Electronic Journal of Statistics
Vol. 16 (2022) 2346–2405
ISSN: 1935-7524
https://doi.org/10.1214/22-EJS2002

Dependent Bayesian nonparametric modeling of compositional data using random Bernstein polynomials*

Claudia Wehrhahn

Department of Statistics, University of California Santa Cruz, Santa Cruz, U.S.A.
e-mail: cwehrhah@ucsc.edu

Andrés F. Barrientos

Department of Statistics, Florida State University, Florida, U.S.A.
e-mail: abarrientos@fsu.edu

and

Alejandro Jara

MiDaS-Center for the Discovery of Structures in Complex Data and Department of Statistics, Pontificia Universidad Católica de Chile, Santiago, Chile
e-mail: atjara@uc.cl

Abstract: We discuss Bayesian nonparametric procedures for the regression analysis of compositional responses, that is, data supported on a multivariate simplex. The procedures are based on a modified class of multivariate Bernstein polynomials and on the use of dependent stick-breaking processes. A general model and two simplified versions of the general model are discussed. Appealing theoretical properties such as continuity, association structure, support, and consistency of the posterior distribution are established. Additionally, we exploit the use of spike-and-slab priors for choosing the version of the model that best adapts to the complexity of the underlying true data-generating distribution. The performance of the proposed model is illustrated in a simulation study and in an application to solid waste data from Colombia.

Keywords and phrases: Fully nonparametric regression, density regression, Dirichlet process, dependent Dirichlet processes.

Received August 2021.

1. Introduction

The statistical modeling of compositional data plays a key role in many scientific areas, economics, political sciences, and engineering, among others, with

*C. Wehrhahn's research was supported by the "Programa de Becas de Postgrado de Chile, CONICYT", NSF-DMS 1738053 and ATD-DMS 1441433. A. Jara's work was supported by the 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).

applications ranging from human microbiome analyses to topic modeling in large text corpora, where the focus is on the modeling of vectors containing information on the relative frequencies in which the different components occur. The need for appropriate methodologies for the analysis of these type of data can be originated by the nature of the scientific questions or limitations associated with the measurement methods. For instance, recent advances in biological high-throughput sequencing technologies can only provide relative abundance information because they can only capture a limited total number of transcripts or sequences in a sample or do not control for the total number of microbes entering the measurement process. Thus, the resulting sequencing count data carry only relative abundance information about the different transcripts or taxa in a given sample. On the other hand, in our motivating problem, we are interested in describing and understanding the solid waste composition generated in a residential area of a city in Colombia.

From a mathematical point of view, compositional data can be defined as multivariate data supported on the m–dimensional simplex, Δ_m, given by

$$\Delta_m \quad = \left\{(y_1, \ldots, y_m) \in [0,1]^m : \textstyle\sum_{i=1}^m y_i \leq 1\right\}.$$

Since Aitchison (1982), several parametric regression models for compositional responses have been proposed. Common approaches transform the compositional responses from Δ_m to $\mathbb{R}^m$, and use the well known and familiar battery of statistical models for normally distributed responses (see, e.g., Aitchison, 1982; Atchison & Shen, 1980; Shimizu et al., 2021; Wang et al., 2010). Other proposals use the Dirichlet distribution to model the compositional responses and link the Dirichlet parameters to covariates (see, e.g., Gueorguieva et al., 2008; Hijazi, 2003; Hijazi & Jernigan, 2009; Van der Merwe, 2019). These models can be easily extended to allow for non-parametric functional forms in the relationship between the model parameters and the predictors (see, e.g., Di Marzio et al., 2015; Tsagris et al., 2020). However, they rely on particular parametric distributional forms which limits the type of inferences that can be obtained. Modeling approaches where the complete distribution of the compositional responses can flexibly vary as a function of the predictors are scarce in the literature. We aim to fill this gap by proposing a class of Bayesian nonparametric (BNP) predictor-dependent mixture models that enjoys appealing theoretical properties and is easy to use.

Most BNP approaches for collections of predictor-dependent probability distributions employ mixtures of densities from parametric families (see, e.g., Müller et al., 2015, and references therein). Mixture models are convenient for density estimation because they induce a prior distribution on densities by placing a prior distribution on the mixing measure. Dependent Dirichlet processes (MacEachern, 1999, 2000; Quintana et al., 2022) are often used as priors for the mixing distributions. Other extensions and alternative constructions for dealing with predictor-dependent probability distributions include the ordered-category probit regression model (Karabatsos & Walker, 2012), the dependent beta process (Trippa et al., 2011), the dependent tail-free processes (Jara &

Hanson, 2011), the dependent neutral to the right processes and correlated two-parameter Poisson-Dirichlet processes (Epifani & Lijoi, 2010; Leisen & Lijoi, 2011), and the general class of dependent normalized completely random measures (Lijoi et al., 2014). Due to their flexibility and ease in computation, these models are routinely implemented in a wide variety of applications (see, e.g., Müller et al., 2015, and references therein).

BNP approaches for collections of predictor-dependent probability distributions have mainly focused on responses defined on the real line. Although those approaches can be applied to compositional responses, by transforming the responses from Δ_m to $\mathbb{R}^m$, i) the resulting density in the simplex could not be well defined at the edges or ii) the resulting density in the simplex could be equal to zero at the edges. This can cause important problems if zeros are observed in the data because either the likelihood could not be defined, if i) holds true, or the likelihood would always be equal to zero, if ii) holds true. Also, other problem associated with the use of transformations is that it is not very clear that the resulting density is flexible at the edges of the simplex (please see Appendix A for more details on this).

We propose modeling compositional responses using a particular class of mixtures of Dirichlet probability density functions that naturally emerges from the theoretical properties and extensions of Bernstein polynomials (BP). Motivated by their uniform approximation properties, frequentist and Bayesian methods based on univariate BP have been proposed for the estimation of probability distributions supported on bounded intervals, unit hyper-cubes, and simplex spaces (see, e.g. Petrone, 1999a,b; Petrone & Wasserman, 2002; Tenbusch, 1994; Ouimet, 2021; Babu & Chaubey, 2006; Zheng et al., 2010). For example, Babu & Chaubey (2006) studied a general multivariate version of the bivariate estimator proposed by Tenbusch (1994), while Zheng et al. (2010) constructed a multivariate Bernstein polynomial (MBP) prior for the spectral density of a random field. Key for our approach, Tenbusch (1994) considered multivariate extensions of Bernstein polynomials defined on Δ_2 to propose and study a density estimator. Tenbusch's approach is easy to extend to the m-dimensional case. The approach is based on the class of MBP associated with G, a cumulative distribution function (CDF) on Δ_m, given by

$$\widetilde{B}_{k,G}(\boldsymbol{y}) \quad = \quad \sum_{\boldsymbol{j} \in \mathscr{J}_{k,m}} G\left(\frac{j_1}{k}, \ldots, \frac{j_m}{k}\right) \mathrm{Mult}\left(\mathbf{j} \mid k, \boldsymbol{y}\right), \; \boldsymbol{y} \in \Delta_m, \tag{1}$$

where $k \in \mathbb{N}$ is the degree of the MBP, $\boldsymbol{j} = (j_1, \ldots, j_m)$,

$$\mathscr{J}_{k,m} = \left\{ (j_1, \ldots, j_m) \in \{0, \ldots, k\}^m : \sum_{l=1}^{m} j_l \leq k \right\},$$

and $\mathrm{Mult}\left(\cdot \mid k, \boldsymbol{y}\right)$ stands for the probability mass function of a multinomial distribution with parameters $(k, \boldsymbol{y})$.

Tenbusch's estimator arises by replacing G in (1) by the empirical CDF of observed data and it is not difficult to show that, if G is a CDF on Δ_m, then

$\widetilde{B}_{k,G}(\cdot)$ is not a CDF on Δ_m for a finite k. In this case, $\widetilde{B}_{k,G}(\cdot)$ can be expressed as a linear combination of CDFs of probability measures defined on Δ_m, where the coefficients are nonnegative but do not add up to 1. Tenbusch's estimator is defined as the derivative of $\widetilde{B}_{k,G}(\cdot)$ and, although it is consistent and optimal at the interior points of the simplex, it is not a valid density function for finite k and finite sample size. To avoid this problem, Barrientos et al. (2015) proposed a modified class of MBP by changing the set $\mathscr{J}_{k,m}$. The class retains the well known approximation properties of the original version. Furthermore, when G is a CDF on Δ_m, the modified MBP is a genuine CDF with density function defined by a mixture of Dirichlet densities.

We propose a class of fully nonparametric regression models for compositional responses, by extending the class of MBP priors of Barrientos et al. (2015). The extension relies on predictor-dependent stick-breaking processes (see, Barrientos et al., 2017, for a similar extension for responses defined on the unit-interval). An important property of the considered model class is that the densities are well-defined in scenarios with compositional data containing zero values. When the response vector contains zero values, either models based on the Dirichlet distribution without restrictions on the parameter space or approaches based on the log-ratio transformation are not properly defined and cannot be employed unless a zero value imputation is applied first (please see Appendix A). We study theoretical properties of the proposed model class, such as continuity, association structure, support, and consistency of the posterior distribution. These properties are non-trivial extensions of the results obtained by Barrientos et al. (2017) for the unit-interval and their proofs are provided in the Appendix. The use of the dependent stick-breaking process raises the question of where to introduce the predictor dependency: on weights, atoms, or both. Each selection leading to a different version of the model. Rather than fitting all versions of the model, as done by Barrientos et al. (2017), we use spike-and-slab mixtures (George & McCulloch, 1993) to define a prior that automatically chooses the version of the model that best accommodates to the complexity of the underlying data-generating mechanism. We evaluate the performance of the proposed approach using simulated data. The proposed approach is also applied for the analysis of solid waste data from Colombia.

