(x_1,\omega\right),\ldots,V_j\left(x_T,\omega\right)\right] \in Q^{\omega}_j, j=1,\ldots,k+1\right\} \times \\ & \prod_{j=1}^{k+1} P\left\{\omega \in \Omega : \theta_j\left(\omega\right) \in A_{j-1}\right\} \times \\ & \prod_{j=k+2}^{\infty} P\left\{\omega \in \Omega : \left[V_j\left(x_1,\omega\right),\ldots,V_j\left(x_T,\omega\right)\right] \in [0,1]^T\right\} \times \\ & \prod_{j=k+2}^{\infty} P\left\{\omega \in \Omega : \theta_j\left(\omega\right) \in \Theta\right\}, \end{aligned}$$

where,

$$Q_1^\omega = \prod_{i=1}^{T} \left[w_{(x_i,0)} - \frac{\epsilon}{2}, w_{(x_i,0)} + \frac{\epsilon}{2} \right],$$

$$\begin{aligned} Q_{j+1}^\omega &= Q_{j+1}^\omega \left(V_1 \left(x_1, \omega \right), \ldots, V_j \left(x_T, \omega \right) \right) \\ &= \prod_{i=1}^{T} \left[\frac{w_{(x_i,j)} - \frac{\epsilon}{2}}{\prod_{l<j+1} \left(1 - V_l \left(x_i, \omega \right)\right)}, \frac{w_{(x_i,j)} + \frac{\epsilon}{2}}{\prod_{l<j+1} \left(1 - V_l \left(x_i, \omega \right)\right)} \right], \end{aligned}$$

for $j = 1, \ldots, k-1$, and

$$\begin{aligned} Q_{k+1}^\omega &= Q_{k+1}^\omega \left(V_1 \left(x_1, \omega \right), \ldots, V_k \left(x_T, \omega \right) \right) \\ &= \prod_{i=1}^{T} \left[\frac{1 - \sum_{j=0}^{k-1} W_j \left(x_i, \omega \right) - \frac{\epsilon}{2}}{\prod_{l<k+1} \left(1 - V_l \left(x_i, \omega \right)\right)}, \frac{1 - \sum_{j=0}^{k-1} W_j \left(x_i, \omega \right)}{\prod_{l<k+1} \left(1 - V_l \left(x_i, \omega \right)\right)} \right]. \end{aligned}$$

By the definition of the process, $P\left\{ \omega \in \Omega : \theta_j \left(\omega \right) \in \Theta, j \in \mathbb{N} \right\} = 1$, and

$$P \left\{ \omega \in \Omega : \left[V_j \left(x_1, \omega \right), \ldots, V_j \left(x_T, \omega \right) \right] \in \left[0,1\right]^T, j \in \mathbb{N} \right\} = 1.$$

It follows that

$$\begin{aligned} P\left\{ \omega \in \Omega \right. &: \left[G\left(x_i,\omega\right)\left(A_0\right), \ldots, G\left(x_i,\omega\right)\left(A_k\right) \right] \in B\left(\mathbf{s}_{x_i}, \epsilon\right), i = 1, \ldots, T \left.\right\} \geq \\ & P\left\{ \omega \in \Omega : \left[V_j\left(x_1,\omega\right), \ldots, V_j\left(x_T,\omega\right) \right] \in Q_j^\omega, j = 1, \ldots, k+1 \right\} \times \\ & \prod_{j=1}^{k+1} P\left\{ \omega \in \Omega : \theta_j\left(\omega\right) \in A_{j-1} \right\}. \end{aligned}$$

Since by assumption $\mathscr{C}_{\mathscr{X}}^V$ is a collection of copulas with positive density w.r.t. Lebesgue measure, the non–singularity of the Beta distribution implies that

$$\begin{aligned} & P\left\{ \omega \in \Omega : \left[V_j\left(x_1,\omega\right), \cdots, V_j\left(x_T,\omega\right) \right] \in Q_j^\omega, j = 1, \ldots, k+1 \right\} = \\ & \int_{Q_1} \int_{Q_2(\mathbf{v}_1)} \cdots \int_{Q_{k+1}(\mathbf{v}_1,\cdots,\mathbf{v}_k)} f_{x_1,\ldots,x_T}^V\left(\mathbf{v}_1\right) \cdots f_{x_1,\ldots,x_T}^V\left(\mathbf{v}_{k+1}\right) d\mathbf{v}_{k+1} \cdots d\mathbf{v}_2 d\mathbf{v}_1 > 0. \end{aligned} \tag{6}$$

Finally, since Θ is the topological support of G^0, it follows that

$$P\left\{ \omega \in \Omega : \theta_j\left(\omega\right) \in A_{j-1} \right\} > 0,$$

which completes the proof of the theorem. □

3.3 The support of dependent Dirichlet process mixture models

As in the case of DPs, the discrete nature of DDPs implies that they cannot be used as a probability model for sets of predictor–dependent densities. A standard approach to deal with this problem is to define a mixture of smooth densities based on the DDP. Such an approach was pioneered by Lo (1984) in the context of single density estimation problems. For every $\theta \in \Theta$, let $\psi(\cdot, \theta)$ be a probability density function, where $\Theta \subseteq \mathbb{R}^q$ now denotes a parameter set. A predictor–dependent mixture model is obtained by considering $f_x(\cdot \mid G_x) = \int_\Theta \psi(\cdot, \theta) G_x(d\theta)$. These mixtures can form a very rich family. For instance, the location and scale mixture of the form $\sigma^{-1} k\left(\frac{\cdot - \mu}{\sigma}\right)$, for some fixed density k, may approximate any density in the L^1–sense if σ is allowed to approach to 0. Thus, a prior on the set of predictor–dependent densities $\{f_x : x \in \mathscr{X}\}$ may be induced by placing some of the versions of the DDP prior on the set of related mixing distributions $\{G_x : x \in \mathscr{X}\}$.

The following theorem shows that under simple conditions on the kernel ψ, the full weak support of the different versions of DDPs ensures the large Hellinger support of the corresponding DDP mixture model.

Theorem 4. *Let ψ be a non–negative valued function defined on the product measurable space $(\mathscr{Y} \times \Theta, \mathscr{B}(\mathscr{Y}) \otimes \mathscr{B}(\Theta))$, where $\mathscr{Y} \subseteq \mathbb{R}^n$ is the sample space with corresponding Borel σ–field $\mathscr{B}(\mathscr{Y})$, $\Theta \subseteq \mathbb{R}^q$ is the parameter space with corresponding Borel σ–field $\mathscr{B}(\Theta)$ and $\mathscr{B}(\mathscr{Y}) \otimes \mathscr{B}(\Theta)$ denotes the product σ–field on $\mathscr{Y} \times \Theta$. Assume that ψ satisfies the following conditions:*

(i) $\int_{\mathscr{Y}} \psi(y, \theta) \lambda(dy) = 1$ *for every* $\theta \in \Theta$ *and some* σ*–finite measure* λ *on* $(\mathscr{Y}, \mathscr{B}(\mathscr{Y}))$.

