Electronic Journal of Statistics
Vol. 18 (2024) 2064–2134
ISSN: 1935-7524
https://doi.org/10.1214/24-EJS2245

On dependent Dirichlet processes for general Polish spaces

Andrés Iturriaga[1], **Carlos A. Sing Long** [2] **and Alejandro Jara**[3]

[1] *Departamento de Matemática y Ciencia de la Computación,*
Universidad de Santiago de Chile,
Santiago, Chile
e-mail: andres.iturriaga@usach.cl

[2] *Institute for Mathematical and Computational Engineering, and*
Institute for Biological and Medical Engineering,
Pontificia Universidad Católica de Chile,
Santiago, Chile
e-mail: casinglo@uc.cl

[3] *Department of Statistics,*
Pontificia Universidad Católica de Chile,
Santiago, Chile
e-mail: atjara@uc.cl

Abstract: We study Dirichlet process–based models for sets of predictor–dependent probability distributions, where the domain and predictor space are general Polish spaces. We generalize the definition of dependent Dirichlet processes, originally constructed on Euclidean spaces, to more general Polish spaces. We provide sufficient conditions under which dependent Dirichlet processes and dependent Dirichlet process mixture models have appealing properties regarding continuity (weak and strong), association structure, and support (under different topologies). The results can be easily extended to more general dependent stick-breaking processes.

MSC2020 subject classifications: Primary 62C10; secondary 60G57.
Keywords and phrases: Related random probability distributions, Bayesian nonparametrics, support of random measures, non-standard spaces.

Received August 2022.

Contents

arXiv: 2205.05635

1. Introduction

Standard regression approaches assume that some finite number of the response distribution characteristics, such as location and scale, change as a (parametric or nonparametric) function of predictors. However, it is not always appropriate to assume a location/scale representation, where the error distribution has unchanging shape over the predictor space, $\mathcal{X}$. In fact, it often happens in applied research that the distribution of responses under study changes with predictors in ways that cannot be reasonably represented by a finite dimensional functional form [see, e.g., 26, 29, 25, 44, 42]. Because this can seriously affect the answers to the scientific questions of interest, more general regression approaches have been developed, which are usually referred to as density regression or fully nonparametric regression [see, e.g., 66, 73].

Fully nonparametric regression or nonparametric conditional density estimation can be seen as an extension of traditional regression models, where it is assumed that $\boldsymbol{y}_i \mid F_{\boldsymbol{x}_i} \overset{ind.}{\sim} F_{\boldsymbol{x}_i}$, $i = 1, \ldots, n$, and that the parameter of interest is the set of predictor-dependent probability measures $\mathcal{F} = \{F_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$, where $F_{\boldsymbol{x}}$ is a probability measure defined on the sample space $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$, with $\mathcal{B}(\mathcal{Y})$ being the Borel σ-field of $\mathcal{Y}$. This paper focuses on the definition and the properties of Bayesian nonparametric (BNP) priors for the set $\mathcal{F}$, where the either the sample $\mathcal{Y}$ or the predictor space $\mathcal{X}$ can be general Polish spaces.

Most of the BNP priors used to account for the dependence of predictors on set of probability measures $\mathcal{F}$ are generalizations of the Dirichlet process (DP) [31, 32] and Dirichlet process mixture (DPM) models [58]. Let $\mathcal{D}(\mathcal{Y})$ be the space of all probability measures, with density w.r.t. Lebesgue measure, defined on $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$. A DPM model is a stochastic process, F, defined on an appropriate probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that for almost every $\omega \in \Omega$, the density function of F is given by

$$f(\boldsymbol{y} \mid G(\omega)) = \int_{\Theta} \psi(\boldsymbol{y}, \boldsymbol{\theta}) G(\omega)(d\boldsymbol{\theta}), \quad \boldsymbol{y} \in \mathcal{Y}, \tag{1}$$

where $\psi(\cdot, \boldsymbol{\theta})$ is a continuous density function on $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$, for every $\boldsymbol{\theta} \in \Theta$, and G is a DP, whose sample paths are probability measures defined on $(\Theta, \mathcal{B}(\Theta))$, with $\mathcal{B}(\Theta)$ being the corresponding Borel σ-field. If G is a DP with parameters (M, G_0), where $M \in \mathbb{R}_0^+$ and G_0 is a probability measure on $(\Theta, \mathcal{B}(\Theta))$, written as $G \mid M, G_0 \sim \mathrm{DP}(MG_0)$, then the trajectories of the process can be a.s. represented by the following stick-breaking representation [79]: $G(B) = \sum_{i=1}^{\infty} \pi_i \delta_{\boldsymbol{\theta}_i}(B)$, $B \in \mathcal{B}(\Theta)$, where $\delta_{\boldsymbol{\theta}}(\cdot)$ is the Dirac measure at $\boldsymbol{\theta}$, $\pi_i = V_i \prod_{j<i}(1 - V_j)$, with $V_i \mid M \overset{iid}{\sim} \mathrm{Beta}(1, M)$, and $\boldsymbol{\theta}_i \mid G_0 \overset{iid}{\sim} G_0$. Discussion of properties and applications of DP can be found, for instance, in [66].

Most of the BNP extensions incorporate dependence on predictors via the mixing distribution in (1), by replacing G with $G_{\boldsymbol{x}}$, and the prior specification problem is related to the modeling of the collection of predictor-dependent mixing probability measures $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ [73]. Some of the earliest developments on predictor-dependent DP models appeared in [22], who defined dependence across related random measures by introducing a regression for the baseline measure of marginally DP random measures. A more flexible construction was proposed by [59], called the dependent Dirichlet process (DDP). The key idea behind the DDP is to create a set of marginally DP random measures and to introduce dependence by modifying the stick-breaking representation of each element in the set. Specifically, [59] generalized the stick-breaking representation by assuming $G_{\boldsymbol{x}}(B) = \sum_{i=1}^{\infty} \pi_i(\boldsymbol{x})\delta_{\boldsymbol{\theta}_i(\boldsymbol{x})}(B)$, $B \in \mathcal{B}(\Theta)$, where the point masses $\boldsymbol{\theta}_i(\boldsymbol{x})$, $i \in \mathbb{N}$, are independent stochastic processes with index set $\mathcal{X}$, and the weights take the form $\pi_i(\boldsymbol{x}) = V_i(\boldsymbol{x}) \prod_{j<i}[1 - V_j(\boldsymbol{x})]$, with $V_i(\boldsymbol{x})$, $i \in \mathbb{N}$, being independent stochastic processes with index set $\mathcal{X}$ and $\text{Beta}(1, M)$ marginal distribution. We refer the reader to [6] for a formal definition of the DDP.

Other extensions of the DP for dealing with related probability distributions include the DPM mixture of normals model for the joint distribution of the response and predictors [64], the hierarchical mixture of DPM [67], the predictor-dependent weighted mixture of DP [29], the kernel-stick breaking process [28], the probit-stick breaking processes [21, 76], the cluster-X model [65], the PPMx model [68], and the general class of stick-breaking processes [6], among many others. Dependent neutral to the right processes and correlated two-parameter Poisson-Dirichlet processes have been proposed by [30] and [56], respectively, by considering suitable Lévy copulas. The general class of dependent normalized completely random measures has been discussed, for instance, in [57]. Based on a different formulation of the conditional density estimation problem, [81] and [43] proposed alternatives to convolutions of dependent stick-breaking approaches.

All of the dependent BNP approaches described previously have focused on responses and parameters defined on Euclidean spaces, and are not appropriate for spaces in which the Euclidean geometry is not valid. A relevant example of this situation arises in statistical shape analysis, where one of the main spaces of interest is Kendall's shape space [48], which can be viewed as the quotient of a Riemannian manifold. Kendall's space is a natural underlying space for applications in different areas, including morphometry [23], meteorology [61], archeology [27] and genetics [17]. In these contexts, to employ standard statistical procedures that do not take into account the geometrical properties of the underlying spaces can lead to wrong inferences, which explains the increasing interest in the development of statistical models for more general Polish spaces.

To date, the development of statistical procedures for non Euclidean spaces has focused on the problem of mean estimation [see, e.g., 14, 15, 16], density estimation [see, e.g., 70, 11, 12], and on the regression problem for Euclidean responses based on non Euclidean predictors [see, e.g., 71, 13]. [14, 15, 16] studied the problem of nonparametric estimation of a location parameter on a Riemannian manifold, by means of the concept of the Fréchet mean [37], and derive its asymptotic distribution. [70] studied the density estimation problem on a compact Riemannian manifold, by adapting kernel-type techniques. [11, 12] considered the problem of density estimation for data supported on a complete metric space from a BNP point of view. [71] considered the nonparametric regression problem for a real-valued response and with predictors supported on a closed Riemannian manifold. [13] studied the problem of prediction of a categorical response based on predictors supported on a general manifold. Related to these problems, the authors in [18] construct a suitable prior for real-valued functions defined on compact manifolds, and study the asymptotic behavior of the posterior distribution in problems of mean estimation, density estimation, and standard regression.

This work has three main parts. In the first, we generalize the definition of a DDP, originally proposed on Euclidean spaces, to more general Polish spaces and establish its basic properties. It is important to stress that the existing DDP definitions given in [59, 60] and [6] cannot be directly extended to more general spaces because they make use of the concept of cumulative distribution function. In our definition, the existence of a DDP in a general Polish space is justified by the extension of Kolmogorov's consistency theorem proposed in [69]. In Section 3 we define the DDP for general Polish spaces, and introduce some parsimonious variants that share similar properties. In Section 4.1 we provide sufficient conditions under which the sample paths of a DDP are continuous under the weak, strong and uniform topologies on the space of probability measures. In Section 4.2 we study the support of a DDP under different topologies, aiming to provide sufficient conditions under which a DDP has full support, or at least contains a sufficiently large set of functions of interest. Finally, in Section 4.3 we provide conditions under which a measure of association is continuous.

In the second part, we focus on DDP mixture models, providing sufficient conditions under which they have a continuous density with respect to a base measure, and have appealing properties regarding continuity, support, association structure, and consistency under i.i.d. sampling. Section 5 introduces a framework to study DDP mixtures and its variants, with a focus on the regularity of the probability density of the mixture. In Section 6.1 we provide sufficient conditions for a DDP mixture model to have continuous sample paths. In particular, the strong regularizing properties of a mixture allows us to show that under mild conditions the sample paths are almost surely uniformly continuous. In Section 6.2 we leverage the fact that the mixtures we study have a density to study the support in the topologies induced by the Hellinger and L^∞ distances, and by the Kullback-Leibler divergence. In Section 6.3 we briefly discuss the regularity of some measures of association. In Section 6.4 we discuss the conditions under which the mixture satisfies posterior consistency.

Finally, in Section 7 we discuss nontrivial examples where the response lies on a non Euclidean space. The first example deals with circular data, whereas the second deals with shape analysis in Kendall's shape space. In both cases we present explicit constructions and show how they satisfy our sufficient conditions for the models to have appealing theoretical properties. We conclude the article with some brief remarks.

2. Preliminaries

In this work, we suppose that we observe regression data $\{(\boldsymbol{x}_i, y_i) : i = 1, \dots, n\}$, where $\boldsymbol{x}_i \in \mathcal{X}$ is a p-dimensional vector of exogenous predictors. Notice that the exogeneity assumption allows us to treat the problem of conditional density estimation as a fully non-parametric regression problem, regardless of the data generating mechanism of the predictors, that is, to treat predictors as fixed by design even though they are randomly generated [see, e.g. 4, 5, 36].

We let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space. If S is a measurable space and $X : \Omega \to S$ is a S-valued random variable, we usually write X^ω or $X(\omega)$ to denote its value at ω. If $\{X_t : t \in T\}$ is a process of S-valued random variables on T we write $X_t^\omega := X_t(\omega) := X(t, \omega)$ to denote its values at (t, ω). In particular, outcomes always appear as superscripts and as the last argument of a function.

If $\mathcal{X}, \mathcal{Y}$ are two sets we denote a generic element with the same letter in lowercase boldface. If $F : \mathcal{X} \times \mathcal{Y} \to S$ then we write $F_{\boldsymbol{x}}$ for a fixed $\boldsymbol{x} \in \mathcal{X}$ to denote the function $F_{\boldsymbol{x}} : \mathcal{Y} \to S$ defined as $F_{\boldsymbol{x}}(\boldsymbol{y}) = F(\boldsymbol{x}, \boldsymbol{y})$. Similarly, we denote $F_{\boldsymbol{y}}$ for fixed $\boldsymbol{y} \in \mathcal{Y}$ the function defined as $F_{\boldsymbol{y}}(\boldsymbol{x}) = F(\boldsymbol{x}, \boldsymbol{y})$.

We recall that a Polish space is a separable and completely metrizable topological space. The sets $\mathcal{X}, \mathcal{Y}$, and Θ are always assumed to be Polish spaces, with complete metrics $d_{\mathcal{X}}, d_{\mathcal{Y}}$, and d_Θ, respectively.

The set Θ represents the set of parameters onto which we define a prior. Hence, we endow Θ with the Borel σ-algebra $\mathcal{B}(\Theta)$ and let $\mathcal{P}(\Theta)$ be the space of probability measures on $(\Theta, \mathcal{B}(\Theta))$. Since Θ is Polish, the elements of $\mathcal{P}(\Theta)$ are regular [24], i.e., for any $P \in \mathcal{P}(\Theta)$ and $B \in \mathcal{B}(\Theta)$ we have that

$$\begin{aligned} P(B) &= \inf\{P(U) : B \subset U, U \text{ open}\}, \\ P(B) &= \sup\{P(K) : B \supset K, K \text{ compact}\}. \end{aligned}$$

In particular, finite collections of elements in $\mathcal{P}(\Theta)$ are always tight. We let $C_b(\Theta)$ be the space of real-valued, bounded continuous functions on Θ endowed with the norm $\|f\|_C := \sup\{|f(\boldsymbol{\theta})| : \boldsymbol{\theta} \in \Theta\}$.

If n is a positive integer, we let $[n] := \{1, \ldots, n\}$. Finally, we use the abbreviations "a.s." for *almost surely*, "a.e." for *almost everywhere*, and "i.i.d." for *independent and identically distributed*.

3. Dependent Dirichlet processes on Polish spaces

Dependent Dirichlet processes (DDP) are a class of $\mathcal{P}(\Theta)$-valued stochastic processes on $\mathcal{X}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$. If $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ is a DDP then

$$G_{\boldsymbol{x}}^\omega = \sum_{i=1}^{\infty} \pi_{i,\boldsymbol{x}}^\omega \delta_{\boldsymbol{\theta}_{i,\boldsymbol{x}}^\omega},$$

a.s. for a sequence $(\{\pi_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i \in \mathbb{N}}$ of processes such that $\pi_{i,\boldsymbol{x}} \geq 0$ for every $i \in \mathbb{N}$ and $\boldsymbol{x} \in \mathcal{X}$ and $\sum_{i \in \mathbb{N}} \pi_{i,\boldsymbol{x}} \equiv 1$ a.s., and a sequence $(\{\boldsymbol{\theta}_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i \in \mathbb{N}}$ of i.i.d. Θ-valued processes. The distinctive feature of the DDP is that the processes $(\{\pi_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i \in \mathbb{N}}$ are defined in terms of a stick-breaking process. Let $(\{V_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i \in \mathbb{N}}$ be a sequence of i.i.d. processes with $V_{i,\boldsymbol{x}} \sim \text{Beta}(1, \alpha_{\boldsymbol{x}})$ for some $\alpha_{\boldsymbol{x}} \in \mathbb{R}_+$ and any $\boldsymbol{x} \in \mathcal{X}$. Then, the stick-breaking process asociated to $(\{V_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i \in \mathbb{N}}$ is

$$\pi_{i,\boldsymbol{x}}^\omega := \begin{cases} V_{i,\boldsymbol{x}}^\omega & i = 1, \\ V_{i,\boldsymbol{x}}^\omega \prod_{j=1}^{i-1}(1 - V_{j,\boldsymbol{x}}^\omega) & i > 1. \end{cases} \tag{2}$$

We now generalize the definition in [6] to Polish spaces.

Definition 3.1. A *dependent Dirichlet process (DDP)* with parameters (Ψ_V, Ψ_Θ), denoted as $G_{\mathcal{X}} \sim \mathrm{DDP}(\Psi_V, \Psi_\Theta)$, is a $\mathcal{P}(\Theta)$-valued stochastic process $G_{\mathcal{X}} = \{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that:

1. There exists a sequence $(\{V_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i\in\mathbb{N}}$ of separable i.i.d. processes, with a law characterized by a finite-dimensional parameter Ψ_V, and with marginal distribution $\mathrm{Beta}(1, \alpha_{\boldsymbol{x}})$ for some $\alpha_{\boldsymbol{x}} \geq 0$ for any $\boldsymbol{x} \in \mathcal{X}$.
2. There exists a sequence $(\{\boldsymbol{\theta}_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i\in\mathbb{N}}$ of i.i.d. processes, with a law characterized by a finite-dimensional parameter Ψ_Θ, and with marginal distribution $G^0_{\boldsymbol{x}} \in \mathcal{P}(\Theta)$ for any $\boldsymbol{x} \in \mathcal{X}$.
3. There exists a null set $N \subset \Omega$ such that for every $\boldsymbol{x} \in \mathcal{X}$, $B \in \mathcal{B}(\Theta)$, and $\omega \in \Omega \setminus N$,
$$G^\omega_{\boldsymbol{x}}(B) = \sum_{i=1}^{\infty} \pi^\omega_{i,\boldsymbol{x}} \delta_{\boldsymbol{\theta}^\omega_{i,\boldsymbol{x}}}(B), \tag{3}$$
where the sequence $(\{\pi_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i\in\mathbb{N}}$ is given by the stick-breaking process in (2).

It is of interest to determine when a DDP can be constructed from a prescribed $\alpha_{\mathcal{X}} := \{\alpha_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ and $G^0_{\mathcal{X}} := \{G^0_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$. First, we can construct a sequence of processes $(\{\boldsymbol{\theta}_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i\in\mathbb{N}}$ satisfying Condition 2 in Definition 3.1 from a prescribed $G^0_{\mathcal{X}}$ by constructing first a process $\{\boldsymbol{\theta}_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ with marginals $\{G^0_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ using an extension of Kolmogorov's consistency theorem to Polish spaces [see Section III.3 in 69]. Remark that in this case the separability of the resulting process is not required.

Second, we can construct a sequence of processes $(\{V_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i\in\mathbb{N}}$ satisfying Condition 1 in Definition 3.1 from a prescribed $\alpha_{\mathcal{X}}$ using Kolmogorov's consistency theorem and a consistent family of copula functions as in [6]. However, Definition 3.1 requires this process to be in addition separable. This ensures that the set of outcomes for which the left-hand side in (3) is not a probability measure is a measurable set. In fact, if for every $\boldsymbol{x} \in \mathcal{X}$ we define the event

$$N_{\boldsymbol{x}} := \left\{\omega \in \Omega : \sum_{i=1}^{\infty} \pi^\omega_{i,\boldsymbol{x}} < 1\right\} = \left\{\omega \in \Omega : \sum_{i=1}^{\infty} \log(\mathbb{E}(1 - V^\omega_{i,\boldsymbol{x}})) = -\infty\right\},$$

then the left-hand side in (3) fails to be a probability measure if

$$\omega \in \bigcup_{\boldsymbol{x}\in\mathcal{X}} N_{\boldsymbol{x}}.$$

However, the above set may not be measurable. The separability of the processes ensures there exists a countable set $\mathcal{X}_V \subset \mathcal{X}$ such that

$$\bigcup_{\boldsymbol{x}\in\mathcal{X}} N_{\boldsymbol{x}} = \bigcup_{\boldsymbol{x}\in\mathcal{X}_V} N_{\boldsymbol{x}}.$$

Therefore,

$$\mathbb{P}\left(\bigcup_{\boldsymbol{x}\in\mathcal{X}} N_{\boldsymbol{x}}\right) \leq \sum_{\boldsymbol{x}\in\mathcal{X}_V} \mathbb{P}(N_{\boldsymbol{x}}),$$

whence it suffices to ensure $\mathbb{P}(N_{\boldsymbol{x}}) = 0$ for every $\boldsymbol{x} \in \mathcal{X}_V$. Since any stochastic process on a separable space with a.s. continuous sample paths is separable, an alternative, standard way to build a separable process on $\mathcal{X}$ with $\text{Beta}(1, \alpha_{\boldsymbol{x}})$ marginal distributions is to transform a real-valued process over $\mathcal{X}$ with a.s. continuous sample paths using the quantile function of the beta distribution [59, 60]. Specifically, let $\{Z_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ be a real-valued stochastic process with a.s. continuous sample paths, and with continuous cumulative distribution function $F_{Z,\boldsymbol{x}}$ at $\boldsymbol{x} \in \mathcal{X}$. Let $F_{B,\boldsymbol{x}}$ denote the cumulative distribution function of a $\text{Beta}(1, \alpha_{\boldsymbol{x}})$ distribution. Then, the process $\{V^{\omega}_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ defined as

$$V^{\omega}_{\boldsymbol{x}} := F^{-1}_{B,\boldsymbol{x}}(F_{Z,\boldsymbol{x}}(Z_{\boldsymbol{x}})),$$

is separable, has a.s. continuous sample paths, and has marginal distribution $\text{Beta}(1, \alpha_{\boldsymbol{x}})$. The choice for the base process $\{Z_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ depends on the structure of $\mathcal{X}$. When it is a Gaussian process, there are known conditions under which it admits a modification with a.s. continuous sample paths. For example, Theorem 2.3.1 in [53] and [72] provide sufficient conditions for the existence of such a modification when $\mathcal{X}$ is compact. Another possibility when $\mathcal{X}$ is a manifold is to use diffusion processes [41] or processes based on heat kernels [18].

Although the DDP is flexible, it is of interest to define parsimonious variants for which either the support points or the weights are independent of $\boldsymbol{x}$. These parsimonious versions should be understood not only as simplifications of the DDP, but also as useful models with comparative advantages over the DDP that make them suitable in specific settings. The first parsimonious version removes the dependence of the weights on $\boldsymbol{x}$.

Definition 3.2. A *single-weights dependent Dirichlet process (wDDP)* with parameters (α, Ψ_Θ), denoted as $G_{\mathcal{X}} \sim w\text{DDP}(\alpha, \Psi_\Theta)$, is a $\mathcal{P}(\Theta)$-valued stochastic process $G_{\mathcal{X}} = \{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ on $\mathcal{X}$ and defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that:

1. There exists a sequence $\{V_i\}_{i\in\mathbb{N}}$ of i.i.d. processes, with a common law $\text{Beta}(1, \alpha)$ for some $\alpha \geq 0$.
2. There exists a sequence $(\{\boldsymbol{\theta}_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i\in\mathbb{N}}$ of i.i.d. processes, with a law characterized by a finite-dimensional parameter Ψ_Θ, and marginal distribution $G^0_{\boldsymbol{x}} \in \mathcal{P}(\Theta)$ for any $\boldsymbol{x} \in \mathcal{X}$.
3. There exists a null set $N \subset \Omega$, such that for every $\boldsymbol{x} \in \mathcal{X}$, $B \in \mathcal{B}(\Theta)$, and $\omega \in \Omega \setminus N$,
$$G^{\omega}_{\boldsymbol{x}}(B) = \sum_{i=1}^{\infty} \pi^{\omega}_i \delta_{\boldsymbol{\theta}^{\omega}_{i,\boldsymbol{x}}}(B),$$
where the sequence $\{\pi_i\}_{i\in\mathbb{N}}$ of random variables is defined as
$$\pi^{\omega}_i := \begin{cases} V^{\omega}_i & i = 1, \\ V^{\omega}_i \prod_{j=1}^{i-1}(1 - V^{\omega}_j) & i > 1. \end{cases} \tag{4}$$

One of the advantages of the single-weights DDP is that it avoids any difficulty that may arise in the construction of a separable process $\{V_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$

with $\text{Beta}(1, \alpha_{\boldsymbol{x}})$ marginals for $\boldsymbol{x} \in \mathcal{X}$. However, it implicitly assumes that we may still be able to construct a process $\{\boldsymbol{\theta}_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ with prescribed marginals $\{G^0_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$. Hence, this variant is desirable when the structure of $\mathcal{X}$ is complex relative to that of Θ.

The second parsimonious variant of the DDP relaxes the dependence of the support points on $\boldsymbol{x}$.

Definition 3.3. A *single-atoms dependent Dirichlet process (θDDP)* with parameters (Ψ_V, G^0), denoted as $G_{\mathcal{X}} \sim \theta\text{DDP}(\Psi_V, G^0)$, is a $\mathcal{P}(\Theta)$-valued stochastic process $G_{\mathcal{X}} = \{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ on $\mathcal{X}$ and defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that

1. There exists a sequence $(\{V_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i \in \mathbb{N}}$ of separable i.i.d. processes, with a law characterized by a finite-dimensional parameter Ψ_V, and with marginal distribution $\text{Beta}(1, \alpha_{\boldsymbol{x}})$ for some $\alpha_{\boldsymbol{x}} \geq 0$ for any $\boldsymbol{x} \in \mathcal{X}$.
2. There exists a sequence $\{\boldsymbol{\theta}_i\}_{i \in \mathbb{N}}$ of i.i.d. Θ-valued random variables, with common law $G^0 \in \mathcal{P}(\Theta)$.
3. There exists a null set $N \subset \Omega$, such that for every $\boldsymbol{x} \in \mathcal{X}$, $B \in \mathcal{B}(\Theta)$, and $\omega \in \Omega \setminus N$,
$$G^{\omega}_{\boldsymbol{x}}(B) = \sum_{i=1}^{\infty} \pi^{\omega}_{i,\boldsymbol{x}} \delta_{\boldsymbol{\theta}^{\omega}_i}(B),$$
where the sequence $(\{\pi_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i \in \mathbb{N}}$ is given by (2).

A single-atoms DDP can be easier to construct in situations where the structure of Θ is complex. As matter of fact, the construction of a stochastic process $\{\boldsymbol{\theta}_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ can be difficult for general spaces $\mathcal{X}$ and Θ, particularly when it needs to satisfy additional properties, such as a.s. continuity of its sample paths. Some specific constructions are available in particular cases of interest. For example, if Θ is Kendall's planar shape space [see, e.g. 49], diffusion processes have been proposed under two different approaches: (i) directly on the landmarks, on the space of configurations, which is referred to as Euclidean diffusion of shape [see, e.g., 48, 50, 51, 54], and (ii) directly on Θ, via infinitesimal generators [see, e.g., 55, 52, 3, 38] and the solution of partial differential equations [see, e.g., 41].

Neither the DDP nor its variants require the continuity of the sample paths of the process $(\{V_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i \in \mathbb{N}}$ nor those of $(\{\boldsymbol{\theta}_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i \in \mathbb{N}}$. However, imposing this additional condition endows the DDP and its variants of some desirable properties that are the focus of this work.

Definition 3.4. We define the following variants.

1. A *continuous parameter DDP* is a DDP such that both the separable i.i.d. processes $(\{V_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i \in \mathbb{N}}$ and the i.i.d. processes $(\{\boldsymbol{\theta}_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i \in \mathbb{N}}$ have a.s. continuous sample paths.
2. A *continuous parameter wDDP* is a wDDP such that the i.i.d. processes $(\{\boldsymbol{\theta}_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i \in \mathbb{N}}$ have a.s. continuous sample paths.
3. A *continuous parameter θDDP* is a θDDP such that the separable i.i.d. processes $(\{V_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i \in \mathbb{N}}$ have a.s. continuous sample paths.

4. Properties of dependent Dirichlet processes on Polish spaces

4.1. Continuity

The continuity of the sample paths of a process is an important property that also plays a critical role in statistical applications. On one hand, it allows us to determine suitable topologies for the spaces containing the sample paths. On the other, it ensures that the process will be able to borrow strength across sparse data sources regarding the predictors. In fact, continuity eliminates the need for replicates of the responses at every value of the predictors to obtain adequate estimates of the predictor-dependent probability distributions [see, e.g., 7, 82].

For the DDP and its variants, the sample paths are functions from $\mathcal{X}$ into $\mathcal{P}(\Theta)$ and their continuity depends on the topologies on these spaces. Although we always assume that $\mathcal{X}$ is endowed with its metric topology, there are several standard choices for the topology on $\mathcal{P}(\Theta)$. Although we mostly focus on the weak topology, we will also study the effect of considering the strong (or weak-*) and uniform (or norm, or total variation) topologies on $\mathcal{P}(\Theta)$.

For the weak topology on $\mathcal{P}(\Theta)$ we denote $C_W(\mathcal{X}, \mathcal{P}(\Theta))$ the space of *weakly continuous functions* from $\mathcal{X}$ into $\mathcal{P}(\Theta)$. These are the functions $P : \mathcal{X} \to \mathcal{P}(\Theta)$ such that for any $f \in C_b(\Theta)$ the function

$$F(\boldsymbol{x}) = \int_{\Theta} f(\boldsymbol{\theta}) dP_{\boldsymbol{x}}(\boldsymbol{\theta}),$$

is continuous on $\mathcal{X}$. The following theorem shows that when the underlying stochastic processes have a.s. continuous sample paths, the DDP and its variants have a.s. weakly continuous sample paths. We defer its proof to Appendix A.1.1.

Theorem 4.1. *Let $G_{\mathcal{X}}$ be a $\mathcal{P}(\Theta)$-valued process. Suppose that $G_{\mathcal{X}}$ is a continuous parameter DDP, a continuous parameter wDDP, or a continuous parameter θDDP. Then, for a.e. $\omega \in \Omega$,*

$$\lim_{\boldsymbol{x} \to \boldsymbol{x}_0} \int f(\boldsymbol{\theta}) dG^{\omega}_{\boldsymbol{x}}(\boldsymbol{\theta}) = \int f(\boldsymbol{\theta})\, dG^{\omega}_{\boldsymbol{x}_0}(\boldsymbol{\theta}),$$

$\forall\, \boldsymbol{x}_0 \in \mathcal{X}$ and $f \in C_b(\Theta)$.

Consequently, to construct a DDP or any of its variants with a.s. weakly continuous sample paths it suffices to construct suitable continuous processes on $\mathcal{X}$. As discussed earlier, a process $\{V_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ with the desired properties can be constructed from a real-valued base process $\{Z_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ on $\mathcal{X}$ with a.s. continuous sample paths which can be, for instance, a suitable Gaussian process.

To our knowledge, there is no similar, standard way to construct a process $\{\boldsymbol{\theta}_{\boldsymbol{x}} : \boldsymbol{x} \in \Theta\}$ with the desired properties for general $\mathcal{X}$ and Θ. When $\mathcal{X} = \mathbb{R}^d$ there are well-known sufficient conditions that ensure the there exists a modification of a Θ-valued process $\{\boldsymbol{\theta}_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ with a.s. continuous sample paths [46, Theorem 2.23]. This modification exists if there exists exponents $\alpha > 0, \gamma > 0$, and a constant $C > 0$, such that, $\forall\, \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$,

$$\mathbb{E}(d_\Theta(\boldsymbol{\theta}_{\boldsymbol{x}}, \boldsymbol{\theta}_{\boldsymbol{x}'})^\alpha) \le C d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}')^{d+\gamma}, \tag{5}$$

where d_Θ is a *complete* metric on Θ. This result can be applied to a Polish space $\mathcal{X}$ that is homeomorphic to $\mathbb{R}^d$ for some d.

When Θ is a Riemannian manifold, processes with a.s. continuous sample paths can be defined through diffusion processes [see, e.g., 41]. When Θ is also a quotient space $\Theta = \mathcal{A}/\mathcal{G}$ for a locally compact space $\mathcal{A}$ and a group $\mathcal{G}$, then a Θ-valued process with a.s. continuous sample paths can be constructed as follows. Let $\mathcal{Q} : \mathcal{A} \to \Theta$ be the canonical quotient map and let $\boldsymbol{A} : \mathcal{X} \times \Omega \to \mathcal{A}$ be a process with a.s. continuous sample paths. Since the canonical quotient map is continuous, the process $\{\mathcal{Q}(\boldsymbol{A}(\boldsymbol{x})) : \boldsymbol{x} \in \mathcal{X}\}$ has a.s. continuous sample paths.

As mentioned earlier, the variants of the DDP should not be thought only as simplifications of the DDP but also as processes with distinct properties. Endowing $\mathcal{P}(\Theta)$ with the strong topology already allows us to distinguish the properties of these processes. Let $C_S(\mathcal{X}, \mathcal{P}(\Theta))$ be the vector space of *strongly continuous functions* from $\mathcal{X}$ into $\mathcal{P}(\Theta)$. These are the functions $P : \mathcal{X} \to \mathcal{P}(\Theta)$ such that for any $f \in L^\infty(\Theta)$ the function

$$F(\boldsymbol{x}) = \int_\Theta f(\boldsymbol{\theta}) dP_{\boldsymbol{x}}(\boldsymbol{\theta}),$$

is continuous on $\mathcal{X}$. The following theorem shows that, under the same hypothesis of Theorem 4.1, the θDDP has a.s. strongly continuous sample paths. Although this is really a corollary of Theorem 4.4, we present this statement independently for clarity.

Theorem 4.2. *Let $G_{\mathcal{X}}$ be a $\mathcal{P}(\Theta)$-valued process. Suppose that $G_{\mathcal{X}}$ is a continuous parameter θDDP. Then, for a.e. $\omega \in \Omega$,*

$$\lim_{\boldsymbol{x} \to \boldsymbol{x}_0} \int f(\boldsymbol{\theta}) dG^\omega_{\boldsymbol{x}}(\boldsymbol{\theta}) = \int f(\boldsymbol{\theta})\, dG^\omega_{\boldsymbol{x}_0}(\boldsymbol{\theta}),$$

$\forall\, \boldsymbol{x}_0 \in \mathcal{X}$ and $f \in L^\infty(\Theta)$.

A natural question is whether the DDP or the wDDP can have a.s. strongly continuous sample paths under similar assumptions. To our knowledge, this cannot be the case unless substantially stronger conditions are imposed on these processes. We defer the proof of the following theorem to Appendix A.1.2.

Theorem 4.3. *Let $G_{\mathcal{X}}$ be a $\mathcal{P}(\Theta)$-valued process. Suppose that $G_{\mathcal{X}}$ is a continuous parameter DDP or a continuous parameter wDDP. Let $\boldsymbol{x}_0 \in \Omega$. If for a.e. $\omega \in \Omega$ we have that*

$$\lim_{\boldsymbol{x} \to \boldsymbol{x}_0} \int f(\boldsymbol{\theta}) dG^\omega_{\boldsymbol{x}}(\boldsymbol{\theta}) = \int f(\boldsymbol{\theta}) dG^\omega_{\boldsymbol{x}_0}(\boldsymbol{\theta}),$$

$\forall\, f \in L^\infty(\Theta)$, then for a.e. $\omega \in \Omega$ there exists an open neighborhood $U^\omega \subset \mathcal{X}$ of $\boldsymbol{x}_0$ and at least one $i^\omega \in \mathbb{N}$ such that $\boldsymbol{\theta}^\omega_{i^\omega}$ is constant on U^ω.