The rest of the paper is organized as follows. The modified class of MBP and its main properties are summarized in Section 2. The proposed model class and its theoretical properties are discussed in Sections 3 and 4, respectively. Section 5 describes the main computational aspects. Section 6 illustrates the performance of the model using simulated data and in an application to solid waste in Colombia. A final discussion concludes the article.

2. Random multivariate Bernstein polynomials

Based on Tenbusch's MBP, Barrientos et al. (2015) defined a modified class of MBP on the $m-$dimensional simplex and proposed a BNP density estimation model for compositional data. The modified class increases the domain of function G and the size of the set $\mathscr{J}_{k,m}$ from the original class of MBP on the $m-$dimensional simplex provided in Equation (1). For a given function $G : \mathbb{R}^m \longrightarrow \mathbb{R}$, the associated modified class of MBP of degree $k \in \mathbb{N}$ on Δ_m is given by

$$B(\boldsymbol{y} \mid k, G) = \sum_{\boldsymbol{j} \in \mathscr{H}_{k,m}} G\left(\frac{j_1}{k}, \ldots, \frac{j_m}{k}\right) \mathrm{Mult}\left(\mathbf{j} \mid k+m-1, \boldsymbol{y}\right), \, \boldsymbol{y} \in \Delta_m,$$

where $\mathscr{H}_{k,m} = \{(j_1, \ldots, j_m) \in \{0, \ldots, k\}^m : \; \sum_{l=1}^m j_l \leq k+m-1\}$.

As shown by Barrientos et al. (2015), this class of MBP retains the appealing approximation properties of univariate BP and the standard class of MBP given in Equation (1). Specifically, if G is a real-valued function defined on $\mathbb{R}^m$ and $G|_{\Delta_m}$ is its restriction on Δ_m, then $B(\cdot \mid k, G)$ converges pointwise to $G|_{\Delta_m}$, as k goes to infinity, and the relation holds uniformly on Δ_m if $G|_{\Delta_m}$ is a continuous function.

It is also possible to show that if G is the CDF of a probability measure defined on Δ_m, then $B(\cdot \mid k, G)$ is also the restriction of the CDF of a probability measure defined on Δ_m. Furthermore, if G is the CDF of a probability measure defined on $\Delta_m^0 = \{\boldsymbol{y} \in \Delta_m : y_j > 0, j = 1, \ldots, m\}$, then $B(\cdot \mid k, G)$ is the restriction of the CDF of a probability measure with density function given by the following mixture of Dirichlet distributions,

$$b(\boldsymbol{y} \mid k, G) = \sum_{\boldsymbol{j} \in \mathscr{H}_{k,m}^0} G\left(\left(\frac{j_1 - 1}{k}, \frac{j_1}{k}\right] \times \ldots \times \left(\frac{j_m - 1}{k}, \frac{j_m}{k}\right]\right) \mathrm{dir}(\boldsymbol{y} \mid \alpha\left(k, \boldsymbol{j}\right)), \quad (2)$$

where $\mathscr{H}_{k,m}^0 = \{(j_1, \ldots, j_m) \in \{1, \ldots, k\}^m : \sum_{l=1}^m j_l \leq k+m-1\}$, $\alpha\left(k, \boldsymbol{j}\right) = (\boldsymbol{j}, \; k+m-\|\mathbf{j}\|_1)$, $\|\cdot\|_1$ denotes the l_1-norm, and $\mathrm{dir}(\cdot \mid (\alpha_1, \ldots, \alpha_{m+1}))$ denotes the density function of an m–dimensional Dirichlet distribution with parameters $(\alpha_1, \ldots, \alpha_{m+1})$.

By considering the density function given by Equation (2), a random function G, and a random degree k, Barrientos et al. (2015) defined a BNP prior for densities defined on Δ_m. The model corresponds to a DP mixture model of specific Dirichlet densities given by

$$\begin{aligned} b(\boldsymbol{y} \mid k, G) &= \int_{\Delta_m} \mathrm{dir}\left(\boldsymbol{y} \mid \alpha(k, \lceil k\boldsymbol{\theta} \rceil)\right) G(d\boldsymbol{\theta}), \\ G \mid M, G_0 &\sim DP(M, G_0), \\ k \mid \lambda &\sim p(\cdot \mid \lambda), \end{aligned} \quad (3)$$

where $DP(\alpha, G_0)$ denotes a Dirichlet process with concentration parameter $M > 0$ and base distribution G_0 on Δ_m^0, $p(\cdot \mid \lambda)$ is the probability mass function of a distribution on $\mathbb{N}$ parameterized by λ, and $\lceil \cdot \rceil$ denotes the ceiling function.

3. The model

Suppose that we observe regression data $\{(\boldsymbol{y}_i, \boldsymbol{x}_i) : i = 1, \ldots, n\}$, where $\boldsymbol{y}_i$ is a continuous Δ_m-valued outcome vector and $\boldsymbol{x}_i \in \mathscr{X} \subseteq \mathbb{R}^p$ is a p-dimensional vector of exogenous predictors. We define the regression model for compositional responses by introducing predictor-dependency in the mixture model given in (3), which allows the complete shape of the conditional densities to flexibly vary with values of $\boldsymbol{x}$. To this end, we replace the mixing measure G by a predictor-dependent mixing measure $G_{\boldsymbol{x}}$. Under this approach, the random conditional densities are given by

$$f_{\boldsymbol{x}}(\boldsymbol{y} \mid k, G_{\boldsymbol{x}}) = \int_{\Delta_m} \mathrm{dir}\left(\boldsymbol{y} \mid \alpha(k, \lceil k\boldsymbol{\theta} \rceil)\right) G_{\boldsymbol{x}}(d\boldsymbol{\theta}), \tag{4}$$

where the set of mixing distributions $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$ follows a dependent stick-breaking process, with elements of the form $G_{\boldsymbol{x}}(\cdot) = \sum_{j=1}^{\infty} w_j(\boldsymbol{x}) \delta_{\boldsymbol{\theta}_j(\boldsymbol{x})}(\cdot)$, where $w_j(\boldsymbol{x}) = V_j(\boldsymbol{x}) \prod_{l<j}[1 - V_l(\boldsymbol{x})]$, and where $V_j(\boldsymbol{x})$ and $\boldsymbol{\theta}_j(\boldsymbol{x})$ are transformations of underlying stochastic processes.

3.1. The formal definition

Let $\mathscr{V} = \{v_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$ and $\mathscr{H} = \{h_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$ be two sets of known bijective continuous functions, such that for every $\boldsymbol{x} \in \mathscr{X}$, $v_{\boldsymbol{x}} : \mathbb{R} \longrightarrow [0, 1]$ and $h_{\boldsymbol{x}} : \mathbb{R}^m \longrightarrow \Delta_m^0$, are such that for every $a \in \mathbb{R}$ and $\boldsymbol{b} \in \mathbb{R}^m$, $v_{\boldsymbol{x}}(a)$ and $h_{\boldsymbol{x}}(\boldsymbol{b})$ are continuous functions of $\boldsymbol{x}$. Let $\mathscr{P}(\Delta_m)$ be the set of all probability measures defined on Δ_m.

Definition 1. *Let $\mathscr{V}$ and $\mathscr{H}$ be two sets of functions as before. Let $F = \{F_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$ be a $\mathscr{P}(\Delta_m)$-valued stochastic process such that:*

(i) $\eta_j = \{\eta_j(\boldsymbol{x}) : \boldsymbol{x} \in \mathscr{X}\}$, $j \geq 1$, *are independent and identically distributed real-valued stochastic processes with law indexed by a finite-dimensional parameter* $\boldsymbol{\Psi}_\eta$.

(ii) $\boldsymbol{z}_j = \{\boldsymbol{z}_j(\boldsymbol{x}) : \boldsymbol{x} \in \mathscr{X}\}$, $j \geq 1$, *are independent and identically distributed real-valued stochastic processes with law indexed by a finite-dimensional parameter* $\boldsymbol{\Psi}_{\boldsymbol{z}}$.

(iii) $k \in \mathbb{N}$ *is a discrete random variable with distribution indexed by a finite-dimensional parameter* λ.

(iv) For every $\boldsymbol{x} \in \mathscr{X}$, *the density function of* $F_{\boldsymbol{x}}$, *w.r.t. Lebesgue measure, is given by the following dependent mixture of Dirichlet densities,*

$$f_{\boldsymbol{x}}(\cdot) = \sum_{j=1}^{\infty} w_j(\boldsymbol{x}) \mathrm{dir}\left(\cdot \mid \alpha(k, \lceil k\boldsymbol{\theta}_j(\boldsymbol{x}) \rceil)\right), \tag{5}$$

where $\boldsymbol{\theta}_j(\boldsymbol{x}) = h_{\boldsymbol{x}}(\boldsymbol{z}_j(\boldsymbol{x}))$, $\lceil k\boldsymbol{\theta}_j(\boldsymbol{x}) \rceil = (\lceil k\theta_{j1}(\boldsymbol{x}) \rceil, \ldots, \lceil k\theta_{jm}(\boldsymbol{x}) \rceil)$, *and* $w_j(\boldsymbol{x}) = V_j(\boldsymbol{x}) \prod_{l<j}[1 - V_l(\boldsymbol{x})]$, *with* $V_j(\boldsymbol{x}) = v_{\boldsymbol{x}}(\eta_j(\boldsymbol{x}))$.