(ii) $\theta \mapsto \psi(y, \theta)$ *is bounded, continuous and* $\mathscr{B}(\Theta)$*–measurable for every* $y \in \mathscr{Y}$.

(iii) At least one of the following conditions hold:

(iii.a) For every $\epsilon > 0$ *and* $y_0 \in \mathscr{Y}$, *there exists* $\delta(\epsilon, y_0) > 0$, *such that*

$$|y - y_0| \leq \delta(\epsilon, y_0),$$

then

$$\sup_{\theta \in \Theta} |\psi(y, \theta) - \psi(y_0, \theta)| < \epsilon.$$

(iii.b) For any compact set $K \subset \mathscr{Y}$ *and* $r > 0$, *the family of mappings*

$$\{\theta \mapsto \psi(y, \theta) : y \in K\},$$

defined on $\overline{B}(0, r)$, *is uniformly equicontinuous, where* $\overline{B}(0, r)$ *denotes a closed* L^1*–norm ball of radius* r *and centered at 0, that is,*

$$\overline{B}(0, r) \equiv \{\theta \in \Theta : \|\theta\|_1 \leq r\}.$$

If $\{G_x : x \in \mathscr{X}\}$ is a DPP*, a* wDDP *or a* θDDP*, satisfying the conditions of Theorem 1, 2 or 3, respectively, then the Hellinger support of the process*

$$\left\{ \int_\Theta \psi\left(\cdot,\theta\right) G_x\left(d\theta\right) : x \in \mathscr{X} \right\},$$

is

$$\prod_{x\in\mathscr{X}} \left\{ \int_\Theta \psi\left(\cdot,\theta\right) P_x\left(d\theta\right) : P_x \in \mathscr{P}(\Theta) \right\},$$

where $\mathscr{P}(\Theta)$ is the space of all probability measures defined on $(\Theta, \mathscr{B}(\Theta))$.

Proof: The proof uses a similar reasoning to the one of Section 3 in Lijoi et al. (2004). In what follows, we consider the Borel σ–field generated by the product topology induced by the Hellinger metric. It is easy to see that the measure of a basic open set for $\left\{f^0_{x_i} : x \in \mathscr{X}\right\}$, where $f^0_{x_i}(\cdot) = \int_\Theta \psi\left(\cdot,\theta\right) P^0_{x_i}\left(d\theta\right)$ and $\left\{P^0_x : x \in \mathscr{X}\right\} \in \mathscr{P}\left(\Theta\right)^{\mathscr{X}}$, is equal to the measure of a set of the form

$$\prod_{i=1}^{T} \left\{ \int_\Theta \psi\left(\cdot,\theta\right) P_{x_i}\left(d\theta\right) : \int_{\mathscr{Y}} \left| \int_\Theta \psi\left(y,\theta\right) P_{x_i}\left(d\theta\right) - f^0_{x_i}\left(y\right) \right| \lambda\left(dy\right) < \epsilon,\ P_{x_i} \in \mathscr{P}\left(\Theta\right) \right\}, \tag{7}$$

where $\epsilon > 0$, $x_1, \ldots, x_T \in \mathscr{X}$, and λ is a σ-finite measure on $(\mathscr{Y}, \mathscr{B}(\mathscr{Y}))$.

To show that the DDP mixture model assigns positive probability mass to sets of the form (7), we construct a weak neighborhood around $\left\{P^0_x : x \in \mathscr{X}\right\} \in \mathscr{P}\left(\Theta\right)^{\mathscr{X}}$ such that every element in it satisfies (7). This is done by appropriately defining the bounded and continuous functions that determine the weak neighborhood.

Let ν, ρ and η be positive constants. Fix a compact set $K_{x_i} \subset \mathscr{B}(\mathscr{Y})$ such that $\int_{K^c_{x_i}} f^0_{x_i}\left(y\right) \lambda\left(dy\right) < \frac{\epsilon}{8}$, and define

$$h^0_{i,1}\left(\theta\right) = \int_{K^c_{x_i}} \psi\left(y,\theta\right) \lambda\left(dy\right),$$

for $i = 1, \ldots, T$. For any ρ and ν, it is possible to define a closed ball of the form $\overline{B}\left(0, r-\nu\right) = \{\theta \in \Theta : \|\theta\|_1 \le r - \nu\}$, for some $r > \nu$ such that $P^0_{x_i}\left[\overline{B}\left(0, r-\nu\right)^c\right] \le \rho$. Now, choose continuous functions $h^0_{i,2}$, such that, for $i = 1, \ldots, T$,

$$I_{\overline{B}(0,r)^c}\left(\theta\right) \le h^0_{i,2}\left(\theta\right) \le I_{\overline{B}(0,r-\nu)^c}\left(\theta\right),$$

for every $\theta \in \Theta$. Note that condition (iii.a) (by continuity) or (iii.b) (by Arzelà–Ascoli's theorem) implies that the family of functions $\{\psi\left(y,\cdot\right) : y \in K_{x_i}\}$ on $\overline{B}\left(0, r\right)$ is a totally bounded set. Thus, given η, we can find a partition $A_{i,1}, \ldots, A_{i,n_i}$ of K_{x_i} and points $z_{i,1} \in A_{i,1}, \ldots, z_{i,n_i} \in A_{i,n_i}$ such that

$$\sup_{y \in A_{i,j}} \sup_{\theta \in \overline{B}(0,r)} \left|\psi\left(y,\theta\right) - \psi\left(z_{i,j},\theta\right)\right| < \eta$$

for each $i = 1, \ldots, T$ and $j = 1, \ldots, n_i$. Finally, for $i = 1, \ldots, T$ and $j = 1, \ldots, n_i$, define

$$h^1_{i,j}(\theta) = k(z_{i,j}, \theta).$$

All the $h^k_{i,j}$ functions considered above are continuous and bounded. Notice also that some of these functions may depend on ν, r, ρ and η. Define now the following family of sets

$$\prod_{i=1}^{T} \{P_{x_i} \in \mathscr{P}(\Theta) : \left| \int h^l_{i,j_l} dP_{x_i} - \int h^l_{i,j_l} dP^0_{x_i} \right| < \nu,\ l = 0, 1,\ j_0 = 1, 2,\ 1 \le j_1 \le n_i \right\}, \quad (8)$$

for $\nu > 0$. We will show that for appropriate choices of η, ν, r, and ρ, every collection $\{P_{x_1}, \ldots, P_{x_T}\}$ in sets of the form (8), satisfies

$$\int_{\mathscr{Y}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f^0_{x_i}(y) \right| \lambda(dy) < \epsilon,$$

for $i = 1, \ldots, T$. Note that

$$\begin{aligned} \int_{\mathscr{Y}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f^0_{x_i}(y) \right| \lambda(dy) &= \int_{K^c_{x_i}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f^0_{x_i}(y) \right| \lambda(dy) \\ &+ \int_{K_{x_i}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f^0_{x_i}(y) \right| \lambda(dy), \end{aligned}$$

for $i = 1, \ldots, T$. Now note that if $\left| \int h^0_{i,1} dP_{x_i} - \int h^0_{i,1} dP^0_{x_i} \right| < \nu$, then

$$\int h^0_{i,1} dP_{x_i} < \nu + \int h^0_{i,1} dP^0_{x_i} \le \nu + \frac{\epsilon}{8},$$

by the definition of $h^0_{i,1}$, and therefore,

$$\begin{aligned} \int_{K^c_{x_i}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f^0_{x_i}(y) \right| \lambda(dy) &\le \int h^0_{i,1} dP_{x_i} + \int_{K^c_{x_i}} f^0_{x_i}(y) \lambda(dy) \\ &\le \nu + \frac{\epsilon}{4}. \end{aligned} \quad (9)$$