Since for the DDP and wDDP the sequence of processes $(\{\boldsymbol{\theta}_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i\in\mathbb{N}}$ is independent and identically distributed, the above implies that when the process has a.s. strongly continuous paths at $\boldsymbol{x}_0$ the process $\{\boldsymbol{\theta}_{1,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ must have a.s. constant sample paths near $\boldsymbol{x}_0$. Although this suggests that the main issue is the behavior of the atoms themselves, the proof shows that the main issue is the independence between the processes $(\{V_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i\in\mathbb{N}}$ and $(\{\boldsymbol{\theta}_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i\in\mathbb{N}}$. We conjecture that for the DDP and wDDP to have a.s. strongly continuous paths, it is necessary to introduce dependence between these processes.

Since the DDP and wDDP do not have a.s. strongly continuous paths, it is clear they will not have a.s. continuous paths with respect to stronger topologies on $\mathcal{P}(\Theta)$. However, for the θDDP we can strengthen the topology on $\mathcal{P}(\Theta)$ while preserving this property. Consider the uniform topology on $\mathcal{P}(\Theta)$ and denote as $C_U(\mathcal{X}, \mathcal{P}(\Theta))$ the set of *uniformly continuous functions* from $\mathcal{X}$ into $\mathcal{P}(\Theta)$. The total variation norm for any signed finite measure Q on $(\Theta, \mathcal{B}(\Theta))$ is defined as

$$\|Q\|_{\mathrm{TV}} := \sup\left\{\int f(\boldsymbol{\theta})dQ(\boldsymbol{\theta}) : \, f \in L^{\infty}(\Theta), \, \|f\|_{L^{\infty}} \leq 1\right\}.$$

Then the elements of $C_U(\mathcal{X}, \mathcal{P}(\Theta))$ are the functions $P : \mathcal{X} \to \mathcal{P}(\Theta)$ such that for any $\boldsymbol{x}_0 \in \mathcal{X}$

$$\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} \|P_{\boldsymbol{x}} - P_{\boldsymbol{x}_0}\|_{\mathrm{TV}} = 0.$$

By choosing indicator functions, it is clear the above is equivalent to

$$\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} \sup_{B\in\mathcal{B}(\Theta)} |P_{\boldsymbol{x}}(B) - P_{\boldsymbol{x}_0}(B)| = 0,$$

$\forall\, \boldsymbol{x}_0 \in \mathcal{X}$, which is an expression that is typically more interpretable in statistical applications. For the uniform topology we can show that, under the same assumptions of Theorem 4.2, the θDDP has a.s. uniformly (or norm) continuous sample paths. This is also known as continuity in total variation. Its proof is deferrered to Appendix A.1.3.

Theorem 4.4. *Let $G_{\mathcal{X}}$ be a continuous parameter θDDP. Then, for a.e. $\omega \in \Omega$,*

$$\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} \|G^{\omega}_{\boldsymbol{x}} - G^{\omega}_{\boldsymbol{x}_0}\|_{\mathrm{TV}} = 0,$$

$\forall\, \boldsymbol{x}_0 \in \mathcal{X}$.

4.2. Support

The sample paths of the DDP and its variants are elements of suitable spaces of functions from $\mathcal{X}$ into $\mathcal{P}(\Theta)$. It is of interest to characterize the size, in a suitable sense, of the set containing the sample paths. This leads us to the concept of support. A large support is an important and basic property that any BNP model should possess. In fact, it is a minimum requirement, and almost a "necessary" property, for a BNP model to be considered "nonparametric." This property is also important because it typically is a necessary condition for the consistency of the posterior distribution. In such settings, the full support of the prior implies that the prior probability model is flexible enough to generate sample paths sufficiently close to any element of the parameter space.

Given a topology $\mathcal{T}$ on $\mathcal{P}(\Theta)^{\mathcal{X}}$ the support of a process is the smallest closed set, in the sense of set inclusion, such that the probability it contains a sample path is equal to one. We say it has full support, or that the support is full, if it is equal to $\mathcal{P}(\Theta)^{\mathcal{X}}$. When the support is not full, its complement is a non-empty open set. In particular, it contains a point with a neighborhood that is disjoint from the support for which the probability of containing a sample path is zero. Consequently, to prove that a process has full support with respect to $\mathcal{T}$ it suffices to show that the probability that any element of a neighborhood basis contains a sample path is positive.

We characterize the support of the DDP and its variant for common choices of $\mathcal{T}$ starting from the weakest. We consider first the (weak) product topology, or pointwise topology [47], on $\mathcal{P}(\Theta)^{\mathcal{X}}$. For reasons that shall be clear soon, we call it the *product-weak topology*. In this topology, a neighborhood basis at $P^0 \in \mathcal{P}(\Theta)^{\mathcal{X}}$ is given by sets of the form

$$\left\{ P \in \mathcal{P}(\Theta)^{\mathcal{X}} : \left| \int_\Theta f_{i,j}(\boldsymbol{\theta}) dP_{\boldsymbol{x}_j}(\boldsymbol{\theta}) - \int_\Theta f_{i,j}(\boldsymbol{\theta}) dP^0_{\boldsymbol{x}_j}(\boldsymbol{\theta}) \right| < \varepsilon_{i,j},\ i, j \in [n] \right\},$$

for $\varepsilon_{1,1}, \dots, \varepsilon_{n,n} > 0$, $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \in \mathcal{X}$ and $f_{1,1}, \dots, f_{n,n} \in C_b(\Theta)$. The following theorem shows the DDP and its variants have full support with respect to this topology. We defer its proof to Appendix A.2.1.

Theorem 4.5. *Let $G_{\mathcal{X}}$ be a $\mathcal{P}(\Theta)$-valued process on $\mathcal{X}$. Suppose that one of the following assertions holds.*

1. *$G_{\mathcal{X}} \sim \mathrm{DDP}(\Psi_V, \Psi_\Theta)$, for any $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \in \mathcal{X}$ the law of the random vector*
$$(\boldsymbol{\theta}_{1,\boldsymbol{x}_1}, \dots, \boldsymbol{\theta}_{1,\boldsymbol{x}_n}),$$
has full support on Θ^n, and the law of the random vector
$$(V_{1,\boldsymbol{x}_1}, \dots, V_{1,\boldsymbol{x}_n}),$$
has full support on $[0, 1]^n$.
2. *$G_{\mathcal{X}} \sim w\mathrm{DDP}(\alpha, \Psi_\Theta)$, for any $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \in \mathcal{X}$ the law of the random vector*
$$(\boldsymbol{\theta}_{1,\boldsymbol{x}_1}, \dots, \boldsymbol{\theta}_{1,\boldsymbol{x}_n}),$$
has full support on Θ^n, and the law of the random variable V_1 has full support on $[0, 1]$.
3. *$G_{\mathcal{X}} \sim \theta\mathrm{DDP}(\Psi_V, G^0)$, G^0 has full support on Θ, and for any $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \in \mathcal{X}$ the law of the random vector*
$$(V_{1,\boldsymbol{x}_1}, \dots, V_{1,\boldsymbol{x}_n}),$$
has full support on $[0, 1]^n$.

Then for any $P^0 \in \mathcal{P}(\Theta)^{\mathcal{X}}$ *we have that*

$$\mathbb{P}\left(\left\{\omega \in \Omega : \left|\int_{\Theta} f_{i,j}(\boldsymbol{\theta}) dG^{\omega}_{\boldsymbol{x}_j}(\boldsymbol{\theta}) - \int_{\Theta} f_{i,j}(\boldsymbol{\theta}) dP^0_{\boldsymbol{x}_j}(\boldsymbol{\theta})\right| < \varepsilon_{i,j},\ i,j \in [n]\right\}\right) > 0,$$

for $\varepsilon_{1,1}, \ldots, \varepsilon_{n,n} > 0$, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathcal{X}$ *and* $f_{1,1}, \ldots, f_{n,n} \in C_b(\Theta)$. *In consequence, the process has full support on* $\mathcal{P}(\Theta)^{\mathcal{X}}$ *endowed with the product-weak topology.*

The product-weak topology is often too coarse in statistical applications. The topology we consider next is the compact-open topology on $\mathcal{P}(\Theta)^{\mathcal{X}}$ [47]. In this topology, a neighborhood basis at $P^0 \in \mathcal{P}(\Theta)^{\mathcal{X}}$ is given by sets of the form

$$\left\{P \in \mathcal{P}(\Theta)^{\mathcal{X}} : \sup_{\boldsymbol{x} \in K} \left|\int_{\Theta} f_i(\boldsymbol{\theta}) dP_{\boldsymbol{x}}(\boldsymbol{\theta}) - \int_{\Theta} f_i(\boldsymbol{\theta}) dP^0_{\boldsymbol{x}}(\boldsymbol{\theta})\right| < \varepsilon_i,\ i \in [n]\right\}, \quad (6)$$

for $\varepsilon_1, \ldots, \varepsilon_n > 0$, $f_1, \ldots, f_n \in C_b(\Theta)$ and $K \subset \mathcal{X}$ compact. As this topology is stronger, it is unlikely the DDP and its variants will still have full support on $\mathcal{P}(\Theta)^{\mathcal{X}}$. For this reason, we determine whether the support contains a subset of $\mathcal{P}(\Theta)^{\mathcal{X}}$ of functions of interest. If the weak topology on $\mathcal{P}(\Theta)$ is of interest, it becomes natural to consider whether the support contains the weakly continuous functions from $\mathcal{X}$ into $\mathcal{P}(\Theta)$.

If the support of a process does not contain $C_W(\mathcal{X}, \mathcal{P}(\Theta))$ then there is at least one $P^0 \in C_W(\mathcal{X}, \mathcal{P}(\Theta))$ in the complement of the support. Since this set is open, it contains at least one set of the form (6). The following result shows that, under mild conditions, the support of both the DDP and θDDP contains $C_W(\mathcal{X}, \mathcal{P}(\Theta))$. We defer its proof to Appendix A.2.2.

Theorem 4.6. *Let* $G_{\mathcal{X}} \sim \mathrm{DDP}(\Psi_V, \Psi_\Theta)$ *or* $G_{\mathcal{X}} \sim \theta\mathrm{DDP}(\Psi_V, G^0)$. *Suppose that the following conditions hold:*

1. *The processes* $(\{V_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i\in\mathbb{N}}$ *have a.s. continuous sample paths.*
2. *For any* $\varepsilon > 0$, *for any continuous function* $h : \mathcal{X} \to [0,1]$, *and for any* $K \subset \mathcal{X}$ *compact we have that*
$$\mathbb{P}\left(\left\{\omega \in \Omega : \sup_{\boldsymbol{x} \in K} |V^{\omega}_{1,\boldsymbol{x}} - h(\boldsymbol{x})| < \varepsilon\right\}\right) > 0. \quad (7)$$
3. *If* $G_{\mathcal{X}} \sim \theta\mathrm{DDP}(\Psi_V, G^0)$, *then* G^0 *has full support on* Θ.
4. *If* $G_{\mathcal{X}} \sim \mathrm{DDP}(\Psi_V, \Psi_\Theta)$, *then for any open* $U \in \mathcal{B}(\Theta)$ *and* $K \subset \mathcal{X}$ *compact we have that*
$$\mathbb{P}(\{\omega \in \Omega : \boldsymbol{\theta}^{\omega}_{1,\boldsymbol{x}} \in U,\ \forall \boldsymbol{x} \in K\}) > 0. \quad (8)$$

Then, for any $P^0 \in C_W(\mathcal{X}, \mathcal{P}(\Theta))$ *we have that*

$$\mathbb{P}\left(\left\{\omega \in \Omega : \sup_{\boldsymbol{x} \in K} \left|\int_{\Theta} f_i(\boldsymbol{\theta}) dG^{\omega}_{\boldsymbol{x}}(\boldsymbol{\theta}) - \int_{\Theta} f_i(\boldsymbol{\theta}) dP^0_{\boldsymbol{x}}(\boldsymbol{\theta})\right| < \varepsilon_i,\ i \in [n]\right\}\right) > 0,$$

for $\varepsilon_1, \ldots, \varepsilon_n > 0$, $f_1, \ldots, f_n \in C_b(\Theta)$ *and* K *compact. In consequence, the support of the process in* $\mathcal{P}(\Theta)^{\mathcal{X}}$ *endowed with the compact-weak topology contains* $C_W(\mathcal{X}, \mathcal{P}(\Theta))$.

To our knowledge, the wDDP does not seem to be flexible enough for Theorem 4.6 to hold. Briefly, this is because our proof method relies on approximating simultaneously a finite collection of continuous functions $h_1, \dots, h_n : \mathcal{X} \to [0,1]$ with $1-\varepsilon < h_1 + \dots + h_n \leq 1$ for some $\varepsilon > 0$ over a compact $K \subset \mathcal{X}$ using continuous sample paths of the weights of the process. In contrast to the DDP and θDDP, the wDDP does not allow this, as every realization of the weights of the process are a.s. constant. This suggests the DDP and θDDP should be preferred when functions of the form $\boldsymbol{x} \to P^0_{\boldsymbol{x}}(A)$ for $A \in \mathcal{B}(\Theta)$ may vary substantially over any compact $K \subset \mathcal{X}$.

To construct a process $\{V_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ satisfying (7) we use the same construction outlined in Section 3. Let $\{Z_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ be a Gaussian process with mean function μ and covariance kernel σ and a.s. continuous sample paths. We define at any given $\boldsymbol{x} \in \mathcal{X}$ the functions

$$F^{-1}_{B,\alpha_{\boldsymbol{x}}}(t) = 1 - (1-t)^{1/\alpha_{\boldsymbol{x}}} \quad \text{and} \quad F_{Z,\boldsymbol{x}}(z) = \Phi\left(\frac{z - \mu_{\boldsymbol{x}}}{\sigma_{\boldsymbol{x},\boldsymbol{x}}}\right).$$

Observe that

$$|F^{-1}_{B,\alpha_{\boldsymbol{x}}}(u) - F^{-1}_{B,\alpha_{\boldsymbol{x}}}(v)| \leq \frac{1}{\alpha_{\boldsymbol{x}}}|u - v|.$$

If we let $V_{\boldsymbol{x}} = F^{-1}_{B,\alpha_{\boldsymbol{x}}} \circ F_{Z,\boldsymbol{x}}(Z_{\boldsymbol{x}})$ then $V_{\boldsymbol{x}}$ has a.s. continuous sample paths and $\text{Beta}(1, \alpha_{\boldsymbol{x}})$ marginal distributions. Let $K \subset \mathcal{X}$ be compact, and let $h : K \to (0,1)$ be continuous. Then

$$|V_{\boldsymbol{x}} - h(\boldsymbol{x})| \leq \frac{1}{\alpha_{\boldsymbol{x}}}|F_{Z,\boldsymbol{x}}(Z_{\boldsymbol{x}}) - F_{B,\alpha_{\boldsymbol{x}}}(h(\boldsymbol{x}))| \leq \frac{1}{\sqrt{2\pi}\alpha_{\boldsymbol{x}}\sigma_{\boldsymbol{x},\boldsymbol{x}}}|Z_{\boldsymbol{x}} - \overline{h}(\boldsymbol{x})|,$$

where $\overline{h} = F^{-1}_{Z,\boldsymbol{x}} \circ F_{B,\alpha_{\boldsymbol{x}}} \circ h$. Note that if $\alpha_{\boldsymbol{x}}, \mu_{\boldsymbol{x}}$ and $\sigma_{\boldsymbol{x},\boldsymbol{x}}$ are continuous, then so is $\overline{h} : \mathcal{X} \to \mathbb{R}$. Hence, if there exists $c > 0$ such that $\sqrt{2\pi}\alpha_{\boldsymbol{x}}\sigma_{\boldsymbol{x},\boldsymbol{x}} \geq c$, then

$$\sup_{\boldsymbol{x} \in K} |V_{\boldsymbol{x}} - h(\boldsymbol{x})| \leq \frac{1}{c} \sup_{\boldsymbol{x} \in K} |Z_{\boldsymbol{x}} - \overline{h}(\boldsymbol{x})|.$$

Therefore, it suffices to choose a process $Z_{\boldsymbol{x}}$ for which the event

$$\mathbb{P}\left\{\omega \in \Omega : \sup_{\boldsymbol{x} \in K} |Z^{\omega}_{\boldsymbol{x}} - \overline{h}(\boldsymbol{x})| < \varepsilon\right\} > 0,$$

for any continuous function $\overline{h} : K \to \mathbb{R}$. In other words, for every compact K the support of the Gaussian process restricted to K in the supremum norm should contain the space of continuous functions on K. A sufficient condition for this is that the reproducing kernel Hilbert space (RKHS) associated to the process is dense on the space of continuous functions. For example, when the space $\mathcal{X}$ can be isometrically embedded in a Hilbert space [77] we can use the covariance kernel

$$\sigma_{\boldsymbol{x}_1,\boldsymbol{x}_2} = \sigma_0 e^{-d_{\mathcal{X}}(\boldsymbol{x}_1,\boldsymbol{x}_2)^2/\tau^2}.$$

The induced RKHS spans the space of all smooth functions if $\tau > 0$ is allowed to vary freely [19]. In addition, although there is evidence no RKHS can contain the space of all continuous functions on a compact set [80], universal kernels generate Hilbert spaces that are dense in the space of continuous functions [62, 20].

It is natural to characterize the support of the DDP or its variants in stronger topologies. One such topology arises when we endow $\mathcal{P}(\Theta)$ with the *strong topology*. In this topology, a neighborhood basis at $P^0 \in \mathcal{P}(\Theta)$ is given by sets of the form

$$\left\{P \in \mathcal{P}(\Theta):\ \left|\int_\Theta f_i(\boldsymbol{\theta}) dP(\boldsymbol{\theta}) - \int_\Theta f_i(\boldsymbol{\theta}) dP^0(\boldsymbol{\theta})\right| < \varepsilon_i,\ i \in [n]\right\},$$

for $\varepsilon_1, \ldots, \varepsilon_n > 0$ and $f_1, \ldots, f_n \in L^\infty(\Theta)$. Hence, we consider the (weak) product topology on $\mathcal{P}(\Theta)^{\mathcal{X}}$ when $\mathcal{P}(\Theta)$ is endowed with the strong topology. We call this the *product-strong topology*. In this topology, a neighborhood basis at $P^0 \in \mathcal{P}(\Theta)^{\mathcal{X}}$ is given by sets of the form

$$\left\{P \in \mathcal{P}(\Theta)^{\mathcal{X}}:\ \left|\int_\Theta f_{i,j}(\boldsymbol{\theta}) dP_{\boldsymbol{x}_j}(\boldsymbol{\theta}) - \int_\Theta f_{i,j}(\boldsymbol{\theta}) dP^0_{\boldsymbol{x}_j}(\boldsymbol{\theta})\right| < \varepsilon_{i,j},\ i,j \in [n]\right\},$$

for $\varepsilon_{1,1}, \ldots, \varepsilon_{n,n} > 0$, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathcal{X}$ and $f_{1,1}, \ldots, f_{n,n} \in L^\infty(\Theta)$. By choosing simple functions, it becomes clear that the sets

$$\left\{P \in \mathcal{P}(\Theta)^{\mathcal{X}}:\ |P_{\boldsymbol{x}_j}(B_i) - P^0_{\boldsymbol{x}_j}(B_i)| \in \varepsilon_{i,j},\ i,j \in [n]\right\},$$

for $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathcal{X}$, $\varepsilon_{1,1}, \ldots, \varepsilon_{n,n} > 0$ open, and $B_1, \ldots, B_n \in \mathcal{B}(\Theta)$ also form a neighborhood basis at P^0.

Theorem 4.2 suggests that neither the DDP nor its variants will have full support on the product-strong topology. However, we are still able to characterize some key features of their support. We first introduce the following technical definition.

Definition 4.7. Let $G_{\mathcal{X}}$ be a $\mathcal{P}(\Theta)$-valued random process on $\mathcal{X}$.

1. For $G_{\mathcal{X}} \sim \mathrm{DDP}(\Psi_V, \Psi_\Theta)$ we let
$$\mathcal{P}(\Theta)^{\mathcal{X}}|_{G_{\mathcal{X}}} := \left\{P \in \mathcal{P}(\Theta)^{\mathcal{X}} : \forall \boldsymbol{x} \in \mathcal{X}:\ P_{\boldsymbol{x}} \ll G^0_{\boldsymbol{x}}\right\}.$$
2. For $G_{\mathcal{X}} \sim w\mathrm{DDP}(\alpha, \Psi_\Theta)$ we let
$$\mathcal{P}(\Theta)^{\mathcal{X}}|_{G_{\mathcal{X}}} := \left\{P \in \mathcal{P}(\Theta)^{\mathcal{X}} : \forall \boldsymbol{x} \in \mathcal{X}:\ P_{\boldsymbol{x}} \ll G^0_{\boldsymbol{x}}\right\}.$$
3. For $G_{\mathcal{X}} \sim \theta\mathrm{DDP}(\Psi_V, G^0)$ we let
$$\mathcal{P}(\Theta)^{\mathcal{X}}|_{G_{\mathcal{X}}} := \left\{P \in \mathcal{P}(\Theta)^{\mathcal{X}} : \forall \boldsymbol{x} \in \mathcal{X}:\ P_{\boldsymbol{x}} \ll G^0\right\}.$$

Therefore, we can associate to a DDP or to any of its variants a specific set of functions from $\mathcal{X}$ into $\mathcal{P}(\Theta)$. The following theorem shows that, in fact, the support of DDP and its variants contain this set. Remark that the hypotheses are essentially the same as those of Theorem 4.5. We defer the proof of the theorem to Appendix A.2.3

Theorem 4.8. *Let $G_{\mathcal{X}}$ be a $\mathcal{P}(\Theta)$-valued process on $\mathcal{X}$. The following assertions are true.*

1. *If $G_{\mathcal{X}} \sim \mathrm{DDP}(\Psi_V, \Psi_\Theta)$ and, for any $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathcal{X}$, the law of the random vector*
$$(\boldsymbol{\theta}_{1,\boldsymbol{x}_1}, \ldots, \boldsymbol{\theta}_{1,\boldsymbol{x}_n}),$$
has full support on Θ^n and the law of the random vector
$$(V_{1,\boldsymbol{x}_1}, \ldots, V_{1,\boldsymbol{x}_n}),$$
has full support on $[0,1]^n$, then the support of $G_{\mathcal{X}}$ in the product-strong topology contains $\mathcal{P}(\Theta)^{\mathcal{X}}|_{G_{\mathcal{X}}}$.
2. *If $G_{\mathcal{X}} \sim w\mathrm{DDP}(\alpha, \Psi_\Theta)$ and, for any $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathcal{X}$, the law of the random vector*
$$(V_{1,\boldsymbol{x}_1}, \ldots, V_{1,\boldsymbol{x}_n}),$$
has full support on $[0,1]^n$, then the support of $G_{\mathcal{X}}$ in the product-strong topology contains $\mathcal{P}(\Theta)^{\mathcal{X}}|_{G_{\mathcal{X}}}$.
3. *If $G_{\mathcal{X}} \sim \theta\mathrm{DDP}(\Psi_V, G^0)$, G^0 has full support on Θ, and, for any $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathcal{X}$, the law of the random vector*
$$(\boldsymbol{\theta}_{1,\boldsymbol{x}_1}, \ldots, \boldsymbol{\theta}_{1,\boldsymbol{x}_n}),$$
has full support on $[0,1]^n$, then the support of $G_{\mathcal{X}}$ in the product-strong topology contains $\mathcal{P}(\Theta)^{\mathcal{X}}|_{G_{\mathcal{X}}}$.

When $\mathcal{P}(\Theta)$ is endowed with the strong topology, we can consider the associated compact-open topology on $\mathcal{P}(\Theta)^{\mathcal{X}}$. In this topology, the neighborhood basis at any $P^0 \in \mathcal{P}(\Theta)^{\mathcal{X}}$ have the form (6) for $f_1, \ldots, f_n \in L^\infty(\Theta)$.

In this case, the functions of interest are strongly continuous functions from $\mathcal{X}$ into $\mathcal{P}(\Theta)$. In contrast to Theorem 4.6 we cannot show that the support of the DDP nor its variants contains this set. However, we can show the support contains the intersection between $C_S(\mathcal{X}, \mathcal{P}(\Theta))$ and the surrogate functions associated to a DDP or θDDP. To our knowledge, this result does not hold for the w DDP for the same reasons discussed after Theorem 4.6. We defer the proof of the following result to Appendix A.2.4.

Theorem 4.9. *Let $G_{\mathcal{X}} \sim \mathrm{DDP}(\Psi_V, \Psi_\Theta)$ or $G_{\mathcal{X}} \sim \theta\mathrm{DDP}(\Psi_V, G^0)$. Suppose that the following conditions hold:*

1. *The processes $(\{V_{i,\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\})_{i \in \mathbb{N}}$ have a.s. continuous sample paths.*
2. *For any $\varepsilon > 0$, continuous function $h : \mathcal{X} \to [0,1]$ and $K \subset \mathcal{X}$ compact we have that*
$$\mathbb{P}\left(\left\{\omega \in \Omega : \sup_{\boldsymbol{x} \in K} |V^\omega_{1,\boldsymbol{x}} - h(\boldsymbol{x})| < \varepsilon\right\}\right) > 0.$$
3. *If $G_{\mathcal{X}} \sim \theta\mathrm{DDP}(\Psi_V, G^0)$, then G^0 has full support on Θ.*

4. *If $G_{\mathcal{X}} \sim \mathrm{DDP}(\Psi_V, \Psi_\Theta)$, then there exists a measure G^0 on Θ such that $G^0_{\boldsymbol{x}} \ll G^0$ for every $\boldsymbol{x} \in \mathcal{X}$ and that for any $A \in \mathcal{B}(\Theta)$ and $K \subset \mathcal{X}$ compact we have that*

$$G^0(A) > 0 \quad \Rightarrow \quad \mathbb{P}(\{\omega \in \Omega : \boldsymbol{\theta}^\omega_{1,\boldsymbol{x}} \in A, \forall \boldsymbol{x} \in K\}) > 0.$$

Then, $C_S(\mathcal{X}, \mathcal{P}(\Theta)) \cap \mathcal{P}(\Theta)^{\mathcal{X}}|_{G_{\mathcal{X}}}$ is in the support of $G_{\mathcal{X}}$ with respect to the compact-strong topology.

4.3. Association structure

In statistical applications, it is of interest to study the behavior of the process $\{G_{\boldsymbol{x}}(B) : \boldsymbol{x} \in \mathcal{X}\}$ for some fixed $B \in \mathcal{B}(\Theta)$. If $G_{\mathcal{X}} \sim \theta\mathrm{DDP}(\Psi_V, G^0)$ the hypothesis of Theorem 4.2 ensure that the process $\{G_{\boldsymbol{x}}(B) : \boldsymbol{x} \in \mathcal{X}\}$ has a.s. continuous sample paths. As a consequence, for any $d \in \mathbb{N}$ and $f : [0,1]^d \to \mathbb{R}$ the process

$$\{f(G^\omega_{\boldsymbol{x}_1}(B), \dots, G^\omega_{\boldsymbol{x}_d}(B)) : (\boldsymbol{x}_1, \dots, \boldsymbol{x}_d) \in \mathcal{X}^d\},$$

has a.s. continuous sample paths. Furthermore, this holds for its expectation

$$F_B(\boldsymbol{x}_1, \dots, \boldsymbol{x}_d) := \mathbb{E}(f(G_{\boldsymbol{x}_1}(B), \dots, G_{\boldsymbol{x}_d}(B))). \tag{9}$$

Some functions of this form that are of statistical interest are the *measures of association*. For instance, the Pearson correlation coefficient is given by

$$\rho(G_{\boldsymbol{x}_1}(B), G_{\boldsymbol{x}_2}(B)) = \frac{\mathbb{E}(G_{\boldsymbol{x}_1}(B)G_{\boldsymbol{x}_2}(B)) - \mathbb{E}(G_{\boldsymbol{x}_1}(B))\mathbb{E}(G_{\boldsymbol{x}_2}(B))}{\mathbb{E}(G_{\boldsymbol{x}_1}(B)^2)^{1/2}\mathbb{E}(G_{\boldsymbol{x}_2}(B)^2)^{1/2}}.$$

It is clear that it is continuous whenever the denominator is non-zero. Continuity implies

$$\lim_{\boldsymbol{x} \to \boldsymbol{x}_0} \rho(G_{\boldsymbol{x}}(B), G_{\boldsymbol{x}_0}(B)) = 1.$$

On the other hand, if

$$\lim_{d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}_0) \to \infty} \mathbb{E}(G_{\boldsymbol{x}}(B), G_{\boldsymbol{x}_0}(B)) = \mathbb{E}(G_{\boldsymbol{x}}(B))\mathbb{E}(G_{\boldsymbol{x}_0}(B)),$$

then it follows that

$$\lim_{d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}_0) \to \infty} \rho(G_{\boldsymbol{x}}(B), G_{\boldsymbol{x}_0}(B)) = 0.$$

Since the DDP and wDDP may not have a.s. strongly continuous paths, the above argument does not hold and, with positive probability, the process $\{G_{\boldsymbol{x}}(B) : \boldsymbol{x} \in \mathcal{X}\}$ may have discontinuous sample paths. In this case, a measure of association can act as a surrogate to study the regularity of this process, on average, at any point. The following theorem states that, under mild conditions, any function of the form (9) is continuous. Its proof is given in Appendix A.3.1.

Theorem 4.10. *Let $G_{\mathcal{X}}$ be a $\mathcal{P}(\Theta)$-valued process on $\mathcal{X}$ and let $B \in \mathcal{B}(\Theta)$. Suppose that one of the following assertions holds.*

1. *$G_{\mathcal{X}} \sim \mathrm{DDP}(\Psi_V, \Psi_\Theta)$ is a continuous sample paths DDP.*
2. *$G_{\mathcal{X}} \sim w\mathrm{DDP}(\alpha, \Psi_\Theta)$ is a continuous sample paths wDDP.*

Furthermore, suppose that for any $d \in \mathbb{N}$ the function

$$(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d) \mapsto \mathbb{P}(\{\omega \in \Omega : \ \boldsymbol{\theta}^{\omega}_{1,\boldsymbol{x}_1} \in B, \ldots, \boldsymbol{\theta}^{\omega}_{1,\boldsymbol{x}_d} \in B\}),$$

is continuous. Then, for any $d \in \mathbb{N}$, and continuous $f : [0,1]^d \to \mathbb{R}$, the function $F : \mathcal{X}^d \to \mathbb{R}$ defined as

$$F(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d) := \mathbb{E}(f(G_{\boldsymbol{x}_1}(B), \ldots, G_{\boldsymbol{x}_d}(B))).$$

is continuous.

5. Dependent Dirichlet process mixture models on Polish spaces

Let $\mathcal{Y}$ be a Polish space and let $\nu_{\mathcal{Y}}$ be a base measure on $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$. To allow for flexible statistical models, we also consider a nonempty Polish space Γ representing mixture parameters. We assume that Γ is endowed with the Borel σ-algebra $\mathcal{B}(\Gamma)$. The mixture models we study are constructed from a fixed measurable function $\psi : \mathcal{Y} \times \Gamma \times \Theta \to \mathbb{R}_+$ such that

$$\forall\, (\boldsymbol{\gamma}, \boldsymbol{\theta}) \in \Gamma \times \Theta : \ \int_{\mathcal{Y}} \psi(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta})\, d\nu_{\mathcal{Y}}(\boldsymbol{y}) = 1.$$

The mixture associated to $P \in \mathcal{P}(\Theta)^{\mathcal{X}}$ is the map $M^P : \Gamma \times \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, formally defined as

$$\forall\, (\boldsymbol{\gamma}, \boldsymbol{\theta}) \in \Gamma \times \Theta,\ B \in \mathcal{B}(\mathcal{Y}) : \ M^P_{\boldsymbol{\gamma},\boldsymbol{x}}(B) := \int_B \int_\Theta \psi(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}) dP_{\boldsymbol{x}}(\boldsymbol{\theta}) d\nu_{\mathcal{Y}}(\boldsymbol{y}).$$

In particular, a *mixture model* is a map

$$M : \mathcal{P}(\Theta)^{\mathcal{X}} \to \mathcal{P}(\mathcal{Y})^{\Gamma \times \mathcal{X}}.$$

By construction, the measure $M^P_{\boldsymbol{\gamma},\boldsymbol{x}}$ is absolutely continuous with respect to $\nu_{\mathcal{Y}}$ for any $(\boldsymbol{\gamma}, \boldsymbol{x}) \in \Gamma \times \mathcal{X}$. For this reason, we distinguish the set $\mathcal{D}(\mathcal{Y}) \subset \mathcal{P}(\mathcal{Y})$ of probability measures on $\mathcal{Y}$ that admit a density with respect to $\nu_{\mathcal{Y}}$. We often use the identification

$$\mathcal{D}(\mathcal{Y}) \cong \left\{ p \in L^1(\mathcal{Y}, \mathcal{B}(\mathcal{Y}), \nu_{\mathcal{Y}}) : \ \int p(\boldsymbol{y})\, d\nu_{\mathcal{Y}}(\boldsymbol{y}) = 1,\, p \geq 0 \right\}.$$

In particular, for the mixtures we study we have an explicit form for their density. For this reason, for $P \in \mathcal{P}(\mathcal{Y})$ we define the function $\rho^P : \mathcal{Y} \times \Gamma \times \mathcal{X} \to \mathbb{R}_+$

$$\rho^P(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{x}) := \int_\Theta \psi(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}) dP_{\boldsymbol{x}}(\boldsymbol{\theta}),$$

representing the density of $M^P_{\boldsymbol{\gamma},\boldsymbol{x}}$ with respect to $\nu_{\mathcal{Y}}$. We sometimes write

$$\rho^P_{\boldsymbol{\gamma},\boldsymbol{x}}(\boldsymbol{y}) := \rho^P(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x}).$$

Hence, the mixture model M induces a map

$$\rho : \mathcal{P}(\Theta)^{\mathcal{X}} \to \mathcal{D}(\mathcal{Y})^{\Gamma\times\mathcal{X}}.$$

Depending on the choice of $\nu_{\mathcal{Y}}$ and ψ the mixture model may have regularizing properties and the density ρ^P may be, for instance, continuous. The following lemma shows that, under mild regularity and decay assumptions on ψ, we can characterize points of continuity of ρ^P when $P : \mathcal{X} \to \mathcal{P}(\Theta)$ is weakly continuous. The proof of the following result is deferred to Appendix A.4.1.

Lemma 5.1. *Let $P : \mathcal{X} \to \mathcal{P}(\Theta)$ be weakly continuous, and suppose that ψ is continuous. Let $(\boldsymbol{y}_0, \boldsymbol{\gamma}_0) \in \mathcal{Y} \times \Gamma$. If for every $\varepsilon > 0$, there exists an open neighborhood $U_{\boldsymbol{y}_0} \subset \mathcal{Y}$ of $\boldsymbol{y}_0$, an open neighborhood $U_{\boldsymbol{\gamma}_0} \subset \Gamma$ of $\boldsymbol{\gamma}_0$, and a compact $K_\Theta \subset \Theta$, such that*

$$\sup\{\psi(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}) : (\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}) \in U_{\boldsymbol{y}_0} \times U_{\boldsymbol{\gamma}_0} \times K^c_\Theta\} < \varepsilon, \tag{10}$$

then, ρ^P is continuous on $U_{\boldsymbol{y}_0} \times U_{\boldsymbol{\gamma}_0} \times \mathcal{X}$.