The process $F = \{F_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$ *will be referred to as dependent MBP process with parameters* $(\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_{\boldsymbol{z}}, \mathscr{V}, \mathscr{H})$, *and denoted by* $\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_{\boldsymbol{z}}, \mathscr{V}, \mathscr{H})$ *and* DMBPP *for short.*

In the search of parsimonious models, it is of interest to study two special cases of the general construction given by Definition 1. The case involving dependent stick-breaking processes with common weights and predictor-dependent support points is referred to as 'single–weights' DMBPP, while the case involving dependent stick-breaking processes with common support points and predictor-dependent weights is referred to as 'single–atoms' DMBPP. In what follows, we briefly discuss the definition of these special cases. Their formal definitions, needed to provide the proofs of the theoretical properties discussed in the following sections, are provided in Appendix B.

In the definition of the 'single–weights' DMBPP, the real-valued stochastic processes of condition (i) in Definition 1, $\eta_j = \{\eta_j(\boldsymbol{x}) : \boldsymbol{x} \in \mathscr{X}\}$, are replaced by $[0,1]$-valued independent and identically distributed random variables, v_j, with common distribution indexed by a finite-dimensional parameter $\boldsymbol{\Psi}_v$. In this special case, the density function of $F_{\boldsymbol{x}}$ is given by

$$f_{\boldsymbol{x}}(\cdot) = \sum_{j=1}^{\infty} w_j \mathrm{dir}\left(\cdot \mid \alpha(k, \lceil k\boldsymbol{\theta}_j(\boldsymbol{x})\rceil)\right), \tag{6}$$

where $\boldsymbol{\theta}_j(\boldsymbol{x})$ and $\lceil k\boldsymbol{\theta}_j(\boldsymbol{x})\rceil$ are defined as in Definition 1 and $w_j = v_j \prod_{l<j} [1 - v_l]$. The process $F = \{F_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$ will be referred to as single–weight dependent MBP process with parameters $(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_z, \mathscr{H})$, and denoted by wDMBPP$(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_z, \mathscr{H})$ and wDMBPP for short.

In the definition of the 'single–atoms' DMBPP, the real-valued stochastic processes of condition (ii) in Definition 1, $\boldsymbol{z}_j = \{\boldsymbol{z}_j(\boldsymbol{x}) : \boldsymbol{x} \in \mathscr{X}\}$, are replaced by independent and identically distributed Δ_m^0-valued random vectors, $\boldsymbol{\theta}_j$, with common distribution indexed by a finite-dimensional parameter $\boldsymbol{\Psi}_{\boldsymbol{\theta}}$. In this case, the density function of $F_{\boldsymbol{x}}$ is given by

$$f_{\boldsymbol{x}}(\cdot) = \sum_{j=1}^{\infty} w_j(\boldsymbol{x}) \mathrm{dir}\left(\cdot \mid \alpha(k, \lceil k\boldsymbol{\theta}_j\rceil)\right), \tag{7}$$

where $w_j(\boldsymbol{x})$ are defined as in Definition 1 and $\lceil k\boldsymbol{\theta}_j\rceil = (\lceil k\theta_{j1}\rceil, \ldots, \lceil k\theta_{jm}\rceil)$. The process $F = \{F_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$ will be referred to as single–atoms dependent MBP process with parameters $(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$, and denoted by θDMBPP$(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$ and θDMBPP for short.

Notice that the DMBPP, including its special cases, is well defined if the mapping induced by (iv) in Definition 1 is measurable, which is discussed in detail in Section 3.2. Notice also that expressions (5), (6), and (7) are indeed a density w.r.t. Lebesgue measure since, for every $\boldsymbol{x} \in \mathscr{X}$,

$$\sum_{j=1}^{\infty} \log\left[1 - \mathrm{E}\left(v_{\boldsymbol{x}}\left\{\eta_j(\boldsymbol{x})\right\}\right)\right] = -\infty, \quad \text{and} \quad \sum_{j=1}^{\infty} \log\left[1 - \mathrm{E}(v_j)\right] = -\infty,$$

which are sufficient and necessary conditions for the corresponding weights to add up to one with probability one. It is important to emphasize that DMBPP generates dependent mixture of Dirichlet densities with constant support points and covariate-dependent weights,

$$f_{\boldsymbol{x}}(\cdot) \quad = \sum_{\boldsymbol{j} \in \mathcal{H}^0_{k,m}} W_{k,\boldsymbol{j},\boldsymbol{x}} \times \operatorname{dir}\left(\cdot \mid \alpha(k, \boldsymbol{j})\right), \tag{8}$$

where

$$W_{k,\boldsymbol{j},\boldsymbol{x}} = \begin{cases} \sum_{l=1}^{\infty} w_l(\boldsymbol{x}) \delta_{\boldsymbol{\theta}_l(\boldsymbol{x})} \left(\left(\frac{j_1 - 1}{k}, \frac{j_1}{k} \right] \times \ldots \times \left(\frac{j_m - 1}{k}, \frac{j_m}{k} \right] \right), \\ \sum_{l=1}^{\infty} w_l \delta_{\boldsymbol{\theta}_l(\boldsymbol{x})} \left(\left(\frac{j_1 - 1}{k}, \frac{j_1}{k} \right] \times \ldots \times \left(\frac{j_m - 1}{k}, \frac{j_m}{k} \right] \right), \\ \sum_{l=1}^{\infty} w_l(\boldsymbol{x}) \delta_{\boldsymbol{\theta}_l} \left(\left(\frac{j_1 - 1}{k}, \frac{j_1}{k} \right] \times \ldots \times \left(\frac{j_m - 1}{k}, \frac{j_m}{k} \right] \right), \end{cases}$$

for the DMBPP, wDMBPP, and θDMBPP, respectively.

3.2. *The measurability of the processes*

In this section we show that the corresponding mappings defining the trajectories of DMBPP, wDMBPP, and θDMBPP are measurable under the Borel σ-field generated by the weak product topology, L_∞ product topology, and L_∞ topology, which correspond to generalizations of standard topologies for spaces of single probability measures. The topologies considered here are formally defined in Appendix C.

Let $\mathscr{D}(\Delta_m) \subset \mathscr{P}(\Delta_m)$ be the space of all probability measures defined on Δ_m that are absolutely continuous w.r.t. Lebesgue measure and with continuous density function and consider the spaces $\mathscr{P}(\Delta_m)^{\mathscr{X}} = \prod_{\boldsymbol{x} \in \mathscr{X}} \mathscr{P}(\Delta_m)$ and $\mathscr{D}(\Delta_m)^{\mathscr{X}} = \prod_{\boldsymbol{x} \in \mathscr{X}} \mathscr{D}(\Delta_m)$. Theorem 1, which proof is provided in Appendix D.1, summarizes the measurability results for the different versions of the proposed model.

Theorem 1. *Let $\mathscr{B}_1$, $\mathscr{B}_2$, and $\mathscr{B}_3$ be the Borel σ-field generated by the weak product topology, L_∞ product topology, and L_∞ topology, respectively. If F is a* DMBPP, wDMBPP *or* θDMBPP, *defined on the appropriate measurable space $(\Omega, \mathscr{A})$, then the following mappings are measurable:*

- $F : (\Omega, \mathcal{A}) \longrightarrow (\mathscr{P}(\Delta_m)^{\mathscr{X}}, \mathscr{B}_1)$.
- $F : (\Omega, \mathcal{A}) \longrightarrow (\mathscr{D}(\Delta_m)^{\mathscr{X}}, \mathscr{B}_2)$.
- $F : (\Omega, \mathcal{A}) \longrightarrow (\mathscr{D}(\Delta_m)^{\mathscr{X}}, \mathscr{B}_3)$.

4. The main properties

We establish basic properties of the proposed class of models in this section. They include the characterization of the topological support, the continuity and association structure of the models, and the asymptotic behavior of the posterior distribution. Detailed proofs of Theorems 2 – 12 are provided in Appendix D.2 – D.12, respectively.

4.1. The support of the processes

Full support is a "necessary" property for a Bayesian model to be considered "nonparametric". In a fully nonparametric regression model setting, full support implies that the prior probability model assigns positive mass to any neighborhood of every collection of probability measures $\{Q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$. Therefore, the definition of support strongly depends on the choice of a "distance" defining the basic neighborhoods. We provide sufficient conditions for $\mathscr{P}(\Delta_m)^{\mathscr{X}}$ and $\mathscr{D}(\Delta_m)^{\mathscr{X}}$ to be the support of DMBPPs under the weak product topology and the L_∞ product topology, respectively.