In addition, note that

$$\int_{K_{x_i}} \left| \int_{\Theta} \psi(y, \theta) P_{x_i}(d\theta) - f^0_{x_i}(y) \right| \lambda(dy) \le B_1 + B_2 + B_3 \quad (10)$$

where,

$$\begin{aligned} B_1 &= \sum_{j=1}^{n_i} \int_{A_{i,j}} \left| \int_{\Theta} \psi(z_{i,j}, \theta) P_{x_i}(d\theta) - \int_{\Theta} \psi(z_{i,j}, \theta) P^0_{x_i}(d\theta) \right| \lambda(dy) \\ &= \sum_{j=1}^{n_i} \int_{A_{i,j}} \left| \int h^1_{i,j} dP_{x_i} - \int h^1_{i,j} dP^0_{x_i} \right| \lambda(dy) \\ &\le \nu \lambda(K_{x_i}), \end{aligned}$$

$$\begin{aligned}
B_2 &= \sum_{j=1}^{n_i} \int_{A_{i,j}} \left| \int_{\Theta} \psi\left(z_{i,j}, \theta\right) P_{x_i}^0\left(d\theta\right) - \int_{\Theta} \psi\left(y, \theta\right) P_{x_i}^0\left(d\theta\right) \right| \lambda\left(dy\right) \\
&\leq \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\overline{B}(0,r-\delta)} \left|\psi\left(z_{i,j}, y\right) - \psi\left(y, \theta\right)\right| P_{x_i}^0\left(d\theta\right) \lambda\left(dy\right) \\
&+ \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\overline{B}(0,r-\delta)^C} \left[\psi\left(z_{i,j}, \theta\right) + \psi\left(y, \theta\right)\right] P_{x_i}^0\left(d\theta\right) \lambda\left(dy\right) \\
&\leq \eta\lambda\left(K_{x_i}\right) + M_{x_i}\rho\lambda\left(K_{x_i}\right) + \rho,
\end{aligned}$$

where, $M_{x_i} = \max_{j\in\{1,\ldots,n_i\}} \sup_\theta \psi\left(z_{i,j}, \theta\right)$, and

$$B_3 = \sum_{j=1}^{n_i} \int_{A_{i,j}} \left| \int_{\Theta} \psi\left(z_{i,j}, \theta\right) P_{x_i}\left(d\theta\right) - \int_{\Theta} \psi\left(y, \theta\right) P_{x_i}\left(d\theta\right) \right| \lambda\left(dy\right).$$

Now, since

$$P_{x_i}\left[\overline{B}\left(0, r\right)^C\right] \leq \nu + \int h_{i,2}^0 dP_{x_i}^0 \leq \nu + P_{x_i}^0\left[\overline{B}\left(0, r-\nu\right)^C\right] \leq \nu + \rho,$$

it follows that

$$\begin{aligned}
B_3 &\leq \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\overline{B}(0,r)} \left|\psi\left(z_{i,j}, \theta\right) - \psi\left(y, \theta\right)\right| P_{x_i}\left(d\theta\right) \lambda\left(dy\right) \\
&+ \sum_{j=1}^{n_i} \int_{A_{i,j}} \int_{\overline{B}(0,r)^C} \left[\psi\left(z_{i,j}, \theta\right) + \psi\left(y, \theta\right)\right] P_{x_i}\left(d\theta\right) \lambda\left(dy\right) \\
&\leq \eta\lambda\left(K_{x_i}\right) + M_{x_i}\left(\nu + \rho\right)\lambda\left(K_{x_i}\right) + \nu + \rho.
\end{aligned}$$

Finally, by (9) and (10), if

$$\eta = \frac{\epsilon}{8\max_{1\leq i\leq T}\left\{\lambda\left(K_{x_i}\right)\right\}},$$

$$\nu = \frac{\epsilon}{4\left(2 + \max_{1\leq i\leq T}\left\{M_{x_i}\lambda\left(K_{x_i}\right)\right\}\right)},$$

and

$$\rho = \frac{\epsilon}{8\max_{1\leq i\leq T}\left\{\left(1 + M_{x_i}\lambda\left(K_{x_i}\right)\right)\right\}},$$

then $\int_{\mathscr{Y}} \left|\int_\Theta \psi\left(y, \theta\right) P_{x_i}\left(d\theta\right) - f_{x_i}^0\left(y\right)\right| \lambda\left(dy\right) \leq \epsilon$. Since $\epsilon > 0$ is arbitrary, the proof is complete. □

If stronger assumptions are placed on ψ, it is possible to show that DDP mixture models have large Kullback–Leibler support. Specifically, we consider the case where ψ belongs to an n–dimensional location–scale family of the form $\psi(\cdot, \theta) = \sigma^{-n} k\left(\frac{\cdot - \mu}{\sigma}\right)$, where $k(\cdot)$ is a probability density function defined on $\mathbb{R}^n$, $\mu = (\mu_1, \ldots, \mu_n)$ is an n–dimensional location vector, and $\sigma \in \mathbb{R}^+$. The following result characterizes the Kullback–Leibler support of the resulting DDP mixture models.