The hypotheses imply that near $\boldsymbol{y}_0$ and $\boldsymbol{\gamma}_0$ the function ψ tends to zero “at infinity” in $\boldsymbol{\theta}$. To gain insight into the consequences of these assumptions, we consider the following example. Let $\mathcal{Y} = [0, 1]$ be endowed with the standard topology, and let $\nu_{\mathcal{Y}}$ be the Lebesgue measure restricted to $[0, 1]$. Let

$$\Theta := \{(\alpha, \beta) \in \mathbb{R}^2 : \alpha, \beta \geq 1\},$$

be endowed with the standard subspace topology, and let $\Gamma = \emptyset$. If we consider the function

$$\psi(y, \alpha, \beta) = \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha, \beta)},$$

associated to a family of $\text{Beta}(\alpha, \beta)$ probability distributions on $[0, 1]$, then the mixture model would not satisfy the properties of the lemma. In fact, if $y_0 = 1/2$ we can choose $\alpha = \beta = t$ to see that, from Stirling's approximation,

$$\psi(y_0, t, t) = \frac{1}{2^{2t-2}B(t, t)} \sim \frac{1}{2^{2t-2}} \frac{2^{2t-1/2}t^{2t-1/2}}{\sqrt{2\pi}t^{2t-1}} = \frac{2^{3/2}}{\sqrt{2\pi}} t^{1/2},$$

for $t \gg 1$. Hence, ψ does not decay over Θ near y_0. This can be mitigated by restricting the values of both α and β to a compact set. A middle ground can be achieved if, for example, one parameter is constrained to a compact set, whereas the other becomes a mixture parameter. For example,

$$\Gamma := \{\alpha \in \mathbb{R} : 1 \leq \alpha\},$$

and

$$\Theta := \{\beta \in \mathbb{R} : 1 \leq \beta \leq \beta_{\max}\}.$$

In this case, the resulting mixture model satisfies the desired properties. Finally, note that failure to satisfy this condition is not always due to a lack of compactness. For instance, we could consider the model

$$\Theta := \{(\alpha, \beta) \in \mathbb{R}^2 : \alpha, \beta \in [1/4, 1/2]\}.$$

In this case, not only ψ is discontinuous, but we also have that

$$\lim_{y \to 0} \psi(y, \alpha, \beta) = \infty,$$

for any choice of α, β.

Due to the continuity properties of a DDP and its variants, the conclusions of Lemma 5.1 follow from milder hypotheses. In fact, in this case the same conclusion follows by only imposing boundedness. We defer the proof of this result to Appendix A.4.2

Lemma 5.2. *Let $G_{\mathcal{X}}$ be a $\mathcal{P}(\Theta)$-valued process on $\mathcal{X}$ and let $\psi : \mathcal{Y} \times \Gamma \times \Theta \to \mathbb{R}_+$. Suppose that one of following conditions hold:*

1. *$G_{\mathcal{X}}$ is a continuous parameter DDP with $G_{\mathcal{X}} \sim \mathrm{DDP}(\Psi_V, \Psi_\Theta)$.*
2. *$G_{\mathcal{X}}$ is a continuous parameter wDDP with $G_{\mathcal{X}} \sim w\mathrm{DDP}(\alpha, \Psi_\Theta)$.*
3. *$G_{\mathcal{X}}$ is a continuous parameter θDDP with $G_{\mathcal{X}} \sim \theta\mathrm{DDP}(\Psi_V, G^0)$.*

Furthermore, suppose that ψ is continuous. Let $(\boldsymbol{y}_0, \boldsymbol{\gamma}_0) \in \mathcal{Y} \times \Gamma$. If there exists an open neighborhood $U_{\boldsymbol{y}_0} \subset \mathcal{Y}$ of $\boldsymbol{y}_0$, and an open neighborhood $U_{\boldsymbol{\gamma}_0} \subset \Gamma$ of $\boldsymbol{\gamma}_0$, such that

$$\sup\{\psi(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}) : (\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}) \in U_{\boldsymbol{y}_0} \times U_{\boldsymbol{\gamma}_0} \times \Theta\} < \infty,$$

then, for a.e. $\omega \in \Omega$ the function ρ^{G^ω} is continuous at any $(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{x}) \in \mathcal{Y} \times \Gamma \times \mathcal{X}$.

6. Properties of dependent Dirichlet process mixture models on Polish spaces

6.1. Continuity

Mixture models have a regularizing effect. Under the same assumptions of Lemma 5.2 the mixture model M maps weakly continuous into uniformly continuous functions from $\mathcal{X}$ into $\mathcal{P}(\Theta)$. We defer the proof of this result to Appendix A.5.1.

Theorem 6.1. *Suppose that $\nu_{\mathcal{Y}}$ is locally finite and that ψ is continuous. Then, for every $P \in C_W(\mathcal{X}, \mathcal{P}(\Theta))$, the mixture M^P is uniformly continuous, i.e.,*

$$\lim_{(\boldsymbol{\gamma}, \boldsymbol{x}) \to (\boldsymbol{\gamma}_0, \boldsymbol{x}_0)} \|M^P_{\boldsymbol{\gamma}, \boldsymbol{x}} - M^P_{\boldsymbol{\gamma}_0, \boldsymbol{x}_0}\|_{\mathrm{TV}} = 0,$$

for any $\boldsymbol{\gamma}_0 \in \Gamma$ and $\boldsymbol{x}_0 \in \mathcal{X}$.

As a consequence of this result, the mixture of a continuous parameter DDP or its variants has uniformly continuous sample paths.

Corollary 6.2. Let $G_{\mathcal{X}}$ be a $\mathcal{P}(\Theta)$-valued process on $\mathcal{X}$ and let $\psi : \mathcal{Y} \times \Gamma \times \Theta \to \mathbb{R}_+$. Suppose that $G_{\mathcal{X}}$ is a continuous parameter DDP, a continuous parameter wDDP, or a continuous parameter θDDP. Then, for a.e. $\omega \in \Omega$, the mixture M^{G^ω} is uniformly continuous, i.e.,

$$\lim_{(\boldsymbol{\gamma},\boldsymbol{x})\to(\boldsymbol{\gamma}_0,\boldsymbol{x}_0)} \|M^{G^\omega}_{\boldsymbol{\gamma},\boldsymbol{x}} - M^{G^\omega}_{\boldsymbol{\gamma}_0,\boldsymbol{x}_0}\|_{\mathrm{TV}} = 0,$$

for any $\boldsymbol{\gamma}_0 \in \Gamma$ and $\boldsymbol{x}_0 \in \mathcal{X}$.

6.2. Support

As in the case of a DDP or any of its variants, it is of interest to determine the effect that a mixture has on the support. As for the mixture models that we study the probability measures on $\mathcal{Y}$ admit a density with respect to $\nu_{\mathcal{Y}}$, we may interpret the sample paths of the mixture as elements of $\mathcal{D}(\mathcal{Y})^{\Gamma\times\mathcal{X}}$. This allows us to consider other topologies defined in terms of the density of the mixture model.

On $\mathcal{D}(\mathcal{Y})$ we consider the topology induced by the Hellinger distance

$$d_H(p_1,p_2)^2 := \frac{1}{2}\int_{\mathcal{Y}} (\sqrt{p_1(\boldsymbol{y})} - \sqrt{p_2(\boldsymbol{y})})^2 d\nu_{\mathcal{Y}}(\boldsymbol{y}) = 1 - \int_{\mathcal{Y}} \sqrt{p_1(\boldsymbol{y})p_2(\boldsymbol{y})}d\nu_{\mathcal{Y}}(\boldsymbol{y}),$$

by the L^∞ distance

$$d_{L^\infty}(p_1,p_2) := \sup_{\boldsymbol{y}\in\mathcal{Y}} |p_1(\boldsymbol{y}) - p_2(\boldsymbol{y})|,$$

and by the Kullback-Leibler (KL) divergence

$$\mathrm{KL}(p_1\| \, p_2) := \int_{\mathcal{Y}} p_1(\boldsymbol{y}) \log\left(\frac{p_1(\boldsymbol{y})}{p_2(\boldsymbol{y})}\right) d\nu_{\mathcal{Y}}(\boldsymbol{y}).$$

for $p_1, p_2 \in \mathcal{D}(\mathcal{Y})$.

6.2.1. The Hellinger distance

We define the *product-Hellinger* topology on $\mathcal{D}(\mathcal{Y})^{\Gamma\times\mathcal{X}}$ as follows. In this topology, a neighborhood basis at $P^0 \in \mathcal{D}(\mathcal{Y})^{\Gamma\times\mathcal{X}}$ is given by sets of the form

$$\left\{P \in \mathcal{D}(\mathcal{Y})^{\Gamma\times\mathcal{X}} : \; d_H(\rho^P_{\boldsymbol{\gamma}_i,\boldsymbol{x}_i}, \rho^{P^0}_{\boldsymbol{\gamma}_i,\boldsymbol{x}_i}) < \varepsilon_i, \, i \in [n]\right\},$$

for some $\varepsilon_1,\ldots,\varepsilon_n > 0$, $\boldsymbol{\gamma}_1,\ldots,\boldsymbol{\gamma}_n \in \Gamma$ and $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n \in \mathcal{X}$. The following result shows that any neighborhood of the image of $P \in \mathcal{P}(\Theta)^{\mathcal{X}}$ under the mixture on the product-Hellinger topology contains, with positive probability, the image of a sample path of a DDP or its variants under the same mixture. We defer the proof of the following result to Appendix A.6.1.

Theorem 6.3. *Suppose that $\nu_{\mathcal{Y}}$ is locally finite, that ψ is continuous and satisfies (10) for any $(\boldsymbol{y}, \boldsymbol{\gamma}) \in \mathcal{Y} \times \Gamma$, and that the hypotheses of Theorem 4.5 hold. Then, for any $P^0 \in \mathcal{P}(\Theta)^{\mathcal{X}}$ the event*

$$\left\{\omega \in \Omega : \ d_H(\rho^{P^0}_{\boldsymbol{\gamma}_i, \boldsymbol{x}_i}, \rho^{G^\omega}_{\boldsymbol{\gamma}_i, \boldsymbol{x}_i}) < \varepsilon_i, \ i \in [n]\right\},$$

has positive probability for any $\varepsilon_1, \ldots, \varepsilon_n > 0$, $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_n \in \Gamma$, and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathcal{X}$.

A stronger topology induced by the Hellinger distance is what we call the *compact-Hellinger* topology on $\mathcal{D}(\mathcal{Y})^{\Gamma \times \mathcal{X}}$. In this topology, a neighborhood basis at $P^0 \in \mathcal{D}(\mathcal{Y})^{\Gamma \times \mathcal{X}}$ is given by sets of the form

$$\left\{P \in \mathcal{D}(\mathcal{Y})^{\Gamma \times \mathcal{X}} : \sup_{(\boldsymbol{\gamma}, \boldsymbol{x}) \in K_\Gamma \times K_{\mathcal{X}}} d_H(\rho^{P}_{\boldsymbol{\gamma}, \boldsymbol{x}}, \rho^{P^0}_{\boldsymbol{\gamma}, \boldsymbol{x}}) < \varepsilon\right\},$$

for some $\varepsilon > 0$, $K_\Gamma \subset \Gamma$ compact, and $K_{\mathcal{X}} \subset \mathcal{X}$ compact. Note these neighborhoods include sets of the form

$$\left\{P \in \mathcal{D}(\mathcal{Y})^{\Gamma \times \mathcal{X}} : \sup_{\boldsymbol{x} \in K_{\mathcal{X}}} d_H(\rho^{P}_{\boldsymbol{\gamma}_i, \boldsymbol{x}}, \rho^{P^0}_{\boldsymbol{\gamma}_i, \boldsymbol{x}}) < \varepsilon_i, \ i \in [n]\right\},$$

for $\varepsilon_1, \ldots, \varepsilon_n > 0$ and $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_n \in \Gamma$. The following result shows that any neighborhood of the image of $P \in \mathcal{P}(\Theta)^{\mathcal{X}}$ under the mixture on the product-Hellinger topology also contains, with positive probability, the image of a sample path of the DDP or its variants under the same mixture. We defer the proof of the following result to Appendix A.6.1.

Theorem 6.4. *Suppose that $\nu_{\mathcal{Y}}$ is locally finite, that ψ is continuous and satisfies (10) for any $(\boldsymbol{y}, \boldsymbol{\gamma}) \in \mathcal{Y} \times \Gamma$, and that the hypotheses of Theorem 4.5 hold. Then, for any $P^0 \in \mathcal{P}(\Theta)^{\mathcal{X}}$ the event*

$$\left\{\omega \in \Omega : \sup_{(\boldsymbol{\gamma}, \boldsymbol{x}) \in K_\Gamma \times K_{\mathcal{X}}} d_H(\rho^{P^0}_{\boldsymbol{\gamma}, \boldsymbol{x}}, \rho^{G^\omega}_{\boldsymbol{\gamma}, \boldsymbol{x}}) < \varepsilon\right\},$$

has positive probability for any $\varepsilon > 0$, and compact $K_\Gamma \subset \Gamma$ and $K_{\mathcal{X}} \subset \mathcal{X}$.

6.2.2. The L^∞ distance

We define the *product-L^∞* topology on $\mathcal{D}(\mathcal{Y})^{\Gamma \times \mathcal{X}}$ as follows. In this topology, a neighborhood basis at $P^0 \in \mathcal{D}(\mathcal{Y})^{\Gamma \times \mathcal{X}}$ is given by sets of the form

$$\left\{P \in \mathcal{D}(\mathcal{Y})^{\Gamma \times \mathcal{X}} : \ d_{L^\infty}(\rho^{P}_{\boldsymbol{\gamma}_i, \boldsymbol{x}_i}, \rho^{P^0}_{\boldsymbol{\gamma}_i, \boldsymbol{x}_i}) < \varepsilon_i, \ i \in [n]\right\},$$

for some $\varepsilon_1, \ldots, \varepsilon_n > 0$, $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_n \in \Gamma$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathcal{X}$. Similarly to the Hellinger distance, any neighborhood of the image of $P \in \mathcal{P}(\Theta)^{\mathcal{X}}$ under the mixture on the product-L^∞ topology contains, with positive probability, the image of a sample path of a DDP or its variants under the same mixture. However, we require the additional hypothesis of compactness of $\mathcal{Y}$. We defer the proof of the following result to Appendix A.6.2.

Theorem 6.5. *Suppose that $\mathcal{Y}$ is compact, that ψ is continuous and satisfies* (10) *for any $(\boldsymbol{y}, \boldsymbol{\gamma}) \in \mathcal{Y} \times \Gamma$, and that the hypotheses of Theorem 4.5 hold. Then, for any $P^0 \in \mathcal{P}(\Theta)^{\mathcal{X}}$ the event*

$$\left\{\omega \in \Omega : \; d_{L^\infty}(\rho^{P^0}_{\boldsymbol{\gamma}_i, \boldsymbol{x}_i}, \rho^{G^\omega}_{\boldsymbol{\gamma}_i, \boldsymbol{x}_i}) < \varepsilon_i, \, i \in [n]\right\},$$

has positive probability for any $\varepsilon_1, \ldots, \varepsilon_n > 0$, $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_n \in \Gamma$, and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathcal{X}$.

The compactness of $\mathcal{Y}$ allows the uniform control of the density $\rho^{P^0}_{\boldsymbol{\gamma}_i, \boldsymbol{x}_i}$ over $\mathcal{Y}$. In the absence of this assumption, the density $\rho^{P^0}_{\boldsymbol{\gamma}_i, \boldsymbol{x}_i}$ may be unbounded on complements of compact subsets of $\mathcal{Y}$. Our proof strategy can be extended for non-compact $\mathcal{Y}$ with minor modifications if we assume the densities decay at infinity. We have not stated this extension as we believe it imposes somewhat artificial constraints on P^0 that may be hard to verify in practice.

The stronger *compact-L^∞* topology on $\mathcal{D}(\mathcal{Y})^{\Gamma \times \mathcal{X}}$ can be defined similarly as for the Hellinger distance. In this topology, a neighborhood basis at $P^0 \in \mathcal{D}(\mathcal{Y})^{\Gamma \times \mathcal{X}}$ is given by sets of the form

$$\left\{P \in \mathcal{D}(\mathcal{Y})^{\Gamma \times \mathcal{X}} : \sup_{(\boldsymbol{\gamma}, \boldsymbol{x}) \in K_\Gamma \times K_{\mathcal{X}}} d_{L^\infty}(\rho^{P}_{\boldsymbol{\gamma}, \boldsymbol{x}}, \rho^{P^0}_{\boldsymbol{\gamma}, \boldsymbol{x}}) < \varepsilon\right\},$$

for some $\varepsilon > 0$, $K_\Gamma \subset \Gamma$ compact, and $K_{\mathcal{X}} \subset \mathcal{X}$ compact. Note these neighborhoods include sets of the form

$$\left\{P \in \mathcal{D}(\mathcal{Y})^{\Gamma \times \mathcal{X}} : \sup_{\boldsymbol{x} \in K_{\mathcal{X}}} d_{L^\infty}(\rho^{P}_{\boldsymbol{\gamma}_i, \boldsymbol{x}}, \rho^{P^0}_{\boldsymbol{\gamma}_i, \boldsymbol{x}}) < \varepsilon_i, \, i \in [n]\right\},$$

for $\varepsilon_1, \ldots, \varepsilon_n > 0$ and $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_n \in \Gamma$. The following result shows that any neighborhood of the image of $P \in \mathcal{P}(\Theta)^{\mathcal{X}}$ under the mixture on the product-L^∞ topology also contains, with positive probability, the image of a sample path of a DDP or its variants under the same mixture. In this case, we also assume that $\mathcal{Y}$ is compact. We defer the proof of the following result to Appendix A.6.2.

Theorem 6.6. *Suppose that $\mathcal{Y}$ is compact, that ψ is continuous and satisfies* (10) *for any $(\boldsymbol{y}, \boldsymbol{\gamma}) \in \mathcal{Y} \times \Gamma$, and that the hypotheses of Theorem 4.5 hold. Then, for any $P^0 \in \mathcal{P}(\Theta)^{\mathcal{X}}$ the event*

$$\left\{\omega \in \Omega : \sup_{(\boldsymbol{\gamma}, \boldsymbol{x}) \in K_\Gamma \times K_{\mathcal{X}}} d_{L^\infty}(\rho^{P^0}_{\boldsymbol{\gamma}, \boldsymbol{x}}, \rho^{G^\omega}_{\boldsymbol{\gamma}, \boldsymbol{x}}) < \varepsilon\right\},$$

has positive probability for any $K_\Gamma \subset \Gamma$ compact, $K_{\mathcal{X}} \subset \mathcal{X}$ compact, and $\varepsilon > 0$.

6.2.3. The Kullback-Leibler divergence

The KL-divergence defines a premetric on $\mathcal{D}(\mathcal{Y})$ that induces a locally convex topology on $\mathcal{D}(\mathcal{Y})$. This topology depends on which argument is used to define the neighborhood. Due to its connection to the consistency of Bayesian procedures, we consider the neighborhood basis at $P^0 \in \mathcal{D}(\mathcal{Y})$ given by sets of the form

$$\{P \in \mathcal{D}(\mathcal{Y}) : \ \mathrm{KL}(\rho^{P^0} \| \rho^P) < \varepsilon\},$$

for $\varepsilon > 0$. The *product-KL* topology on $\mathcal{D}(\mathcal{Y})^{\Gamma\times\mathcal{X}}$ is defined as follows. A neighborhood basis for $P^0 \in \mathcal{D}(\mathcal{Y})^{\Gamma\times\mathcal{X}}$ on the product-KL topology is given by sets of the form

$$\left\{P \in \mathcal{D}(\mathcal{Y})^{\Gamma\times\mathcal{X}} : \ \mathrm{KL}(\rho^{P^0}_{\boldsymbol{\gamma}_i,\boldsymbol{x}_i} \| \rho^{P}_{\boldsymbol{\gamma}_i,\boldsymbol{x}_i}) < \varepsilon_i,\, i \in [n]\right\},$$

for some $\varepsilon_1, \dots, \varepsilon_n > 0$, $\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_n \in \Gamma$ and $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \in \mathcal{X}$. In this case, we obtain a result similar to that obtained for the Hellinger distance. In this case, we also assume that $\mathcal{Y}$ is compact. We defer the proof of the following result to Appendix A.6.3.

Theorem 6.7. *Suppose that $\mathcal{Y}$ is compact, and that ψ is continuous, strictly positive, and satisfies (10) for any $(\boldsymbol{y}, \gamma) \in \mathcal{Y}\times\Gamma$. Furthermore, suppose that the hypotheses of Theorem 4.5 hold. Then, for any $P^0 \in \mathcal{P}(\Theta)^{\mathcal{X}}$ the event*

$$\left\{\omega \in \Omega : \ \mathrm{KL}(\rho^{P^0}_{\boldsymbol{\gamma}_i,\boldsymbol{x}_i} \| \rho^{G^\omega}_{\boldsymbol{\gamma}_i,\boldsymbol{x}_i}) < \varepsilon_i,\, i \in [n]\right\},$$

has positive probability for any $\varepsilon_1, \dots, \varepsilon_n > 0$, $\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_n \in \Gamma$ and $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \in \mathcal{X}$.

The compactness of $\mathcal{Y}$ ensures that $\rho^{P^0}_{\boldsymbol{\gamma}_i,\boldsymbol{x}_i}$ is bounded above and below on $\mathcal{Y}$. This allows the direct control of the logarithm in the definition of the KL-divergence. Our proof strategy would not hold for non-compact $\mathcal{Y}$ as the density $\rho^{P^0}_{\boldsymbol{\gamma}_i,\boldsymbol{x}_i}$ could become unbounded, and so would be the integrand in the KL-divergence. Additional assumptions could preclude this from happening, i.e., one could assume $\rho^{P^0}_{\boldsymbol{\gamma}_i,\boldsymbol{x}_i}$ is bounded above and below, so that our proof strategy still yields the desired result. However, these assumptions seem artificial, and difficult to prove in practice.

The stronger *compact-KL* topology on $\mathcal{D}(\mathcal{Y})^{\Gamma\times\mathcal{X}}$ can be defined similarly as for the Hellinger and L^∞ distances. In this topology, a neighborhood basis at $P^0 \in \mathcal{D}(\mathcal{Y})^{\Gamma\times\mathcal{X}}$ is given by

$$\left\{P \in \mathcal{D}(\mathcal{Y})^{\Gamma\times\mathcal{X}} : \ \sup_{(\boldsymbol{\gamma},\boldsymbol{x})\in K_\Gamma\times K_{\mathcal{X}}} \mathrm{KL}(\rho^{P^0}_{\boldsymbol{\gamma},\boldsymbol{x}} \| \rho^{P}_{\boldsymbol{\gamma},\boldsymbol{x}}) < \varepsilon\right\},$$

for some $\varepsilon > 0$, $K_\Gamma \subset \Gamma$ compact, and $K_{\mathcal{X}} \subset \mathcal{X}$ compact. Note these neighborhoods include sets of the form

$$\left\{P \in \mathcal{D}(\mathcal{Y})^{\Gamma\times\mathcal{X}} : \ \sup_{\boldsymbol{x}\in K_{\mathcal{X}}} \mathrm{KL}(\rho^{P^0}_{\boldsymbol{\gamma}_i,\boldsymbol{x}} \| \rho^{P}_{\boldsymbol{\gamma}_i,\boldsymbol{x}}) < \varepsilon_i,\, i \in [n]\right\},$$

for $\varepsilon_1, \dots, \varepsilon_n > 0$ and $\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_n \in \Gamma$. The following result shows that any neighborhood of the image of $P \in \mathcal{P}(\Theta)^{\mathcal{X}}$ under the mixture on the product-L^∞

topology also contains, with positive probability, the image of a sample path of a DDP or its variants under the same mixture. We defer the proof of the following result to Appendix A.6.3.

Theorem 6.8. *Suppose that $\mathcal{Y}$ is compact, and that ψ is continuous, strictly positive, and satisfies (10) for any $(\boldsymbol{y}, \boldsymbol{\gamma}) \in \mathcal{Y} \times \Gamma$. Furthermore, suppose that the hypotheses of Theorem 4.5 hold. Then, for any $P^0 \in \mathcal{P}(\Theta)^{\mathcal{X}}$ the event*

$$\left\{\omega \in \Omega : \sup_{(\boldsymbol{\gamma},\boldsymbol{x}) \in K_\Gamma \times K_{\mathcal{X}}} \mathrm{KL}(\rho^{P^0}_{\boldsymbol{\gamma},\boldsymbol{x}} \| \, \rho^{G^\omega}_{\boldsymbol{\gamma},\boldsymbol{x}}) < \varepsilon \right\},$$

has positive probability for any $K_\Gamma \subset \Gamma$ compact, $K_{\mathcal{X}} \subset \mathcal{X}$ compact, and $\varepsilon > 0$.

6.3. Association structure

As a consequence of Theorem 6.1, when ψ is continuous and the DDP or any of its variants have a.s. weakly continuous sample paths, then their mixture will have a.s. strongly continuous sample paths. Therefore, the process $\{M^{G^\omega}_{\boldsymbol{\gamma},\boldsymbol{x}}(B) : \boldsymbol{\gamma} \in \Gamma, \, \boldsymbol{x} \in \mathcal{X}\}$ for some fixed $B \in \mathcal{B}(\Theta)$ has a.s. continuous paths. Following the arguments in Section 4.3, for any $d \in \mathbb{N}$ and $f : [0,1]^d \to \mathbb{R}$ the process

$$\{f(M^G_{\boldsymbol{\gamma}_1,\boldsymbol{x}_1}(B), \ldots, M^G_{\boldsymbol{\gamma}_d,\boldsymbol{x}_d}(B)) : \; (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d) \in \mathcal{X}^d\},$$

has a.s. continuous sample paths. This also holds for its expectation

$$F_B(\boldsymbol{\gamma}_1, \boldsymbol{x}_1, \ldots, \boldsymbol{\gamma}_d, \boldsymbol{x}_d) := \mathbb{E}(f(M^G_{\boldsymbol{\gamma}_1,\boldsymbol{x}_1}(B), \ldots, M^G_{\boldsymbol{\gamma}_d,\boldsymbol{x}_d}(B))).$$

In some applications, it is useful to consider the parameter $\boldsymbol{\gamma}$ as random. Let $(\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_d)$ be a random vector defined on $(\Omega, \mathcal{F}, \mathbb{P})$. Then the process

$$\{F_B(\boldsymbol{\gamma}_1, \boldsymbol{x}_1, \ldots, \boldsymbol{\gamma}_d, \boldsymbol{x}_d) : \; (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d) \in \mathcal{X}^d\},$$

has a.s. continuous sample paths. By the same arguments as before,

$$\mathbb{E}(F_B(\boldsymbol{\gamma}_1, \boldsymbol{x}_1, \ldots, \boldsymbol{\gamma}_d, \boldsymbol{x}_d)) = \int_{\Gamma^d} F_B(\boldsymbol{\gamma}_1, \boldsymbol{x}_1, \ldots, \boldsymbol{\gamma}_d, \boldsymbol{x}_d) \, d\mathcal{P}_\Gamma(\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_d),$$

where $\mathcal{P}_\Gamma$ is the probability law of $(\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_d)$, is continuous.

6.4. Posterior consistency

An important property of mixture models induced by the DDP is their posterior consistency. To study the asymptotic behavior of these mixture models, we consider a random sample of size n given by pairs $(\boldsymbol{y}_i, \boldsymbol{x}_i)$ for $i \in [n]$. As is common in regression settings, we assume that $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ contain only exogenous covariates, which allows us to focus on the problem of conditional density estimation, regardless of the mechanism generating the predictors. Let P^0 be the true probability measure generating the predictors admitting a density p^0 with respect to a measure $\nu_{\mathcal{X}}$. By the exogeneity assumption, the true probability model for the response variable and predictors takes the form $m^0(\boldsymbol{y},\boldsymbol{x}) = p^0(\boldsymbol{x})q^0(\boldsymbol{y} \mid \boldsymbol{x})$. In this case, both p^0 and $\{q^0_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ are in free variation, where $q^0_{\boldsymbol{x}}(\boldsymbol{y}) := q^0(\boldsymbol{y} \mid \boldsymbol{x})$ denoting a conditional density defined on $\mathcal{Y}$ for every $\boldsymbol{x} \in \mathcal{X}$.

Let $m^\omega(\boldsymbol{y},\boldsymbol{x}) := p^0(\boldsymbol{x})g^\omega_{\boldsymbol{x}}(\boldsymbol{y})$ be the random joint distribution for the response and predictors arising when $\{g_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ is a mixture model induced by a θDDP. Since the KL divergence between m^0 and the implied joint distribution m^ω can be bounded as

$$\begin{aligned}
\mathrm{KL}(m^0 \| m^\omega) &= \int_{\mathcal{X}} \int_{\mathcal{Y}} m^0(\boldsymbol{y},\boldsymbol{x}) \log\left(\frac{m^0(\boldsymbol{y},\boldsymbol{x})}{m^\omega(\boldsymbol{y},\boldsymbol{x})}\right) d\nu_{\mathcal{Y}}(\boldsymbol{y}) d\nu_{\mathcal{X}}(\boldsymbol{x}),\\
&= \int_{\mathcal{X}} p^0(\boldsymbol{x}) \int_{\mathcal{Y}} q^0(\boldsymbol{y} \mid \boldsymbol{x}) \log\left(\frac{q^0(\boldsymbol{y} \mid \boldsymbol{x})}{g^\omega_{\boldsymbol{x}}(\boldsymbol{y})}\right) d\nu_{\mathcal{Y}}(\boldsymbol{y}) d\nu_{\mathcal{X}}(\boldsymbol{x}),\\
&\le \sup_{\boldsymbol{x}\in\mathcal{X}} \int_{\mathcal{Y}} q^0(\boldsymbol{y} \mid \boldsymbol{x}) \log\left(\frac{q^0(\boldsymbol{y} \mid \boldsymbol{x})}{g^\omega_{\boldsymbol{x}}(\boldsymbol{y})}\right) d\nu_{\mathcal{Y}}(\boldsymbol{y}),
\end{aligned}$$

when $\boldsymbol{x}$ contains only continuous predictors, it follows that, under the assumptions of Theorem 6.8 when $\mathcal{X}$ is compact, for every $\varepsilon > 0$,

$$\begin{aligned}
&\mathbb{P}(\{\omega \in \Omega : \mathrm{KL}(m^0 \| m^\omega) < \varepsilon\})\\
&\quad \ge \mathbb{P}\left(\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathcal{X}} \int_{\mathcal{Y}} q^0(\boldsymbol{y} \mid \boldsymbol{x}) \log\left(\frac{q^0(\boldsymbol{y} \mid \boldsymbol{x})}{g^\omega_{\boldsymbol{x}}(\boldsymbol{y})}\right) d\nu_{\mathcal{Y}}(\boldsymbol{y}) < \varepsilon\right\}\right) > 0.
\end{aligned}$$

Thus, by Schwartz's theorem [78], it follows that the posterior distribution associated with the random joint distribution induced by any of the proposed models is weakly consistent, that is, the posterior measure of any weak neighborhood, of any joint distribution of the form $m^0(\boldsymbol{y},\boldsymbol{x}) = p^0(\boldsymbol{x})q^0(\boldsymbol{y} \mid \boldsymbol{x})$, converges to 1 as the sample size goes to infinity. This result is summarized in the following theorem.

Theorem 6.9. *Suppose that assumptions of Theorem 6.8 hold. Then, the random distribution* $m^\omega(\boldsymbol{y},\boldsymbol{x}) = p^0(\boldsymbol{x})f(\boldsymbol{y}|\boldsymbol{x},G^\omega_{\boldsymbol{x}})$ *associated to the random joint distribution induced by the process* $G_{\mathcal{X}}$ *and the density* p^0 *generating the predictors, where*

$$f(\boldsymbol{y} \mid \boldsymbol{x}, G^\omega_{\boldsymbol{x}}) = \int_\Theta \psi(\boldsymbol{y},\boldsymbol{\theta}) dG^\omega_{\boldsymbol{x}}(\boldsymbol{\theta}),$$

is weakly consistent, under independent sampling, at any joint distribution of the form $m^0(\boldsymbol{y},\boldsymbol{x}) = p^0(\boldsymbol{x})f^0(\boldsymbol{y} \mid \boldsymbol{x})$ *with*

$$f^0(\boldsymbol{y} \mid \boldsymbol{x}) = \int_\Theta \psi(\boldsymbol{y},\boldsymbol{\theta}) dP^0_{\boldsymbol{x}}(\boldsymbol{\theta}),$$

and $P^0 \in C_S(\mathcal{X},\nu_{\mathcal{Y}})$*.*

Although Theorem 6.9 assumes that $\boldsymbol{x}$ contains only continuous predictors, a similar result can be obtained when $\boldsymbol{x}$ contains only predictors with finite support (e.g., categorical, ordinal and discrete predictors) or mixed continuous and predictors with finite support. The theorem can be also extended for more general DDP mixture models.

Theorem 6.10. *Suppose that assumptions in Theorem 6.8 hold and that $\mathcal{X}$ is compact. Then, the posterior distribution associated to the random joint distribution*

$$m(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\gamma}^{\omega}) = p^0(\boldsymbol{x}) f(\boldsymbol{y} \mid \boldsymbol{x}, G^{\omega}_{\boldsymbol{x}}, \boldsymbol{\gamma}^{\omega}),$$

where p^0 is the density generating the predictors and

$$f(\boldsymbol{y} \mid \boldsymbol{x}, G^{\omega}_{\boldsymbol{x}}, \boldsymbol{\gamma}^{\omega}) = \int_{\Theta} \psi(\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\gamma}^{\omega}) dG^{\omega}_{\boldsymbol{x}}(\boldsymbol{\theta}),$$

is weakly consistent, under independent sampling, at any joint distribution of the form

$$m^0(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\gamma}_0) = p^0(\boldsymbol{x}) f^0(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\gamma}_0),$$

with

$$f^0(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\gamma}_0) = \int_{\Theta} \psi(\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\gamma}_0) dP^0_{\boldsymbol{x}}(\boldsymbol{\theta}),$$

where $P^0 \in C_S(\mathcal{X}, \nu_{\mathcal{Y}})$ and $\boldsymbol{\gamma}_0 \in \Gamma$.

7. Illustrations

In this section we include two concrete and nontrivial applications where the data lies on a non Euclidean space and the covariates may or may not lie on an Euclidean space. The first example deals with circular data, whereas the second deals with shape analysis in Kendall's shape space. We present explicit constructions for both, and we show how these satisfy the conditions required by our theoretical results.

7.1. *Circular data*

In several applications there is a natural interest in the relation between circular data and Euclidean covariates. Examples include wild fire occurrences in ecology [2], the propagation of waves in oceanography [45], sudden infant death syndrome in medicine [63], and the relationship between the distance an animal has traveled and the direction of its movement; additional covariates, such as the type of species and the travel time, could be also included [34].