Theorem 2. *Let F be a* $\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_z, \mathscr{V}, \mathscr{H})$*, a* $\theta\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$*, or a* $w\mathrm{DMBPP}$ $(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_z, \mathscr{H})$*. If F is defined such that:*

(i) for every $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L) \in \mathscr{X}^L$, $L \geq 1$, the joint distribution of $(\eta_j(\boldsymbol{x}_1), \ldots, \eta_j(\boldsymbol{x}_L))$, $j \geq 1$, has full support on $\mathbb{R}^L$,
(ii) for every $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L) \in \mathscr{X}^L$, $L \geq 1$, the joint distribution of $(\boldsymbol{z}_j(\boldsymbol{x}_1), \ldots, \boldsymbol{z}_j(\boldsymbol{x}_L))$, $j \geq 1$, has full support on $\mathbb{R}^{m \times L}$,
(iii) k has full support on $\mathbb{N}$,
(iv) v_j, $j \geq 1$, has full support on $[0, 1]$,
(v) $\boldsymbol{\theta}_j$, $j \geq 1$, has full support on Δ_m^0,

then $\mathscr{P}(\Delta_m)^{\mathscr{X}}$ and $\mathscr{D}(\Delta_m)^{\mathscr{X}}$ is the support of F under the weak product topology and the L_∞ product topology, respectively.

If stronger assumptions on the parameter space are imposed, a stronger support property can be obtained. Specifically, consider the sub-space $\tilde{\mathscr{D}}(\Delta_m)^{\mathscr{X}} \subset \mathscr{D}(\Delta_m)^{\mathscr{X}}$, where

$$\tilde{\mathscr{D}}(\Delta_m)^{\mathscr{X}} = \left\{ \{Q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \mathscr{D}(\Delta_m)^{\mathscr{X}} : (\boldsymbol{y}, \boldsymbol{x}) \longmapsto q_{\boldsymbol{x}}(\boldsymbol{y}) \text{ is continuous} \right\},$$

and $q_{\boldsymbol{x}}$ denotes the density function of $Q_{\boldsymbol{x}}$ w.r.t. Lebesgue measure. The following theorem provides sufficient conditions for $\tilde{\mathscr{D}}(\Delta_m)^{\mathscr{X}}$ to be in the support of DMBPPs under the L_∞ topology.

Theorem 3. *Let F be a* $\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_z, \mathscr{V}, \mathscr{H})$*, a* $\theta\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$ *or a* $w\mathrm{DMBPP}$ $(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_z, \mathscr{H})$*. Assume that $\boldsymbol{x} \in \mathscr{X}$ contains only continuous components and that $\mathscr{X}$ is compact. If F is defined such that:*

(i) for every $B \in \mathscr{B}(\Delta_m)$, every Δ_m^0-valued continuous mapping $\boldsymbol{x} \mapsto f(\boldsymbol{x})$, and every $j \geq 1$,

$$\Pr\left\{ \sup_{\boldsymbol{x} \in \mathscr{X}} |h_{\boldsymbol{x}}(\boldsymbol{z}_j(\boldsymbol{x})) - f_{\boldsymbol{x}}| \in B \right\} > 0,$$

(ii) for every $\epsilon > 0$, every $[0, 1]$-valued continuous mapping $\boldsymbol{x} \mapsto f(\boldsymbol{x})$, and every $j \geq 1$,

$$\Pr\left\{ \sup_{\boldsymbol{x} \in \mathscr{X}} |v_{\boldsymbol{x}}(\eta_j(\boldsymbol{x})) - f_{\boldsymbol{x}}| < \epsilon \right\} > 0,$$

(iii) k *has full support on* $\mathbb{N}$,
(iv) v_j, $j \geq 1$, *has full support on* $[0,1]$,
(v) $\boldsymbol{\theta}_j$, $j \geq 1$, *has full support on* Δ_m^0,

then $\tilde{\mathscr{D}}(\Delta_m)^{\mathscr{X}}$ *is contained in the support of* F *under the* L_∞ *topology.*

An important consequence of the previous theorem is that the proposed processes can assign positive mass to arbitrarily small neighborhoods of any collection of probability measures $\{Q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \tilde{\mathscr{D}}(\Delta_m)^{\mathscr{X}}$, based on the supremum over the predictor space of Kullback-Leibler (KL) divergences between the predictor-dependent probability measures.

Theorem 4. *Let* F *be a* $\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_z, \mathscr{V}, \mathscr{H})$, *a* $\theta\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$ *or a* $w\mathrm{DMBPP}$ $(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_z, \mathscr{H})$. *Under the same assumptions of Theorem 3, it follows that*

$$\Pr\left\{\sup_{\boldsymbol{x}\in\mathscr{X}} \int_{\Delta_m} q_{\boldsymbol{x}}(\boldsymbol{y}) \log\left(\frac{q_{\boldsymbol{x}}(\boldsymbol{y})}{f_{\boldsymbol{x}}(\boldsymbol{y})}\right) d\boldsymbol{y} < \epsilon\right\} > 0,$$

for every $\epsilon > 0$, *and every* $\{Q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \tilde{\mathscr{D}}(\Delta_m)^{\mathscr{X}}$ *with density functions* $\{q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$.

4.2. The continuity and association structure of the processes

The characteristics of the stochastic processes used in the definitions of aDMBPP determine important properties of the resulting model. Regardless of the specific choice of the stochastic processes used in its definition, the use of almost surely (a.s.) continuous stochastic processes ensures that DMBPP and wDMBPP have a.s. a limit.

Theorem 5. *Let* F *be* $\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_z, \mathscr{V}, \mathscr{H})$ *or* $w\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_z, \mathscr{H})$, *defined such that* $\mathscr{V}$ *and* $\mathscr{H}$ *are sets of equicontinuous functions of* $\boldsymbol{x}$, *and for every* $i \geq 1$, *the stochastic processes* η_i *and* $\boldsymbol{z}_i$ *have a.s. continuous trajectories. Then, for every* $\{\boldsymbol{x}_l\}_{l=0}^{\infty}$, *with* $\boldsymbol{x}_l \in \mathscr{X}$, *such that* $\lim_{l\to\infty} \boldsymbol{x}_l = \boldsymbol{x}_0$, $F_{\boldsymbol{x}}$ *has a.s. a limit with the total variation norm.*

An interesting property of the θDMBPP compared to the other version, and the general model, is that the use of a.s. continuous stochastic processes in the weights guarantees a.s. continuity of the 'single–atoms' DMBPP.

Theorem 6. *Let* F *be a* $\theta\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$, *defined such that* $\mathscr{V}$ *is a set of equicontinuous functions, and such that for every* $j \geq 1$, *the stochastic process* η_j *is a.s. continuous. Then, for every* $\{\boldsymbol{x}_l\}_{l=0}^{\infty}$, *with* $\boldsymbol{x}_l \in \mathscr{X}$, *such that* $\lim_{l\to\infty} \boldsymbol{x}_l = \boldsymbol{x}_0$,

$$\lim_{l\to\infty} \sup_{B\in\mathscr{B}(\Delta_m)} |F_{\boldsymbol{x}_l}(B) - F_{\boldsymbol{x}_0}(B)| = 0, \ \text{a.s.}.$$

That is, $F_{\boldsymbol{x}_l}$ *converges a.s. in total variation norm to* $F_{\boldsymbol{x}_0}$, *when* $\boldsymbol{x}_l \longrightarrow \boldsymbol{x}_0$.

The dependence structure of DMBPPs is completely determined by the association structure of the stochastic processes used in their definition. For instance, under mild conditions on the stochastic processes defining the DMBPPs, the correlation between the corresponding random measures approaches to one as the predictor values get closer.

Theorem 7. *Let F be a* $\text{DMBPP}(\lambda, \mathbf{\Psi}_\eta, \mathbf{\Psi}_z, \mathscr{V}, \mathscr{H})$*, a* $\theta\text{DMBPP}(\lambda, \mathbf{\Psi}_\eta, \mathscr{V}, \mathbf{\Psi}_{\boldsymbol{\theta}})$ *or a* $w\text{DMBPP}$ $(\lambda, \mathbf{\Psi}_v, \mathbf{\Psi}_z, \mathscr{H})$*, defined such that $\mathscr{V}$ and $\mathscr{H}$ are sets of equicontinuous functions, and such that for every $\{\boldsymbol{x}_l\}_{l=0}^{\infty}$, with $\boldsymbol{x}_l \in \mathscr{X}$, such that $\lim_{l\to\infty} \boldsymbol{x}_l = \boldsymbol{x}_0$, we have $\eta_j(\boldsymbol{x}_l) \xrightarrow{\mathscr{L}} \eta_j(\boldsymbol{x}_0)$ and $\boldsymbol{z}_j(\boldsymbol{x}_l) \xrightarrow{\mathscr{L}} \boldsymbol{z}_j(\boldsymbol{x}_0)$, as $l \to \infty$, $j \geq 1$. Then, for every $\boldsymbol{y} \in \tilde{\Delta}_m$,*

$$\lim_{l\to\infty} \rho\left[F_{\boldsymbol{x}_l}(B_{\boldsymbol{y}}), F_{\boldsymbol{x}_0}(B_{\boldsymbol{y}})\right] = 1,$$

where $\rho(A, B)$ denotes the Pearson correlation between A and B, $B_{\boldsymbol{y}} = [0, y_1] \times \ldots \times [0, y_m]$.

If the stochastic processes defining the DMBPP and wDMBPP are such that the pairwise finite-dimensional distributions converge to the product of the corresponding marginal distributions as the Euclidean distance between the predictors grows larger, then under mild conditions the correlation between the corresponding random measures can approach zero. The following theorem shows that under the assumptions previously discussed, the marginal covariance between the random measures is equal to the covariance between the conditional expectations of the random measures, given the degree of the MBP.