Theorem 5. *Assume that ψ belongs to a location–scale family, $\psi(\cdot, \theta) = \sigma^{-n} k\left(\frac{\cdot - \mu}{\sigma}\right)$, where $\mu = (\mu_1, \ldots, \mu_n)$ is an n–dimensional vector, and $\sigma \in \mathbb{R}^+$. Let k be a non–negative valued function defined on $(\mathscr{Y} \times \Theta, \mathscr{B}(\mathscr{Y}) \otimes \mathscr{B}(\Theta))$, where $\mathscr{Y} \subseteq \mathbb{R}^n$ is the sample space with corresponding Borel σ–field $\mathscr{B}(\mathscr{Y})$ and $\Theta \subseteq \mathbb{R}^n \times \mathbb{R}^+$ is the parameter space with corresponding Borel σ–field $\mathscr{B}(\Theta)$. Assume k satisfies the following conditions:*

(i) $k(\cdot)$ *is bounded, continuous and strictly positive,*

(ii) *there exists $l_1 > 0$ such that $k(z)$ decreases as z moves away from 0 outside the ball $\{z : \|z\| < l_1\}$, where $||\cdot||$ is the L_2–norm,*

(iii) *there exists $l_2 > 0$ such that $\sum_{j=1}^{n} z_j \left(\left. \frac{\partial k(t)}{\partial t_j} \right|_{t=z} \right) k(z)^{-1} < -1$, for $\|z\| \geq l_2$,*

(iv) *when $n \geq 2$, $k(z) = o(\|z\|)$ as $\|z\| \longrightarrow \infty$.*

Furthermore, assume the elements in $\{f^0_{x_i} : i = 1, \ldots, T\}$ satisfy the following conditions:

(v) *for some $M \in \mathbb{R}^+$, $0 < f^0_{x_i}(y) \leq M$, for every $y \in \mathbb{R}^n$,*

(vi) $\int f^0_{x_i}(y) \log\left(f^0_{x_i}(y)\right) dy < \infty$,

(vii) *for some $\delta > 0$, $\int f^0_{x_i}(y) \log\left(\frac{f^0_{x_i}(y)}{\inf_{||y-t||<\delta}\left\{f^0_{x_i}(t)\right\}}\right) dy < \infty$,*

(viii) *there exists $\eta > 0$, such that $\left|\int f^0_{x_i}(y) \log k\left(2y\|y\|^{\eta}\right) dy\right| < \infty$ and such that for any $a \in \mathbb{R}^n$ and $b \in \mathbb{R}^+$, we have $\int f^0_{x_i}(y) \left|\log k\left(\frac{y-a}{b}\right)\right| dy < \infty$.*

If $\{G_x : x \in \mathscr{X}\}$ is a DPP, *a* wDDP *or a* θDDP, *where $\mathbb{R}^n \times \mathbb{R}^+$ is the support of the corresponding centering distributions, and satisfying the conditions of Theorem 1, 2 or 3, respectively, then*

$$P\left\{\omega \in \Omega : d_{KL}\left[\int_{\mathbb{R}^n \times \mathbb{R}^+} \psi(\cdot, \theta)\, G(x_i, \omega)(d\theta), f^0_{x_i}\right] < \epsilon,\ i = 1, \ldots, T\right\} > 0,$$

for $\epsilon > 0$.

Proof: A direct application of Theorem 2 in Wu and Ghosal (2008), implies that there exist a probability measure $P^{\epsilon}_{x_i}$ and a weak neighborhood $\mathscr{W}_{x_i}$ such that

$$\int_{\mathscr{Y}} f^0_{x_i}(y) \log\left[\frac{f^0_{x_i}(y)}{\int_{\mathbb{R}^n \times \mathbb{R}^+} \psi(y, \theta)\, P^{\epsilon}_{x_i}(d\theta)}\right] dy < \frac{\epsilon}{2},$$

and

$$\int_{\mathscr{Y}} f^0_{x_i}(y) \log\left[\frac{\int_{\mathbb{R}^n\times\mathbb{R}^+} \psi(y,\theta)\, P^{\epsilon}_{x_i}(d\theta)}{\int_{\mathbb{R}^n\times\mathbb{R}^+} \psi(y,\theta)\, P_{x_i}(d\theta)}\right] dy < \frac{\epsilon}{2},$$

for every $P_{x_i} \in \mathscr{W}_{x_i}$ and $i = 1, \ldots, T$. Next note that

$$d_{KL}\left[\int_{\mathbb{R}^n\times\mathbb{R}^+} \psi(\cdot,\theta)\, P_{x_i}(d\theta)\,;\; f^0_{x_i}\right] < \int_{\mathscr{Y}} f^0_{x_i}(y) \log\left[\frac{f^0_{x_i}(y)}{\int_{\mathbb{R}^n\times\mathbb{R}^+} \psi(y,\theta)\, P^{\epsilon}_{x_i}(d\theta)}\right] dy$$
$$+ \int_{\mathscr{Y}} f^0_{x_i}(y) \log\left[\frac{\int_{\mathbb{R}^n\times\mathbb{R}^+} \psi(y,\theta)\, P^{\epsilon}_{x_i}(d\theta)}{\int_{\mathbb{R}^n\times\mathbb{R}^+} \psi(y,\theta)\, P_{x_i}(d\theta)}\right] dy,$$

and from Theorems 1, 2 and 3, it follows that

$$P\left\{\omega \in \Omega : d_{KL}\left[\int_{\mathbb{R}^n\times\mathbb{R}^+} \psi(\cdot,\theta)\, G(x_i,\omega)(d\theta)\,, f^0_{x_i}\right] < \epsilon,\; i = 1, \ldots, T\right\} \geq$$
$$P\left\{\omega \in \Omega : (G(x_1,\omega), \ldots, G(x_T,\omega)) \in \mathscr{W}_{x_1} \times \ldots \times \mathscr{W}_{x_T}\right\} > 0,$$

which completes the proof. □

Notice that the conditions of Theorem 5 are satisfied for most of the important location–scale kernels. In fact, Wu and Ghosal (2008) show that conditions (i) – (iv) are satisfied by the normal, skew–normal, double–exponential, logistic and t-Student kernels.

3.4 Extensions to more general dependent processes

Although the previous results about the support of models for collections of probability distributions are focused on MacEachern's DDP, similar results can be obtained for more general dependent processes. Natural candidates for the definition of dependent processes include the general class of stick–breaking (SB) processes, which includes the DP, the two–parameter Poisson–Dirichlet processes (Pitman and Yor 1997), the beta two–parameter processes (Ishwaran and James 2001) and the geometric stick–breaking processes (Mena et al. 2011), as important special cases. A SB probability measure is given by expression (1), but where the beta distribution associated with the SB construction of the weights can be replaced by any collection of distributions defined on the unit interval $[0,1]$ such that the resulting weights add up to one almost surely. Specifically, the weights are given by $W_i = V_i \prod_{j<i}(1-V_j)$, for every $i \geq 1$, where $V_i \mid H_i \stackrel{ind}{\sim} H_i$, with H_i being a probability measure on $[0,1]$, for every $i \in \mathbb{N}$, and such that

$$\sum_{i=1}^{\infty} W_i \stackrel{a.s.}{=} 1. \tag{11}$$

Notice that, under an SB prior, it can be shown that a necessary and sufficient condition for expression (11) to hold is that $\sum_{i=1}^{\infty} \log(1 - E_{H_i}(V_i)) = -\infty$.

For every $i \in \mathbb{N}$, let $\mathscr{C}_{\mathscr{X}}^{V_i}$ be a set of copulas satisfying the consistency conditions of Corollary 1 and set $\mathscr{C}_{\mathscr{X},\mathbb{N}}^{V} = \left\{\mathscr{C}_{\mathscr{X}}^{V_i} : i \in \mathbb{N}\right\}$. For every $i \in \mathbb{N}$, let $\mathscr{V}_{\mathscr{X}}^{V_i} = \{H_{(i,x)} : x \in \mathscr{X}\}$ be a collection of probability distributions defined on $([0,1], \mathscr{B}([0,1]))$ and set $\mathscr{V}_{\mathscr{X},\mathbb{N}} = \left\{\mathscr{V}_{\mathscr{X}}^{V_i} : i \in \mathbb{N}\right\}$.