The majority of regression models for this type of data are parametric [34], which limits their application, for instance, in scenarios involving multimodal distributions [1]. Although frequentist non-parametric approaches have been recently proposed in the literature [1], it is challenging to assess uncertainty or to perform predictions with them. Hence, we show how to construct a suitable

DDP mixture model to illustrate the applicability of BNP models to this type of problem.

Circular data can be represented as points on the unit circle $\mathbb{S}^1 = \{\boldsymbol{y} \in \mathbb{R}^2 : \|\boldsymbol{y}\| = 1\}$. By choosing an arbitrary but fixed reference point, the elements of $\mathbb{S}^1$ can be represented as angles. Thus, without loss of generality, we let $\mathcal{Y} = [0, 2\pi)$ denote $\mathbb{R}$ endowed with the usual equivalence relation $\boldsymbol{y} \sim \boldsymbol{y} + 2\pi k$, for $k \in \mathbb{Z}$. A natural parametric candidate for constructing a BNP model in this context is the von Mises distribution, which density is given by

$$\psi(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}) = \frac{\exp(\boldsymbol{\gamma}\cos(\boldsymbol{y} - \boldsymbol{\theta}))}{2\pi I_0(\boldsymbol{\gamma})}, \quad \boldsymbol{y} \in \mathcal{Y},\, \boldsymbol{\theta} \in \Theta,\, \boldsymbol{\gamma} \in \Gamma,$$

where I_0 is the 0-th modified Bessel function of the first kind,

$$I_0(\boldsymbol{\gamma}) = \frac{1}{2\pi}\int_0^{2\pi} \exp(\boldsymbol{\gamma}\cos(\boldsymbol{y}))\, d\boldsymbol{y},$$

$\boldsymbol{\theta} \in \Theta = [0, 2\pi)$ is the location parameter, and $\boldsymbol{\gamma} \in \Gamma = \mathbb{R}_+$ is the dispersion parameter.

We assume Euclidean covariates, $\mathcal{X} = \mathbb{R}^p$, and use the DDP to build a $\mathcal{P}(\Theta)$-valued stochastic process. To construct the stochastic process associated to the weights, we consider Gaussian random fields, $\{Z_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$, with covariance functions given by

$$k_1(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-a\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2),$$

for $a > 0$, and define

$$V_{\boldsymbol{x}} = F^{-1}_{B,\boldsymbol{x}}(F_{Z,\boldsymbol{x}}(Z_{\boldsymbol{x}})),$$

where $F_{Z,\boldsymbol{x}}$ is the continuous cumulative distribution function of $Z_{\boldsymbol{x}}$ and

$$F^{-1}_{B,\boldsymbol{x}}(u) = 1 - (1-u)^{1/\alpha_{\boldsymbol{x}}},$$

where $\boldsymbol{x} \to \alpha_{\boldsymbol{x}}$ is continuous.

Several options to construct the stochastic process associated to the atoms can be considered, such as SinCos-GP and Angle-GP [39]. We study here the construction based on the Angle-GP. Let $\{Z^{\sin}_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ and $\{Z^{\cos}_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ be Gaussian random fields with the same covariance functions

$$k_2(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-b\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2),$$

for $b > 0$. The stochastic process $\{\boldsymbol{\theta}_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ is defined as the solution to the equations

$$\sin(\boldsymbol{\theta}_{\boldsymbol{x}}) = 2\Phi(Z^{\sin}_{\boldsymbol{x}}) - 1,$$

and

$$\cos(\boldsymbol{\theta}_{\boldsymbol{x}}) = 2\Phi(Z^{\cos}_{\boldsymbol{x}}) - 1,$$

using the usual equivalence relation between $\boldsymbol{\theta}$ and $\boldsymbol{\theta} + 2\pi k$ for $k \in \mathbb{Z}$.

This DDP and its induced DDP mixture model satisfy the conditions for several of our results. First, as the Gaussian random fields with kernels k_1 and k_2

are separable and have a.s. continuous sample paths, it follows by Definition 3.4 that this DDP is a continuous parameter DDP. Hence, Theorem 4.1 holds and it has a.s. weakly continuous sample paths. It is straightforward to verify that, for this choice of kernels, for any $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathcal{X}$ the law of the random vector

$$(\boldsymbol{\theta}_{1,\boldsymbol{x}_1}, \ldots, \boldsymbol{\theta}_{1,\boldsymbol{x}_n}),$$

has full support on Θ^n, and the law of the random vector

$$(V_{1,\boldsymbol{x}_1}, \ldots, V_{1,\boldsymbol{x}_n}),$$

has full support on $[0,1]^n$. Hence, Theorem 4.5 holds in this case and this DDP has full support on the product-weak topology. However, this conclusion can be strengthened. First, the kernel $k_1(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is universal [62], whence the comments in Section 4.2 imply that the stochastic process $\{V_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ satisfies the condition given in expression (7). Second, as Θ is compact and Angle-GP has full support, the condition given in expression (8) is satisfied. As a consequence, the stronger Theorem 4.6 holds in this case. In fact, this can be further refined as these results imply that Theorem 4.8 also holds in this case.

Regarding this DDP mixture, the compactness of $\mathbb{S}^1$ together with the continuity of ψ imply that ψ satisfies condition (10). Furthermore, since we may choose $\nu_{\mathcal{Y}}$ as a scaling of the Haar measure on $\mathbb{S}^1$, it is locally finite. Hence, Theorem 6.1 holds, and the DDP mixture is a.s. uniformly continuous. Finally, these conditions imply that Theorem 6.3 and Theorem 6.4 hold, and the DDP mixture has full support on the topologies induced by the Hellinger distance. As in addition $\mathbb{S}^1$ is compact, Theorem 6.5 and Theorem 6.6 hold, and the DDP mixture has full support on the topologies induced by the L^∞ distance, and, as both $\mathbb{S}^1$ is compact and ψ is strictly positive, Theorem 6.7 and Theorem 6.8 hold, and the DDP mixture has full support on the topologies induced by the KL-divergence.

To conclude this example, we briefly comment on regression models using angular covariates, known as circular-circular regression [33]. It this case $\mathcal{X} = [0, 2\pi)$ and our previous construction applies to this case as soon as we construct a Gaussian random field on $\mathbb{S}^1$ whose kernel is universal. In [62] it is shown that if the kernel has the form

$$k_1(\boldsymbol{x}_i, \boldsymbol{x}_j) = g(d(\boldsymbol{x}_i, \boldsymbol{x}_j)),$$

where $d(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is the geodesic distance on $\mathbb{S}^1$, and

$$g(t) = \sum_{k \in \mathbb{Z}_+} a_k P_k(\cos(t)),$$

for $a_k > 0$ and where P_k is the Chebyshev polynomial of degree k, then the kernel is universal. In this case, we may use $G_{\mathcal{X}} \sim \theta\text{DDP}(\Psi_V, G^0)$ for $G^0 \equiv \text{vM}(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)$, with $\boldsymbol{\theta}_0 \in [0, 2\pi)$ and $\boldsymbol{\gamma}_0 > 0$. This θDDP is constructed from a Gaussian random field with kernel k_1 that is separable and has a.s. continuous sample paths. Hence it is a continuous parameter θDDP whence Theorem 4.1 holds and this θDDP has a.s. weakly continuous sample paths. In fact, the stronger Theorem 4.4 holds in this case, and its sample paths are a.s. uniformly continuous. The choice of G^0 has full support on Θ whereas for $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathcal{X}$ we have, as before, that the law of the random vector

$$(V_{1,\boldsymbol{x}_1}, \ldots, V_{1,\boldsymbol{x}_n}),$$

has full support on $[0,1]^n$. Hence, Theorem 4.5 holds in this case. Similarly to our previous construction, the universality of the kernel $k_1(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and the full support of G^0 imply that both Theorem 4.6 and Theorem 4.8 hold in this case. Since this process is a θDDP, these conditions imply that the refinement in Theorem 4.9 also holds. Finally, since the hypotheses of Theorem 4.5 hold for this construction, the same results that held for the DDP mixture in the previous example also hold in this case.

7.2. Landmark-based planar shape data

There are many applications in which the shape of an object contains the information of interest. Hence, comparing data reduces to comparing differences between shapes. This kind of application arises frequently in computer vision, in object recognition, and in medical imaging. As an example, several authors have studied the changes in the shape of the corpus callosum under different pathologies [9, 10, 8]. In this case, it is usually of interest to model the effect that Euclidean covariates can have on its shape: the relationship between the age and the shape of the corpus callosum has been studied in [35] and [83], whereas he effect of age and sex on the shape of the corpus callosum in people with and without autism has been studied in [40]. To analyze shape data, some authors have proposed regression models that make parametric distributional assumptions which have limited interpretation [40]. Others have proposed geodesic regression models for which the probability distribution of the errors is not supported directly on the manifold [35]. Finally, some authors explore kernel-based regression models, but these do not account for data uncertainty [74]. A non-parametric Bayesian model, such as the dependent Dirichlet process, can be used to overcome these challenges when the probability distribution for the response variable is supported directly on the manifold.

Following Kendall [48], the shape of an object is the geometric information that remains once the effects of rotation, translation and scale are removed. To study planar shapes, usually a fixed number $k \in \mathbb{N}$ of landmarks are selected on each shape, and the differences in shapes are assumed to be those in their respective landmarks. Landmark-based planar shape data consists of k landmarks representing a planar shape. The landmarks can be interpreted as k points in the complex plane, and the shape this data represents can be modeled as a point in Kendall's shape space Σ_2^k. To construct this space we proceed as follows. Let $\boldsymbol{z}^0 = (z_1^0, \ldots, z_k^0)^T \in \mathbb{C}^k$ be the k landmarks in two dimensions. We define

$\boldsymbol{z}^H = (z_1^H, \ldots, z_{k-1}^H)^T := H\boldsymbol{z}^0$, where H is the $(k-1)\times k$ Helmert sub-matrix defined by taking the $k\times k$ Helmert matrix and removing the first row [27]. Notice that $\boldsymbol{z} = (z_1, \ldots, z_{k-1})^T := \boldsymbol{z}^H/\|\boldsymbol{z}^H\| \in$ where $\mathbb{CS}^{k-2} := \{\boldsymbol{z} \in \mathbb{C}^{k-1} : \boldsymbol{z}^*\boldsymbol{z} = 1\}$ is the unit complex sphere and where $\boldsymbol{z}^*$ denotes the complex conjugate of $\boldsymbol{z}^T$. The space $\mathbb{CS}^{k-2}$ is called the pre-shape space and corresponds to the space of all possible landmark-based data after eliminating the effects of translation and scaling. We endow this space with the following equivalence relation. For $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathbb{CS}^{k-2}$ we say $\boldsymbol{z}_1 \sim \boldsymbol{z}_2$ if and only if $\boldsymbol{z}_1 = e^{i\varrho}\boldsymbol{z}_2$ for some $\varrho \in [0, 2\pi)$ where $i^2 = -1$ is the imaginary unit. Thus, Kendall's shape space is the quotient $\Sigma_2^k := \mathbb{CS}^{k-2}/\sim$. Consequently, the equivalence class for a pre-shape $\boldsymbol{z} \in \mathbb{CS}^{k-2}$ is given by

$$[\boldsymbol{z}] := \{e^{i\varrho}\boldsymbol{z} : \varrho \in [0, 2\pi)\} \in \Sigma_2^k.$$

The complexity of analyzing objects in space Σ_2^k becomes apparent, as this space inherits the non-Euclidean structure of the space $\mathbb{CS}^{k-2}$ and it is also a quotient space. Furthermore, in practice k can be very large, requiring the use of statistical approaches to high-dimensional data.

We now construct a DDP variant for analyzing data in Kendall's shape space. First, let $\varphi(\boldsymbol{z}, \boldsymbol{\zeta}, \boldsymbol{\vartheta})$ be the probability density function of the complex Watson distribution on the pre-shape sphere $\mathbb{CS}^{k-2}$

$$\varphi(\boldsymbol{z}, \boldsymbol{\zeta}, \boldsymbol{\vartheta}) = c_1(\boldsymbol{\zeta})^{-1}\exp\{\boldsymbol{\zeta}|\boldsymbol{z}^*\boldsymbol{\vartheta}|^2\}, \qquad \boldsymbol{z}, \boldsymbol{\vartheta} \in \mathbb{CS}^{k-2},\ \boldsymbol{\zeta} > 0,$$

where $\boldsymbol{\zeta}$ is the concentration parameter, $\boldsymbol{\vartheta}$ is the modal vector on the pre-shape sphere, and the integrating constant $c_1(\boldsymbol{\zeta})$ is

$$c_1(\boldsymbol{\zeta}) = \frac{2\pi^{k-1}}{(k-2)!}\,{}_1F_1(1; k-1; \boldsymbol{\zeta}),$$

where ${}_1F_1$ is the confluent hypergeometric function of the first kind. By setting $\boldsymbol{\zeta} = 2\boldsymbol{\gamma}$ we can write [27, p. 227]

$$\varphi(\boldsymbol{z}, \boldsymbol{\zeta}, \boldsymbol{\vartheta}) = c_1(2\boldsymbol{\gamma})^{-1}e^{\boldsymbol{\gamma}}\exp\{\boldsymbol{\gamma}\cos(2\rho([H^T\boldsymbol{z}], [H^T\boldsymbol{\vartheta}]))\},$$

where $\rho([H^T\boldsymbol{z}], [H^T\boldsymbol{\vartheta}])$ is the Riemannian distance between the shapes $[H^T\boldsymbol{z}]$ and $[H^T\boldsymbol{\vartheta}]$. Let $\boldsymbol{y} = [\boldsymbol{z}]$ and $\boldsymbol{\theta} = [\boldsymbol{\vartheta}]$ be the shapes associated with $\boldsymbol{z}$ and $\boldsymbol{\vartheta}$ respectively. Then, a shape distribution with respect to the uniform measure can be obtained by making a change of variables to Kent's polar shape coordinates [27, p. 232]

$$\psi(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}) = c_1(2\boldsymbol{\gamma})^{-1}2\pi e^{\boldsymbol{\gamma}}\exp\{\boldsymbol{\gamma}\cos(2\rho(\boldsymbol{y}, \boldsymbol{\theta}))\}, \quad \boldsymbol{y}, \boldsymbol{\theta} \in \Sigma_2^k, \boldsymbol{\gamma} > 0.$$

Denote this probability distribution on Σ_2^k for fixed $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ as $\mathrm{D}(\boldsymbol{\theta}, \boldsymbol{\gamma})$. Let $\mathcal{X} = \mathbb{R}^p$ be the covariate space, let $\Gamma = \{\boldsymbol{\gamma} \in \mathbb{R} : \boldsymbol{\gamma} > 0\}$ be a dispersion parameter, let $\Theta = \Sigma_2^k$ be the space of location parameters and let $G_{\mathcal{X}} \sim \theta\mathrm{DDP}(\Psi_V, G^0)$ for $G^0 = \mathrm{D}(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)$ where $\boldsymbol{\theta}_0 \in \Sigma_2^k$ and $\boldsymbol{\gamma}_0 > 0$ are known. For the construction of the stochastic process associated with the weights, we proceed as in the previous example. Let $\{Z_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ be a Gaussian random field with covariance

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-a\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2),$$

with $a > 0$, and define

$$V_{\boldsymbol{x}} = F_{B,\boldsymbol{x}}^{-1}(F_{Z,\boldsymbol{x}}(Z_{\boldsymbol{x}})),$$

where $F_{Z,\boldsymbol{x}}$ is the continuous cumulative distribution function of $Z_{\boldsymbol{x}}$ and

$$F_{B,\boldsymbol{x}}^{-1}(u) = 1 - (1-u)^{1/\alpha_{\boldsymbol{x}}},$$

with $\boldsymbol{x} \to \alpha_{\boldsymbol{x}}$ a continuous function.

As in the previous example, it can be shown that the resulting θDDP and DDP mixture model satisfy the conditions of several of our results. This θDDP is constructed from a Gaussian random field with kernel k_1 on $\mathcal{X}$ that is separable and has a.s. continuous sample paths. Hence, it is a continuous parameter θDDP whence Theorem 4.1 holds and this θDDP has a.s. weakly continuous sample paths. Similarly as in the last construction in the previous section, the stronger Theorem 4.4 holds in this case, and the sample paths of the process are a.s. uniformly continuous. The choice of G^0 has full support on Θ whereas for $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathcal{X}$ we once again have that the law of the random vector

$$(V_{1,\boldsymbol{x}_1}, \ldots, V_{1,\boldsymbol{x}_n}),$$

has full support on $[0,1]^n$. Therefore, Theorem 4.5 holds in this case and this process has full support on the product-weak topology. Finally, the universality of the kernel $k_1(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and the full support of G^0 imply that both Theorem 4.6 and Theorem 4.8 hold, and, as this process is a θDDP, it follows that the refinement in Theorem 4.9 also holds for this process.

Regarding the mixture, the conditions satisfied by ψ are analogous to those shown in the first example in the previous section. The compactness of Σ_2^k and the continuity of ψ imply that condition (10) holds. Furthermore, as Σ_2^k can be identified with a subset of the unit complex sphere, may choose $\nu_{\mathcal{Y}}$ as the measure induced by a scaling of the Haar measure on the complex unit sphere, whence it is locally finite. Hence, Theorem 6.1 holds, and the DDP mixture is a.s. uniformly continuous. Finally, these conditions imply that Theorem 6.3 and Theorem 6.4 hold, and the DDP mixture has full support on the topologies induced by the Hellinger distance; as in addition Σ_2^k is compact, Theorem 6.5 and Theorem 6.6 hold, and the DDP mixture has full support on the topologies induced by the L^∞ distance; and, as both Σ_2^k is compact and ψ is strictly positive, Theorem 6.7 and Theorem 6.8 hold, and the DDP mixture has full support on the topologies induced by the KL-divergence.

8. Concluding remarks

We have defined a DDP for general Polish spaces and introduced some parsimonious variants that may be desirable on specific applications. Furthermore, we provided sufficient conditions for different versions of DDP defined on general Polish spaces to have appealing prior theoretical properties regarding the continuity of their sample paths under different topologies, and the continuity of their autocovariance function and other, more general measures of association.

These properties are of practical importance because they ensure that different versions of the model can combine and borrow strength across sparse data sources regarding the predictors and, therefore, avoid the need of replicates of the responses for every value of the predictors to obtain adequate estimates of the predictor-dependent probability distributions.

Furthermore, we studied mixture models arising from a DDP or any of its variants. We provided sufficient conditions that ensure the resulting mixture has a continuous density. This case is of practical interest in statistical applications. In addition, we provided sufficient conditions under which DDP mixture models have large full or large support, considering different topologies, and study the behavior of the posterior distribution under i.i.d. joint sampling of responses and predictors. The study of stronger consistency results and concentration rates is the subject of ongoing research.

Finally, the results provided in this article can be easily extended to more general dependent stick-breaking processes.

Proof of the main results

A.1. The continuity of the DDP and its variants

A.1.1. Proof of Theorem 4.1

By hypothesis, there is a set $\Omega_0 \subset \Omega$ of full measure such that for any $\omega \in \Omega_0$ we have that: (i) $\pi_i^\omega \geq 0$; (ii) $\sum_{i\in\mathbb{N}} \pi_i^\omega \equiv 1$; (iii) π_i^ω and $\boldsymbol{\theta}_i^\omega$ are continuous on $\mathcal{X}$ in the case of a DDP, $\boldsymbol{\theta}_i^\omega$ is continuous on $\mathcal{X}$ in the case of a wDDP, and π_i^ω is continuous on $\mathcal{X}$ in the case of a θDDP.

First, we prove the statement in the case of a DDP. Fix $\omega \in \Omega_0$ and $\boldsymbol{x}_0 \in \mathcal{X}$. Let $f \in C_b(\Theta)$ and suppose, without loss, that $\|f\|_C \leq 1$. Fix $\varepsilon > 0$ and let $N_\varepsilon \in \mathbb{N}$ be such that

$$\sum_{i>N_\varepsilon} \pi_{i,x_0}^\omega < \frac{1}{4}\varepsilon.$$

Define the set U_ε as

$$U_\varepsilon := \bigcap_{i=1}^{N_\varepsilon} \left\{ \boldsymbol{x} \in \mathcal{X} : \; |\pi_{i,\boldsymbol{x}}^\omega - \pi_{i,x_0}^\omega| < \frac{1}{4N_\varepsilon}\varepsilon \quad \text{and} \quad |f(\boldsymbol{\theta}_{i,\boldsymbol{x}}^\omega) - f(\boldsymbol{\theta}_{i,x_0}^\omega))| < \frac{1}{4N_\varepsilon}\varepsilon \right\}.$$

Since π_i^ω and $f \circ \boldsymbol{\theta}_i^\omega$ are continuous on Θ we conclude that U_ε is an open neighborhood of $\boldsymbol{x}_0$. Observe that for any $\boldsymbol{x} \in U_\varepsilon$ we have that

$$\sum_{i>N_\varepsilon} \pi_{i,\boldsymbol{x}}^\omega = 1 - \sum_{i=1}^{N_\varepsilon} \pi_{i,\boldsymbol{x}}^\omega = \sum_{i>N_\varepsilon} \pi_{i,x_0}^\omega - \sum_{i=1}^{N_\varepsilon} (\pi_{i,\boldsymbol{x}}^\omega - \pi_{i,x_0}^\omega) < \frac{1}{2}\varepsilon.$$

Let $\boldsymbol{x} \in U_\varepsilon$ and consider the decomposition

$$\int f(\boldsymbol{\theta}) dG_{\boldsymbol{x}}^\omega(\boldsymbol{\theta}) - \int f(\boldsymbol{\theta}) dG_{\boldsymbol{x}_0}^\omega(\boldsymbol{\theta}) = \left(\sum_{i=1}^{N_\varepsilon} + \sum_{i>N_\varepsilon} \right) (\pi_{i,\boldsymbol{x}}^\omega f(\boldsymbol{\theta}_{i,\boldsymbol{x}}^\omega) - \pi_{i,x_0}^\omega f(\boldsymbol{\theta}_{i,x_0}^\omega)).$$

The first sum on the right-hand side can be bounded as

$$\left|\sum_{i=1}^{N_\varepsilon}(\pi_{i,\boldsymbol{x}}^\omega f(\boldsymbol{\theta}_{i,\boldsymbol{x}}^\omega)-\pi_{i,x_0}^\omega f(\boldsymbol{\theta}_{i,x_0}^\omega))\right| \le \sum_{i=1}^{N_\varepsilon}|\pi_{i,\boldsymbol{x}}^\omega-\pi_{i,x_0}^\omega|+\sum_{i=1}^{N_\varepsilon}|f(\boldsymbol{\theta}_{i,\boldsymbol{x}}^\omega)-f(\boldsymbol{\theta}_{i,x_0}^\omega)|, \\ < \frac{1}{2}\varepsilon,$$

whereas the second sum can be bounded as

$$\left|\sum_{i>N_\varepsilon}(\pi_{i,\boldsymbol{x}}^\omega f(\boldsymbol{\theta}_{i,\boldsymbol{x}}^\omega)-\pi_{i,x_0}^\omega f(\boldsymbol{\theta}_{i,x_0}^\omega))\right| \le \sum_{i>N_\varepsilon}\pi_{i,\boldsymbol{x}}^\omega+\sum_{i>N_\varepsilon}\pi_{i,x_0}^\omega < \frac{1}{2}\varepsilon.$$

Consequently,

$$\left|\int f(\boldsymbol{\theta})dG_{\boldsymbol{x}}^\omega(\boldsymbol{\theta})-\int f(\boldsymbol{\theta})dG_{\boldsymbol{x}_0}^\omega(\boldsymbol{\theta})\right|<\varepsilon,$$

for any $\boldsymbol{x}\in U_\varepsilon$. Since U_ε is open, the claim follows in the case of a DDP.

Remark that in the case of a wDDP or the case of a θDDP the previous arguments can be adapted with minor modifications to prove the claim. We omit the details for brevity. □

A.1.2. Proof of Theorem 4.3

By hypothesis, there is a set $\Omega_0\subset\Omega$ of full measure such that for any $\omega\in\Omega_0$ we have that: (i) $\pi_i^\omega\ge 0$; (ii) $\sum_{i\in\mathbb{N}}\pi_i^\omega\equiv 1$; (iii) π_i^ω and $\boldsymbol{\theta}_i^\omega$ are continuous on $\mathcal{X}$ in the case of a DDP, $\boldsymbol{\theta}_i^\omega$ is continuous on $\mathcal{X}$ in the case of a wDDP, and π_i^ω is continuous on $\mathcal{X}$ in the case of a θDDP.

First, we prove the statement in the case of a DDP. Fix $\omega\in\Omega_0$ and let $\varepsilon>0$. Let N_ε be such that

$$\sum_{i>N_\varepsilon}\pi_{i,x_0}^\omega<\frac{1}{8}\varepsilon.$$

Define the measurable set

$$B:=\{\boldsymbol{\theta}_{i,x_0}^\omega:\ i\in[N_\varepsilon]\},$$

and let $f=\mathbb{I}_B$. Then

$$F(\boldsymbol{x}):=\int f(\boldsymbol{\theta})dG_{\boldsymbol{x}}^\omega(\boldsymbol{\theta}),$$

is continuous at $\boldsymbol{x}_0$ by hypothesis. Hence, the set

$$U_\varepsilon:=\left\{\boldsymbol{x}\in\mathcal{X}:\ |F(\boldsymbol{x})-F(\boldsymbol{x}_0)|<\frac{1}{4}\varepsilon\right\}\cap\bigcap_{i=1}^{N_\varepsilon}\left\{\boldsymbol{x}\in\mathcal{X}:\ |\pi_{i,\boldsymbol{x}}^\omega-\pi_{i,x_0}^\omega|<\frac{1}{8N_\varepsilon}\varepsilon\right\},$$

is open. For any $\boldsymbol{x}\in U_\varepsilon$ we have that

$$\sum_{i>N_\varepsilon}\pi_{i,\boldsymbol{x}}^\omega<\frac{1}{4}\varepsilon,$$

and that

$$\begin{aligned}\left|\sum_{i=1}^{N_\varepsilon}\pi^\omega_{i,x_0}(\delta_{\boldsymbol{\theta}^\omega_{i,\boldsymbol{x}}}(B)-1)\right| &\leq \sum_{i=1}^{N_\varepsilon}|\pi^\omega_{i,\boldsymbol{x}}-\pi^\omega_{i,x_0}|+\left|\sum_{i=1}^{N_\varepsilon}(\pi^\omega_{i,\boldsymbol{x}}\delta_{\boldsymbol{\theta}^\omega_{i,\boldsymbol{x}}}(B)-\pi^\omega_{i,x_0}\delta_{\boldsymbol{\theta}^\omega_{i,x_0}}(B))\right|,\\ &< \frac{1}{8}\varepsilon+\sum_{i>N_\varepsilon}(\pi^\omega_{i,x_0}+\pi^\omega_{i,\boldsymbol{x}})+|F(\boldsymbol{x})-F(\boldsymbol{x}_0)|,\\ &< \varepsilon.\end{aligned}$$

It is clear that at least one $\boldsymbol{\theta}^\omega_{i,\boldsymbol{x}}$ belongs to B. Note that B is a finite set. Hence, by continuity, such $\boldsymbol{\theta}^\omega_{i,\boldsymbol{x}}$ must be constant on U_ε. This proves the claim in the case of a DDP. For the wDDP the previous arguments can be adapted with minor modifications to prove the claim. We omit the details for brevity. □

A.1.3. Proof of Theorem 4.4

In the case of a θDDP, for every $\omega\in\Omega$ we have that

$$\sup_{B\in\mathcal{B}(\Theta)}|G^\omega_{\boldsymbol{x}}(B)-G^\omega_{\boldsymbol{x}_0}(B)|\leq\sum_{i=1}^{\infty}|\pi^\omega_{i,\boldsymbol{x}}-\pi^\omega_{i,x_0}|.$$

Let $\Omega_0\subset\Omega$ be a set of full measure such that $\omega\in\Omega_0$ implies V_i^ω is continuous for every $i\in\mathbb{N}$. Fix $\omega\in\Omega_0$. Then π_i^ω is continuous for every $i\in\mathbb{N}$. We follow a similar argument as in the proof of Theorem 4.1 in Appendix A.1.1. Let $\varepsilon>0$ and let $N_\varepsilon\in\mathbb{N}$ be such that

$$\sum_{i>N_\varepsilon}\pi^\omega_{i,x_0}<\frac{1}{4}\varepsilon.$$

Define the open neighborhood U_ε of $\boldsymbol{x}_0$ as

$$U_\varepsilon:=\bigcap_{i=1}^{N_\varepsilon}\left\{\boldsymbol{x}\in\mathcal{X}:\ |\pi^\omega_{i,\boldsymbol{x}}-\pi^\omega_{i,x_0}|<\frac{1}{4N_\varepsilon}\varepsilon\right\}.$$

For any $\boldsymbol{x}\in U_\varepsilon$ we have that

$$\sum_{i>N_\varepsilon}\pi^\omega_{i,\boldsymbol{x}}=\sum_{i>N_\varepsilon}\pi^\omega_{i,x_0}-\sum_{i=1}^{N_\varepsilon}(\pi^\omega_{i,\boldsymbol{x}}-\pi^\omega_{i,x_0})<\frac{1}{2}\varepsilon,$$

whence

$$\sum_{i=1}^{\infty}|\pi^\omega_{i,\boldsymbol{x}}-\pi^\omega_{i,x_0}|\leq\sum_{i=1}^{N_\varepsilon}|\pi^\omega_{i,\boldsymbol{x}}-\pi^\omega_{i,x_0}|+\sum_{i>N_\varepsilon}(\pi^\omega_{i,\boldsymbol{x}}+\pi^\omega_{i,x_0})<\varepsilon.$$

Consequently, for any $\boldsymbol{x}\in U_\varepsilon$ we have that

$$\sup_{B\in\mathcal{B}(\Theta)}|G^\omega_{\boldsymbol{x}}(B)-G^\omega_{\boldsymbol{x}_0}(B)|<\varepsilon,$$

from where the theorem follows. □

A.2. The support of the DDP and its variants

We first prove the following auxiliary result.

Lemma A.1. *Let $P : \mathcal{X} \to \mathcal{P}(\Theta)$ and $f_1, \dots, f_n \in L^\infty(\Theta)$. Define*

$$F_i(\boldsymbol{x}) := \int_\Theta f_i(\boldsymbol{\theta})\, dP_{\boldsymbol{x}}(\boldsymbol{\theta}).$$

Let $K \subset \mathcal{X}$ be compact. If $F_1, \dots, F_n : K \to \mathbb{R}$ are continuous, then for any $\varepsilon > 0$ there exists $\bar{P} : K \to \mathcal{P}(\Theta)^{\mathcal{X}}$ such that:

1. *For any $\boldsymbol{x} \in K$ we have that*

$$\left| \int_\Theta f_i(\boldsymbol{\theta})\, dP_{\boldsymbol{x}}(\boldsymbol{\theta}) - \int_\Theta f_i(\boldsymbol{\theta})\, d\bar{P}_{\boldsymbol{x}}(\boldsymbol{\theta}) \right| < \varepsilon.$$

2. *For any $B \in \mathcal{B}(\Theta)$ the map $\boldsymbol{x} \mapsto \bar{P}_{\boldsymbol{x}}(B)$ is Lipschitz continuous on K.*
3. *The collection $\{\bar{P}_{\boldsymbol{x}} : \boldsymbol{x} \in K\}$ is tight.*

Proof of Lemma A.1. Let $\varepsilon > 0$. We begin by constructing a suitable partition of unity in K. Since $F_1, \dots, F_n$ are uniformly continuous on K there exists $\delta > 0$ such that

$$\forall\, \boldsymbol{x}, \boldsymbol{x}' \in K,\ i \in [n] :\ d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}') < \delta \ \Rightarrow\ |F_i(\boldsymbol{x}) - F_i(\boldsymbol{x}')| < \varepsilon.$$

Let $r < \delta/2$. From the open cover $\{B(\boldsymbol{x}, r)\}_{\boldsymbol{x} \in K}$ we can extract the finite subcover $\{B(\boldsymbol{x}_k, r)\}_{k=1}^{N_K}$. We construct a continuous partition of the unity on K surrogate to this cover as follows. Define the continuous functions

$$\bar{\varphi}_k(\boldsymbol{x}) = \max\left(0, 1 - \frac{d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}_k)}{r}\right),$$

for $k \in [N_K]$. Note that $\boldsymbol{x} \notin B(\boldsymbol{x}_k, r)$ implies $\bar{\varphi}_k(\boldsymbol{x}) = 0$. It can be verified that

$$\sum_{k=1}^{N_K} \bar{\varphi}_k(\boldsymbol{x}) \geq c_{\min} > 0,$$

for any $\boldsymbol{x} \in K$. Therefore, we define the continuous functions

$$\varphi_k(\boldsymbol{x}) = \frac{\bar{\varphi}_k(\boldsymbol{x})}{\sum_{k=1}^{N_K} \bar{\varphi}_k(\boldsymbol{x})},$$

which satisfy $\sum_{k=1}^{N_K} \varphi_k \equiv 1$ over K. Define

$$\bar{F}_i(\boldsymbol{x}) := \sum_{k=1}^{N_K} F_i(\boldsymbol{x}_k)\varphi_k(\boldsymbol{x}) \quad \text{and} \quad \bar{P}_{\boldsymbol{x}} := \sum_{k=1}^{N_K} \varphi_k(\boldsymbol{x}) P_{\boldsymbol{x}_k}.$$

These functions satisfy

$$\bar{F}_i(\boldsymbol{x}) := \int_\Theta f_i(\boldsymbol{\theta}) \left(\sum_{k=1}^{N_K} \varphi_k(\boldsymbol{x}) dP_{\boldsymbol{x}_k}(\boldsymbol{\theta}) \right) = \int_\Theta f_i(\boldsymbol{\theta}) d\bar{P}_{\boldsymbol{x}}(\boldsymbol{\theta}).$$

To prove (a) note that $\bar{F}_i$ satisfies

$$|\bar{F}_i(\boldsymbol{x}) - F_i(\boldsymbol{x})| \leq \sum_{k:\, \boldsymbol{x} \in B(\boldsymbol{x}_k, r)} |F_i(\boldsymbol{x}_k) - F_i(\boldsymbol{x})| \varphi_k(\boldsymbol{x}) < \varepsilon,$$

for any $\boldsymbol{x} \in K$. To prove (b) note that for any $B \in \mathcal{B}(\Theta)$ we have that

$$\bar{P}_{\boldsymbol{x}}(B) = \sum_{k=1}^{N_K} \varphi_i(\boldsymbol{x}) \bar{P}_{\boldsymbol{x}_k}(B),$$

which is continuous by construction. To prove it is Lipschitz, note that

$$\begin{aligned} |\varphi_i(\boldsymbol{x}) - \varphi_i(\boldsymbol{x}')| \leq & \frac{|\bar{\varphi}_i(\boldsymbol{x}) - \bar{\varphi}_i(\boldsymbol{x}')|}{\sum_{j=1}^{N_K} \bar{\varphi}_j(\boldsymbol{x})} + \\ & \frac{\bar{\varphi}_j(\boldsymbol{x}')}{\sum_{j=1}^{N_K} \bar{\varphi}_j(\boldsymbol{x}) \sum_{k=1}^{N_K} \bar{\varphi}_k(\boldsymbol{x}')} \sum_{\ell=1}^{N_K} |\bar{\varphi}_\ell(\boldsymbol{x}) - \bar{\varphi}_\ell(\boldsymbol{x}')|, \\ \leq & c_{\min}^{-1}(N_K c_{\min}^{-1} + 1) d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}'). \end{aligned}$$

Finally, to prove (c) note that Θ is Polish. Hence, the collection $P_{\boldsymbol{x}_1}, \ldots, P_{\boldsymbol{x}_{N_K}}$ is tight. For any $\varepsilon > 0$ there exists $K_\Theta \subset \Theta$ such that

$$P_{\boldsymbol{x}_k}(K_\Theta) > 1 - \varepsilon.$$

However, since $\varphi_i \geq 0$ we have that

$$\bar{P}_{\boldsymbol{x}}(B) = \sum_{k=1}^{N_K} \varphi_i(\boldsymbol{x}) \bar{P}_{\boldsymbol{x}_k}(B) > 1 - \varepsilon,$$

proving the claim. □

A.2.1. Proof of Theorem 4.5

By possibly modifying the sequence $\varepsilon_{1,1}, \ldots, \varepsilon_{n,n}$ we can assume, without loss, that $\|f_{i,j}\|_C \leq 1$. Let $\Omega_0 \subset \Omega$ be a set of full measure such that for any $\omega \in \Omega_0$ we have that: (i) $\pi_i^\omega \geq 0$; (ii) $\sum_{i \in \mathbb{N}} \pi_i^\omega \equiv 1$; (iii) π_i^ω and $\boldsymbol{\theta}_i^\omega$ are continuous on $\mathcal{X}$ in the case of a DDP, $\boldsymbol{\theta}_i^\omega$ are continuous on $\mathcal{X}$ in the case of a wDDP, or π_i^ω are continuous on $\mathcal{X}$ in the case of a θDDP.