Theorem 8. *Let F be a* $\text{DMBPP}(\lambda, \mathbf{\Psi}_\eta, \mathbf{\Psi}_z, \mathscr{V}, \mathscr{H})$ *or a* $w\text{DMBPP}$ $(\lambda, \mathbf{\Psi}_v, \mathbf{\Psi}_z, \mathscr{H})$*, defined such that $\mathscr{V}$ and $\mathscr{H}$ are sets of equicontinuous functions and there exists a constant $\gamma > 0$ such that if $(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathscr{X}^2$ and $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| > \gamma$, then $Cov\left[\mathbb{I}_{\{\eta_j(\boldsymbol{x}_1)\in A_1\}}, \mathbb{I}_{\{\eta_j(\boldsymbol{x}_2)\in A_2\}}\right] = 0$, for every $A_1, A_2 \in \mathscr{B}(\mathbb{R})$, and $Cov\left[\mathbb{I}_{\{\boldsymbol{z}_j(\boldsymbol{x}_1)\in A_3\}}, \mathbb{I}_{\{\boldsymbol{z}_j(\boldsymbol{x}_2)\in A_4\}}\right] = 0$, for every $A_3, A_4 \in \mathscr{B}(\mathbb{R}^m)$, $j \geq 1$. Assume also that for every $(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathscr{X}^2$ and for every sequence $\{(\boldsymbol{x}_{1l}, \boldsymbol{x}_{2l})\}_{l=1}^{\infty}$, with $(\boldsymbol{x}_{1l}, \boldsymbol{x}_{2l}) \in \mathscr{X}^2$ and such that $\lim_{l\to\infty}(\boldsymbol{x}_{1l}, \boldsymbol{x}_{2l}) = (\boldsymbol{x}_1, \boldsymbol{x}_2)$, we have that*

$$(\eta_j(\boldsymbol{x}_{1l}), \eta_j(\boldsymbol{x}_{2l})) \xrightarrow{\mathscr{L}} (\eta_j(\boldsymbol{x}_1), \eta_j(\boldsymbol{x}_2)),$$

and

$$(\boldsymbol{z}_j(\boldsymbol{x}_{1l}), \boldsymbol{z}_j(\boldsymbol{x}_{2l})) \xrightarrow{\mathscr{L}} (\boldsymbol{z}_j(\boldsymbol{x}_1), \boldsymbol{z}_j(\boldsymbol{x}_2)),$$

$j \geq 1$, as $l \to \infty$. Then, for every $\boldsymbol{y} \in \Delta_m$,

$$\lim_{l\to\infty} Cov\left[F_{\boldsymbol{x}_{1l}}(B_{\boldsymbol{y}}), F_{\boldsymbol{x}_{2l}}(B_{\boldsymbol{y}})\right] = Cov\left[E\left\{F_{\boldsymbol{x}_1}(B_{\boldsymbol{y}}) \mid k\right\}, E\left\{F_{\boldsymbol{x}_2}(B_{\boldsymbol{y}}) \mid k\right\}\right],$$

with

$$E\left\{F_{\boldsymbol{x}}(B_{\boldsymbol{y}})|k\right\} = \sum_{\mathbf{j}\in\mathcal{H}_{k,m}} G_{0,\boldsymbol{x}}\left(A_{\mathbf{j},k}\right) \text{Mult}(\mathbf{j} \mid k+m-1, \boldsymbol{y}),$$

where $B_{\boldsymbol{y}} = [0, y_1] \times \ldots \times [0, y_m]$, $A_{\mathbf{j},k} = [0, j_1/k] \times \ldots \times [0, j_m/k]$ and $G_{0,\boldsymbol{x}}$ is the marginal probability measure of $\boldsymbol{\theta}_j(\boldsymbol{x})$ defined on Δ_m^0.

Notice that the assumption $Cov\left[\mathbb{I}_{\{\eta_j(\boldsymbol{x}_1)\in A_1\}}, \mathbb{I}_{\{\eta_j(\boldsymbol{x}_2)\in A_2\}}\right] = 0$, for every A_1, $A_2 \in \mathscr{B}(\mathbb{R})$ is equivalent to assuming that $\eta_j(\boldsymbol{x}_1)$ and $\eta_j(\boldsymbol{x}_2)$ are independent. This also applies for the process $\boldsymbol{z}_j$. Notice also that an example of a process meeting the conditions of Theorem 8 is the Gaussian process with spherical covariance function (see Banerjee et al., 2003, Chapter 2). From Theorem 8 it is easy to see that if DMBPP or wDMBPP are specified such that the marginal distribution of k is degenerate, then the correlation between the corresponding random measures goes to zero, since $\lim_{l\to\infty} Cov\left[F_{\boldsymbol{x}_{1l}}(B_{\boldsymbol{y}}), F_{\boldsymbol{x}_{2l}}(B_{\boldsymbol{y}})\right] = 0$. For θDMBPP the correlation between the associated random measures when the predictor values are far apart reaches a different limit. In such case, it is difficult to establish conditions on the prior specification ensuring that the limit is zero.

Theorem 9. *Let F be a $\theta\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$. Assume that $\mathscr{V}$ is a set of equicontinuous functions and that there exists a constant $\gamma > 0$, such that if $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathscr{X}$ and $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| > \gamma$, then $Cov\left[\mathbb{I}_{\{\eta_j(\boldsymbol{x}_1)\in A_1\}}, \mathbb{I}_{\{\eta_j(\boldsymbol{x}_2)\in A_2\}}\right] = 0$, for every $A_1, A_2 \in \mathscr{B}(\mathbb{R})$, $j \geq 1$. Assume also that for every $(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathscr{X}^2$ and for every sequence $\{(\boldsymbol{x}_{1l}, \boldsymbol{x}_{2l})\}_{l=1}^{\infty}$, with $(\boldsymbol{x}_{1l}, \boldsymbol{x}_{2l}) \in \mathscr{X}^2$, such that $\lim_{l\to\infty}(\boldsymbol{x}_{1l}, \boldsymbol{x}_{2l}) = (\boldsymbol{x}_1, \boldsymbol{x}_2)$, we have $(\eta_j(\boldsymbol{x}_{1l}), \eta_j(\boldsymbol{x}_{2l})) \xrightarrow{\mathscr{L}} (\eta_j(\boldsymbol{x}_1), \eta_j(\boldsymbol{x}_2))$, $j \geq 1$, as $l \to \infty$. Then, for every $\boldsymbol{y} \in \Delta_m$,*

$$\lim_{l\to\infty} Cov\left[F_{\boldsymbol{x}_{1l}}(B_{\boldsymbol{y}}), F_{\boldsymbol{x}_{2l}}(B_{\boldsymbol{y}})\right] = \sum_{k_1=1}^{\infty} \Pr\{k = k_1\} \sum_{\mathbf{j}_1,\ \mathbf{j}_2 \in \mathcal{H}_{k_1,m}} \bar{M}(\mathbf{j}_1, \mathbf{j}_2 \mid k_1 + m - 1, \boldsymbol{y})$$
$$\times \sum_{j=1}^{\infty} E\left[w_j(\boldsymbol{x}_1)\right] E[w_j(\boldsymbol{x}_2)]\, Cov\left[\mathbb{I}_{\{\boldsymbol{\theta}_j \in A_{\mathbf{j}_1,k_1}\}}, \mathbb{I}_{\{\boldsymbol{\theta}_j \in A_{\mathbf{j}_2,k_1}\}}\right]$$
$$+ Cov\left[E\left\{F_{\boldsymbol{x}_1}(B_{\boldsymbol{y}}) \mid k\right\}, E\left\{F_{\boldsymbol{x}_2}(B_{\boldsymbol{y}}) \mid k\right\}\right],$$

with $E\{F_{\boldsymbol{x}}(B_{\boldsymbol{y}})|k\}$, $B_{\boldsymbol{y}}$, $A_{\mathbf{j},k}$, and $G_{0,\boldsymbol{x}}$ as defined in Theorem 8 and $\bar{M}(\mathbf{j}, \mathbf{j}_1 \mid k + m - 1, \boldsymbol{y}) = \mathrm{Mult}(\mathbf{j} \mid k + m - 1, \boldsymbol{y}) \times \mathrm{Mult}(\mathbf{j}_1 \mid k + m - 1, \boldsymbol{y})$.

Finally, although the trajectories of the DMBPP and wDMBPP have a.s. a limit only, the autocorrelation function of all versions of the model are continuous under mild conditions on the elements defining the processes.