Definition 4. *Let $\{G_x : x \in \mathscr{X}\}$ be a $\mathscr{P}(S)$–valued stochastic process on an appropriate probability space $(\Omega, \mathscr{A}, P)$ such that:*

(i) *$V_1, V_2, \ldots$ are independent stochastic processes of the form $V_i : \mathscr{X} \times \Omega \to [0,1]$, $i \geq 1$, with finite dimensional distributions determined by the set of copulas $\mathscr{C}_{\mathscr{X}}^{V_i}$ and the set of marginal distributions $\mathscr{V}_{\mathscr{X}}^{V_i}$, such that, for every $x \in \mathscr{X}$,*

$$\sum_{i=1}^{\infty} \log\left[1 - E_{H_{(i,x)}}(V_i(x,\cdot))\right] = -\infty.$$

(ii) *$\theta_1, \theta_2, \ldots$ are independent stochastic processes of the form $\theta_i : \mathscr{X} \times \Omega \to S$, $i \geq 1$, with common finite dimensional distributions determined by the set of copulas $\mathscr{C}_{\mathscr{X}}^{\theta}$ and the set of marginal distributions $G_{\mathscr{X}}^{0}$.*

(iii) *For every $x \in \mathscr{X}$, $B \in \mathscr{S}$ and almost every $\omega \in \Omega$,*

$$G(x,\omega)(B) = \sum_{i=1}^{\infty} \left\{ V_i(x,\omega) \prod_{j<i} [1 - V_j(x,\omega)] \right\} \delta_{\theta_i(x,\omega)}(B).$$

Such a process $\mathscr{H} = \{G_x \doteq G(x,\cdot) : x \in \mathscr{X}\}$ will be referred to as a dependent stick–breaking process with parameters $\left(\mathscr{C}_{\mathscr{X},\mathbb{N}}^{V}, \mathscr{C}_{\mathscr{X}}^{\theta}, \mathscr{V}_{\mathscr{X},\mathbb{N}}^{V}, G_{\mathscr{X}}^{0}\right)$, and denoted by $\text{DSBP}\left(\mathscr{C}_{\mathscr{X},\mathbb{N}}^{V}, \mathscr{C}_{\mathscr{X}}^{\theta}, \mathscr{V}_{\mathscr{X},\mathbb{N}}^{V}, G_{\mathscr{X}}^{0}\right)$.

As in the DDP case, two simplifications of the general definition of the DSBP can be considered. If the stochastic processes in (i) of Definition 4 are replaced by independent random variables with label–specific distribution H_i, then the process will be referred to as "single weights" DSBP, to emphasize the fact that the weights in the stick–breaking representation (iii) of Definition 4, are not indexed by predictors x. In this case, the process is parameterized by $\left(\mathscr{C}_{\mathscr{X}}^{\theta}, \mathscr{V}_{\mathbb{N}}^{V}, G_{\mathscr{X}}^{0}\right)$, and denoted by $\text{wDSBP}\left(\mathscr{C}_{\mathscr{X}}^{\theta}, \mathscr{V}_{\mathbb{N}}^{V}, G_{\mathscr{X}}^{0}\right)$, where $\mathscr{V}_{\mathbb{N}}^{V} = \{H_i : i \in \mathbb{N}\}$ is a collection of probability distributions on $[0,1]$, such that condition (11) holds. If the stochastic processes in (ii) of Definition 4 are replaced by independent random vectors with common distribution G^0, where G^0 is supported on the measurable space $(S, \mathscr{S})$, then the process will be referred to as "single atoms" DSBP, to emphasize the fact that the support points in the stick–breaking representation are not indexed by predictors x. This version of the process is parameterized by $\left(\mathscr{C}_{\mathscr{X},\mathbb{N}}^{V}, \mathscr{V}_{\mathscr{X},\mathbb{N}}^{V}, G^0\right)$, and denoted by $\theta\text{DSBP}\left(\mathscr{C}_{\mathscr{X},\mathbb{N}}^{V}, \mathscr{V}_{\mathscr{X},\mathbb{N}}^{V}, G^0\right)$.

Theorem 6. *Let* $\{G_x : x \in \mathscr{X}\}$ *be a* DSBP$\left(\mathscr{C}^{V}_{\mathscr{X},\mathbb{N}}, \mathscr{C}^{\theta}_{\mathscr{X}}, \mathscr{V}^{V}_{\mathscr{X},\mathbb{N}}, G^{0}_{\mathscr{X}}\right)$. *If* $\Theta \subseteq S$ *is the support of* G^0_x, *for every* $x \in \mathscr{X}$, $\mathscr{C}^{V}_{\mathscr{X},\mathbb{N}}$ *and* $\mathscr{C}^{\theta}_{\mathscr{X}}$ *are collections of copulas with positive density w.r.t. Lebesgue measure, on the appropriate unitary hyper–cubes, and, for every* $i \in \mathbb{N}$, *the elements in* $\mathscr{V}^{V_i}_{\mathscr{X}}$ *have positive density on* $[0,1]$, *then* $\mathscr{P}(\Theta)^{\mathscr{X}}$ *is the weak support of the process, i.e., the* DSBP *has full weak support.*

Proof: The proof follows similar arguments to the ones of Theorem 1. Specifically, it is only needed to replace

$$\int_{Q_1}\int_{Q_2(\mathbf{v}_1)}\cdots\int_{Q_{k+1}(\mathbf{v}_1,\ldots,\mathbf{v}_k)} f^{V}_{x_1,\ldots,x_T}(\mathbf{v}_1)\cdots f^{V}_{x_1,\ldots,x_T}(\mathbf{v}_{k+1})\,d\mathbf{v}_{k+1}\cdots d\mathbf{v}_2 d\mathbf{v}_1,$$

in expression (4) by

$$\int_{Q_1}\int_{Q_2(\mathbf{v}_1)}\cdots\int_{Q_{k+1}(\mathbf{v}_1,\ldots,\mathbf{v}_k)} f^{V_1}_{x_1,\ldots,x_T}(\mathbf{v}_1)\cdots f^{V_{k+1}}_{x_1,\ldots,x_T}(\mathbf{v}_{k+1})\,d\mathbf{v}_{k+1}\cdots d\mathbf{v}_2 d\mathbf{v}_1,$$

where $f^{V_i}_{x_1,\ldots,x_T}(\mathbf{v}_j)$, $j = 1,\ldots,k+1$, is the density function of

$$C^{V_i}_{x_1,\ldots,x_T}\left(H_{i,x_1}((0,v_1]),\ldots,H_{i,x_T}((0,v_T])\right).$$