Let $\varepsilon > 0$ be such that $\varepsilon < \varepsilon_0 := \min(\varepsilon_{1,1}, \ldots, \varepsilon_{n,n})$ and define the event

$$\left\{ \omega \in \Omega : \left| \int_\Theta f_{i,j}(\boldsymbol{\theta}) dG_{\boldsymbol{x}_j}^\omega(\boldsymbol{\theta}) - \int_\Theta f_{i,j}(\boldsymbol{\theta}) dP_{\boldsymbol{x}_j}^0(\boldsymbol{\theta}) \right| < \varepsilon_0,\ i, j \in [n] \right\}.$$

By inspection, the event is measurable. It suffices to show that this event has positive probability. The first step of the proof is the same regardless of the variant.

Since Θ is Polish, the collection $P^0_{\boldsymbol{x}_1}, \ldots, P^0_{\boldsymbol{x}_n}$ is tight. Let $K_\Theta \subset \Theta$ be a compact set such that $P^0_{\boldsymbol{x}_j}(\Theta \setminus K_\Theta) < \varepsilon$ for $j \in [n]$. Since $f_{1,1}, \ldots, f_{n,n}$ are uniformly continuous on K_Θ, there exists $\delta > 0$ such that

$$\forall\ \boldsymbol{\theta}, \boldsymbol{\theta}' \in K_\Theta,\ i, j \in [n]:\ d_\Theta(\boldsymbol{\theta}, \boldsymbol{\theta}') < \delta \quad \Rightarrow \quad |f_{i,j}(\boldsymbol{\theta}) - f_{i,j}(\boldsymbol{\theta}')| < \varepsilon.$$

Let $r < \delta/2$. Then $\{B(\boldsymbol{\theta}, r)\}_{\boldsymbol{\theta} \in K_\Theta}$ is an open cover of K_Θ and we can extract a finite subcover $\{B(\boldsymbol{\theta}_k, r)\}_{k=1}^{N_\Theta}$. By possibly removing elements, we may assume that no ball is covered by the union of the remanining ones. This subcover induces the partition

$$\begin{aligned} A_1 &:= K_\Theta \cap B(\boldsymbol{\theta}_k, r), \\ A_k &:= K_\Theta \cap \left(B(\boldsymbol{\theta}_k, r) \setminus \bigcup_{\ell=1}^{k-1} B(\boldsymbol{\theta}_\ell, r) \right), \qquad k \in \{2, \ldots, N_\Theta\}. \end{aligned} \tag{11}$$

of K_Θ. Remark that no A_k is empty and that $A_k \subset B(\boldsymbol{\theta}_k, r)$. Therefore, no f_i varies by more than 2ε over any A_k. Note that

$$\left| \int_{\Theta \setminus K_\Theta} f_{i,j}(\boldsymbol{\theta}) dP^0_{\boldsymbol{x}_j}(\boldsymbol{\theta}) \right| \leq P^0_{\boldsymbol{x}_j}(\Theta \setminus K_\Theta) < \varepsilon,$$

and

$$\left| \int_{K_\Theta} f_{i,j}(\boldsymbol{\theta}) dP^0_{\boldsymbol{x}_i}(\boldsymbol{\theta}) - \sum_{k=1}^{N_\Theta} f_{i,j}(\boldsymbol{\theta}_k) P^0_{\boldsymbol{x}_j}(A_k) \right| < 2\varepsilon \sum_{k=1}^{N_\Theta} P_{\boldsymbol{x}_j}(A_k) \leq 2\varepsilon.$$

Therefore,

$$\begin{aligned} &\left| \int_\Theta f_{i,j}(\boldsymbol{\theta}) dG^\omega_{\boldsymbol{x}_j}(\boldsymbol{\theta}) - \int_\Theta f_{i,j}(\boldsymbol{\theta}) dP^0_{\boldsymbol{x}_j}(\boldsymbol{\theta}) \right| \\ &\qquad < 3\varepsilon + \left| \int_\Theta f_{i,j}(\boldsymbol{\theta}) dG^\omega_{\boldsymbol{x}_j}(\boldsymbol{\theta}) - \sum_{k=1}^{N_\Theta} f_{i,j}(\boldsymbol{\theta}_k) P^0_{\boldsymbol{x}_j}(A_k) \right|. \end{aligned}$$

Since $f_{1,1}, \ldots, f_{n,n}$ are continuous, define the open subsets

$$U_k := \bigcap_{j=1}^n \bigcap_{i=1}^n \left\{ \boldsymbol{\theta} \in \Theta:\ |f_{i,j}(\boldsymbol{\theta}) - f_{i,j}(\boldsymbol{\theta}_k)| < \frac{1}{N_\Theta} \varepsilon \right\},$$

and the event

$$\Omega_\pi := \bigcap_{k=1}^{N_\Theta} \bigcap_{j=1}^n \left\{ \omega \in \Omega_0:\ |\pi^\omega_{k,\boldsymbol{x}_j} - P^0_{\boldsymbol{x}_j}(A_k)| < \frac{1}{N_\Theta} \varepsilon \right\},$$

which has positive measure by hypothesis. Note that for any $\omega \in \Omega_\pi$ we have that

$$\sum_{k>N_\Theta} \pi^\omega_{k,\boldsymbol{x}_j} = 1 - \sum_{k=1}^{N_\Theta} P^0_{\boldsymbol{x}_j}(A_k) - \sum_{k=1}^{N_\Theta} (\pi^\omega_{k,\boldsymbol{x}_j} - P^0_{\boldsymbol{x}_j}(A_k)) < 2\varepsilon,$$

for $i \in [N_\Theta]$.

The second step of the proof changes slightly in the case of a DDP, a wDDP, and the θDDP. We present it first in the case of a DDP. For $\omega \in \Omega_\pi$ we have that

$$\begin{aligned}
&\left| \int_\Theta f_{i,j}(\boldsymbol{\theta}) dG^\omega_{\boldsymbol{x}_j}(\boldsymbol{\theta}) - \sum_{k=1}^{N_\Theta} f_{i,j}(\boldsymbol{\theta}_k) P^0_{\boldsymbol{x}_j}(A_k) \right| \\
&\qquad \leq \sum_{k>N_\Theta} \pi^\omega_{k,\boldsymbol{x}_j} + \left| \sum_{k=1}^{N_\Theta} (\pi^\omega_{k,\boldsymbol{x}_j} f_{i,j}(\boldsymbol{\theta}^\omega_{k,\boldsymbol{x}_j}) - P^0_{\boldsymbol{x}_j}(A_k) f_{i,j}(\boldsymbol{\theta}_k)) \right| \\
&\qquad\qquad \leq 3\varepsilon + \sum_{k=1}^{N_\Theta} P^0_{\boldsymbol{x}_j}(A_k) |f_i(\boldsymbol{\theta}^\omega_{k,\boldsymbol{x}_j}) - f_{i,j}(\boldsymbol{\theta}_k)|.
\end{aligned}$$

Consider the event

$$\Omega_\theta := \bigcap_{k=1}^{N_\Theta} \bigcap_{j=1}^{n} \{\omega \in \Omega_0 : \ \boldsymbol{\theta}^\omega_{k,\boldsymbol{x}_j} \in U_k\},$$

which has positive measure by hypothesis as each U_k is open. Since the events Ω_π and Ω_θ are independent, the event $\Omega_\pi \cap \Omega_\theta$ has positive measure. Hence, for any $\omega \in \Omega_\pi \cap \Omega_\theta$ we have that

$$\left| \int_\Theta f_{i,j}(\boldsymbol{\theta}) dG^\omega_{\boldsymbol{x}_j}(\boldsymbol{\theta}) - \sum_{k=1}^{N_\Theta} f_{i,j}(\boldsymbol{\theta}_k) P^0_{\boldsymbol{x}_j}(A_k) \right| < 4\varepsilon,$$

whence

$$\left| \int_\Theta f_{i,j}(\boldsymbol{\theta}) dG^\omega_{\boldsymbol{x}_j}(\boldsymbol{\theta}) - \int_\Theta f_{i,j}(\boldsymbol{\theta}) dP^0_{\boldsymbol{x}_j}(\boldsymbol{\theta}) \right| < 7\varepsilon,$$

for any $i, j \in [n]$. This proves the theorem in the case of a DDP.

To prove the theorem in the case of a θDDP, the key inequality is

$$\left| \int_\Theta f_{i,j}(\boldsymbol{\theta}) dG^\omega_{\boldsymbol{x}_j}(\boldsymbol{\theta}) - \sum_{k=1}^{N_\Theta} f_{i,j}(\boldsymbol{\theta}_k) P^0_{\boldsymbol{x}_j}(A_k) \right| < 3\varepsilon + \sum_{k=1}^{N_\Theta} P^0_{\boldsymbol{x}_j}(A_k) |f_i(\boldsymbol{\theta}^\omega_k) - f_{i,j}(\boldsymbol{\theta}_k)|.$$

Note that it suffices to consider the event

$$\Omega_\theta := \bigcap_{k=1}^{N_\Theta} \{\omega \in \Omega_0 : \ \boldsymbol{\theta}^\omega_k \in U_k\},$$

which has positive measure as the U_k are open. The rest of the argument is essentially the same as in the case of a DDP.

The proof in the case of a wDDP follows first the same step as in the other cases. However, the remainder of the proof is different and slighly more involved. Let $M \in \mathbb{N}$ be such that there are integers $m_{j,k} \in [M]$ such that

$$\left| P^0_{\boldsymbol{x}_j}(A_k) - \frac{m_{j,k}}{M} \right| < \frac{\varepsilon}{2N_\Theta}.$$

Note that for this choice, for any $j \in [n]$,

$$\sum_{k=1}^{N_\Theta} \frac{m_{j,k}}{M} < \sum_{k=1}^{N_\Theta} P^0_{\boldsymbol{x}_j}(A_k) + \frac{1}{2}\varepsilon \leq 1 + \frac{1}{2}\varepsilon,$$

and

$$\sum_{k=1}^{N_\Theta} \frac{m_{j,k}}{M} \geq \sum_{k=1}^{N_\Theta} P^0_{\boldsymbol{x}_j}(A_k) - \frac{\varepsilon}{2} = P^0_{\boldsymbol{x}_j}(K_\Theta) - \frac{\varepsilon}{2} \geq 1 - \frac{3}{2}\varepsilon.$$

Define the event

$$\Omega_\pi := \bigcap_{i=1}^{M} \left\{ \omega \in \Omega_0 : \pi_i^\omega \in \left(\frac{1 - \varepsilon/2N_\Theta}{M}, \frac{1}{M} \right] \right\},$$

which has positive measure by hypothesis. Note that for $\omega \in \Omega_\pi$ we have that

$$\left(1 - \frac{\varepsilon}{2N_\Theta}\right) \frac{M'}{M} < \sum_{i=1}^{M'} \pi_i^\omega \leq \frac{M'}{M},$$

for any $M' \in [M]$. Hence, for every j we have that

$$\begin{aligned} \left| P^0_{\boldsymbol{x}_j}(A_k) - \sum_{i \in I_{j,k}} \pi_i^\omega \right| &\leq \left| P^0_{\boldsymbol{x}_j}(A_k) - \frac{m_{j,k}}{M} \right| + \left| \frac{m_{j,k}}{M} - \sum_{i \in I_{j,k}} \pi_i^\omega \right| \\ &< \frac{\varepsilon}{2N_\Theta} + \frac{\varepsilon}{2N_\Theta} = \frac{\varepsilon}{N_\Theta}, \end{aligned}$$

where $I_{j,k} \subset [M]$ is an arbitrary subset such that $|I_{j,k}| = m_{j,k}$. Therefore, for every $j \in [n]$ we let $\{I_{j,k}\}_{k=1}^{N_\Theta}$ be a collection of disjoint sets $[M]$ for which $|I_{j,k}| = m_{j,k}$ and we let I_j^0 be the complement of their union. Note that

$$\sum_{i \in I_j^0} \pi_i^\omega = \sum_{i=1}^{M} \pi_i^\omega - \sum_{k=1}^{N_\Theta} \sum_{i \in I_{j,k}} \pi_i^\omega < 1 + \frac{1}{2}\varepsilon - P^0_{\boldsymbol{x}_j}(K_\Theta) < \frac{3}{2}\varepsilon,$$

and

$$\sum_{i > M} \pi_i^\omega = 1 - \sum_{i=1}^{M} \pi_i^\omega < \frac{1}{2N_\Theta}\varepsilon < \frac{1}{2}\varepsilon.$$

Consequently, in the case of a wDDP we obtain the inequality

$$\begin{aligned}
&\left|\int_\Theta f_{i,j}(\boldsymbol{\theta})dG^\omega_{\boldsymbol{x}_j}(\boldsymbol{\theta}) - \sum_{k=1}^{N_\Theta} f_{i,j}(\boldsymbol{\theta}_k)P^0_{\boldsymbol{x}_j}(A_k)\right| \\
&\quad\le \sum_{\ell>M}\pi^\omega_\ell + \sum_{\ell\in I^0_j}\pi^\omega_\ell + \sum_{k=1}^{N_\Theta}\left|\sum_{\ell\in I_{j,k}}\pi^\omega_\ell f_{i,j}(\boldsymbol{\theta}^\omega_{\ell,\boldsymbol{x}_j}) - P^0_{\boldsymbol{x}_j}(A_k)f_{i,j}(\boldsymbol{\theta}_k)\right| \\
&\quad < 2\varepsilon + \sum_{k=1}^{N_\Theta}\left|\sum_{\ell\in I_{j,k}}\pi^\omega_\ell\left(f_{i,j}(\boldsymbol{\theta}^\omega_{\ell,\boldsymbol{x}_j}) - f_{i,j}(\boldsymbol{\theta}_k)\right)\right| \\
&\qquad + \sum_{k=1}^{N_\Theta}|f_{i,j}(\boldsymbol{\theta}_k)|\left|\sum_{\ell\in I_{j,k}}\pi^\omega_\ell - P^0_{\boldsymbol{x}_j}(A_k)\right| \\
&\quad < 3\varepsilon + \sum_{k=1}^{N_\Theta}\left|\sum_{\ell\in I_{j,k}}\pi^\omega_\ell\left(f_{i,j}(\boldsymbol{\theta}^\omega_{\ell,\boldsymbol{x}_j}) - f_{i,j}(\boldsymbol{\theta}_k)\right)\right|.
\end{aligned}$$

Hence, consider the event

$$\Omega_\theta := \bigcap_{k=1}^{N_\Theta}\bigcap_{j=1}^{n}\bigcap_{i\in I_{j,k}}\{\omega\in\Omega_0 :\ \boldsymbol{\theta}^\omega_i \in U_k\},$$

which has positive measure by hypothesis. By independence, $\Omega_\pi\cap\Omega_\theta$ has positive measure, and for $\omega\in\Omega_\pi\cap\Omega_\theta$ we have that

$$\begin{aligned}
&\sum_{k=1}^{N_\Theta}\left|\sum_{\ell\in I_{j,k}}\pi^\omega_\ell\left(f_{i,j}(\boldsymbol{\theta}^\omega_{\ell,\boldsymbol{x}_j}) - f_{i,j}(\boldsymbol{\theta}_k)\right)\right| \\
&\qquad < \frac{\varepsilon}{N_\Theta}\sum_{k=1}^{N_\Theta}\sum_{\ell\in I_{j,k}}\pi^\omega_\ell < \frac{\varepsilon}{N_\Theta}\left(\varepsilon + \sum_{k=1}^{N_\Theta}P^0_{\boldsymbol{x}_j}(A_k)\right) < \varepsilon.
\end{aligned}$$

Therefore,

$$\left|\int_\Theta f_{i,j}(\boldsymbol{\theta})dG^\omega_{\boldsymbol{x}_j}(\boldsymbol{\theta}) - \sum_{k=1}^{N_\Theta} f_{i,j}(\boldsymbol{\theta}_k)P^0_{\boldsymbol{x}_j}(A_k)\right| < 4\varepsilon,$$

proving the theorem in the case of a wDDP. □

A.2.2. Proof of Theorem 4.6

By possibly modifying the sequence $\varepsilon_1,\dots,\varepsilon_n$ we can assume, without loss, that $\|f_i\|_C \le 1$. Let $\Omega_0\subset\Omega$ be a set of full measure such that for any $\omega\in\Omega_0$ we have that: (i) $\pi^\omega_i \ge 0$; (ii) $\sum_{i\in\mathbb{N}}\pi^\omega_i \equiv 1$; (iii) π^ω_i and $\boldsymbol{\theta}^\omega_i$ are continuous on $\mathcal{X}$ in the case of a DDP, or π^ω_i are continuous on $\mathcal{X}$ in the case of a θDDP.

For simplicity, we write for $\omega \in \Omega$

$$F_i^\omega(\boldsymbol{x}) := \int_\Theta f_i(\boldsymbol{\theta})\, dG_{\boldsymbol{x}}^\omega(\boldsymbol{\theta}).$$

By Theorem 4.1, F_i^ω is continuous on $\mathcal{X}$ for almost every $\omega \in \Omega$. By possibly removing from Ω_0 a null set, we may assume that F_i^ω is continuous for every $\omega \in \Omega_0$ and every $i \in [n]$. Furthermore, we write

$$F_i(\boldsymbol{x}) := \int_\Theta f_i(\boldsymbol{\theta}) dP_{\boldsymbol{x}}(\boldsymbol{\theta}),$$

which, by hypothesis, is a continuous function on $\mathcal{X}$. Since $\|f_i\|_C \leq 1$ we have that $|F_i^\omega| \leq 1$ and $|F_i| \leq 1$.

Let $\varepsilon_0 < \min(\varepsilon_1, \ldots, \varepsilon_n)$ and define the event

$$\Omega_{K,\varepsilon_0} := \bigcap_{i=1}^{N} \left\{ \omega \in \Omega_0 : \sup_{\boldsymbol{x} \in K} |F_i^\omega(\boldsymbol{x}) - F_i(\boldsymbol{x})| < \varepsilon_0 \right\}.$$

Since Θ is separable, a compact $K \subset \Theta$ is also separable. By hypothesis F_i is continuous and by construction F_i^ω is continuous for every $\omega \in \Omega_0$. Hence, the event is measurable, and the theorem follows if we show that this event has positive measure. Let $\varepsilon > 0$ with $\varepsilon < \varepsilon_0$. Since K is compact, by Lemma A.1 there is a function $\bar{P} : K \to \mathcal{P}(\Theta)$ such that for

$$\bar{F}_i(\boldsymbol{x}) := \int_\Theta f_i(\boldsymbol{\theta}) d\bar{P}_{\boldsymbol{x}}(\boldsymbol{\theta}),$$

we have that

$$\sup_{\boldsymbol{x} \in K} |F_i(\boldsymbol{x}) - \bar{F}_i(\boldsymbol{x})| < \varepsilon.$$

Hence,

$$|F_i^\omega(\boldsymbol{x}) - F_i(\boldsymbol{x})| < |F_i^\omega(\boldsymbol{x}) - \bar{F}_i(\boldsymbol{x})| + \varepsilon,$$

for any $i \in [n]$ and $\boldsymbol{x} \in K$. Since $\{\bar{P}_{\boldsymbol{x}} : \boldsymbol{x} \in K\}$ is tight by Lemma A.1, there exists $K_\Theta \subset \Theta$ compact such that $\bar{P}_{\boldsymbol{x}}(\Theta \setminus K_\Theta) < \varepsilon$ for any $\boldsymbol{x} \in K$. Hence,

$$\left| \bar{F}_i(\boldsymbol{x}) - \int_{K_\Theta} f_i(\boldsymbol{\theta}) \bar{P}_{\boldsymbol{x}}(\boldsymbol{\theta}) \right| \leq \bar{P}_{\boldsymbol{x}}(\Theta \setminus K_\Theta) < \varepsilon$$

and

$$|F_i^\omega(\boldsymbol{x}) - F_i(\boldsymbol{x})| \leq |F_i^\omega(\boldsymbol{x}) - \bar{F}_i(\boldsymbol{x})| + \varepsilon \leq \left| F_i^\omega(\boldsymbol{x}) - \int_{K_\Theta} f_i(\boldsymbol{\theta}) d\bar{P}_{\boldsymbol{x}}(\boldsymbol{\theta}) \right| + 2\varepsilon,$$

on K. Finally, since $f_1, \ldots, f_n$ are continuous, we can use the same construction as that in the proof of Theorem 4.5 in Appendix A.2.1 to obtain a partition

$\{A_k\}_{k=1}^{N_\Theta}$ of K_Θ as in (11) where each A_k is measurable, non-empty, and every f_i varies at most by 2ε over any A_k. Then

$$\left|\int_{K_\Theta} f_i(\boldsymbol{\theta})d\bar{P}_{\boldsymbol{x}}(\boldsymbol{\theta}) - \sum_{k=1}^{N_\Theta} f_i(\boldsymbol{\theta}_k)\bar{P}_{\boldsymbol{x}}(A_k)\right| \le \sum_{k=1}^{N_\Theta}\int_{A_k} |f_i(\boldsymbol{\theta}) - f_i(\boldsymbol{\theta}_k)|\, d\bar{P}_{\boldsymbol{x}}(\boldsymbol{\theta}) < \varepsilon,$$

whence

$$|F_i^\omega(\boldsymbol{x}) - F_i(\boldsymbol{x})| < \left|F_i^\omega(\boldsymbol{x}) - \sum_{k=1}^{N_\Theta} f_i(\boldsymbol{\theta}_k)\bar{P}_{\boldsymbol{x}}(A_k)\right| + 3\varepsilon.$$

Since $\boldsymbol{x} \to \bar{P}_{\boldsymbol{x}}(A_k)$ is continuous by Lemma A.1, with values on $[0,1]$, and

$$1-\varepsilon < \sum_{k=1}^{N_\Theta} \bar{P}_{\boldsymbol{x}}(A_k) \le 1,$$

the event

$$\Omega_\pi := \bigcap_{k=1}^{N_\Theta} \left\{\omega \in \Omega_0 :\ \sup_{\boldsymbol{x}\in K} |\pi_{k,\boldsymbol{x}}^\omega - \bar{P}_{\boldsymbol{x}}(A_k)| < \frac{1}{N_\Theta}\varepsilon\right\},$$

is measurable, as $\boldsymbol{x} \mapsto \pi_{k,\boldsymbol{x}}^\omega$ is continuous for $\omega \in \Omega_0$ and $\boldsymbol{x} \mapsto \bar{P}_{\boldsymbol{x}}(A_k)$ is continuous by Lemma A.2, and has positive measure by hypothesis. The proof now proceeds exactly the same as the proof of Theorem 4.5 in Appendix A.2.1 for the DDP and θDDP. We omit the details for brevity. □

A.2.3. Proof of Theorem 4.8

Let $\Omega_0 \subset \Omega$ be a set of full measure such that for any $\omega \in \Omega_0$ we have that: (i) $\pi_i^\omega \ge 0$; (ii) $\sum_{i\in\mathbb{N}} \pi_i^\omega \equiv 1$; (iii) π_i^ω and $\boldsymbol{\theta}_i^\omega$ are continuous on $\mathcal{X}$ for the DDP, $\boldsymbol{\theta}_i^\omega$ are continuous on $\mathcal{X}$ in the case of a wDDP, or π_i^ω is continuous on $\mathcal{X}$ in the case of a θDDP.

We prove in detail the statement in the case of a DDP. If $P^0 \in \mathcal{P}(\Theta)^{\mathcal{X}}|_{G_\mathcal{X}}$ then we consider the event

$$\left\{\omega \in \Omega :\ \left|\int_\Theta f_{i,j}(\boldsymbol{\theta})dG_{\boldsymbol{x}_j}^\omega(\boldsymbol{\theta}) - \int_\Theta f_{i,j}(\boldsymbol{\theta})dP_{\boldsymbol{x}_j}^0(\boldsymbol{\theta})\right| < \varepsilon_{i,j},\ i,j \in [n]\right\},$$

which is measurable by inspection. A standard argument shows that, by possibly reducing $\varepsilon_{1,1},\dots,\varepsilon_{n,n}$, we can assume that $f_{i,j}$ are simple functions, and that there exists a partition $\{A_k\}_{k=1}^{N_f}$ of Θ of measurable sets such that

$$f_{i,j} = \sum_{k=1}^{N_f} c_{i,j,k}\mathbb{I}_{A_k},$$

where $\mathbb{I}_A$ is the indicator function of the set A and $|c_{i,j,k}| \le 1$ for any $i,j \in [n]$ and $k \in [N_f]$.

Let $\varepsilon_0 > 0$ be such that $\varepsilon_0 < \min(\varepsilon_{1,1}, \dots, \varepsilon_{n,n})$. Note that

$$\int_\Theta f_{i,j}(\boldsymbol{\theta}) dG^\omega_{\boldsymbol{x}_j}(\boldsymbol{\theta}) - \int_\Theta f_{i,j}(\boldsymbol{\theta}) dP^0_{\boldsymbol{x}_j}(\boldsymbol{\theta}) = \sum_{k=1}^{N_f} c_{i,j,k}(G^\omega_{\boldsymbol{x}_j}(A_k) - P^0_{\boldsymbol{x}_j}(A_k)).$$

Consider the event

$$\Omega_\pi := \bigcap_{k=1}^{N_f} \bigcap_{j=1}^{n} \left\{ \omega \in \Omega_0 : \, |\pi^\omega_{k,\boldsymbol{x}_j} - P^0_{\boldsymbol{x}_j}(A_k)| < \frac{1}{N_f^2}\varepsilon \right\},$$

which has positive measure by hypothesis. Remark that, in this case, for any $j \in [n]$ we have that

$$\sum_{i > N_f} \pi^\omega_{i,\boldsymbol{x}_j} = 1 - \sum_{i=1}^{N_f} \pi^\omega_{i,\boldsymbol{x}_j} = \sum_{i=1}^{N_f} (P^0_{\boldsymbol{x}_j}(A_k) - \pi^\omega_{i,\boldsymbol{x}_j}) < \frac{1}{N_f}\varepsilon,$$

where we used the fact that $\{A_k\}_{k=1}^{N_f}$ is a partition. Hence,

$$\begin{aligned} \left| \sum_{k=1}^{N_f} c_{i,j,k}(G^\omega_{\boldsymbol{x}_j}(A_k) - P^0_{\boldsymbol{x}_j}(A_k)) \right| &\leq N_f \sum_{i > N_f} \pi^\omega_{i,\boldsymbol{x}_j} \\ &\quad + \sum_{k=1}^{N_f} \left| \sum_{i=1}^{N_f} \pi^\omega_{i,\boldsymbol{x}_j} \delta_{\boldsymbol{\theta}^\omega_{i,\boldsymbol{x}_j}}(A_k) - P^0_{\boldsymbol{x}_j}(A_k) \right| \\ &< \varepsilon + \sum_{k=1}^{N_f} \left| \sum_{i=1}^{N_f} \pi^\omega_{i,\boldsymbol{x}_j} \delta_{\boldsymbol{\theta}^\omega_{i,\boldsymbol{x}_j}}(A_k) - P^0_{\boldsymbol{x}_j}(A_k) \right|. \end{aligned}$$

Now we use the fact that $P^0 \in \mathcal{P}(\Theta)^{\mathcal{X}}|_{G_{\mathcal{X}}}$. In this case, $G^0_{\boldsymbol{x}_j}(A_k) = 0$ implies $P^0_{\boldsymbol{x}_j}(A_k) = 0$ and no such terms contribute to the sum. We can define the event

$$\Omega_\theta := \{\omega \in \Omega_0 : \, \boldsymbol{\theta}^\omega_{k,\boldsymbol{x}_j} \in A_k \text{ and } G^0_{\boldsymbol{x}_j}(A_k) > 0, \, j \in [n], \, k \in [N_f]\},$$

which has positive measure. By independence, $\Omega_\pi \cap \Omega_\theta$ has positive measure. Hence, for $\omega \in \Omega_\pi \cap \Omega_\theta$ we have that

$$\left| \sum_{i=1}^{N_f} \pi^\omega_{i,\boldsymbol{x}_j} \delta_{\boldsymbol{\theta}^\omega_{i,\boldsymbol{x}_j}}(A_k) - P^0_{\boldsymbol{x}_j}(A_k) \right| = |\pi^\omega_{k,\boldsymbol{x}_j} - P^0_{\boldsymbol{x}_j}(A_k)| < \frac{1}{N_f^2}\varepsilon,$$

if $G^0_{\boldsymbol{x}_j}(A_k) > 0$. If $G^0_{\boldsymbol{x}_j}(A_k) = 0$ then $P^0_{\boldsymbol{x}_j}(A_k) = 0$ and

$$\left| \sum_{i=1}^{N_f} \pi^\omega_{i,\boldsymbol{x}_j} \delta_{\boldsymbol{\theta}^\omega_{i,\boldsymbol{x}_j}}(A_k) - P^0_{\boldsymbol{x}_j}(A_k) \right| \leq \sum_{k:\, G^0_{\boldsymbol{x}_j}(A_k) = 0} \pi^\omega_{i,\boldsymbol{x}_j} < \frac{1}{N_f}\varepsilon.$$

Therefore,

$$\left|\int_\Theta f_{i,j}(\boldsymbol{\theta})dG^\omega_{\boldsymbol{x}_j}(\boldsymbol{\theta}) - \int_\Theta f_{i,j}(\boldsymbol{\theta})dP^0_{\boldsymbol{x}_j}(\boldsymbol{\theta})\right| < 2\varepsilon,$$

proving the theorem in the case of a DDP.

To prove the theorem in the case of a θDDP, the argument is the same up to the inequality

$$\left|\sum_{k=1}^{N_f} c_{i,j,k}(G^\omega_{\boldsymbol{x}_j}(A_k) - P^0_{\boldsymbol{x}_j}(A_k))\right| < \varepsilon + \sum_{k=1}^{N_f}\left|\sum_{i=1}^{N_f}\pi^\omega_{i,\boldsymbol{x}_j}\delta_{\boldsymbol{\theta}^\omega_i}(A_k) - P^0_{\boldsymbol{x}_j}(A_k)\right|.$$

In this case, we define the event

$$\Omega_\theta := \{\omega \in \Omega_0 :\ \boldsymbol{\theta}^\omega_k \in A_k \text{ and } G^0(A_k) > 0,\ k \in [N_f]\}.$$

Since $P^0 \in \mathcal{P}(\Theta)^{\mathcal{X}}|_{G_{\mathcal{X}}}$ this event has positive measure. The proof then follows the same steps as those in the case of a DDP.

Finally, in the case of a wDDP we follow a similar argument as that on the proof of Theorem 4.5 in Appendix A.2.1. By choosing a suitable M such that there are integers $m_{j,k} \in [M]$ such that

$$\left|P^0_{\boldsymbol{x}_j}(A_k) - \frac{m_{j,k}}{M}\right| < \frac{\varepsilon}{2N_f},$$

we can define the event

$$\Omega_\pi := \bigcap_{i=1}^{M}\left\{\omega \in \Omega_0 :\ \pi^\omega_i \in \left(\frac{1-\varepsilon/2N_f}{M}, \frac{1}{M}\right]\right\},$$

which has positive measure by hypothesis, and for every $j \in [n]$ we can define a collection $\{I_{j,k}\}_{k=1}^{N_f}$ of disjoint sets $[M]$ for which $|I_{j,k}| = m_{j,k}$. We let I^0_j be the complement of their union. Note that for $\omega \in \Omega_\pi$ we have that

$$\sum_{i\in I^0_j}\pi^\omega_i < \frac{3}{2}\varepsilon \quad \text{and} \quad \sum_{i>M}\pi^\omega_i < \frac{1}{2}\varepsilon.$$

Hence, we obtain the inequality

$$\begin{aligned}
|G^\omega_{\boldsymbol{x}_j}(A_k) - P^0_{\boldsymbol{x}_j}(A_k)| &\le \sum_{\ell>M}\pi^\omega_\ell + \left|\sum_{\ell=1}^{M}\pi^\omega_\ell\delta_{\boldsymbol{\theta}^\omega_{\ell,\boldsymbol{x}_j}}(A_k) - P^0_{\boldsymbol{x}_j}(A_k)\right| \\
&< \frac{3}{2N_f}\varepsilon + \sum_{\ell\in I^0_j} + \left|\sum_{k=1}^{N_f}\sum_{\ell\in I_{j,k}}\pi^\omega_\ell\delta_{\boldsymbol{\theta}^\omega_{\ell,\boldsymbol{x}_j}}(A_k) - P^0_{\boldsymbol{x}_j}(A_k)\right| \\
&< \frac{2}{N_f}\varepsilon + \left|\sum_{k=1}^{N_f}\sum_{\ell\in I_{j,k}}\pi^\omega_\ell\delta_{\boldsymbol{\theta}^\omega_{\ell,\boldsymbol{x}_j}}(A_k) - P^0_{\boldsymbol{x}_j}(A_k)\right|.
\end{aligned}$$

Since $G^0_{\boldsymbol{x}_j}(A_j) = 0$ implies $P^0_{\boldsymbol{x}_j}(A_k) = 0$ it suffices to consider the event

$$\Omega_\theta := \{\omega \in \Omega_0 : \ \boldsymbol{\theta}^\omega_{i,\boldsymbol{x}_j} \in A_k \text{ and } G^0_{\boldsymbol{x}_j}(A_k),\, k \in [N_f], j \in [n],\, i \in I_{j,k}\}.$$

By hypothesis, this has positive measure and, by independence, so does $\Omega_\pi \cap \Omega_\theta$. The proof then follows the same arguments as those in the proof of Theorem 4.5 in Appendix A.2.1 in the case of a wDDP. We omit the details for brevity. □

A.2.4. Proof of Theorem 4.9

The proof is similar to the proof of Theorem 4.6 in Appendix A.2.2 with minor modifications. Let $\Omega_0 \subset \Omega$ be a set of full measure such that for any $\omega \in \Omega_0$ we have that: (i) $\pi^\omega_i \geq 0$; (ii) $\sum_{i\in\mathbb{N}} \pi^\omega_i \equiv 1$; (iii) π^ω_i and $\boldsymbol{\theta}^\omega_i$ are continuous on $\mathcal{X}$ in the case of a DDP, or π^ω_i is continuous on $\mathcal{X}$ in the case of a θDDP. Furthermore, by possibly reducing Ω_0 by a null set, we may assume that

$$\boldsymbol{x} \mapsto \int f_i(\boldsymbol{\theta}) dG^\omega_{\boldsymbol{x}}(\boldsymbol{\theta}),$$

is continuous for every $\omega \in \Omega_0$ and $i \in [n]$.