Theorem 10. *Let F be a $\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_z, \mathscr{V}, \mathscr{H})$, a $\theta\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$ or a wDMBPP $(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_z, \mathscr{H})$, defined such that $\mathscr{V}$ and $\mathscr{H}$ are sets of equicontinuous functions. Assume that for every $(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathscr{X}^2$ and for every sequence $\{(\boldsymbol{x}_{1l}, \boldsymbol{x}_{2l})\}_{l=1}^{\infty}$, with $(\boldsymbol{x}_{1l}, \boldsymbol{x}_{2l}) \in \mathscr{X}^2$, such that $\lim_{l\to\infty}(\boldsymbol{x}_{1l}, \boldsymbol{x}_{2l}) = (\boldsymbol{x}_1, \boldsymbol{x}_2)$, we have that*

$$(\eta_j(\boldsymbol{x}_{1l}), \eta_j(\boldsymbol{x}_{2l})) \xrightarrow{\mathscr{L}} (\eta_j(\boldsymbol{x}_1), \eta_j(\boldsymbol{x}_2)),$$

and

$$(\boldsymbol{z}_j(\boldsymbol{x}_{1l}), \boldsymbol{z}_j(\boldsymbol{x}_{2l})) \xrightarrow{\mathscr{L}} (\boldsymbol{z}_j(\boldsymbol{x}_1), \boldsymbol{z}_j(\boldsymbol{x}_2)),$$

as $l \to \infty$, $j \geq 1$. Then, for every $\boldsymbol{y} \in \Delta_m^0$,

$$\lim_{l\to\infty} \rho\left[F_{\boldsymbol{x}_{1l}}(B_{\boldsymbol{y}}), F_{\boldsymbol{x}_{2l}}(B_{\boldsymbol{y}})\right] = \rho\left[F_{\boldsymbol{x}_1}(B_{\boldsymbol{y}}), F_{\boldsymbol{x}_2}(B_{\boldsymbol{y}})\right],$$

where $B_{\boldsymbol{y}} = [0, y_1] \times \ldots \times [0, y_m]$.

4.3. The asymptotic behavior of the posterior distribution

We study the asymptotic behavior of the posterior distribution of the proposed model class in this section. Here we assume that we observe a random sample $(\boldsymbol{y}_i, \boldsymbol{x}_i)$, $i = 1, \ldots, n$. Recall that, as is common in regression settings, we assume that the predictor vector $\boldsymbol{x}_i$ contains only exogenous predictors. Notice that the exogeneity assumption allows us to focus on the conditional density estimation problem, regardless of the data generating mechanism of the predictors, that is, if they are randomly generated or fixed by design (see, e.g. Barndorff-Nielsen, 1973, 1978; Florens et al., 1990). Let Q be the true probability measure generating the predictors, with density w.r.t. a corresponding σ-additive measure denoted by q. By the exogeneity assumption, the true probability model for the response variable and predictors takes the form $h_0(\boldsymbol{y}, \boldsymbol{x}) = q(\boldsymbol{x}) q_0(\boldsymbol{y} \mid \boldsymbol{x})$, where both q and $\{q_0(\cdot \mid \boldsymbol{x}) : \boldsymbol{x} \in \mathscr{X}\}$ are in free variation, with $q_0(\boldsymbol{y} \mid \boldsymbol{x})$ denoting a conditional density defined on Δ_m, and $\boldsymbol{x} \in \mathscr{X}$.

Theorem 11. *Let F be a* $\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_z, \mathscr{V}, \mathscr{H})$*, a* $\theta\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$ *or a* $w\mathrm{DMBPP}$ $(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_z, \mathscr{H})$*. If the assumptions of Theorem 3 are satisfied, then the posterior distribution associated with the random joint distribution induced by the corresponding* DMBPP *model,* $h(\boldsymbol{y}, \boldsymbol{x}) = q(\boldsymbol{x}) f_{\boldsymbol{x}}(\boldsymbol{y})$*, where q is the density generating the predictors, is weakly consistent at any joint distribution of the form* $h_0(\boldsymbol{y}, \boldsymbol{x}) = q(\boldsymbol{x}) q_0(\boldsymbol{y} \mid \boldsymbol{x})$*, where* $\{q_0(\cdot \mid \boldsymbol{x}) : \boldsymbol{x} \in \mathscr{X}\} \in \tilde{\mathscr{D}}(\Delta_m)^{\mathscr{X}}$*.*

Although Theorem 11 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 a combination of continuous predictors and predictors with finite support.

The following theorem states a stronger posterior consistency result when a specific probit stick-breaking process is assumed in the definition of the θDMBPP.

Theorem 12. *Let F be a* $\theta\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$*. If $\mathscr{X} = [0, 1]^p$ and the* θDMBPP *is defined such that*

(i) for every $j \in \mathbb{N}$, η_j is a Gaussian process with zero mean function and covariance kernel given by $c_j(\boldsymbol{x}, \boldsymbol{x}') = \tau^2 \exp\{-A_j \|\boldsymbol{x} - \boldsymbol{x}'\|^2\}$, where $(\boldsymbol{x}, \boldsymbol{x}') \in \mathscr{X}^2$ and A_j is a random variable, such that for some positive constants κ and κ_0, and some sequence $r_n \uparrow \infty$, such that $r_n^p n^\kappa (\log n)^{p+1} = o(n)$,

$$\Pr\{A_j > \delta_n\} \leq \exp\{-n^{-\kappa_0} j^{(\kappa_0+2)/\kappa} \log j\},$$

and

$$\Pr\{A_n > r_n\} \leq \exp\{-n\},$$

where $\delta_n = O((\log n)^2 / n^{5/2})$,

(ii) for every $v_{\boldsymbol{x}} \in \mathscr{V}$, $v_{\boldsymbol{x}} \equiv \Phi$, where Φ denotes the CDF of a standardnormal distribution.

(iii) G_0 has full support on Δ_m^0, where G_0 is the distribution of $\boldsymbol{\theta}_j$, $j \geq 1$.

(iv) k has full support on $\mathbb{N}$,

(v) there exists a sequence $k_n \in \mathbb{N}$ *such that* $\log\left(\frac{k_n(k_n+m)!}{k_n!(m+1)!}\right) \preceq O(n)$ *and* $\Pr\{k > k_n\} \preceq O(\exp\{-n\})$*, where* $\preceq$ *stands for inequality up to a constant.*

Then, the posterior distribution associated with the random joint distribution induced by the θDMBPP *model,* $h(\boldsymbol{y}, \boldsymbol{x}) = q(\boldsymbol{x}) f_{\boldsymbol{x}}(\boldsymbol{y})$*, where* q *is the density generating the predictors, is* L_1*–consistent at any joint distribution of the form* $h_0(\boldsymbol{y}, \boldsymbol{x}) = q(\boldsymbol{x}) q_0(\boldsymbol{y} \mid \boldsymbol{x})$*, where* $\{q_0(\cdot \mid \boldsymbol{x}) : \boldsymbol{x} \in \mathscr{X}\} \in \tilde{\mathscr{D}}(\Delta_m)^{\mathscr{X}}$.

For an example of how to construct the sequence of random variables A_j, see Remark 5.12 in Pati et al. (2013).

5. Computational aspects

As can be noted from the definitions of the proposed models, predictors can be included in different manners. In what follows, we consider special definitions by exploiting the relation between Gaussian processes and Bayesian linear regression models. We also make use of spike-and-slab prior distributions on the regression coefficients that allow us for an automatic selection of the version of the model that best accommodates to the complexity of the underlying data-generating mechanism. Note that such a prior avoids the need to fit each version of the model, as done by Barrientos et al. (2017).

We specify the predictor dependent weights and atoms of the dependent stick-breaking process in the DMBPP by means of transformations of a linear predictor. To define the weights of the DMBPP we consider $v_{\boldsymbol{x}}(a) = e^a/(1+e^a)$, $a \in \mathbb{R}$, and the stochastic process $\eta_j(\boldsymbol{x}) = \beta_{0j}^{\eta} + \boldsymbol{x}^t \boldsymbol{\beta}_j^{\eta}$, where $\beta_{0j}^{\eta} \in \mathbb{R}$ and $\boldsymbol{\beta}_j^{\eta} \in \mathbb{R}^p$ are independent and identically distributed for $j \geq 1$, and $\boldsymbol{x} = (x_1, \ldots, x_p) \in \mathscr{X}^p$ denotes the vector of covariates. Similarly, to define the atoms of the dependent stick-breaking process in the DMBPP we consider the transformation

$$h_{\boldsymbol{x}}(\mathbf{b}) = \left(e^{b_1}, \ldots, e^{b_m}\right) / \left(1 + \sum_{l=1}^{m} e^{b_l}\right), \quad \mathbf{b} \in \mathbb{R}^m,$$

and the stochastic process $\boldsymbol{z}_j(\boldsymbol{x}) = (\boldsymbol{z}_{j1}(\boldsymbol{x}), \ldots, \boldsymbol{z}_{jm}(\boldsymbol{x}))$, where $\boldsymbol{z}_{jl}(\boldsymbol{x}) = \beta_{0jl}^{\boldsymbol{z}} + \boldsymbol{x}^t \boldsymbol{\beta}_{jl}^{\boldsymbol{z}}$ and $\beta_{0jl}^{\boldsymbol{z}} \in \mathbb{R}$ and $\boldsymbol{\beta}_{jl}^{\boldsymbol{z}} \in \mathbb{R}^p$ are independent and identically distributed for $j \geq 1$, $l = 1, \ldots, m$.