The non–singularity of the $H_{(i,x)}$'s and of the associated copula functions imply that, for every $i \in \mathbb{N}$,

$$P\left\{\omega \in \Omega : [V_j(x_1,\omega),\cdots,V_j(x_T,\omega)] \in Q^{\omega}_j, j = 1,\ldots,k+1\right\} = \\ \int_{Q_1}\int_{Q_2(\mathbf{v}_1)}\cdots\int_{Q_{k+1}(\mathbf{v}_1,\cdots,\mathbf{v}_k)} f^{V_1}_{x_1,\ldots,x_T}(\mathbf{v}_1)\cdots f^{V_{k+1}}_{x_1,\ldots,x_T}(\mathbf{v}_{k+1})\,d\mathbf{v}_{k+1}\cdots d\mathbf{v}_2 d\mathbf{v}_1 > 0.$$

□

Theorem 7. *Let* $\{G_x : x \in \mathscr{X}\}$ *be a* wDSBP$\left(\mathscr{C}^{\theta}_{\mathscr{X}}, \mathscr{V}^{V}_{\mathbb{N}}, G^{0}_{\mathscr{X}}\right)$. *If* $\Theta \subseteq S$ *is the support of* G^0_x, *for every* $x \in \mathscr{X}$, $\mathscr{C}^{\theta}_{\mathscr{X}}$ *is a collection of copulas with positive density w.r.t. Lebesgue measure, on the appropriate unitary hyper–cubes, and, for every* $i \in \mathbb{N}$, H_i *has positive density on* $[0,1]$, *then* $\mathscr{P}(\Theta)^{\mathscr{X}}$ *is the weak support of the process, i.e., the* wDSBP *has full weak support.*

Proof: The non–singularity of the $H_{(i)}$'s implies that condition (5) holds, for every $i \in \mathbb{N}$. The rest of the proof remains the same as for Theorem 2. □

Theorem 8. *Let* $\{G_x : x \in \mathscr{X}\}$ *be a* θDSBP$\left(\mathscr{C}^{V}_{\mathscr{X},\mathbb{N}}, \mathscr{V}^{V}_{\mathscr{X},\mathbb{N}}, G^{0}\right)$, *where* Θ *is the support of* G_0. *If* $\mathscr{C}^{V}_{\mathscr{X},\mathbb{N}}$ *is a collection of copulas with positive density w.r.t. to Lebesgue measure,* *on the appropriate unitary hyper–cubes, and, for every $i \in \mathbb{N}$, the elements in $\mathscr{V}_{\mathscr{X}}^{V_i}$ have positive density on $[0,1]$, then the support of the process is $\mathscr{P}(\Theta)^{\mathscr{X}}$.*

Proof: The proof follows similar arguments to the ones of Theorem 3. It is only needed to replace

$$\int_{Q_1} \int_{Q_2(\mathbf{v}_1)} \cdots \int_{Q_{k+1}(\mathbf{v}_1,\ldots,\mathbf{v}_k)} f_{x_1,\ldots,x_T}^{V}(\mathbf{v}_1) \cdots f_{x_1,\ldots,x_T}^{V}(\mathbf{v}_{k+1})\, d\mathbf{v}_{k+1} \cdots d\mathbf{v}_2 d\mathbf{v}_1,$$

in expression (6) by

$$\int_{Q_1} \int_{Q_2(\mathbf{v}_1)} \cdots \int_{Q_{k+1}(\mathbf{v}_1,\ldots,\mathbf{v}_k)} f_{x_1,\ldots,x_T}^{V_1}(\mathbf{v}_1) \cdots f_{x_1,\ldots,x_T}^{V_{k+1}}(\mathbf{v}_{k+1})\, d\mathbf{v}_{k+1} \cdots d\mathbf{v}_2 d\mathbf{v}_1,$$

where $f_{x_1,\ldots,x_T}^{V_i}(\mathbf{v}_j)$, $j = 1, \ldots, k+1$, is the density function of

$$C_{x_1,\ldots,x_T}^{V_i}\left(H_{i,x_1}\left((0,v_1]\right), \ldots, H_{i,x_T}\left((0,v_T]\right)\right).$$

The non–singularity of the $H_{(i,x)}$'s and of the associated copula functions imply that, for every $i \in \mathbb{N}$,

$$P\left\{\omega \in \Omega : \left[V_j(x_1,\omega), \cdots, V_j(x_T,\omega)\right] \in Q_j^{\omega}, j = 1, \ldots, k+1\right\} =$$
$$\int_{Q_1} \int_{Q_2(\mathbf{v}_1)} \cdots \int_{Q_{k+1}(\mathbf{v}_1,\cdots,\mathbf{v}_k)} f_{x_1,\ldots,x_T}^{V_1}(\mathbf{v}_1) \cdots f_{x_1,\ldots,x_T}^{V_{k+1}}(\mathbf{v}_{k+1})\, d\mathbf{v}_{k+1} \cdots d\mathbf{v}_2 d\mathbf{v}_1 > 0.$$

□

Since the proofs of Theorems 4 and 5 depend on the dependent mixing distributions through their weak support only, the results are also valid for the different versions of the DSBP. Thus, the following theorems are stated without any proof.

Theorem 9. *Let ψ be a non–negative valued function defined on the product measurable space $(\mathscr{Y} \times \Theta, \mathscr{B}(\mathscr{Y}) \otimes \mathscr{B}(\Theta))$, where $\mathscr{Y} \subseteq \mathbb{R}^n$ is the sample space with corresponding Borel σ–field $\mathscr{B}(\mathscr{Y})$ and $\Theta \subseteq \mathbb{R}^q$ is the parameter space with corresponding Borel σ–field $\mathscr{B}(\Theta)$. Assume that ψ satisfies conditions (i) – (iii) of Theorem 4. If $\{G_x : x \in \mathscr{X}\}$ is a DSBP, a wDSBP or a θDSBP, satisfying the conditions of Theorem 6, 7 or 8, respectively, then the Hellinger support of the process $\left\{\int_\Theta \psi(\cdot,\theta)\, G_x(d\theta) : x \in \mathscr{X}\right\}$ is*

$$\prod_{x \in \mathscr{X}} \left\{\int_\Theta \psi(\cdot,\theta)\, P_x(d\theta) : P_x \in \mathscr{P}(\Theta)\right\},$$

where $\mathscr{P}(\Theta)$ is the space of all probability measures defined on $(\Theta, \mathscr{B}(\Theta))$.