We prove the theorem in detail in the case of a DDP. Suppose that $C_S(\mathcal{X}, \mathcal{P}(\Theta)) \cap \mathcal{P}(\Theta)^{\mathcal{X}}|_{G_{\mathcal{X}}}$ is non-empty, as otherwise there is nothing to prove. Our goal is to show that for $P^0 \in C_S(\mathcal{X}, \mathcal{P}(\Theta)) \cap \mathcal{P}(\Theta)^{\mathcal{X}}|_{G_{\mathcal{X}}}$ the event

$$\left\{\omega \in \Omega : \ \sup_{\boldsymbol{x}\in K} \left| \int_\Theta f_i(\boldsymbol{\theta}) dG^\omega_{\boldsymbol{x}}(\boldsymbol{\theta}) - \int_\Theta f_i(\boldsymbol{\theta}) dP^0_{\boldsymbol{x}}(\boldsymbol{\theta}) \right| < \varepsilon_i,\, i \in [n] \right\},$$

has positive probability. As argued in the proof of Theorem 4.6 in Appendix A.2.2, we can assume that there exists a partition $\{A_k\}_{k=1}^{N_f}$ of Θ of measurable sets such that

$$f_i = \sum_{k=1}^{N_f} c_{i,k} \mathbb{I}_{A_k},$$

where $|c_{i,k}| \leq 1$ for any $i \in [n]$ and $k \in [N_f]$. Let $\varepsilon_0 > 0$ be such that $\varepsilon_0 < \min(\varepsilon_1, \dots, \varepsilon_n)$. Note that

$$\int_\Theta f_i(\boldsymbol{\theta}) dG^\omega_{\boldsymbol{x}}(\boldsymbol{\theta}) - \int_\Theta f_i(\boldsymbol{\theta}) dP^0_{\boldsymbol{x}}(\boldsymbol{\theta}) = \sum_{k=1}^{N_f} c_{i,j,k}(G^\omega_{\boldsymbol{x}}(A_k) - P^0_{\boldsymbol{x}}(A_k)).$$

By hypothesis, if $G^0(A_k) = 0$ then $P^0_{\boldsymbol{x}}(A_k) = 0$ for any $\boldsymbol{x} \in K$. Hence, without loss, we can assume that $G^0(A_k) > 0$. Note that in this case, we have that

$$\sum_{k=1}^{N_f} P^0_{\boldsymbol{x}}(A_k) = 1.$$

Consider the event

$$\Omega_\pi := \bigcap_{k=1}^{N_f}\bigcap_{j=1}^{n}\left\{\omega \in \Omega_0 : \sup_{\boldsymbol{x}\in K} |\pi^\omega_{k,\boldsymbol{x}} - P^0_{\boldsymbol{x}}(A_k)| < \frac{1}{N_f^2}\varepsilon\right\}.$$

Since $P^0 \in C_S(\mathcal{X}, \mathcal{P}(\Theta))$ we see that $\boldsymbol{x} \mapsto P^0_{\boldsymbol{x}}(A_k)$ is continuous whence Ω_π has positive measure by hypothesis. Remark that, in this case, for any $j \in [n]$ we have that

$$\sum_{i>N_f} \pi^\omega_{i,\boldsymbol{x}} = 1 - \sum_{i=1}^{N_f} \pi^\omega_{i,\boldsymbol{x}} = \sum_{i=1}^{N_f} (P^0_{\boldsymbol{x}}(A_k) - \pi^\omega_{i,\boldsymbol{x}_j}) < \frac{1}{N_f}\varepsilon,$$

where we used the fact that $\{A_k\}_{k=1}^{N_f}$ is a partition. Hence,

$$\left|\sum_{k=1}^{N_f} c_{i,j,k}(dG^\omega_{\boldsymbol{x}_j}(A_k) - P^0_{\boldsymbol{x}_j}(A_k))\right| < \varepsilon + \sum_{k=1}^{N_f}\left|\sum_{i=1}^{N_f} \pi^\omega_{i,\boldsymbol{x}_j}\delta_{\boldsymbol{\theta}^\omega_{i,\boldsymbol{x}_j}}(A_k) - P^0_{\boldsymbol{x}_j}(A_k)\right|.$$

Since $P^0 \in \mathcal{P}(\Theta)^{\mathcal{X}}|_{G_{\mathcal{X}}}$, the event

$$\Omega_\theta := \{\omega \in \Omega_0 : \ \boldsymbol{\theta}^\omega_{k,\boldsymbol{x}} \in A_k \text{ and } G^0_{\boldsymbol{x}}(A_k) > 0, \, \boldsymbol{x} \in K, \, k \in [N_f]\},$$

has positive measure. The proof then proceeds exactly as the proof of Theorem 4.6 in Appendix A.2.2. We omit the proof in the case of a θDDP for brevity. □

A.3. Association structure of the DDP and its variants

A.3.1. Proof of Theorem 4.10

We prove the theorem in the case of a DDP as the arguments are the same in the case of a wDDP. To prove the theorem we proceed as follows. We first show that for any $d \in \mathbb{N}$

$$(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d) \mapsto \mathbb{E}\left(\prod_{k=1}^{d} G_{\boldsymbol{x}_k}(B)\right),$$

is continuous. Then, by leveraging Stone-Weierstrass's theorem [75], we can approximate any continuous f uniformly over the hypercube by a polynomial. Since the expectation of this polynomial is continuous by our first claim, the theorem follows.

Define the sequence of functions $h_n : \mathcal{X}^d \times \Omega \to \mathbb{R}$ as

$$h_n(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d, \omega) = \prod_{k=1}^{d}\sum_{i_k=1}^{n} \pi^\omega_{i_k,\boldsymbol{x}_k}\delta_{\boldsymbol{\theta}^\omega_{i_k,\boldsymbol{x}_k}}(B) = \sum_{i_1,\ldots,i_d=1}^{n}\prod_{k=1}^{d} \pi^\omega_{i_k,\boldsymbol{x}_k}\delta_{\boldsymbol{\theta}^\omega_{i_k,\boldsymbol{x}_k}}(B).$$

Since the summands are a.e. non-negative, $h_n \leq h_{n+1}$. Furthermore, $h_n \in [0,1]$ for any $n \in \mathbb{N}$, and

$$\lim_{n\to\infty} h_n(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_d,\omega) = \prod_{k=1}^{d} G^{\omega}_{\boldsymbol{x}_k}(B),$$

for a.e. $\omega \in \Omega$ and $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_d \in \mathcal{X}$.

We first control the expectation for fixed $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_d \in \mathcal{X}$. Write $h_n^{\boldsymbol{x}}(\omega) = h_n(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_d,\omega)$ for simplicity, and define the function $g_n : \mathcal{X}^d \to \mathbb{R}$ as

$$\begin{aligned} g_n(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_d) &= \mathbb{E}(h_n^{\boldsymbol{x}}) \\ &= \sum_{i_1,\ldots,i_m=1}^{n} \mathbb{E}\left(\prod_{k=1}^{m} \pi_{i_k,\boldsymbol{x}_k}\right) \mathbb{P}\left(\{\omega\in\Omega:\, \boldsymbol{\theta}_{i_1,\boldsymbol{x}_1}\in B,\ldots,\boldsymbol{\theta}_{i_d,\boldsymbol{x}_d}\in B\}\right), \end{aligned} \tag{12}$$

where in the last equality we used the hypothesis of independence. Since the sequence $\{h_n^{\boldsymbol{x}}\}$ is monotone non-decreasing, by the monotone convergence theorem we have that

$$\lim_{n\to\infty} g_n(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_d) = g_\infty(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_d) := \mathbb{E}\left(\prod_{k=1}^{d} G_{\boldsymbol{x}_k}(B)\right).$$

Hence, the sequence $\{g_n\}$ converges pointwise to g_∞ over $\mathcal{X}^d$. To conclude g_∞ is continuous, we will show that $\{g_n\}$ is a sequence of continuous functions, to then use a uniform approximation argument to show the continuity of g_∞ near any $(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_d) \in \mathcal{X}^d$.

To show g_n is continuous, note that for a.e. ω the functions $\boldsymbol{x} \to \pi^{\omega}_{i,\boldsymbol{x}}$ are continuous for any $i \in \mathbb{N}$. Therefore, for any $i_1,\ldots,i_d \in \mathbb{N}$ the function

$$(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_d) \to \mathbb{E}\left(\prod_{k=1}^{d} \pi_{i_k,\boldsymbol{x}_k}\right),$$

is continuous on $\mathcal{X}^d$. Similarly,

$$(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_d) \to \mathbb{P}\left(\{\omega\in\Omega:\, \boldsymbol{\theta}_{i_1,\boldsymbol{x}_1}\in B,\ldots,\boldsymbol{\theta}_{i_d,\boldsymbol{x}_d}\in B\}\right),$$

is continuous by hypothesis.

We now show continuity of g_∞ near any $(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_d) \in \mathcal{X}^d$ using a uniform approximation. Fix $(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_d) \in \mathcal{X}^d$ and let $\varepsilon > 0$. Remark that for $n > m$ we have the bound

$$|h_n(\boldsymbol{x}'_1,\ldots,\boldsymbol{x}'_d,\omega) - h_m(\boldsymbol{x}'_1,\ldots,\boldsymbol{x}'_d,\omega)| \leq \sum_{k=1}^{d}\sum_{i_k=m+1}^{n} \pi^{\omega}_{i_k,\boldsymbol{x}'_k},$$

whence

$$|g_n(\boldsymbol{x}'_1,\ldots,\boldsymbol{x}'_d) - g_m(\boldsymbol{x}'_1,\ldots,\boldsymbol{x}'_d)| \leq \sum_{k=1}^{d}\sum_{i_k=m+1}^{n} \mathbb{E}(\pi_{i_k,\boldsymbol{x}'_k}).$$

Hence, we can control the differences on the left-hand side by controlling the expectations on the right-hand side. By similar arguments as those used in the proof of Theorem 4.1 in Appendix A.1.1, for any $\delta < \varepsilon/3$ we can find $N_\delta \in \mathbb{N}$ such that

$$\sum_{i_k > N_\delta} \mathbb{E}(\pi_{i_k,\boldsymbol{x}_k}) < \frac{1}{2d}\delta,$$

for every $k \in [d]$. We define the open neighborhood

$$U_\delta := \bigcap_{k=1}^{d} \bigcap_{i_k=1}^{N_\varepsilon} \left\{ (\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_d) \in \mathcal{X}^d : \ |\mathbb{E}(\pi_{i_k,\boldsymbol{x}'_k}) - \mathbb{E}(\pi_{i_k,\boldsymbol{x}_k})| < \frac{1}{2dN_\delta}\delta \right\},$$

of $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d)$. For $(\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_d) \in U_\delta$ we have that

$$\sum_{i_k > N_\delta} \mathbb{E}(\pi_{i_k,\boldsymbol{x}'_k}) = \sum_{i_k > N_\delta} \mathbb{E}(\pi_{i_k,\boldsymbol{x}_k}) - \sum_{i_k=1}^{N_\delta} (\mathbb{E}(\pi_{i_k,\boldsymbol{x}_k}) - \mathbb{E}(\pi_{i_k,\boldsymbol{x}'_k})) < \frac{1}{d}\delta.$$

Therefore, for $n > m > N_\delta$ we have that

$$\sup_{(\boldsymbol{x}'_1,\ldots,\boldsymbol{x}'_d) \in U_\delta} |g_n(\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_d) - g_m(\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_d)| \leq \sup_{(\boldsymbol{x}'_1,\ldots,\boldsymbol{x}'_d) \in U_\delta} \sum_{k=1}^{d} \sum_{i_k > N_\delta} \mathbb{E}(\pi_{i_k,\boldsymbol{x}'_k}) < \delta.$$

In particular, this implies

$$\sup_{(\boldsymbol{x}'_1,\ldots,\boldsymbol{x}'_d) \in U_\varepsilon} |g_n(\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_d) - g_\infty(\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_d)| < \delta,$$

for any $n > N_\delta$. Hence, it suffices to choose $n > N_\varepsilon$ to define the open neighborhood

$$W_\varepsilon := \left\{ (\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_d) \in U_\delta : \ |g_n(\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_d) - g_n(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d)| < \frac{1}{3}\varepsilon \right\},$$

of $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d)$. Then, for any $(\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_d) \in W_\varepsilon$ we have that

$$|g_\infty(\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_d) - g_\infty(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d)| < 2\delta + \frac{1}{3}\varepsilon < \varepsilon,$$

whence g_∞ is continuous.

Now, let $f : [0,1]^d \to \mathbb{R}$ be a continuous function. Since $[0,1]^d$ is compact, f is bounded, and the function

$$F(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d) = \mathbb{E}(f(G_{\boldsymbol{x}_1}(B), \ldots, G_{\boldsymbol{x}_d}(B))),$$

is well-defined. We now show it is continuous. Let $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d) \in \mathcal{X}^d$ and fix $\varepsilon > 0$. Since $[0,1]^d$ is compact, f can be approximated uniformly by a polynomial $p : [0,1]^d \to \mathbb{R}$. If

$$p(z_1, \ldots, z_d) = \sum_{|\boldsymbol{\gamma}|=0}^{N} c_{\boldsymbol{\gamma}} \boldsymbol{z}^{\boldsymbol{\gamma}} \quad \text{and} \quad \sup_{\boldsymbol{z} \in [0,1]^d} |f(\boldsymbol{z}) - p(\boldsymbol{z})| < \frac{1}{3}\varepsilon,$$

then

$$P(\boldsymbol{x}_1,\dots,\boldsymbol{x}_d) := \mathbb{E}(p(G_{\boldsymbol{x}_1}(B),\dots,G_{\boldsymbol{x}_d}(B))) = \sum_{|\boldsymbol{\gamma}|=0}^{N} c_{\boldsymbol{\gamma}}\mathbb{E}\left(\prod_{k=1}^{d}\prod_{i_k=1}^{\gamma_k} G_{\boldsymbol{x}_k}(B)\right),$$

is a continuous function by our previous result. Furthermore, if we define the open neighborhood

$$U_\varepsilon = \left\{(\boldsymbol{x}_1',\dots,\boldsymbol{x}_d') : |P(\boldsymbol{x}_1',\dots,\boldsymbol{x}_d') - P(\boldsymbol{x}_1,\dots,\boldsymbol{x}_d)| < \frac{1}{3}\varepsilon\right\},$$

of $(\boldsymbol{x}_1,\dots,\boldsymbol{x}_d)$, then for any $(\boldsymbol{x}_1',\dots,\boldsymbol{x}_d') \in U_\varepsilon$ we have that

$$\begin{aligned}|F(\boldsymbol{x}_1',\dots,\boldsymbol{x}_d') - F(\boldsymbol{x}_1,\dots,\boldsymbol{x}_d)| \\ \le \mathbb{E}(|f(G_{\boldsymbol{x}_1'}(B),\dots,G_{\boldsymbol{x}_d'}(B)) - f(G_{\boldsymbol{x}_1}(B),\dots,G_{\boldsymbol{x}_d}(B))|) < \varepsilon,\end{aligned}$$

proving the claim. □

A.4. The regularity of the probability density function of DDP mixture models

A.4.1. Proof of Lemma 5.1

We assume that P is fixed and we drop the superscript on ρ^P. Let $(\boldsymbol{y}_0,\boldsymbol{\gamma}_0,\boldsymbol{x}_0) \in \mathcal{Y}\times\Gamma\times\mathcal{X}$ and $\varepsilon > 0$. By hypothesis, there exists an open neighborhood $U_{\boldsymbol{y}_0}\times U_{\boldsymbol{\gamma}_0}$ of $(\boldsymbol{y}_0,\boldsymbol{\gamma}_0)$ and a compact set $K_{\boldsymbol{\theta}_0} \subset \Theta$ such that

$$(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}) \in U_{\boldsymbol{y}_0}\times U_{\boldsymbol{\gamma}_0}\times K_{\boldsymbol{\theta}_0}^c :\ \psi(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}) < \frac{1}{4}\varepsilon.$$

Furthermore, since Θ is Polish and $P_{\boldsymbol{x}_0}$ is finite, there exists $K_{\boldsymbol{x}_0} \subset \Theta$ compact such that

$$P_{\boldsymbol{x}_0}(\Theta \setminus K_{\boldsymbol{x}_0}) < \varepsilon.$$

Define the compact set $K_\Theta := K_{\boldsymbol{\theta}_0} \cup K_{\boldsymbol{x}_0}$. For $\boldsymbol{\theta} \in K_\Theta$ let $\delta_{\boldsymbol{\theta}} > 0$ be such that

$$\begin{aligned}\max\{d_{\mathcal{Y}}(\boldsymbol{y}',\boldsymbol{y}_0), d_\Gamma(\boldsymbol{\gamma}',\boldsymbol{\gamma}_0), d_\Theta(\boldsymbol{\theta}',\boldsymbol{\theta})\} < \delta_{\boldsymbol{\theta}} \\ \Rightarrow \quad |\psi(\boldsymbol{y}',\boldsymbol{\gamma}',\boldsymbol{\theta}') - \psi(\boldsymbol{y}_0,\boldsymbol{\gamma}_0,\boldsymbol{\theta})| < \frac{1}{8}\varepsilon.\end{aligned}$$

Let $r_{\boldsymbol{\theta}} > 0$ be such that $2r_{\boldsymbol{\theta}} < \delta_{\boldsymbol{\theta}}$. Then $\{B(\boldsymbol{\theta}, r_{\boldsymbol{\theta}})\}_{\boldsymbol{\theta}\in K_\Theta}$ is an open cover of K_Θ. Hence, we can extract a finite subcover $\{B(\boldsymbol{\theta}_k, r_{\boldsymbol{\theta}_k})\}_{k=1}^N$. Let $r_0 > 0$ be such that

$$r_0 < \min\{r_{\boldsymbol{\theta}_k}\}_{k=1}^N \quad \text{and} \quad B(\boldsymbol{y}_0, r_0)\times B(\boldsymbol{\gamma}_0, r_0) \subset U_{\boldsymbol{y}_0}\times U_{\boldsymbol{\gamma}_0},$$

and define

$$U_\Theta := \bigcup_{k=1}^N B(\boldsymbol{\theta}_k, r_{\boldsymbol{\theta}_k}).$$

Then there exists a continuous function $h : \Theta \to [0,1]$ such that $h \equiv 1$ on K_Θ and $h \equiv 0$ on U_Θ^c. Let

$$\psi^0(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x}) = h(\boldsymbol{\theta})\psi(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}) \quad \text{and} \quad \psi^R(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x}) = (1-h(\boldsymbol{\theta}))\psi(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}),$$

and define ρ^0 and ρ^R similarly. Then, for any

$$(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}) \in B(\boldsymbol{y}_0, r_0) \times B(\boldsymbol{\gamma}_0, r_0) \times U_\Theta,$$

there exists $\boldsymbol{\theta}_k \in K_\Theta$ such that

$$\begin{aligned}|\psi^0(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}) &- \psi^0(\boldsymbol{y}_0,\boldsymbol{\gamma}_0,\boldsymbol{\theta})| \\ &< |\psi(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}) - \psi(\boldsymbol{y}_0,\boldsymbol{\gamma}_0,\boldsymbol{\theta}_k)| + |\psi(\boldsymbol{y}_0,\boldsymbol{\gamma}_0,\boldsymbol{\theta}_k) - \psi(\boldsymbol{y}_0,\boldsymbol{\gamma}_0,\boldsymbol{\theta})| < \frac{1}{4}\varepsilon.\end{aligned}$$

For $(\boldsymbol{y},\boldsymbol{\gamma}) \in B(\boldsymbol{y}_0, r_0) \times B(\boldsymbol{\gamma}_0, r_0)$ we have that

$$\begin{aligned}|\rho(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x}) &- \rho(\boldsymbol{y}_0,\boldsymbol{\gamma}_0,\boldsymbol{x}_0)| \\ &< |\rho(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x}) - \rho(\boldsymbol{y}_0,\boldsymbol{\gamma}_0,\boldsymbol{x})| + |\rho(\boldsymbol{y}_0,\boldsymbol{\gamma}_0,\boldsymbol{x}) - \rho(\boldsymbol{y}_0,\boldsymbol{\gamma}_0,\boldsymbol{x}_0)|.\end{aligned}$$

The second term can be controlled using the weak continuity of P. In fact, it suffices to consider the open set

$$U_{\boldsymbol{x}_0} := \left\{\boldsymbol{x} \in \mathcal{X} : \ |\rho(\boldsymbol{y}_0,\boldsymbol{\gamma}_0,\boldsymbol{x}) - \rho(\boldsymbol{y}_0,\boldsymbol{\gamma}_0,\boldsymbol{x}_0)| < \frac{1}{4}\varepsilon\right\}.$$

For the first term, consider first

$$|\rho^0(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x}) - \rho^0(\boldsymbol{y}_0,\boldsymbol{\gamma}_0,\boldsymbol{x})| < \int_{\overline{U_\Theta}} (\psi^0(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}) - \psi^0(\boldsymbol{y}_0,\boldsymbol{\gamma}_0,\boldsymbol{\theta}))dP_{\boldsymbol{x}}(\boldsymbol{\theta}) < \frac{1}{4}\varepsilon.$$

For the second, note that

$$\rho^R(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x}) = \int \psi^R(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta})dP_{\boldsymbol{x}}(\boldsymbol{\theta}) \leq \int_{K_\Theta^c} \psi(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta})dP_{\boldsymbol{x}}(\boldsymbol{\theta}) < \frac{1}{8}\varepsilon.$$

Consequently, for $(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x}) \in B(\boldsymbol{y}_0, r_0) \times B(\boldsymbol{\gamma}_0, r_0) \times U_{\boldsymbol{x}_0}$ we have that

$$|\rho(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x}) - \rho(\boldsymbol{y}_0,\boldsymbol{\gamma}_0,\boldsymbol{x}_0)| < \rho^R(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x}) + \rho^R(\boldsymbol{y}_0,\boldsymbol{\gamma}_0,\boldsymbol{x}) + \frac{1}{4}\varepsilon + \frac{1}{4}\varepsilon < \varepsilon,$$

proving the claim. □

A.4.2. Proof of Lemma 5.2

Consider the sequence of functions $\{\rho_n\}_{n\in\mathbb{N}}$ for $\rho_n : \mathcal{Y} \times \Gamma \times \mathcal{X} \times \Omega \to \mathbb{R}_+$ given by

$$\rho^\omega_{n,\boldsymbol{x}}(\boldsymbol{y}) = \sum_{i=1}^n \pi^\omega_{i,\boldsymbol{x}} \psi(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}^\omega_{i,\boldsymbol{x}}).$$

Let $\Omega_0 \subset \Omega$ be a set of full measure such that: (i) $\pi_i^\omega \geq 0$; (ii) $\sum_{i\in\mathbb{N}} \pi_i^\omega \equiv 1$; and (iii) both π_i^ω and $\boldsymbol{\theta}_i^\omega$ are continuous on $\mathcal{X}$ for any $i \in \mathbb{N}$. If we restrict the functions to $\mathcal{Y}\times\Gamma\times\mathcal{X}\times\Omega_0$ then $\rho_n \geq 0$ and $\{\rho_n\}$ is a monotone non-decreasing sequence. Hence,

$$\rho_{\infty,\boldsymbol{x}}^\omega(\boldsymbol{y}) := \lim_{n\to\infty} \rho_{n,\boldsymbol{x}}^\omega(\boldsymbol{y}),$$

is well-defined for a.e. ω. By the monotone convergence theorem

$$\int \rho_{\infty,\boldsymbol{x}}^\omega(\boldsymbol{y}) d\nu_{\mathcal{Y}}(\boldsymbol{y}) \equiv 1.$$

To prove continuity for a.e. ω, fix $\omega \in \Omega_0$. Let $\boldsymbol{y}_0 \in \mathcal{Y}$, $\boldsymbol{x}_0 \in \mathcal{X}$ and $\boldsymbol{\gamma}_0 \in \Gamma$. By hypothesis, there exists an open neighborhood $U_{\boldsymbol{y}_0}$ of $\boldsymbol{y}_0$ and $U_{\boldsymbol{\gamma}_0}$ of $\boldsymbol{\gamma}_0$ such that

$$\forall\, (\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}) \in U_{\boldsymbol{y}_0} \times U_{\boldsymbol{\gamma}_0} \times \Theta : \quad \psi(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}) \leq c_f < \infty,$$

for some $c_f > 0$. Let $\varepsilon > 0$ and let $N_\varepsilon \in \mathbb{N}$ be such that

$$\sum_{i>N_\varepsilon} \pi_{i,x_0}^\omega < \frac{1}{8c_f}\varepsilon.$$

Define the open set $U_\pi \subset \mathcal{X}$

$$U_\pi := \bigcap_{i=1}^{N_\varepsilon} \left\{ \boldsymbol{x} \in \mathcal{X} :\ |\pi_{i,\boldsymbol{x}}^\omega - \pi_{i,x_0}^\omega| < \frac{1}{8c_f N_\varepsilon}\varepsilon \right\}.$$

For $\boldsymbol{x} \in U_\pi$ we have that

$$\sum_{i>N_\varepsilon} \pi_{i,\boldsymbol{x}}^\omega = 1 - \sum_{i=1}^{N_\varepsilon} \pi_{i,x_0}^\omega + \sum_{i=1}^{N_\varepsilon} (\pi_{i,\boldsymbol{x}}^\omega - \pi_{i,x_0}^\omega) < \frac{1}{8c_f}\varepsilon + \frac{1}{8c_f}\varepsilon < \frac{1}{4c_f}\varepsilon.$$

Now, let the open set $U_{f,x} \subset \mathcal{Y}\times\Gamma\times\mathcal{X}$ be

$$U_{f,x} := \bigcap_{i=1}^{N_\varepsilon} \left\{ (\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x}) \in U_{\boldsymbol{y}_0} \times U_{\boldsymbol{\gamma}_0} \times \mathcal{X} :\ |\psi(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}_{i,\boldsymbol{x}}^\omega) - \psi(\boldsymbol{y}_0,\boldsymbol{\gamma}_0,\boldsymbol{\theta}_{i,x_0}^\omega)| < \frac{1}{4c_f N_\varepsilon} \right\}.$$

This set is open by continuity of $\boldsymbol{\theta}_i^\omega$. Then, for any $(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x}) \in U_{f,x}$

$$\sum_{i>N_\varepsilon} \pi_{i,\boldsymbol{x}}^\omega \psi(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}_{i,\boldsymbol{x}}^\omega) < c_f \sum_{i>N_\varepsilon} \pi_{i,\boldsymbol{x}}^\omega < \frac{1}{4}\varepsilon.$$

Therefore,

$$\begin{aligned} &|\rho_{\infty,\boldsymbol{x},\boldsymbol{\gamma}}^\omega(\boldsymbol{y}) - \rho_{\infty,\boldsymbol{x}_0,\boldsymbol{\gamma}_0}^\omega(\boldsymbol{y}_0)| \\ &\leq \frac{1}{2}\varepsilon + \sum_{i=1}^{N_\varepsilon} |\pi_{i,\boldsymbol{x}}^\omega - \pi_{i,x_0}^\omega||\psi(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}_{i,\boldsymbol{x}}^\omega)| + \sum_{i=1}^{N_\varepsilon} \pi_{i,x_0}^\omega |\psi(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}_{i,\boldsymbol{x}}^\omega) - \psi(\boldsymbol{y}_0,\boldsymbol{\gamma}_0,\boldsymbol{\theta}_{i,x_0}^\omega)| \\ &\qquad < \frac{1}{2}\varepsilon + \frac{1}{4}\varepsilon + \frac{1}{4}\varepsilon = \varepsilon, \end{aligned}$$

proving the lemma. □

A.5. The continuity of DDP mixture models

A.5.1. Proof of Theorem 6.1

Fix $P \in C_W(\mathcal{X}, \mathcal{P}(\Theta))$. To simplify notation, we drop the superscript on M^P and ρ^P. Let $\varepsilon > 0$. We will prove that there exist a neighborhood of $(\boldsymbol{\gamma}_0, \boldsymbol{x}_0)$ such that

$$|M_{\boldsymbol{\gamma},\boldsymbol{x}}(B) - M_{\boldsymbol{\gamma}_0,\boldsymbol{x}_0}(B)| < \varepsilon,$$

uniformly on B. The strategy is to find suitable compact subsets of $\mathcal{Y}$ and Θ where the measures M and $P_{\boldsymbol{x}_0}$ are concentrated near $(\boldsymbol{\gamma}_0, \boldsymbol{x}_0)$. Then, we leverage the continuity of ψ and the weak continuity of P.

First, since $\mathcal{Y}$ is Polish there exists $K_{\mathcal{Y}} \subset \mathcal{Y}$ compact such that

$$M_{\boldsymbol{\gamma}_0,\boldsymbol{x}_0}(\mathcal{Y} \setminus K_{\mathcal{Y}}) < \frac{1}{16}\varepsilon.$$

Since $\nu_{\mathcal{Y}}$ is locally finite, there exists $c_{\mathcal{Y}} > 0$ such that

$$\max\{1, \nu_{\mathcal{Y}}(K_{\mathcal{Y}})\} < c_{\mathcal{Y}} < \infty.$$

Second, since Θ is Polish, there exists $K_\Theta \subset \Theta$ compact such that

$$P_{\boldsymbol{x}_0}(\Theta \setminus K_\Theta) < \frac{1}{16}\varepsilon.$$

We now construct a suitable cover for $K_{\mathcal{Y}} \times \{\boldsymbol{\gamma}_0\} \times K_\Theta$. Let $\delta_{\boldsymbol{y},\boldsymbol{\theta}} > 0$ be such that

$$\max\{d_{\mathcal{Y}}(\boldsymbol{y}', \boldsymbol{y}), d_\Gamma(\boldsymbol{\gamma}', \boldsymbol{\gamma}_0), d_\Theta(\boldsymbol{\theta}', \boldsymbol{\theta})\} < \delta_{\boldsymbol{y},\boldsymbol{\theta}}$$
$$\Rightarrow \quad |\psi(\boldsymbol{y}', \boldsymbol{\gamma}', \boldsymbol{\theta}') - \psi(\boldsymbol{y}, \boldsymbol{\gamma}_0, \boldsymbol{\theta})| < \frac{1}{64 c_{\mathcal{Y}}}\varepsilon.$$

Similarly, let $r_{\boldsymbol{y},\boldsymbol{\theta}} > 0$ be $2r_{\boldsymbol{y},\boldsymbol{\theta}} < \delta_{\boldsymbol{y},\boldsymbol{\theta}}$. Then, for every $\boldsymbol{y} \in K_{\mathcal{Y}}$ the collection

$$\{B(\boldsymbol{y}, r_{\boldsymbol{y},\boldsymbol{\theta}}) \times B(\boldsymbol{\gamma}_0, r_{\boldsymbol{y},\boldsymbol{\theta}}) \times B(\boldsymbol{\theta}, r_{\boldsymbol{y},\boldsymbol{\theta}})\}_{\boldsymbol{\theta} \in K_\Theta},$$

is an open cover of the compact set $\{\boldsymbol{y}\} \times \{\boldsymbol{\gamma}_0\} \times K_\Theta$. Hence, there exists a finite subcover

$$\{B(\boldsymbol{y}, r_{\boldsymbol{y},\boldsymbol{\theta}_\ell}) \times B(\boldsymbol{\gamma}_0, r_{\boldsymbol{y},\boldsymbol{\theta}_\ell}) \times B(\boldsymbol{\theta}_\ell, r_{\boldsymbol{y},\boldsymbol{\theta}_\ell})\}_{\ell=1}^{N_{\boldsymbol{y}}}.$$

Let $r_{\boldsymbol{y}} > 0$ be such that

$$r_{\boldsymbol{y}} < \min\{r_{\boldsymbol{y},\boldsymbol{\theta}_1}, \ldots, r_{\boldsymbol{y},\boldsymbol{\theta}_{N_{\boldsymbol{y}}}}\}.$$

Define the open neighborhood of K_Θ

$$K_\Theta^{r_{\boldsymbol{y}}} := \{\boldsymbol{\theta} \in \Theta : \ d_\Theta(\boldsymbol{\theta}, K_\Theta) < r_{\boldsymbol{y}}\},$$

and the open set

$$W_{\boldsymbol{y}} := B(\boldsymbol{y}, r_{\boldsymbol{y}}) \times B(\boldsymbol{\gamma}_0, r_{\boldsymbol{y}}) \times K_\Theta^{r_{\boldsymbol{y}}}.$$

Note that

$$\{\boldsymbol{y}\} \times \{\boldsymbol{\gamma}_0\} \times K_\Theta \subset W_{\boldsymbol{y}}.$$

Furthermore, for any $(\boldsymbol{y}', \boldsymbol{\gamma}', \boldsymbol{\theta}') \in W_{\boldsymbol{y}}$ there exists, by construction, $\boldsymbol{\theta}'' \in K_\Theta$ such that

$$d_\Theta(\boldsymbol{\theta}', \boldsymbol{\theta}'') < r_{\boldsymbol{y}}.$$

Hence, for some $\ell \in [N_{\boldsymbol{y}}]$ we have that

$$d_\Theta(\boldsymbol{\theta}', \boldsymbol{\theta}_\ell) \leq d_\Theta(\boldsymbol{\theta}', \boldsymbol{\theta}'') + d_\Theta(\boldsymbol{\theta}'', \boldsymbol{\theta}_\ell) < r_{\boldsymbol{y}} + r_{\boldsymbol{y},\boldsymbol{\theta}_\ell} < \delta_{\boldsymbol{y},\boldsymbol{\theta}_\ell}.$$