In order to choose the version of the DMBPP model that best adapts to the data and following George & McCulloch (1993), we consider a two-components mixture of normal distributions with different variances as a prior distribution on the coefficients of the linear predictor associated with the covariates, that is, on $\boldsymbol{\beta}_j^{\eta}$ and $\boldsymbol{\beta}_{jl}^{\boldsymbol{z}}$. For the intercepts of the linear predictors we assume $\beta_{0j}^{\eta} \overset{iid}{\sim} N(0, \sigma_{\eta}^2)$ and $\beta_{0jl}^{\boldsymbol{z}} \overset{iid}{\sim} N(0, \sigma_{\boldsymbol{z}}^2)$. For $\boldsymbol{\beta}_j^{\eta}$ and $\boldsymbol{\beta}_{jl}^{\boldsymbol{z}}$ we introduce latent binary variables γ^{η} and $\gamma^{\boldsymbol{z}}$ and assume

$$\boldsymbol{\beta}_j^{\eta} \mid \gamma^{\eta} \overset{iid}{\sim} N_p(\mathbf{0}, \boldsymbol{\Sigma}_1^{\eta})^{1-\gamma^{\eta}} \times N_p(\mathbf{0}, \boldsymbol{\Sigma}_2^{\eta})^{\gamma^{\eta}}, \quad j \geq 1, \tag{9}$$

$$\boldsymbol{\beta}_{jl}^{\boldsymbol{z}} \mid \gamma^{\boldsymbol{z}} \overset{iid}{\sim} N_p(\mathbf{0}, \boldsymbol{\Sigma}_1^{\boldsymbol{z}})^{1-\gamma^{\boldsymbol{z}}} \times N_p(\mathbf{0}, \boldsymbol{\Sigma}_2^{\boldsymbol{z}})^{\gamma^{\boldsymbol{z}}}, \quad j \geq 1, \quad l = 1, \ldots, m, \tag{10}$$

where $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the $p-$dimensional multivariate normal distribution with mean vector $\boldsymbol{\mu} \in \mathbb{R}^p$ and $p \times p$ positive definite covariance matrix $\boldsymbol{\Sigma}$. The covariance matrices $\boldsymbol{\Sigma}_1^{\eta}$ and $\boldsymbol{\Sigma}_1^{\boldsymbol{z}}$ define the "spike" component of the prior and are set such that define normal distributions that are highly concentrated around zero, while $\boldsymbol{\Sigma}_2^{\eta}$ and $\boldsymbol{\Sigma}_2^{\boldsymbol{z}}$ define the "slab" component of the prior and are set such that the resulting normal distributions are less concentrated around zero. Therefore, binary parameters γ^{η} and $\gamma^{\boldsymbol{z}}$, which are common for every $\boldsymbol{\beta}_j^{\eta}$ and $\boldsymbol{\beta}_j^{\boldsymbol{z}}$, control the predictor dependency structure of the model.

When the vector of binary variables $(\gamma^{\eta}, \gamma^{\boldsymbol{z}})$ is equal to $(1,1)$, $(0,1)$, $(1,0)$, or $(0,0)$, then the chosen model is fully dependent, single-weight, single-atom, or predictor independent, respectively. To complete the prior for $\boldsymbol{\beta}_j^{\eta}$ and $\boldsymbol{\beta}_{jl}^{\boldsymbol{z}}$, we consider

$$(\gamma^{\eta}, \gamma^{\boldsymbol{z}}) \sim \pi_1 \delta_{(1,1)} + \pi_2 \delta_{(0,1)} + \pi_3 \delta_{(1,0)} + \pi_4 \delta_{(0,0)}, \tag{11}$$

where $\pi_i \geq 0$ and $\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1$. Finally, to complete the prior specification for the DMBPP model we assume $k \mid \lambda \sim Poisson(\lambda)\mathbb{I}_{\{k \geq 1\}}$.

Posterior sampling of the DMBPP model can be based on any conditional algorithm designed for BNP mixture models. The specific implementation employed in the simulation study and in the application was based on a finite representation of the dependent stick-breaking process to a level N (Ishwaran & James, 2001). We use Gibbs sampling algorithms to generate samples from the posterior distribution. To sample the non conjugate full conditional distributions of the coefficients in the linear predictors we use the slice sampler algorithm (Neal, 2003). We use a Metropolis-Hastings step (Tierney, 1994) to update the degree of the polynomial. The binary parameters are sampled from their conjugate categorical posterior distribution. More details are provided in Appendix E. The code employed here to fit the proposed model is available in GitHub.

6. Illustrations

In this section we illustrate the performance of the model in a simulation study and in an application to solid waste recycling in the city of Santiago de Cali, Colombia. In the simulation study, we show the ability of the model to estimate the true conditional densities as well as its capacity to choose the version of the DMBPP model (fully–dependent, single–atoms, single–weights, or independent) that best accommodates to the complexity of the true data-generating mechanism. In the application, we compare the performance of our proposed model with the performance of a parametric Dirichlet regression model on a transformed version of the data.

6.1. Simulation study

We consider four simulation scenarios representing varying degrees of complexity and shapes as the predictor varies, based on mixtures of predictor-dependent Dirichlet densities that are not particular cases of the implemented model. In all cases, the predictor is univariate and uniformly distributed on the $(0, 1)$ interval. For Scenario I, both the weights and the parameters of the Dirichlet distributions depend on the predictor. Under this scenario, for small values of x the conditional density has one mode which splits into two and later merges into one again as the value of the predictor increases. For Scenario II, only the parameters of the Dirichlet densities depend on the predictor. Under this scenario, for small values of x the conditional density has three well separated modes, one at each corner of the simplex, which merge into two and later into only one irregularly shaped as the value of x increases. For Scenario III only weights depend on the predictor. Under this scenario, for small values of x the conditional density has only one mode which splits into two and later merges into one mode centered roughly in the middle of the simplex as the value of x increases. Finally, Scenario IV is given by a predictor-independent Dirichlet density. The specification of true conditional densities for Scenarios I – IV is given in Table 1.

Table 1

Simulation Study: true conditional density functions considered in the simulation study. Here $w_1(x) = \frac{x}{4-3x}$, $\boldsymbol{\theta}_1(x) = (25 - 20x, 5 + 25x, 3)$, $\boldsymbol{\theta}_2(x) = (5, 5 + 15x, 30 - 17x)$, $\boldsymbol{\theta}_3(x) = (5 + 9x, 30 + 9x, 3 + 9x)$, *and* $x \in (0, 1)$.

Scenario	$f_0(\boldsymbol{y} \mid x)$
I	$w_1(x)\mathrm{dir}(\boldsymbol{y} \mid \boldsymbol{\theta}_1(x)) + (1 - w_1(x))\mathrm{dir}(\boldsymbol{y} \mid \boldsymbol{\theta}_2(x))$
II	$0.6\ \mathrm{dir}(\boldsymbol{y} \mid \boldsymbol{\theta}_1(x)) + 0.2\ \mathrm{dir}(\boldsymbol{y} \mid \boldsymbol{\theta}_2(x)) + 0.2\ \mathrm{dir}(\boldsymbol{y} \mid \boldsymbol{\theta}_3(x))$
III	$w_1(x)\mathrm{dir}(\boldsymbol{y} \mid (10, 12, 12)) + (1 - w_1(x))\mathrm{dir}(\boldsymbol{y} \mid (24, 6, 6))$
IV	$\mathrm{dir}(\boldsymbol{y} \mid (35, 25, 40))$

We consider three sample sizes, $n = 250$, $n = 500$, and $n = 1,000$ for each scenario. A Monte Carlo sample size of 100 was considered for each scenario and sample size. Following Zellner (1983), we consider $\boldsymbol{\Sigma}_l^\eta = \tau_l^\eta(\mathbb{X}^t\mathbb{X})^{-1}$ and $\boldsymbol{\Sigma}_l^z = \tau_l^z(\mathbb{X}^t\mathbb{X})^{-1}$, for $l = 1, 2$, where $\mathbb{X}$ denotes the design matrix without including the intercept, τ_1^η and τ_1^z are small positive values, while τ_2^η and τ_2^z are large positive values. In Appendix F we provide the justification for the particular choices. Finally, we set $\sigma_\eta^2 = \sigma_{\boldsymbol{z}}^2 = 100$. For the prior of the binary latent variables, $(\gamma^\eta, \gamma^{\boldsymbol{z}})$, we set $\pi_1 = 1/t^2$, $\pi_2 = \pi_3 = (t-1)/2t^2$, and $\pi_4 = (t-1)/t$, with $t > 1$. A priori, larger values of t favor more parsimonious models. Two prior specifications were employed by setting $t = 2$ (Prior I) and $t = 10$ (Prior II). Under Prior I, the prior probability of the covariate independent model is $\pi_4 = 0.50$, followed by the prior probability of the fully covariate dependent model, which is $\pi_1 = 0.25$. Prior II strongly favors parsimonious models. Under this specification, the prior probability of the covariate independent model is $\pi_4 = 0.90$, while the prior probability of its fully covariate dependent counterpart is only $\pi_1 = 0.01$. Finally, to complete the prior specification we consider $\lambda = 25$.

A single Markov chain was generated for each simulated data set. For $n = 250$ and $n = 500$ a chain of length $110,000$ was generated and the posterior inference was based on a reduced chain of $10,000$ samples obtained after a burn-in period of $10,000$ and keeping 1 every 10 samples. A similar specification was considered for $n = 1,000$, but considered a burn-in period of $50,000$ samples in such cases. To assess the performance of the proposed model in estimating the true data generating mechanism, we compute an estimate of the integrated–L_1 and L_∞ distances, denoted by $\widehat{IL}_1$ and $\widehat{L}_\infty$, respectively. Specifically, we compute

$$\widehat{IL}_1 = \frac{1}{L}\frac{1}{M}\sum_{j=1}^{L}\sum_{i=1}^{M}|\hat{f}(\boldsymbol{y}_i \mid \boldsymbol{x}_j) - f_0(\boldsymbol{y}_i \mid \boldsymbol{x}_j)|,$$
$$\widehat{L}_\infty = \max_i \max_j |\hat{f}(\boldsymbol{y}_i \mid \boldsymbol{x}_j) - f_0(\boldsymbol{y}_i \mid \boldsymbol{x}_j)|,$$

where $\widehat{f}(\cdot \mid x)$ denotes the posterior mean of the conditional density, $f_0(\cdot \mid x)$ denotes the true conditional density, and $\{\boldsymbol{y}_i\}_{i=1}^{M}$ and $\{\boldsymbol{x}_j\}_{j=1}^{L}$ define an equally spaced grid of Δ_m and $\mathscr{X}$, respectively.