Theorem 10. *Assume that ψ belongs to a location–scale family, $\psi(\cdot,\theta) = \sigma^{-n} k\left(\frac{\cdot-\mu}{\sigma}\right)$, where $\mu = (\mu_1, \ldots, \mu_n)$ is an n–dimensional vector, and $\sigma \in \mathbb{R}^+$. Let k be a non–negative* *valued function defined on $(\mathscr{Y} \times \Theta, \mathscr{B}(\mathscr{Y}) \otimes \mathscr{B}(\Theta))$, where $\mathscr{Y} \subseteq \mathbb{R}^n$ is the sample space with corresponding Borel σ–field $\mathscr{B}(\mathscr{Y})$ and $\Theta \subseteq \mathbb{R}^n \times \mathbb{R}^+$ is the parameter space with corresponding Borel σ–field $\mathscr{B}(\Theta)$. Assume k satisfies conditions (i) – (iv) of Theorem 5 and that the elements in $\{f_{x_i}^0 : i = 1, \ldots, T\}$ satisfy conditions (v) – (viii) of Theorem 5. If $\{G_x : x \in \mathscr{X}\}$ is a DSBP, a wDSBP or a θDSBP, where $\mathbb{R}^n \times \mathbb{R}^+$ is the support of the corresponding centering distributions, and satisfying the conditions of Theorem 6, 7 or 8, respectively, then*

$$P\left\{\omega \in \Omega : d_{KL}\left[\int_{\mathbb{R}^n \times \mathbb{R}^+} \psi\left(\cdot, \theta\right) G\left(x_i, \omega\right)\left(d\theta\right), f_{x_i}^0\right] < \epsilon,\ i = 1, \ldots, T\right\} > 0,$$

for $\epsilon > 0$.

4 Concluding remarks and future research

We have studied the support properties of DDP and DDP mixture models, as well as those of more general dependent stick–breaking processes. By exploiting the connection between copulas and stochastic processes, we have provided sufficient conditions for weak and Hellinger support of models based on DDP's. We have also characterized the Kullback–Leibler support of mixture models induced by DDP's and showed that the results can be generalized for the class of dependent stick–breaking processes. Several versions of the DDP were considered, in particular a version where only the weights are indexed by the predictors. The results suggest that we may consider parsimonious models that index only the weights or only the support points by the predictors, while retaining the appealing support properties of a full DDP model. This opens new possibilities for the development of single–atoms DDP models, for which there is a scarcity of literature. In particular, a back–to–back comparison of these simplified models is of interest.

The results on the support of MacEachern's DDP, DSBP and their induced mixture models provided here can be useful for studying frequentist asymptotic properties of the posterior distribution in these models. In fact, using the same strategy adopted in Norets and Pelenis (2011) and Pati et al. (2011), the weak and strong consistency of the different versions of MacEachern's DDP and DSBP mixture models could be anticipated. These authors study the frequentist consistency of the posterior distribution of the induced joint model for responses and predictors, (y, x), under iid sampling. Therefore, the asymptotic properties provided by these authors are based on the consistency results for single density estimation problems. Our approach differs from these works in that we adopt a conditional framework (of the responses given the predictors), which implies the need to work with product spaces. The study of the asymptotic behavior in the conditional context is also of interest and is the subject of ongoing research.

References

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York, USA: Dover. 282

Caron, F., Davy, M., Doucet, A., Duflos, E., and Vanheeghe, P. (2008). "Bayesian inference for linear dynamic models with Dirichlet process mixtures." *IEEE Transactions on Signal Processing*, 56: 71–84. 278

Chung, Y. and Dunson, D. B. (2009). "Nonparametric Bayes conditional distribution modeling with variable selection." *Journal of the American Statistical Association*, 104: 1646–1660. 278, 279

Cifarelli, D. and Regazzini, E. (1978). "Problemi statistici non parametrici in condizioni di scambiabilità parziale e impiego di medie associative." Technical report, Quaderni Istituto Matematica Finanziaria, Torino. 278

Darsow, W. F., Nguyen, B., and Olsen, E. T. (1992). "Copulas and Markov processes." *Illinois Journal of Mathematics*, 36: 600–642. 281

De Iorio, M., Johnson, W. O., Müller, P., and Rosner, G. L. (2009). "Bayesian nonparametric non-proportional hazards survival modelling." *Biometrics*, 65: 762–771. 278, 289

De Iorio, M., Müller, P., Rosner, G. L., and MacEachern, S. N. (2004). "An ANOVA model for dependent random measures." *Journal of the American Statistical Association*, 99: 205–215. 278, 289

De la Cruz, R., Quintana, F. A., and Müller, P. (2007). "Semiparametric Bayesian classification with longitudinal markers." *Applied Statistics*, 56(2): 119–137. 278, 289

Dunson, D. B. and Herring, A. H. (2006). "Semiparametric Bayesian latent trajectory models." Technical report, ISDS Discussion Paper 16, Duke University, Durham, NC, USA. 278

Dunson, D. B. and Park, J. H. (2008). "Kernel stick-breaking processes." *Biometrika*, 95: 307–323. 278, 279

Dunson, D. B., Pillai, N., and Park, J. H. (2007). "Bayesian density regression." *Journal of the Royal Statistical Society, Series B*, 69: 163–183. 278

Epifani, I. and Lijoi, A. (2010). "Nonparametric priors for vectors of survival functions." *Statistica Sinica*, 20: 1455–1484. 281

Escobar, M. D. and West, M. (1995). "Bayesian density estimation and inference using mixtures." *Journal of the American Statistical Association*, 90: 577–588. 277

Ferguson, T. S. (1973). "A Bayesian analysis of some nonparametric problems." *Annals of Statistics*, 1: 209–230. 277, 286

— (1974). "Prior distribution on the spaces of probability measures." *Annals of Statistics*, 2: 615–629. 277

— (1983). "Bayesian density estimation by mixtures of normal distributions." In Siegmund, D., Rustage, J., and Rizvi, G. G. (eds.), *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on His Sixtieth Birthday*, 287–302. Bibliohound. 277

Gelfand, A. E., Kottas, A., and MacEachern, S. N. (2005). "Bayesian nonparametric spatial modeling with Dirichlet process mixing." *Journal of the American Statistical Association*, 100: 1021–1035. 278, 289

Giudici, P., Mezzetti, M., and Muliere, P. (2003). "Mixtures of Dirichlet process priors for variable selection in survival analysis." *Journal of Statistical Planning and Inference*, 111: 101–115. 278

Ishwaran, H. and James, L. F. (2001). "Gibbs sampling methods for stick-breaking priors." *Journal of the American Statistical Association*, 96: 161–173. 300

Jara, A., Lesaffre, E., De Iorio, M., and Quintana, F. A. (2010). "Bayesian semiparametric inference for multivariate doubly-interval-censored data." *Annals of Applied Statistics*, 4: 2126–2149. 278, 289

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. New York, USA: Chapman and Hall/CRC. 280

Kolmogorov, A. N. (1933). *Foundations of the theory of probability*. New York, USA: Nathan Morrison, 2nd edition. 280

Leisen, F. and Lijoi, A. (2011). "Vectors of two–parameter Poisson–Dirichlet processes." *Journal of Multivariate Analysis*, 102: 482–495. 281

Lijoi, A., Prünster, I., and Walker, S. (2004). "Extending Doob's consistency theorem to nonparametric densities." *Bernoulli*, 10: 651–663. 296

Lo, A. Y. (1984). "On a class of Bayesian nonparametric estimates I: Density estimates." *The Annals of Statistics*, 12: 351–357. 277, 295