Furthermore, for the same choice of $\boldsymbol{\theta}_\ell$ we also have that $d_{\mathcal{Y}}(\boldsymbol{y}', \boldsymbol{y}) < \delta_{\boldsymbol{y},\boldsymbol{\theta}_\ell}$ and $d_\Gamma(\boldsymbol{\gamma}', \boldsymbol{\gamma}_0) < r_{\boldsymbol{y},\boldsymbol{\theta}_\ell}$. Hence, we deduce that

$$\begin{aligned} &|\psi(\boldsymbol{y}', \boldsymbol{\gamma}', \boldsymbol{\theta}') - \psi(\boldsymbol{y}, \boldsymbol{\gamma}_0, \boldsymbol{\theta}')| \\ &\quad \leq |\psi(\boldsymbol{y}', \boldsymbol{\gamma}', \boldsymbol{\theta}') - \psi(\boldsymbol{y}, \boldsymbol{\gamma}_0, \boldsymbol{\theta}_\ell)| + |\psi(\boldsymbol{y}, \boldsymbol{\gamma}_0, \boldsymbol{\theta}_\ell) - \psi(\boldsymbol{y}, \boldsymbol{\gamma}_0, \boldsymbol{\theta}')| < \frac{1}{32 c_{\mathcal{Y}}} \varepsilon. \end{aligned}$$

The collection $\{W_{\boldsymbol{y}}\}_{\boldsymbol{y} \in K_{\mathcal{Y}}}$ is an open cover of the compact set $K_{\mathcal{Y}} \times \{\boldsymbol{\gamma}_0\} \times K_\Theta$. Hence, we can extract a subcover $\{W_{\boldsymbol{y}_k}\}_{k=1}^{N_{\mathcal{Y}}}$. By possibly removing elements, we may assume that no ball is covered by the union of the remanining ones. Note that $\{B(\boldsymbol{y}_k, r_{\boldsymbol{y}_k})\}_{k=1}^{N_{\mathcal{Y}}}$ is a cover for $K_{\mathcal{Y}}$ with the same property. We partition $K_{\mathcal{Y}}$ into the sets

$$\begin{aligned} A_1 &:= K_{\mathcal{Y}} \cap B(\boldsymbol{y}_k, r_{\boldsymbol{y}_k}), \\ A_k &:= K_{\mathcal{Y}} \cap \left(B(\boldsymbol{y}_k, r_{\boldsymbol{y}_k}) \setminus \bigcup_{\ell=1}^{k-1} B(\boldsymbol{y}_\ell, r_{\boldsymbol{y}_\ell}) \right), \qquad k \in \{2, \dots, N_{\mathcal{Y}}\}. \end{aligned}$$

Additionally, we let

$$U_{\boldsymbol{\gamma}_0} := \bigcap_{k=1}^{N_{\mathcal{Y}}} B(\boldsymbol{\gamma}_0, r_{\boldsymbol{y}_k}).$$

Finally, consider the open set

$$U_\Theta := \{\boldsymbol{\theta} \in \Theta : \ d_\Theta(\boldsymbol{\theta}, K_\Theta) < \min\{r_{\boldsymbol{y}_1}, \dots, r_{\boldsymbol{y}_{N_{\mathcal{Y}}}}\}\} \subset \bigcap_{k=1}^{N_{\mathcal{Y}}} K_\Theta^{r_{\boldsymbol{y}_k}}.$$

Then, there exists a continuous function $h : \Theta \to [0, 1]$ such that $h|_{K_\Theta} \equiv 1$ and $h|_{U_\Theta^c} \equiv 0$. From K_Θ and U_Θ we obtain the decomposition,

$$\psi^R(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}) := (1 - h(\boldsymbol{\theta}))\psi(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}) \quad \text{and} \quad \psi^0(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}) = h(\boldsymbol{\theta})\psi(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}).$$

We define ρ^R, ρ^0, M^R and M^0 similarly. By construction ψ^0 is supported on $\overline{U_\Theta}$ and

$$P_{\boldsymbol{x}_0}(K_\Theta \setminus \overline{U_\Theta}) \leq P_{\boldsymbol{x}_0}(K_\Theta \setminus K_\Theta) < \frac{1}{16} \varepsilon.$$

Furthermore, for $(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}) \in B(\boldsymbol{y}_k, r_{\boldsymbol{y}_k}) \times U_{\boldsymbol{\gamma}_0} \times K_\Theta^{r_{\boldsymbol{y}_k}}$

$$|\psi_0(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}) - \psi_0(\boldsymbol{y}_k,\boldsymbol{\gamma}_0,\boldsymbol{\theta})| \leq h(\boldsymbol{\theta})|\psi(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}) - \psi(\boldsymbol{y}_k,\boldsymbol{\gamma}_0,\boldsymbol{\theta})| < \frac{1}{32c_{\mathcal{Y}}}\varepsilon.$$

We now show that

$$M^0_{\boldsymbol{\gamma},\boldsymbol{x}}(K_{\mathcal{Y}} \cap B) := \int_{K_{\mathcal{Y}}} \int \psi_0(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}) dP_{\boldsymbol{x}}(\boldsymbol{\theta}) d\nu_{\mathcal{Y}}(\boldsymbol{y}),$$

can be approximated by a continuous function near $(\boldsymbol{\gamma}_0,\boldsymbol{x}_0)$. Consider the decomposition

$$\begin{aligned} M^0_{\boldsymbol{\gamma},\boldsymbol{x}}(B) = &\sum_{k=1}^{N_{\mathcal{Y}}} \nu_{\mathcal{Y}}(A_k \cap B) \int \psi_0(\boldsymbol{y}_k,\boldsymbol{\gamma}_0,\boldsymbol{\theta}) dP_{\boldsymbol{x}}(\boldsymbol{\theta}) \\ &- \sum_{k=1}^{N_{\mathcal{Y}}} \int_{A_k \cap B} \int_{\overline{U_\Theta}} (\psi^0(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}) - \psi^0(\boldsymbol{y}_k,\boldsymbol{\gamma}_0,\boldsymbol{\theta})) dP_{\boldsymbol{x}}(\boldsymbol{\theta}) d\nu_{\mathcal{Y}}(\boldsymbol{y}). \end{aligned}$$

Since

$$\left| \int_{A_k \cap B} \int_{\overline{U_\Theta}} (\psi^0(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}) - \psi^0(\boldsymbol{y}_k,\boldsymbol{\gamma}_0,\boldsymbol{\theta})) dP_{\boldsymbol{x}}(\boldsymbol{\theta}) d\nu_{\mathcal{Y}}(\boldsymbol{y}) \right| < \frac{1}{32c_{\mathcal{Y}}}\varepsilon \nu_{\mathcal{Y}}(A_k \cap B),$$

whence

$$\left| M^0_{\boldsymbol{\gamma},\boldsymbol{x}}(K_{\mathcal{Y}} \cap B) - \sum_{k=1}^{N_{\mathcal{Y}}} \nu_{\mathcal{Y}}(A_k \cap B) \int \psi^0(\boldsymbol{y}_k,\boldsymbol{\gamma}_0,\boldsymbol{\theta}) dP_{\boldsymbol{x}}(\boldsymbol{\theta}) \right| < \frac{1}{32}\varepsilon.$$

Therefore

$$\begin{aligned} &|M^0_{\boldsymbol{\gamma},\boldsymbol{x}}(K_{\mathcal{Y}} \cap B) - M^0_{\boldsymbol{\gamma}_0,\boldsymbol{x}_0}(K_{\mathcal{Y}} \cap B)| \\ &\qquad < \frac{1}{16}\varepsilon + \sum_{k=1}^{N_{\mathcal{Y}}} \nu_{\mathcal{Y}}(A_k \cap B) |\rho^0(\boldsymbol{y}_k,\boldsymbol{\gamma}_0,\boldsymbol{x}) - \rho^0(\boldsymbol{y}_k,\boldsymbol{\gamma}_0,\boldsymbol{x}_0)|. \end{aligned}$$

It suffices to choose

$$U_{\boldsymbol{x}_0} := \bigcap_{k=1}^{N_{\mathcal{Y}}} \left\{ \boldsymbol{x} \in \mathcal{X} : \left| \int \psi^0(\boldsymbol{y}_k,\boldsymbol{\gamma}_0,\boldsymbol{\theta}) dP_{\boldsymbol{x}}(\boldsymbol{\theta}) - \int \psi^0(\boldsymbol{y}_k,\boldsymbol{\gamma}_0,\boldsymbol{\theta}) dP_{\boldsymbol{x}_0}(\boldsymbol{\theta}) \right| < \frac{1}{16}\varepsilon \right\},$$

as an open neighborhood for $\boldsymbol{x}_0$. Hence, we obtain the bound

$$|M^0_{\boldsymbol{\gamma},\boldsymbol{x}}(K_{\mathcal{Y}} \cap B) - M^0_{\boldsymbol{\gamma}_0,\boldsymbol{x}_0}(K_{\mathcal{Y}} \cap B)| < \frac{1}{8}\varepsilon,$$

which is uniform over B for $(\boldsymbol{\gamma},\boldsymbol{x}) \in U_{\boldsymbol{\gamma}_0} \times U_{\boldsymbol{x}_0}$.

We now show that the remaining terms do not concentrate too much mass. By construction,

$$M_{\boldsymbol{\gamma}_0,\boldsymbol{x}_0}(K_{\mathcal{Y}}) = M^0_{\boldsymbol{\gamma}_0,\boldsymbol{x}_0}(K_{\mathcal{Y}}) + M^R_{\boldsymbol{\gamma}_0,\boldsymbol{x}_0}(K_{\mathcal{Y}}) \geq 1 - \frac{1}{16}\varepsilon.$$

Furthermore, by Fubini's theorem,

$$\begin{aligned}|M_{\boldsymbol{\gamma}_0,\boldsymbol{x}_0}(K_{\mathcal{Y}}) - M^0_{\boldsymbol{\gamma}_0,\boldsymbol{x}_0}(K_{\mathcal{Y}})| &< \int_{K_{\mathcal{Y}}} \int_{\Theta\setminus K_\Theta} \psi^R(\boldsymbol{y},\boldsymbol{\gamma}_0,\boldsymbol{\theta}) dP_{\boldsymbol{x}_0}(\boldsymbol{\theta}) d\nu_{\mathcal{Y}}(\boldsymbol{y}),\\ &\leq \int_{\Theta\setminus K_\Theta} \int_{K_{\mathcal{Y}}} \psi^R(\boldsymbol{y},\boldsymbol{\gamma}_0,\boldsymbol{\theta}) d\nu_{\mathcal{Y}}(\boldsymbol{y}) dP_{\boldsymbol{x}_0}(\boldsymbol{\theta}),\\ &\leq P_{\boldsymbol{x}_0}(\Theta \setminus K_\Theta),\\ &< \frac{1}{16}\varepsilon,\end{aligned}$$

where we used the fact that $1-h$ is supported on $\Theta \setminus K_\Theta$. It follows that

$$M^0_{\boldsymbol{\gamma}_0,\boldsymbol{x}_0}(K_{\mathcal{Y}}) > 1 - \frac{1}{8}\varepsilon \quad \Rightarrow \quad M^R_{\boldsymbol{\gamma}_0,\boldsymbol{x}_0}(K_{\mathcal{Y}}) \leq \frac{1}{8}\varepsilon.$$

Note then that

$$M^0_{\boldsymbol{\gamma},\boldsymbol{x}}(K_{\mathcal{Y}}) > M^0_{\boldsymbol{\gamma}_0,\boldsymbol{x}_0}(K_{\mathcal{Y}}) - \frac{1}{8}\varepsilon > 1 - \frac{1}{4}\varepsilon,$$

whence

$$1 - M^0_{\boldsymbol{\gamma},\boldsymbol{x}}(K_{\mathcal{Y}}) = M^R_{\boldsymbol{\gamma},\boldsymbol{x}}(K_{\mathcal{Y}}) + M_{\boldsymbol{\gamma},\boldsymbol{x}}(\mathcal{Y} \setminus K_{\mathcal{Y}}) < \frac{1}{4}\varepsilon.$$

Finally, from the decomposition

$$\begin{aligned}M_{\boldsymbol{\gamma},\boldsymbol{x}}(B) &= M_{\boldsymbol{\gamma},\boldsymbol{x}}(K_{\mathcal{Y}} \cap B) + M_{\boldsymbol{\gamma},\boldsymbol{x}}(B \setminus K_{\mathcal{Y}})\\ &= M^0_{\boldsymbol{\gamma},\boldsymbol{x}}(K_{\mathcal{Y}} \cap B) + M^R_{\boldsymbol{\gamma},\boldsymbol{x}}(K_{\mathcal{Y}} \cap B) + M_{\boldsymbol{\gamma},\boldsymbol{x}}(B \setminus K_{\mathcal{Y}}),\end{aligned}$$

we deduce that

$$\begin{aligned}&|M_{\boldsymbol{\gamma},\boldsymbol{x}}(B) - M_{\boldsymbol{\gamma}_0,\boldsymbol{x}_0}(B)| \leq |M^0_{\boldsymbol{\gamma},\boldsymbol{x}}(K_{\mathcal{Y}} \cap B) - M^0_{\boldsymbol{\gamma}_0,\boldsymbol{x}_0}(K_{\mathcal{Y}} \cap B)|\\ &\quad + M^R_{\boldsymbol{\gamma},\boldsymbol{x}}(K_{\mathcal{Y}} \cap B) + M_{\boldsymbol{\gamma},\boldsymbol{x}}(B \setminus K_{\mathcal{Y}}) + M^R_{\boldsymbol{\gamma}_0,\boldsymbol{x}_0}(K_{\mathcal{Y}} \cap B) + M_{\boldsymbol{\gamma}_0,\boldsymbol{x}_0}(B \setminus K_{\mathcal{Y}})\\ &\qquad < \frac{1}{8}\varepsilon + \frac{1}{4}\varepsilon + \frac{1}{4}\varepsilon < \varepsilon,\end{aligned}$$

for $(\boldsymbol{\gamma},\boldsymbol{x}) \in U_{\boldsymbol{\gamma}_0} \times U_{\boldsymbol{x}_0}$ uniformly in B. This proves the theorem. □

A.6. The support of DDP mixture models

To characterize the support of mixtures in different topologies, we will use repeatedly the following lemma. It provides uniform control on the behavior of mixtures uniformly over weakly continuous $P : \mathcal{X} \to \mathcal{P}(\Theta)$.

Lemma A.2. *Suppose that ψ is continuous and that for every $\boldsymbol{y}_0 \in \mathcal{Y}$, $\boldsymbol{\gamma}_0 \in \Gamma$ and $\varepsilon > 0$ there exists $K_0 \subset \Theta$ compact such that*

$$(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}) \in U_{\boldsymbol{y}_0} \times U_{\boldsymbol{\gamma}_0} \times K_0^c \quad \Rightarrow \quad \psi(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}) < \varepsilon.$$

The following assertions are true.

1. *Let $K_{\mathcal{Y}\times\Gamma} \subset \mathcal{Y}\times\Gamma$ be compact. Then, for every $\varepsilon > 0$ there exists $U_{K_{\mathcal{Y}\times\Gamma}} \supset K_{\mathcal{Y}\times\Gamma}$ open, and $\delta > 0$ such that for any $(\boldsymbol{y}, \boldsymbol{\gamma}), (\boldsymbol{y}', \boldsymbol{\gamma}') \in U_{K_{\mathcal{Y}\times\Gamma}}$ we have that*

$$d_{\mathcal{Y}\times\Gamma}((\boldsymbol{y}, \boldsymbol{\gamma}), (\boldsymbol{y}', \boldsymbol{\gamma}')) < \delta \\ \Rightarrow \sup_{\boldsymbol{x}\in\mathcal{X}} \left| \int_\Theta \psi(\boldsymbol{y}', \boldsymbol{\gamma}', \boldsymbol{\theta}) dP_{\boldsymbol{x}}(\boldsymbol{\theta}) - \int_\Theta \psi(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}) dP_{\boldsymbol{x}}(\boldsymbol{\theta}) \right| < \varepsilon,$$

uniformly over $P : \mathcal{X} \to \mathcal{P}(\Theta)$ weakly continuous.

2. *Let $K_{\Gamma\times\mathcal{X}} \subset \Gamma \times \mathcal{X}$ be compact, and let $P : \mathcal{X} \to \mathcal{P}(\Theta)$ be weakly continuous. Then, for every $\varepsilon > 0$ there exists $K_{\mathcal{Y}} \subset \mathcal{Y}$ compact and $U_{K_{\Gamma\times\mathcal{X}}} \supset K_{\Gamma\times\mathcal{X}}$ open such that*

$$\forall\, (\boldsymbol{\gamma}, \boldsymbol{x}) \in U_{K_{\Gamma\times\mathcal{X}}} : \quad Q^P_{\boldsymbol{\gamma},\boldsymbol{x}}(\mathcal{Y} \setminus K_{\mathcal{Y}}) < \varepsilon.$$

Proof of Lemma A.2. Proof of 1. Let $\varepsilon' \in (0, \varepsilon)$. We first prove the result for $K_{\mathcal{Y}\times\Gamma} = \{(\boldsymbol{y}_0, \boldsymbol{\gamma}_0)\}$. By hypothesis, there exists neighborhoods $U_{\boldsymbol{y}_0}$ and $U_{\boldsymbol{\gamma}_0}$ of $\boldsymbol{y}_0$ and $\boldsymbol{\gamma}_0$ respectively, and $K_\Theta \subset \Theta$ compact such that

$$(\boldsymbol{y}', \boldsymbol{\gamma}', \boldsymbol{\theta}') \in U_{\boldsymbol{y}_0} \times U_{\boldsymbol{\gamma}_0} \times K_\Theta^c : \quad \psi(\boldsymbol{y}', \boldsymbol{\gamma}', \boldsymbol{\theta}') < \frac{1}{8}\varepsilon'.$$

For $\boldsymbol{\theta}_0 \in K_\Theta$ let $\delta_{\boldsymbol{\theta}_0} > 0$ be such that

$$d_{\mathcal{Y}\times\Gamma\times\Theta}((\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}), (\boldsymbol{y}_0, \boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)) < \delta_{\boldsymbol{\theta}_0} \;\Rightarrow\; |\psi(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}) - \psi(\boldsymbol{y}_0, \boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)| < \frac{1}{8}\varepsilon'.$$

Let $2r_{\boldsymbol{\theta}_0} < \delta_{\boldsymbol{\theta}_0}$. Then $\{B(\boldsymbol{\theta}, r_{\boldsymbol{\theta}})\}_{\boldsymbol{\theta}\in K_\Theta}$ is an open cover of K_Θ from which we can extract a finite subcover $\{B(\boldsymbol{\theta}_k, r_{\boldsymbol{\theta}_k})\}_{k=1}^N$. Define the neighborhoods

$$\begin{aligned} U'_{\boldsymbol{y}_0} &:= U_{\boldsymbol{y}_0} \cap \bigcup_{k=1}^N B(\boldsymbol{y}_0, r_0), \\ U'_{\boldsymbol{\gamma}_0} &:= U_{\boldsymbol{\gamma}_0} \cap \bigcup_{k=1}^N B(\boldsymbol{\gamma}_0, r_{\boldsymbol{\theta}_k}), \\ U_{K_\Theta} &:= \bigcup_{k=1}^N B(\boldsymbol{\theta}_k, r_{\boldsymbol{\theta}_k}), \end{aligned}$$

of $\boldsymbol{y}_0, \boldsymbol{\gamma}_0$ and K_Θ respectively, and let $h : \Theta \to [0, 1]$ be a continuous map such that $h \equiv 1$ on K_Θ and $h \equiv 0$ on $U^c_{K_\Theta}$. Then, for any $\boldsymbol{y} \in U'_{\boldsymbol{y}_0}$ and $\boldsymbol{\gamma} \in U'_{\boldsymbol{\gamma}_0}$ we have that

$$\int_\Theta (1 - h(\boldsymbol{\theta}))\psi(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta})\, dP_{\boldsymbol{x}}(\boldsymbol{\theta}) \le \int_{K^c_\Theta} \psi(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta})\, dP_{\boldsymbol{x}}(\boldsymbol{\theta}) < \frac{1}{8}\varepsilon P_{\boldsymbol{x}}(K^c_\Theta) < \frac{1}{8}\varepsilon',$$

for any weakly continuous $P : \mathcal{X} \to \mathcal{P}(\Theta)$ and $\boldsymbol{x} \in \mathcal{X}$. Hence, define

$$\psi_{K_\Theta}(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}) := h(\boldsymbol{\theta})\psi(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}).$$

Then, for any $(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}) \in U'_{\boldsymbol{y}_0} \times U'_{\boldsymbol{\gamma}_0} \times U_{K_\Theta}$ there exists $\boldsymbol{\theta}_k$ such that

$$|\psi_{K_\Theta}(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}) - \psi_{K_\Theta}(\boldsymbol{y}_0, \boldsymbol{\gamma}_0, \boldsymbol{\theta})| < |\psi(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}) - \psi(\boldsymbol{y}_0, \boldsymbol{\gamma}_0, \boldsymbol{\theta}_k)| + |\psi(\boldsymbol{y}_0, \boldsymbol{\gamma}_0, \boldsymbol{\theta}_k) - \psi(\boldsymbol{y}_0, \boldsymbol{\gamma}_0, \boldsymbol{\theta})| < \frac{1}{4}\varepsilon'.$$

It follows that for any $(\boldsymbol{y}, \boldsymbol{\gamma}) \in U'_{\boldsymbol{y}_0} \times U'_{\boldsymbol{\gamma}_0}$ we have that

$$\left|\int_\Theta \psi(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}) dP_{\boldsymbol{x}}(\boldsymbol{\theta}) - \int_\Theta \psi(\boldsymbol{y}_0, \boldsymbol{\gamma}_0, \boldsymbol{\theta}) dP_{\boldsymbol{x}}(\boldsymbol{\theta})\right| < \frac{1}{4}\varepsilon' + \int_{U_{K_\Theta}} |\psi_{K_\Theta}(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}) - \psi_{K_\Theta}(\boldsymbol{y}_0, \boldsymbol{\gamma}_0, \boldsymbol{\theta})| dP_{\boldsymbol{x}}(\boldsymbol{\theta}) < \frac{1}{2}\varepsilon',$$

for any weakly continuous $P : \mathcal{X} \to \mathcal{P}(\Theta)$ and $\boldsymbol{x} \in \mathcal{X}$. From now on, we let $r_{(\boldsymbol{y}_0, \boldsymbol{\gamma}_0)} > 0$ be such that

$$B((\boldsymbol{y}_0, \boldsymbol{\gamma}_0), r_{(\boldsymbol{y}_0, \boldsymbol{\gamma}_0)}) \subset U'_{\boldsymbol{y}_0} \times U'_{\boldsymbol{\gamma}_0}.$$

We now consider the case for an arbitrary compact set $K_{\mathcal{Y}\times\Gamma}$. From the open cover $\{B((\boldsymbol{y}, \boldsymbol{\gamma}), r_{(\boldsymbol{y}, \boldsymbol{\gamma})}/2)\}_{(\boldsymbol{y}, \boldsymbol{\gamma}) \in K_{\mathcal{Y}\times\Gamma}}$, we can extract a finite subcover $\{B((\boldsymbol{y}_k, \boldsymbol{\gamma}_k), r_{(\boldsymbol{y}_k, \boldsymbol{\gamma}_k)}/2)\}_{k=1}^N$. Remark the radius of the cover is half of that obtained in the previous step. Define

$$U_{K_{\mathcal{Y}\times\Gamma}} := \bigcup_{k=1}^N B((\boldsymbol{y}_k, \boldsymbol{\gamma}_k), r_{(\boldsymbol{y}_k, \boldsymbol{\gamma}_k)}/2),$$

and

$$\delta := \frac{1}{4} \min\{r_{(\boldsymbol{y}_k, \boldsymbol{\gamma}_k)} : k \in [N]\}.$$

Let $(\boldsymbol{y}', \boldsymbol{\gamma}'), (\boldsymbol{y}, \boldsymbol{\gamma}) \in U_{K_{\mathcal{Y}\times\Gamma}}$ be such that

$$d_{\mathcal{Y}\times\Gamma}((\boldsymbol{y}', \boldsymbol{\gamma}'), (\boldsymbol{y}, \boldsymbol{\gamma})) < \delta.$$

Then, there exists $(\boldsymbol{y}_k, \boldsymbol{\gamma}_k) \in K_{\mathcal{Y}\times\Gamma}$ such that

$$d_{\mathcal{Y}\times\Gamma}((\boldsymbol{y}, \boldsymbol{\gamma}), (\boldsymbol{y}_k, \boldsymbol{\gamma}_k)) < \frac{1}{2} r_{(\boldsymbol{y}_k, \boldsymbol{\gamma}_k)}.$$

This implies that

$$d_{\mathcal{Y}\times\Gamma}((\boldsymbol{y}', \boldsymbol{\gamma}'), (\boldsymbol{y}_k, \boldsymbol{\gamma}_k)) < d_{\mathcal{Y}\times\Gamma}((\boldsymbol{y}', \boldsymbol{\gamma}'), (\boldsymbol{y}, \boldsymbol{\gamma})) + \frac{1}{2} r_{(\boldsymbol{y}_k, \boldsymbol{\gamma}_k)} < r_{(\boldsymbol{y}_k, \boldsymbol{\gamma}_k)},$$

from where it follows that $(\boldsymbol{y}', \boldsymbol{\gamma}'), (\boldsymbol{y}, \boldsymbol{\gamma}) \in B((\boldsymbol{y}_k, \boldsymbol{\gamma}_k), r_{\boldsymbol{y}_k, \boldsymbol{\gamma}_k})$. In particular,

$$\begin{aligned}\left|\int_\Theta \psi(\boldsymbol{y}',\boldsymbol{\gamma}',\boldsymbol{\theta})dP_{\boldsymbol{x}}(\boldsymbol{\theta}) - \int_\Theta \psi(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta})dP_{\boldsymbol{x}}(\boldsymbol{\theta})\right| \\ \leq \left|\int_\Theta \psi(\boldsymbol{y}',\boldsymbol{\gamma}',\boldsymbol{\theta})dP_{\boldsymbol{x}}(\boldsymbol{\theta}) - \int_\Theta \psi(\boldsymbol{y}_k,\boldsymbol{\gamma}_k,\boldsymbol{\theta})dP_{\boldsymbol{x}}(\boldsymbol{\theta})\right| \\ + \left|\int_\Theta \psi(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta})dP_{\boldsymbol{x}}(\boldsymbol{\theta}) - \int_\Theta \psi(\boldsymbol{y}_k,\boldsymbol{\gamma}_k,\boldsymbol{\theta})dP_{\boldsymbol{x}}(\boldsymbol{\theta})\right| < \frac{1}{2}\varepsilon' + \frac{1}{2}\varepsilon' = \varepsilon',\end{aligned}$$

for any $P : \mathcal{X} \to \mathcal{P}(\Theta)$ and $\boldsymbol{x} \in \mathcal{X}$. Since the supremum over $\mathcal{X}$ can be at most ε' with $\varepsilon' < \varepsilon$, this proves the lemma.
Proof of 2. The hypothesis allows us to conclude from Theorem 6.1 that for any weakly continuous $P : \mathcal{X} \to \mathcal{P}(\Theta)$ the map $M^P : \Gamma \times \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ is strongly continuous. Hence, for every $(\boldsymbol{\gamma}, \boldsymbol{x}) \in K_{\Gamma\times\mathcal{X}}$ let $K_{\boldsymbol{\gamma},\boldsymbol{x}} \subset \mathcal{Y}$ be a compact set such that

$$M^P_{\boldsymbol{\gamma},\boldsymbol{x}}(\mathcal{Y} \setminus K_{\boldsymbol{\gamma},\boldsymbol{x}}) < \frac{1}{2}\varepsilon.$$

Then, let $U_{\boldsymbol{\gamma},\boldsymbol{x}} \subset \Gamma \times \mathcal{X}$ be an open neighborhood of $(\boldsymbol{\gamma}, \boldsymbol{x})$ such that

$$\forall\, (\boldsymbol{\gamma}', \boldsymbol{x}') \in U_{\boldsymbol{\gamma},\boldsymbol{x}} : \quad |M^P_{\boldsymbol{\gamma}',\boldsymbol{x}'}(\mathcal{Y} \setminus K_{\boldsymbol{\gamma},\boldsymbol{x}}) - M^P_{\boldsymbol{\gamma}.\boldsymbol{x}}(\mathcal{Y} \setminus K_{\boldsymbol{\gamma},\boldsymbol{x}})| < \frac{1}{2}\varepsilon.$$

Hence, $\{U_{\boldsymbol{\gamma},\boldsymbol{x}}\}_{(\boldsymbol{\gamma},\boldsymbol{x})\in K_\Gamma\times K_\mathcal{X}}$ is an open cover of $K_{\Gamma\times\mathcal{X}}$ from which we can extract a finite subcover $\{U_{\boldsymbol{\gamma}_k,\boldsymbol{x}_k}\}_{k=1}^N$. Let

$$U_{K_{\Gamma\times\mathcal{X}}} := \bigcup_{k=1}^N U_{\boldsymbol{\gamma}_k,\boldsymbol{x}_k}.$$

Then that for every $(\boldsymbol{\gamma}, \boldsymbol{x}) \in U_{K_{\Gamma\times\mathcal{X}}}$ there exists $(\boldsymbol{\gamma}_k, \boldsymbol{x}_k)$ such that

$$M^P_{\boldsymbol{\gamma},\boldsymbol{x}}(\mathcal{Y} \setminus K_{\boldsymbol{\gamma}_k,\boldsymbol{x}_k}) \leq \frac{1}{2}\varepsilon + |M^P_{\boldsymbol{\gamma},\boldsymbol{x}}(\mathcal{Y} \setminus K_{\boldsymbol{\gamma}_k,\boldsymbol{x}_k}) - M^P_{\boldsymbol{\gamma}_k,\boldsymbol{x}_k}(\mathcal{Y} \setminus K_{\boldsymbol{\gamma}_k,\boldsymbol{x}_k})| < \varepsilon.$$

Hence, we can choose the compact set

$$K_\mathcal{Y} := \bigcup_{k=1}^N K_{\boldsymbol{\gamma}_k,\boldsymbol{x}_k},$$

proving the claim. □

As a consequence of the lemma, if the hypotheses of Theorem 4.1 hold, then for every ω on a set of full measure we have that

$$\left|\int_\Theta \psi(\boldsymbol{y}',\boldsymbol{\gamma}',\boldsymbol{\theta})dG^\omega_{\boldsymbol{x}}(\boldsymbol{\theta}) - \int_\Theta \psi(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta})dG^\omega_{\boldsymbol{x}}(\boldsymbol{\theta})\right| < \varepsilon,$$

for any $\boldsymbol{x} \in \mathcal{X}$.

The lemma allows us to use essentially the same argument to characterize the support both in the product and compact-open topologies as follows. The statements in the case of the compact-open topology involve the supremum over a compact set $K_\Gamma \times K_{\mathcal{X}} \subset \Gamma \times \mathcal{X}$ and $P^0 : \mathcal{X} \to \mathcal{P}(\Theta)$ weakly continuous. The statements in the case of the product topology can be reduced to this as follows. First, in the case of the product topology we need to consider a finite set $\{(\boldsymbol{\gamma}_i, \boldsymbol{x}_i) : i \in [n]\}$. We will see this is equivalent to bounding a supremum over the compact set

$$K_{\Gamma\times\mathcal{X}} := \{(\boldsymbol{\gamma}_i, \boldsymbol{x}_i) : i \in [n]\},$$

or by considering first the compact sets

$$K_\Gamma := \{\boldsymbol{\gamma}_i : i \in [n]\} \quad \text{and} \quad K_{\mathcal{X}} := \{\boldsymbol{x}_i : i \in [n]\},$$

and defining $K_{\Gamma\times\mathcal{X}} := K_\Gamma \times K_{\mathcal{X}}$. Second, in the case of the product topology we make no assumptions about the continuity of $P^0 : \mathcal{X} \to \mathcal{P}(\Theta)$. However, since only its values on the finite set $\{\boldsymbol{x}_i : i \in [n]\}$ are relevant, we can leverage Lemma A.1 to replace P^0 by its weakly continuous interpolant $\overline{P^0}$. Once we have performed this reduction, Lemma A.2 will allow us to prove the desired results using analogous arguments for the product and compact-open topologies.

A.6.1. Proof of Theorems 6.3 and 6.4

In the case of the product-Hellinger topology, we define the compact sets $K_\Gamma := \{\boldsymbol{\gamma}_i : i \in [n]\}$ and $K_{\mathcal{X}} := \{\boldsymbol{x}_i : i \in [n]\}$. Furthermore, we let $K_{\Gamma\times\mathcal{X}} = K_\Gamma \times K_{\mathcal{X}}$ and we let $\varepsilon_0 < \min\{\varepsilon_1, \dots, \varepsilon_n\}$. As indicated before, over the finite set $K_{\mathcal{X}}$ we can assume without loss that P^0 is weakly continuous. For the compact-Hellinger topology, we define $K_{\Gamma\times\mathcal{X}} = K_\Gamma \times K_{\mathcal{X}}$ and let $\varepsilon_0 < \varepsilon$.

This reduction allows us to consider the event

$$\left\{\omega \in \Omega : \sup_{(\boldsymbol{\gamma},\boldsymbol{x})\in K_{\Gamma\times\mathcal{X}}} d_H(\rho^{G^\omega}_{\boldsymbol{\gamma},\boldsymbol{x}}, \rho^{P^0}_{\boldsymbol{\gamma},\boldsymbol{x}}) < \varepsilon_0\right\}.$$

Both Theorem 6.3 and 6.4 follow if we show the above event has positive probability.

Note that

$$d_H(\rho^{G^\omega}_{\boldsymbol{\gamma},\boldsymbol{x}}, \rho^{P^0}_{\boldsymbol{\gamma},\boldsymbol{x}}) < \varepsilon_0 \;\Leftrightarrow\; 1 - \int_{\mathcal{Y}} \rho^{G^\omega}(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{x})^{1/2} \rho^{P^0}(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{x})^{1/2} d\nu_{\mathcal{Y}}(\boldsymbol{y}) < 2\varepsilon_0^2.$$

Let $\varepsilon \in (0,1)$ be such that $\varepsilon < 2\varepsilon_0^2$. From Lemma A.2 there exists a compact set $K_{\mathcal{Y}} \subset \mathcal{Y}$ such that

$$\sup_{(\boldsymbol{\gamma},\boldsymbol{x})\in K_\Gamma\times K_{\mathcal{X}}} M^P_{\boldsymbol{\gamma},\boldsymbol{x}}(\mathcal{Y} \setminus K_{\mathcal{Y}}) < \frac{1}{4}\varepsilon.$$

Since $\nu_{\mathcal{Y}}$ is locally finite, we can assume without loss that $\nu_{\mathcal{Y}}(K_{\mathcal{Y}}) < \infty$. Define the compact set $K_{\mathcal{Y}\times\Gamma} = K_{\mathcal{Y}} \times K_\Gamma$. By Lemma 5.1 there exists $\delta > 0$ such that for any $(\boldsymbol{y}', \boldsymbol{\gamma}'), (\boldsymbol{y}, \boldsymbol{\gamma}) \in K_{\mathcal{Y}\times\Gamma}$ we have that

$$d_{\mathcal{Y}\times\Gamma}((\boldsymbol{y}',\boldsymbol{\gamma}'),(\boldsymbol{y},\boldsymbol{\gamma}))<\delta$$
$$\Rightarrow \quad \left|\int_\Theta \psi(\boldsymbol{y}',\boldsymbol{\gamma}',\boldsymbol{\theta})dP_{\boldsymbol{x}}(\boldsymbol{\theta})-\int_\Theta \psi(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta})dP_{\boldsymbol{x}}(\boldsymbol{\theta})\right|<\frac{1}{16\nu_{\mathcal{Y}}(K_{\mathcal{Y}})}\varepsilon^2,$$

uniformly over weakly continuous $P:\mathcal{X}\to\mathcal{P}(\Theta)$ and $\boldsymbol{x}\in\mathcal{X}$. In particular,

$$\left|\left(\int_\Theta \psi(\boldsymbol{y}',\boldsymbol{\gamma}',\boldsymbol{\theta})dP_{\boldsymbol{x}}(\boldsymbol{\theta})\right)^{1/2}-\left(\int_\Theta \psi(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta})dP_{\boldsymbol{x}}(\boldsymbol{\theta})\right)^{1/2}\right|<\frac{1}{4\nu_{\mathcal{Y}}(K_{\mathcal{Y}})^{1/2}}\varepsilon.$$

Let $2r<\delta$. We construct a finite open cover for $K_{\mathcal{Y}\times\Gamma}$ as follows. First, from the open cover $\{B(\boldsymbol{y},r)\}_{\boldsymbol{y}\in K_{\mathcal{Y}}}$ we can extract a finite subcover $\{B(\boldsymbol{y}_k,r)\}_{k=1}^N$. Without loss, we can assume that it is minimal, and we can partition $K_{\mathcal{Y}}$ in terms of the measurable sets of positive measure

$$A_1:=K_{\mathcal{Y}}\cap B(\boldsymbol{y}_k,r),$$
$$A_k:=K_{\mathcal{Y}}\cap\left(B(\boldsymbol{y}_k,r)\setminus\bigcup_{\ell=1}^{k-1}B(\boldsymbol{y}_\ell,r)\right),\qquad k\in\{2,\dots,N\}.$$

Second, from the open cover $\{B(\boldsymbol{\gamma},r)\}_{\boldsymbol{\gamma}\in K_\Gamma}$ we can extract a finite subcover $\{B(\boldsymbol{\gamma}_\ell,r)\}_{\ell=1}^M$. Note that $\{B(\boldsymbol{y}_k,r)\times B(\boldsymbol{\gamma}_\ell,r):\,k\in[N],\,\ell\in[M]\}$ is an open cover for $K_{\mathcal{Y}\times\Gamma}$. Hence, for any $(\boldsymbol{\gamma},\boldsymbol{x})\in K_{\Gamma\times\mathcal{X}}$ we can write

$$\begin{aligned}
1-\frac{1}{4}\varepsilon &\le \int_{K_{\mathcal{Y}}}\rho^{P^0}(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x})\,d\nu_{\mathcal{Y}}(\boldsymbol{y}),\\
&=\sum_{k=1}^N\int_{A_k}(\rho^{P^0}(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x})^{1/2}-\rho^{P^0}(\boldsymbol{y}_k,\boldsymbol{\gamma}_\ell,\boldsymbol{x})^{1/2})\rho^{P^0}(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x})^{1/2}d\nu_{\mathcal{Y}}(\boldsymbol{y}),\\
&+\sum_{k=1}^N\int_{A_k}(\rho^{P^0}(\boldsymbol{y}_k,\boldsymbol{\gamma}_\ell,\boldsymbol{x})^{1/2}-\rho^{G^\omega}(\boldsymbol{y}_k,\boldsymbol{\gamma}_\ell,\boldsymbol{x})^{1/2})\rho^{P^0}(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x})^{1/2}d\nu_{\mathcal{Y}}(\boldsymbol{y}),\\
&+\sum_{k=1}^N\int_{A_k}(\rho^{G^\omega}(\boldsymbol{y}_k,\boldsymbol{\gamma}_\ell,\boldsymbol{x})^{1/2}-\rho^{G^\omega}(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x})^{1/2})\rho^{P^0}(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x})^{1/2}d\nu_{\mathcal{Y}}(\boldsymbol{y}),\\
&+\int_{K_{\mathcal{Y}}}\rho^{G^\omega}(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x})^{1/2}\rho^{P^0}(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x})^{1/2}d\nu_{\mathcal{Y}}(\boldsymbol{y}),
\end{aligned}$$

where $\boldsymbol{\gamma}\in B(\boldsymbol{\gamma}_\ell,r)$. The first sum can be bounded as

$$\begin{aligned}
&\int_{A_k}(\rho^{P^0}(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x})^{1/2}-\rho^{P^0}(\boldsymbol{y}_k,\boldsymbol{\gamma}_\ell,\boldsymbol{x})^{1/2})\rho^{P^0}(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x})^{1/2}d\nu_{\mathcal{Y}}(\boldsymbol{y}),\\
&\qquad\le\frac{1}{4\nu_{\mathcal{Y}}(K_{\mathcal{Y}})^{1/2}}\varepsilon\int_{K_{\mathcal{Y}}}\rho^{P^0}(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x})^{1/2}d\nu_{\mathcal{Y}}(\boldsymbol{y}),\\
&\qquad\qquad\le\frac{1}{4\nu_{\mathcal{Y}}(K_{\mathcal{Y}})^{1/2}}\varepsilon\nu_{\mathcal{Y}}(K_{\mathcal{Y}})^{1/2}Q^P(K_{\mathcal{Y}})^{1/2}\le\frac{1}{4}\varepsilon.
\end{aligned}$$

The third sum is a.s. bounded by the same arguments. Hence,

$$
\begin{aligned}
&1 - \int_{\mathcal{Y}} \rho^{G^\omega}(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x})^{1/2}\rho^{P^0}(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x})^{1/2} d\nu_{\mathcal{Y}}(\boldsymbol{y}) \\
&< \frac{3}{4}\varepsilon + \sum_{k=1}^{N}\int_{A_k} (\rho^{P}(\boldsymbol{y}_k,\boldsymbol{\gamma}_\ell,\boldsymbol{x})^{1/2} - \rho^{G^\omega}(\boldsymbol{y}_k,\boldsymbol{\gamma}_\ell,\boldsymbol{x})^{1/2})\rho^{P^0}(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x})^{1/2} d\nu_{\mathcal{Y}}(\boldsymbol{y}).
\end{aligned}
$$

To prove the theorems, it remains to bound the integral in the right-hand side.

The hypotheses of Theorem 6.3 allow us to apply Theorem 4.5 to show that the event

$$
\begin{aligned}
\Bigg\{\omega \in \Omega : \left| \int_\Theta \psi(\boldsymbol{y}_k,\boldsymbol{\gamma}_\ell,\boldsymbol{\theta}) dG^\omega_{\boldsymbol{x}_i}(\boldsymbol{\theta}) - \int_\Theta \psi(\boldsymbol{y}_k,\boldsymbol{\gamma}_i,\boldsymbol{\theta}) dP^0_{\boldsymbol{x}_i}(\boldsymbol{\theta})\right| < \frac{\varepsilon}{4\nu_{\mathcal{Y}}(K_{\mathcal{Y}})^{1/2}}, \\
k \in [N],\ \ell \in [M],\ i \in [n]\Bigg\},
\end{aligned}
$$

has positive probability. This proves Theorem 6.3.

The hypotheses of Theorem 6.4 allow us to apply Theorem 4.8 to show that the event

$$
\begin{aligned}
\Bigg\{\omega \in \Omega\colon \sup_{\boldsymbol{x}\in K_{\mathcal{X}}} \left| \int_\Theta \psi(\boldsymbol{y}_k,\boldsymbol{\gamma}_\ell,\boldsymbol{\theta}) dG^\omega_{\boldsymbol{x}}(\boldsymbol{\theta}) - \int_\Theta \psi(\boldsymbol{y}_k,\boldsymbol{\gamma}_\ell,\boldsymbol{\theta}) dP^0_{\boldsymbol{x}}(\boldsymbol{\theta})\right| < \frac{\varepsilon}{8\nu_{\mathcal{Y}}(K_{\mathcal{Y}})^{1/2}}, \\
k \in [N],\ \ell \in [M]\Bigg\},
\end{aligned}
$$

has positive probability. This proves Theorem 6.3. □

A.6.2. Proof of Theorems 6.5 and 6.6

In the case of the product-L^∞ topology we can define the compact set $K_{\Gamma\times\mathcal{X}} \subset K_\Gamma \times K_{\mathcal{X}}$ as in Appendix A.6.1 and let $\varepsilon_0 < \min\{\varepsilon_1,\dots,\varepsilon_n\}$. As indicated before, over this finite set we can assume without loss that P^0 is weakly continuous. For the compact-L^∞ topology, we define $K_{\Gamma\times\mathcal{X}} = K_\Gamma \times K_{\mathcal{X}}$ and let $\varepsilon_0 < \varepsilon$.

This reduction allows us to consider the event

$$
\left\{\omega \in \Omega : \sup_{(\boldsymbol{\gamma},\boldsymbol{x})\in K_{\Gamma\times\mathcal{X}}} \|\rho^{G^\omega}_{\boldsymbol{\gamma},\boldsymbol{x}} - \rho^{P^0}_{\boldsymbol{\gamma},\boldsymbol{x}}\|_{L^\infty} < \varepsilon_0\right\}.
$$

Both Theorem 6.5 and 6.6 and follow is we show the above event has positive probability. Let $\varepsilon \in [0,1)$ be such that $\varepsilon < \varepsilon_0$.

Let $K_{\mathcal{Y}\times\Gamma} = \mathcal{Y} \times K_\Gamma$ which, by hypothesis, is compact. From Lemma A.2 there exists $\delta > 0$ such that for any $(\boldsymbol{y}',\boldsymbol{\gamma}'), (\boldsymbol{y},\boldsymbol{\gamma}) \in K_{\mathcal{Y}\times\Gamma}$ we have that

$$
\begin{aligned}
&d_{\mathcal{Y}\times\Gamma}((\boldsymbol{y}',\boldsymbol{y}),(\boldsymbol{\gamma}',\boldsymbol{\gamma})) < \delta \\
&\qquad \Rightarrow \quad \left| \int_\Theta \psi(\boldsymbol{y}',\boldsymbol{\gamma}',\boldsymbol{\theta}) dP_{\boldsymbol{x}}(\boldsymbol{\theta}) - \int_\Theta \psi(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}) dP_{\boldsymbol{x}}(\boldsymbol{\theta})\right| < \frac{1}{4}\varepsilon,
\end{aligned}
$$

uniformly over weakly continuous $P : \mathcal{X} \to \mathcal{P}(\Theta)$ and $\boldsymbol{x} \in \mathcal{X}$. Let $r > 0$ be such that $2r < \delta$. From the open cover $\{B((\boldsymbol{y}, \boldsymbol{\gamma}), r)\}_{(\boldsymbol{y},\boldsymbol{\gamma}) \in K_{\mathcal{Y}\times\Gamma}}$ we can extract a finite subcover $\{B((\boldsymbol{y}_k, \boldsymbol{\gamma}_k), r)\}_{k=1}^N$. Hence, for any $(\boldsymbol{y}, \boldsymbol{\gamma}) \in K_{\mathcal{Y}\times\Gamma}$ there exists $k \in [N]$ such that

$$\left| \int_\Theta \psi(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}) dP_{\boldsymbol{x}}(\boldsymbol{\theta}) \right| < \frac{1}{4}\varepsilon + \left| \int_\Theta \psi(\boldsymbol{y}_k, \boldsymbol{\gamma}_k, \boldsymbol{\theta}) dP_{\boldsymbol{x}}(\boldsymbol{\theta}) \right|,$$

uniformly over $P : \mathcal{X} \to \mathcal{P}(\Theta)$ weakly continuous and $\boldsymbol{x} \in \mathcal{X}$. In particular,

$$\begin{aligned} &\left| \int_\Theta \psi(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}) dG^\omega_{\boldsymbol{x}}(\boldsymbol{\theta}) - \int_\Theta \psi(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\theta}) dP_{\boldsymbol{x}}(\boldsymbol{\theta}) \right| \\ &\qquad \le \frac{1}{2}\varepsilon + \left| \int_\Theta \psi(\boldsymbol{y}_k, \boldsymbol{\gamma}_k, \boldsymbol{\theta}) dG^\omega_{\boldsymbol{x}}(\boldsymbol{\theta}) - \int_\Theta \psi(\boldsymbol{y}_k, \boldsymbol{\gamma}_k, \boldsymbol{\theta}) dP_{\boldsymbol{x}}(\boldsymbol{\theta}) \right|. \end{aligned}$$

The hypotheses of Theorem 6.5 and the fact that $\boldsymbol{\gamma}_k \in \{\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_n\}$ allow us to apply Theorem 4.5 to show that the event

$$\left\{ \omega \in \Omega: \left| \int_\Theta \psi(\boldsymbol{y}_k, \boldsymbol{\gamma}_i, \boldsymbol{\theta}) dG^\omega_{\boldsymbol{x}_i}(\boldsymbol{\theta}) - \int_\Theta \psi(\boldsymbol{y}_k, \boldsymbol{\gamma}_i, \boldsymbol{\theta}) dP^0_{\boldsymbol{x}_i}(\boldsymbol{\theta}) \right| < \frac{1}{2}\varepsilon, \ i \in [n], \, k \in [N] \right\},$$

has positive probability. This proves Theorem 6.5. The hypotheses of Theorem 6.6 allow us to apply Theorem 4.8 to show that the event

$$\begin{aligned} \Bigg\{ \omega \in \Omega : \sup_{\boldsymbol{x} \in K_{\mathcal{X}}} &\left| \int_\Theta \psi(\boldsymbol{y}_k, \boldsymbol{\gamma}_k, \boldsymbol{\theta}) dG^\omega_{\boldsymbol{x}}(\boldsymbol{\theta}) - \int_\Theta \psi(\boldsymbol{y}_k, \boldsymbol{\gamma}_k, \boldsymbol{\theta}) dP^0_{\boldsymbol{x}}(\boldsymbol{\theta}) \right| \\ &< \frac{1}{2}\varepsilon, \ k \in [N] \Bigg\}, \end{aligned}$$

has positive probability. This proves Theorem 6.6. □

A.6.3. Proof of Theorems 6.7 and 6.8

In the case of the product-KL topology we proceed as for the product-L^∞ topology. We define the compact sets $K_\Gamma := \{\boldsymbol{\gamma}_i : i \in [n]\}$ and $K_{\mathcal{X}} := \{\boldsymbol{x}_i : i \in [n]\}$. Furthermore, we let $K_{\Gamma\times\mathcal{X}} = K_\Gamma \times K_{\mathcal{X}}$ and we let $\varepsilon_0 < \min\{\varepsilon_1, \dots, \varepsilon_n\}$. As indicated before, over this finite set we can assume without loss that P^0 is weakly continuous. For the compact-KL topology, we define $K_{\Gamma\times\mathcal{X}} = K_\Gamma \times K_{\mathcal{X}}$ and let $\varepsilon_0 < \varepsilon$.

The hypothesis $\psi > 0$ implies that $\rho^{P^0} > 0$. Furthermore, since $\mathcal{Y} \times K_\Gamma \times K_{\mathcal{X}}$ is compact, there exists $c_{\max} > 0$ such that $\rho^{P^0} \le c_{\max}$. Let $\varepsilon' > 0$ be such that $\varepsilon' < \varepsilon_0/(1 + \varepsilon_0)$ and consider the event

$$\left\{ \omega \in \Omega : \sup_{(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{x}) \in K_{\mathcal{Y}} \times K_\Gamma \times K_{\mathcal{X}}} |\rho^{G^\omega}(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{x}) - \rho^{P^0}(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{x})| < \varepsilon' c_{\max} \right\}.$$

The hypotheses allow us to apply Theorem 6.5 or Theorem 6.6 respectively to prove this event has positive probability. Furthermore, on this event we have that

$$\left| \frac{\rho^{G^\omega}(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{x})}{\rho^{P^0}(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{x})} - 1 \right| < \varepsilon'.$$

Since for $t > -1$ we have that

$$\frac{t}{1+t} \leq \log(1+t) \leq t,$$

we deduce that

$$\left| \log \left(\frac{\rho^{P^0}(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{x})}{\rho^{G^\omega}(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{x})} \right) \right| < \frac{\varepsilon'}{1-\varepsilon'} < \varepsilon_0.$$

In particular,

$$\mathrm{KL}(q^{P^0} \,\|\, \rho^{G^\omega}) = \int_{\mathcal{Y}} \rho^{P^0}(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{x}) \log \left(\frac{\rho^{P^0}(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{x})}{\rho^{G^\omega}(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{x})} \right) d\nu_{\mathcal{Y}}(\boldsymbol{y}) < \varepsilon_0.$$

Consequently, the event

$$\left\{ \omega \in \Omega : \sup_{(\boldsymbol{\gamma}, \boldsymbol{x}) \in K_\Gamma \times K_{\mathcal{X}}} \mathrm{KL}(\rho^{P^0}_{\boldsymbol{\gamma}, \boldsymbol{x}} \,\|\, \rho^{G^\omega}_{\boldsymbol{\gamma}, \boldsymbol{x}}) < \varepsilon_0 \right\},$$

has positive probability, proving the theorem. □

Acknowledgments

The authors would like to thank the anonymous referees for their constructive comments that improved the quality of this paper. The authors are grateful for helpful discussions with Prof. Judith Rousseau.

Funding

A. Iturriaga was supported by "Becas de Doctorado Nacional de CONICYT" and by Agencia Nacional de Investigación y Desarrollo (ANID) through the Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) grant No 1200221.

C. A. Sing Long was supported by Agencia Nacional de Investigación y Desarrollo (ANID) through the Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) grant No 1211643 and through grant NCN17 059 from Millennium Science Initiative Program, Millennium Nucleus Center for the Discovery of Structures in Complex Data (MIDAS).

A. Jara's work was supported by Agencia Nacional de Investigación y Desarrollo (ANID) through the Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) grant No 1220907 and through grant NCN17 059 from Millennium Science Initiative Program, Millennium Nucleus Center for the Discovery of Structures in Complex Data (MIDAS).

References

[1] ALONSO-PENA, M. and CRUJEIRAS, R. M. (2023). Analyzing animal escape data with circular nonparametric multimodal regression. *The Annals of Applied Statistics* **17** 130 – 152. https://doi.org/10.1214/22-AOAS1619. MR4539025

[2] AMEIJEIRAS-ALONSO, J., LAGONA, F., RANALLI, M. and CRUJEIRAS, R. M. (2019). A circular nonhomogeneous hidden Markov field for the spatial segmentation of wildfire occurrences. *Environmetrics* **30** e2501. https://doi.org/10.1002/env.2501. MR3919903

[3] BALL, F. G., DRYDEN, I. L. and GOLALIZADEH, M. (2008). Brownian motion and Ornstein-Uhlenbeck processes in planar shape space. *Methodology and Computing in Applied Probability* **10** 1–22. https://doi.org/10.1007/s11009-007-9042-6. MR2394032

[4] BARNDORFF-NIELSEN, O. (1973). On M-ancillarity. *Biometrika* **60** 447–455. https://doi.org/10.2307/2334993. MR0345255

[5] BARNDORFF-NIELSEN, O. (1978). *Information and exponential families in statistical theory*. John Wiley & Sons, Ltd., Chichester Wiley Series in Probability and Mathematical Statistics. https://doi.org/10.1002/9781118857281. MR0489333

[6] BARRIENTOS, A. F., JARA, A. and QUINTANA, F. A. (2012). On the support of MacEachern's dependent Drichlet processes and extensions. *Bayesian Analysis* **7** 277– 310. https://doi.org/10.1214/12-BA709. MR2934952

[7] BARRIENTOS, A. F., JARA, A. and QUINTANA, F. A. (2017). Fully nonparametric regression for bounded data using dependent Bernstein polynomials. *Journal of the American Statistical Association* **112** 806–825. https://doi.org/10.1080/01621459.2016.1180987. MR3671772

[8] BAYKARA, M., BAYKARA, S. and ATMACA, M. (2022). Statistical shape analysis of corpus callosum in functional neurological disorder. *Neuropsychiatria i Neuropsychologia* **17** 1-8. https://doi.org/10.1016/j.pscychresns.2023.111695.

[9] BAYKARA, S., BAYKARA, M. and ATMACA, M. (2023). Statistical Shape Analysis of Corpus Callosum in Vaginismus. *Archives of Psychiatry and Psychotherapy* **25** 37-44. https://doi.org/10.12740/APP/155218.

[10] BAYKARA, S., BAYKARA, M. and ATMACA, M. (2023). Regular Islamic prayers have different corpus callosum: a shape analysis study. *Egyptian Journal of Neurology Psychiatry and Neurosurgery* **59**. https://doi.org/10.1186/s41983-023-00683-x.

[11] BHATTACHARYA, A. and DUNSON, D. B. (2010). Nonparametric Bayesian density estimation on manifolds with applications to planar shapes. *Biometrika* **97** 851–865. https://doi.org/10.1093/biomet/asq044. MR2746156

[12] BHATTACHARYA, A. and DUNSON, D. B. (2012). Strong consistency of nonparametric Bayes density estimation on compact metric spaces with applications to specific manifolds. *Annals of the Institute of Statistical Math-*

ematics **64** 687–714. https://doi.org/10.1007/s10463-011-0341-x. MR2927766

[13] Bhattacharya, A. and Dunson, D. B. (2012). Nonparametric Bayes classification and hypothesis testing on manifolds. *Journal of Multivariate Analysis* **111** 1–19. https://doi.org/10.1016/j.jmva.2012.02.020. MR2944402

[14] Bhattacharya, R. and Patrangenaru, V. (2002). Nonparametric estimation of location and dispersion on Riemannian manifolds. *Journal of Statistical Planning and Inference* **108** 23–35. https://doi.org/10.1016/S0378-3758(02)00268-9. MR1947389

[15] Bhattacharya, R. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *The Annals of Statistics* **31** 1–29. https://doi.org/10.1214/aos/1046294456. MR1962498

[16] Bhattacharya, R. and Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *The Annals of statistics* **33** 1225–1259. https://doi.org/10.1214/009053605000000093. MR2195634

[17] Billera, L. J., Holmes, S. P. and Vogtmann, K. (2001). Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics* **27** 733–767. https://doi.org/10.1006/aama.2001.0759. MR1867931

[18] Castillo, I., Kerkyacharian, G. and Picard, D. (2014). Thomas Bayes' walk on manifolds. *Probability Theory and Related Fields* **158** 665–710. https://doi.org/10.1007/s00440-013-0493-0. MR3176362

[19] Choudhuri, N., Ghosal, S. and Roy, A. (2007). Nonparametric binary regression using a Gaussian process prior. *Statistical Methodology* **4** 227–243. https://doi.org/10.1016/j.stamet.2006.07.003. MR2368147

[20] Christmann, A. and Steinwart, I. (2010). Universal Kernels on Non-Standard Input Spaces. In *Advances in Neural Information Processing Systems* (J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel and A. Culotta, eds.) **23**. Curran Associates, Inc.

[21] Chung, Y. and Dunson, D. B. (2009). Nonparametric Bayes conditional distribution modeling with variable selection. *Journal of the American Statistical Association* **104** 1646–1660. https://doi.org/10.1198/jasa.2009.tm08302. MR2750582

[22] Cifarelli, D. and Regazzini, E. (1978). Problemi statistici non parametrici in condizioni di scambialbilita parziale e impiego di medie associative Technical Report, Quaderni Istituto Matematica Finanziaria, Torino.

[23] Claude, J. (2008). *Morphometrics with R. Use R!* Springer New York. https://doi.org/10.1007/978-0-387-77789-4.

[24] Cohn, D. L. (2013). *Measure Theory*, 2nd ed. Springer New York. https://doi.org/10.1007/978-1-4614-6956-8. MR3098996

[25] De Iorio, M., Johnson, W. O., Müller, P. and Rosner, G. L. (2009). Bayesian nonparametric non-proportional hazards survival modelling. *Biometrics* **65** 762–771. https://doi.org/10.1111/j.1541-0420.

2008.01166.x. MR2649849

[26] 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. https://doi.org/10.1198/016214504000000205. MR2054299

[27] Dryden, I. L. and Mardia, K. V. (2016). *Statistical Shape Analysis: With Applications in R*. *Wiley Series in Probability and Statistics*. Wiley. https://doi.org/10.1002/9781119072492. MR3559734

[28] Dunson, D. B. and Park, J. H. (2008). Kernel stick-breaking processes. *Biometrika* **95** 307–323. https://doi.org/10.1093/biomet/asn012. MR2521586

[29] Dunson, D. B., Pillai, N. and Park, J. H. (2007). Bayesian density regression. *Journal of the Royal Statistical Society, Series B* **69** 163–183. https://doi.org/10.1111/j.1467-9868.2007.00582.x. MR2325270

[30] Epifani, I. and Lijoi, A. (2010). Nonparametric priors for vectors of survival functions. *Statistica Sinica* **20** 1455–1484. MR2777332

[31] Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1** 209–230. https://doi.org/10.1214/aos/1176342360. MR0350949

[32] Ferguson, T. S. (1974). Prior distribution on the spaces of probability measures. *The Annals of Statistics* **2** 615–629. https://doi.org/10.1214/aos/1176342752. MR0438568

[33] Fisher, N. I. (1995). *Statistical analysis of circular data*. Cambridge University Press. https://doi.org/10.1017/CBO9780511564345. MR1251957

[34] Fisher, N. I. and Lee, A. J. (1992). Regression models for an angular response. *Biometrics* **48** 665–677. https://doi.org/10.2307/2532334. MR1187598

[35] Fletcher, P. T. (2013). Geodesic Regression and the Theory of Least Squares on Riemannian Manifolds. *International Journal of Computer Vision* **105** 171-185. https://doi.org/10.1007/s11263-012-0591-y. MR3104017

[36] Florens, J.-P., Mouchart, M. and Rolin, J.-M. (1990). *Elements of Bayesian statistics*. *Monographs and Textbooks in Pure and Applied Mathematics* **134**. Marcel Dekker, Inc., New York. https://doi.org/10.1201/9780203758526. MR1051656

[37] Fréchet, M. (1948). Les élements aléatoires de nature quelconque dans un espace distancié. *Annales de l'institut Henri Poincaré* **10** 215–310. MR0027464

[38] Golalizadeh, M. (2010). A useful family of stochastic processes for modeling shape diffusions. *Journal of Statistical Research of Iran* **7** 21–36.

[39] Guerrero, P. and Ruiz del Solar, J. (2014). Circular Regression Based on Gaussian Processes. In *2014 22nd International Conference on Pattern Recognition* 3672-3677. https://doi.org/10.1109/ICPR.2014.631.

[40] Gutierrez, L., Gutierrez-Pena, E. and Mena, R. H. (2019). A Bayesian Approach to Statistical Shape Analysis via the Projected Nor-

mal Distribution. *Bayesian Analysis* **14** 427-447. https://doi.org/10.1214/18-BA1113. MR3934092

[41] Hsu, E. P. (2002). *Stochastic Analysis on Manifolds*. *Contemporary Mathematics*. American Mathematical Society.

[42] Iñacio, V., Jara, A., Hanson, T. E. and de Carvalho, M. (2013). Bayesian nonparametric ROC regression modeling. *Bayesian Analysis* **8** 623–646. https://doi.org/10.1214/13-BA825. MR3102228

[43] Jara, A. and Hanson, T. (2011). A class of mixtures of dependent tail-free processes. *Biometrika* **98** 553–566. https://doi.org/10.1093/biomet/asq082. MR2836406

[44] Jara, A., Lesaffre, E., De Iorio, M. and Quintana, F. A. (2010). Bayesian semiparametric inference for multivariate doubly-interval-censored data. *The Annals of Applied Statistics* **4** 2126–2149. https://doi.org/10.1214/10-AOAS368. MR2829950

[45] Jona-Lasinio, G., Gelfand, A. and Jona-Lasinio, M. (2012). Spatial analysis of wave direction data using wrapped Gaussian processes. *The Annals of Applied Statistics* **6** 1478 – 1498. https://doi.org/10.1214/12-AOAS576. MR3058672

[46] Kallenberg, O. (1997). *Foundations of Modern Probability*. *Probability and its Applications A Series of the Applied Probability Trust*. Springer. https://doi.org/10.1007/978-3-030-61871-1. MR1464694

[47] Kelley, J. L. (1975). *General Topology*. Springer New York. MR0370454

[48] Kendall, D. G. (1977). The diffusion of shape. *Advances in Applied Probability* **9** 428–430. https://doi.org/10.2307/1426091.

[49] Kendall, D. G. (1984). Shape manifolds, Procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society* **16** 81–121. https://doi.org/10.1112/blms/16.2.81. MR0737237

[50] Kendall, W. S. (1988). Symbolic computation and the diffusion of shapes of triads. *Advances in Applied Probability* **20** 775–797. https://doi.org/10.2307/1427360. MR0967998

[51] Kendall, W. S. (1990). The diffusion of Euclidean shape. In *In Disorder in Physical Systems* 428–430. Oxford University Press. MR1064562

[52] Kendall, W. S. (1998). A diffusion model for Bookstein triangle shape. *Advances in Applied Probability* **30** 317–334. https://doi.org/10.1239/aap/1035228071. MR1642841

[53] Khoshnevisan, D. (2002). *Multiparameter Processes: An Introduction to Random Fields*. *Monographs in Mathematics*. Springer. https://doi.org/10.1007/b97363. MR1914748

[54] Le, H. (1991). A stochastic calculus approach to the shape distribution induced by a complex normal model. *Mathematical Proceedings of the Cambridge Philosophical Society* **109** 221–228. https://doi.org/10.1017/S0305004100069681. MR1075133

[55] Le, H. (1994). Brownian motions on shape and size-and-shape spaces. *Journal of Applied Probability* **31** 101–113. https://doi.org/10.2307/3215238. MR1260574

[56] Leisen, F. and Lijoi, A. (2011). Vectors of two–parameter Poisson–

Dirichlet processes. *Journal of Multivariate Analysis* **102** 482–495. https://doi.org/10.1016/j.jmva.2010.10.008. MR2755010

[57] Lijoi, A., Nipoti, B. and Prünster, I. (2014). Bayesian inference with dependent normalized completely random measures. *Bernoulli* **20** 1260–1291. https://doi.org/10.3150/13-BEJ521. MR3217444

[58] Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates I: Density estimates. *The Annals of Statistics* **12** 351–357. https://doi.org/10.1214/aos/1176346412. MR0733519

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

[60] MacEachern, S. N. (2000). Dependent Dirichlet processes Technical Report, Department of Statistics, The Ohio State University.

[61] Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*. *Wiley Series in Probability and Statistics*. Wiley. https://doi.org/10.1002/9780470316979. MR1828667

[62] Miccheli, C. A., Xu, Y. and Zhang, H. (2006). Universal Kernels. *The Journal of Machine Learning Research* **7** 2651-2667. MR2274454

[63] Mooney, J. A., Helms, P. J. and Jolliffe, I. T. (2003). Fitting mixtures of von Mises distributions: a case study involving sudden infant death syndrome. *Computational Statistics and Data Analysis* **41** 505 – 513. https://doi.org/10.1016/S0167-9473(02)00181-0. MR1973727

[64] Müller, P., Erkanli, A. and West, M. (1996). Bayesian curve fitting using multivariate normal mixtures. *Biometrika* **83** 67–79. https://doi.org/10.1093/biomet/83.1.67. MR1399156

[65] Müller, P. and Quintana, F. A. (2010). Random partition models with regression on covariates. *Journal of Statistical Planning and Inference* **140** 2801–2808. https://doi.org/10.1016/j.jspi.2010.03.002. MR2651966

[66] Müller, P., Quintana, F. A., Jara, A. and E, H. T. (2015). *Bayesian Nonparametric Data Analysis*. Springer, New York, USA. https://doi.org/10.1007/978-3-319-18968-0. MR3738474

[67] Müller, P., Quintana, F. A. and Rosner, G. (2004). A method for combining inference across related nonparametric Bayesian models. *Journal of the Royal Statistical Society, Series B* **66** 735–749. https://doi.org/10.1111/j.1467-9868.2004.05564.x. MR2088779

[68] Müller, P., Quintana, F. A. and Rosner, G. L. (2011). A product partition model with regression on covariates. *Journal of Computational and Graphical Statistics* **20** 260–278. https://doi.org/10.1198/jcgs.2011.09066. MR2816548

[69] Neveu, J. (1965). *Mathematical Foundations of the Calculus of Probability*. Holden-Day. MR0198505

[70] Pelletier, B. (2005). Kernel density estimation on Riemannian manifolds. *Statistics and Probability Letters* **73** 297–304. https://doi.org/10.1016/j.spl.2005.04.004. MR2179289

[71] Pelletier, B. (2006). Non-parametric regression estimation on closed

Riemannian manifolds. *Journal of Nonparametric Statistics* **18** 57–67. https://doi.org/10.1080/10485250500504828. MR2214065

[72] Preston, C. (1972). Continuity Properties of Some Gaussian Processes. *The Annals of Mathematical Statistics* **43** 285-292. https://doi.org/10.1214/aoms/1177692721. MR0307316

[73] Quintana, F., Müller, P., Jara, A. and MacEachern, S. (2022). The Dependent Dirichlet Process and Related Models. *Statistical Science* **37** 24–41. https://doi.org/10.1214/20-STS819. MR4371095

[74] Quintana-Orti, G. and Simo, A. (2019). A Kernel Regression Procedure in the 3D Shape Space with an Application to Online Sales of Children's Wear. *Statistical Science* **34** 236-252. https://doi.org/10.1214/18-STS675. MR3983327

[75] Reed, M. C. and Simon, B. (1980). *Methods of Modern Mathematical Physics I: Functional Analysis*, Revised edition ed. Academic Press. MR0493421

[76] Rodriguez, A. and Dunson, D. B. (2011). Nonparametric Bayesian models through probit stick-breaking processes. *Bayesian Analysis* **6** 145–178. https://doi.org/10.1214/11-BA605. MR2781811

[77] Schoenberg, I. J. (1938). Metric spaces and positive definite functions. *Transactions of the American Mathematical Society* **44** 522–536. https://doi.org/10.2307/1989894. MR1501980

[78] Schwartz, L. (1965). On Bayes procedures. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **4** 10–26. MR0184378

[79] Sethuraman, J. (1994). A constructive definition of Dirichlet prior. *Statistica Sinica* **2** 639–650. MR1309433

[80] Steinwart, I. (2024). Reproducing kernel Hilbert spaces cannot contain all continuous functions on a compact metric space. *Archiv der Mathematik* **122** 553-557. https://doi.org/10.1007/s00013-024-01976-0. MR4734568

[81] Tokdar, S. T., Zhu, Y. M. and Ghosh, J. K. (2010). Bayesian density regression with logistic Gaussian process and subspace projection. *Bayesian Analysis* **5** 1–26. https://doi.org/10.1214/10-BA605. MR2719655

[82] Wehrhahn, C., Barrientos, A. F. and Jara, A. (2022). Dependent Bayesian nonparametric modeling of compositional data using random Bernstein polynomials. *Electronic Journal of Statistics* **16** 2346 – 2405. https://doi.org/10.1214/22-EJS2002. MR4402566

[83] Xuanxuan Sheng, D. X. and Ying, S. (2022). Intrinsic semi-parametric regression model on Grassmannian manifolds with applications. *Communications in Statistics – Simulation and Computation* **0** 1–20. https://doi.org/10.1080/03610918.2022.2112961.