MacEachern, S. N. (1999). "Dependent nonparametric processes." In *ASA Proceedings of the Section on Bayesian Statistical Science, Alexandria, VA*. American Statistical Association. 278, 279

— (2000). "Dependent Dirichlet processes." Technical report, Department of Statistics, The Ohio State University. 278, 279

Mena, R. H., Ruggiero, M., and Walker, S. G. (2011). "Geometric stick-breaking processes for continuous-time Bayesian nonparametric modeling." *Journal of Statistical Planning and Inference*, 141: 3217–3230. 300

Mira, A. and Petrone, S. (1996). "Bayesian hierarchical nonparametric inference for change-point problems." In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M. (eds.), *Bayesian Statistics 5*. Oxford University Press. 278

Muliere, P. and Petrone, S. (1993). "A Bayesian predictive approach to sequential search for an optimal dose: parametric and nonparametric models." *Journal of the Italian Statistical Society*, 2: 349–364. 278

Müller, P., Erkanli, A., and West, M. (1996). "Bayesian curve fitting using multivariate normal mixtures." *Biometrika*, 83: 67–79. 278

Müller, P., Rosner, G. L., De Iorio, M., and MacEachern, S. (2005). "A nonparametric Bayesian model for inference in related longitudinal studies." *Journal of the Royal Statistical Society, Series C*, 54: 611–626. 278, 289

Norets, A. (2010). "Approximation of conditional densities by smooth mixtures of regressions." *The Annals of Statistics*, 38: 1733–1766. 279

Norets, A. and Pelenis, J. (2011). "Posterior consistency in conditional density estimation by covariate dependent mixtures." Technical report, Princeton University, Economics Department. 279, 304

Parthasarathy, K. R. (1967). *Probability Measures in Metric Spaces*. Providence, USA: AMS Chelsea Publishing. 308

Pati, D., Dunson, D., and Tokdar, S. (2011). "Posterior consistency in conditional distribution estimation." Technical report, Duke University, Department of Statistical Science. 279, 304

Pitman, J. and Yor, M. (1997). "The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator." *The Annals of Probability*, 25: 855–900. 281, 300

Sethuraman, J. (1994). "A constructive definition of Dirichlet priors." *Statistica Sinica*, 2: 639–650. 278

Sklar, A. (1959). "Fonctions de répartition à n dimensions et leurs marges." *Publications de l'Institut de Statistique de L'Université de Paris*, 8: 229–231. 280

Wu, Y. and Ghosal, S. (2008). "Kullback Leibler property of kernel mixture priors in Bayesian density estimation." *Electronic Journal of Statistics*, 2: 298–331. 299, 300

Appendix A

Lemma 1. *Let $\mathscr{P}(\Theta)$ be the space of all probability measures defined on $(\Theta, \mathscr{B}(\Theta))$. Let G_0 be an absolutely continuous probability measure w.r.t. Lebesgue measure, with support Θ. Let*

$$U(P_0, f_1, \ldots, f_k, \epsilon) = \left\{ P \in \mathscr{P}(\Theta) : \left| \int f_i dP - \int f_i dP_0 \right| < \epsilon,\ i = 1 \ldots k \right\}$$

be a weak neighborhood of $P_0 \in \mathscr{P}(\Theta)$, where ϵ is a positive constant and f_i, $i = 1,\dots,k$, are bounded continuous functions. Then there exists a probability measure in $U(P_0, f_1,\dots,f_k,\epsilon)$ which is absolutely continuous w.r.t. G_0.

Proof: Since the set of all probability measures whose supports are finite subsets of a dense set in Θ is dense in $\mathscr{P}(\Theta)$ (Parthasarathy 1967, page 44), there exists a probability measure $Q^*(\cdot) = \sum_{j=1}^N W_j \delta_{\theta_j}(\cdot)$, where $N \in \mathbb{N}$, $(W_1,\dots,W_N) \in \Delta_N$, with $\Delta_N = \{w_1,\dots,w_N : w_i \geq 0, i = 1,\dots,N, \sum_{i=1}^N w_i = 1\}$ denoting the N–simplex, and different support points $\theta_1,\dots,\theta_N \in \Theta$, such that

$$\left| \int f_i dQ^* - \int f_i dP_0 \right| < \frac{\epsilon}{2},\ i = 1,\dots,k.$$

In addition, there exists an open ball of radius $\delta > 0$, denoted by $B(\theta_j,\delta)$, such that for every $\theta \in B(\theta_j,\delta)$, with $B(\theta_l,\delta) \bigcap B(\theta_j,\delta) = \emptyset$, for every $l \neq j$, $f_i(\theta)$ satisfies the following relation

$$f_i(\theta_j) - \frac{\epsilon}{2N} < f_i(\theta) < f_i(\theta_j) + \frac{\epsilon}{2N}.$$

Now, let Q be a probability measure with density function given by

$$q(\theta) = \sum_{j=1}^N \frac{W_j}{c_{\theta_j,\delta}} I_{B(\theta_j,\delta)\bigcap\Theta}(\theta),$$

where $c_{\theta_j,\delta}$ denotes the Lebesgue measure of $B(\theta_j,\delta)\bigcap\Theta$ and $I_A(\cdot)$ is the indicator function of the set A. It follows that

$$W_j f_i(\theta_j) - W_j\left(f_i(\theta_j) + \frac{\epsilon}{2N}\right) < \\ W_j f_i(\theta_j) - \int_{B(\theta_j,\delta)} f_i(\theta) q(\theta) d\theta < W_j f_i(\theta_j) - W_j\left(f_i(\theta_j) - \frac{\epsilon}{2N}\right),$$

and

$$\left| W_j f_i(\theta_j) - \int_{B(\theta_j,\delta)} f_i(\theta) q(\theta) d\theta \right| < \frac{\epsilon}{2N},$$

which implies that

$$\left| \int f_i dQ^* - \int f_i(\theta) q(\theta) d\theta \right| < \sum_{j=1}^N \left| W_j f_i(\theta_j) - \int_{B(\theta_j,\delta)} f_i(\theta) q(\theta) d\theta \right| < \frac{\epsilon}{2}.$$

Thus,

$$\left| \int f_i dQ - \int f_i dP_0 \right| \leq \left| \int f_i dQ^* - \int f_i dP_0 \right| + \left| \int f_i dQ - \int f_i dQ^* \right| \leq \epsilon,$$

and therefore, $Q \in U(P_0, f_1,\dots,f_k,\epsilon)$. Moreover, the support of Q is contained in Θ, i.e., Q is an absolutely continuous probability measure w.r.t. G_0. $\square$

Acknowledgments

The first author was supported by the "Programa Becas Latinoamericanas de CONICYT". The second author was supported by Fondecyt 11100144 grant. The third author was supported by Fondecyt 1100010 grant. The authors are grateful to the associate editor and reviewer for many detailed suggestions and helpful comments that have led to a considerable improvement in the manuscript.