To assess the model's ability to choose the version that best accommodates to the complexity of the underlying true data-generating distribution, we select the combination of $(\gamma^\eta, \gamma^{\boldsymbol{z}})$ that concentrates the highest posterior probability and compare it to the true predictor dependency structure of the simulation scenario. Recall that $(\gamma^\eta, \gamma^{\boldsymbol{z}})$ control which part of the model depends on the predictor and that each of the simulation scenarios depend on the predictor in different ways. Scenario I involves predictors in weights and in Dirichlet densities, Scenario II only in Dirichlet densities, Scenario III only in weights, and Scenario IV does not depend on predictors at all.

Table 2 shows the mean, across replicates, of the integrated–L_1 distance between the true and the posterior mean for each simulation scenario, sample size, and spike-and-slab prior. As expected, the integrated–L_1 distance decreases as the sample size increases for each simulation scenario under both spike-and-slab priors. For small samples sizes ($n = 250, 500$), the smallest integrated–L_1 distances are observed for Scenario III, the single–atoms true model, while for $n = 1000$, the smallest integrated–L_1 distance is observed for Scenario IV, the predictor independent true model. The largest integrated–L_1 distances for small sample sizes are observed for Scenario IV, while for $n = 1000$ the largest distance is observed for Scenario II, the single–weights true model. The model seems to be robust regarding the choice of the spike-and-slab prior. Similar results are obtained when the L_∞ distance is considered, which are shown in Appendix G.

Table 3 shows the proportion of times, across Monte Carlo replicates, that the selected predictor dependency structure of the fit model agrees with the one of the true model. The proportion increases as the sample size increases for each simulation scenario and spike-and-slab prior specification. Remarkably, for Scenarios I and IV, the DMBPP model is able to choose the version of the model that is in agreement with the predictor dependency structure of the true model

TABLE 2
Mean, across Monte Carlo replicates, of the integrated L_1 distance between the truth and the posterior mean of the conditional densities for each simulation scenario, spike-and-slab prior (Prior I and Prior II), and sample size (n).

	Prior I			Prior II		
Scenario	$n = 250$	$n = 500$	$n = 1,000$	$n = 250$	$n = 500$	$n = 1,000$
I	0.413	0.345	0.326	0.414	0.349	0.322
II	0.479	0.426	0.411	0.482	0.434	0.413
III	0.411	0.301	0.271	0.410	0.301	0.267
IV	0.599	0.380	0.234	0.599	0.380	0.230

TABLE 3
Proportion of times, across Monte Carlo replicates, in which the true predictor dependency structure is selected for each simulation scenario, spike-and-slab prior (Prior I and Prior II), and sample size (n).

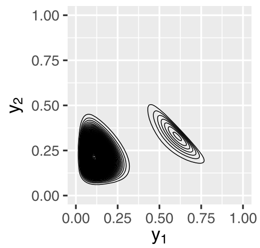
	Prior I			Prior II		
Scenario	$n = 250$	$n = 500$	$n = 1000$	$n = 250$	$n = 500$	$n = 1000$
I	1.000	1.000	1.000	1.000	1.000	1.000
II	0.440	0.670	0.870	0.990	1.000	0.980
III	0.960	0.960	0.980	0.960	0.970	0.980
IV	1.000	1.000	1.000	1.000	1.000	1.000

for every replicated data set and sample size. For Scenario III, the proportion of times that the chosen version of the model agrees with the true model increases from 0.96 to 0.98 as the sample size increases from 250 to 1000. The true model for which it is most difficult to choose the version of the model that agrees with the predictor dependency structure of the true model, is Scenario II (single–weights model) and under Prior I. Interestingly, it seems that the ability of the model to choose the version of the model that best fits the data is not completely related to the capacity of the model to estimate the conditional densities. For example, for sample sizes 250 and 500, the smallest integrated L_1 mean distances are observed for Scenario III, while the binary latent variable estimates agree with the predictor dependency structure of the true model the most for Scenarios I and IV. Again, the results are robust regarding the model selection prior distribution.

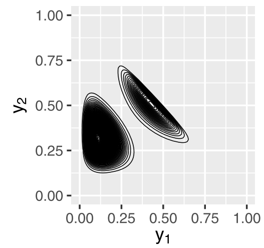
Figures 1 to 4 display the contour plot of the mean, across simulation, of the posterior mean of the conditional density for selected values of the predictor, each sample size and simulation scenario, under Prior I for $(\gamma^{\eta}, \gamma^{\boldsymbol{z}})$. The results are consistent with the previous discussion. For the different values of the predictor, the figures show how close the estimates are to the true model and how they improve as the sample size increases.

6.2. Application to solid waste in Colombia

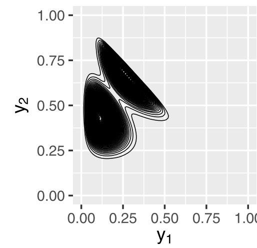
In this section, we analyze data about solid waste generated in a residential area in the city of Santiago de Cali, Colombia. The data set was collected to



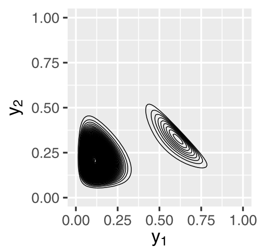
(a) True, $x = 0.25$



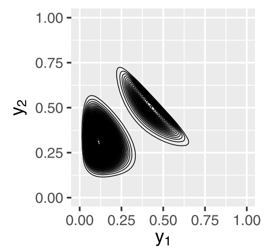
(b) True, $x = 0.50$



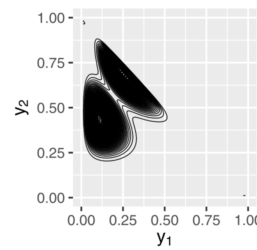
(c) True, $x = 0.75$



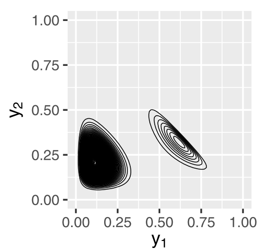
(d) $n = 250$, $x = 0.25$



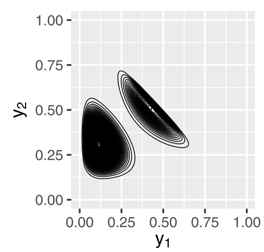
(e) $n = 250$, $x = 0.50$



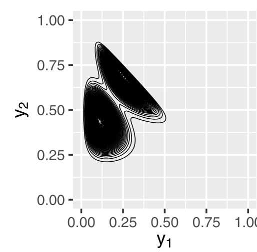
(f) $n = 250$, $x = 0.75$



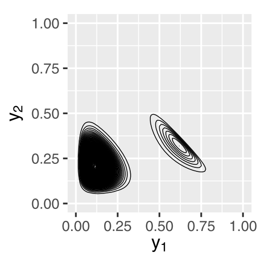
(g) $n = 500$, $x = 0.25$



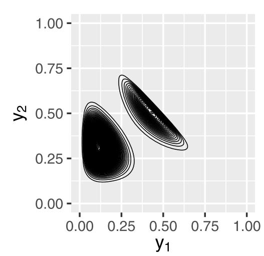
(h) $n = 500$, $x = 0.50$



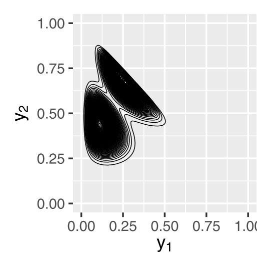
(i) $n = 500$, $x = 0.75$



(j) $n = 1000$, $x = 0.25$

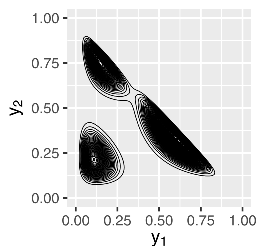


(k) $n = 100$, $x = 0.50$

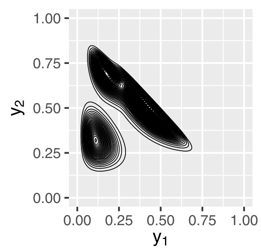


(l) $n = 1000$, $x = 0.75$

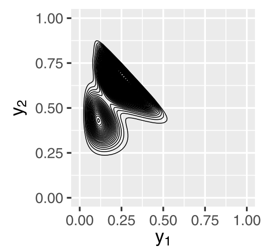
Fig 1. *Simulation study - Scenario I: contour plots of the true density (first row) and mean across replicates of the posterior mean of the conditional density for* $n = 250$ *(second row),* $n = 500$ *(third row), and* $n = 1000$ *(fourth row). The results are shown under Prior I for* $(\gamma^{\eta}, \gamma^{\mathbf{z}})$. *Results are displayed for* $x = 0.25$ *(first column),* $x = 0.50$ *(second column), and* $x = 0.75$ *(third column).*



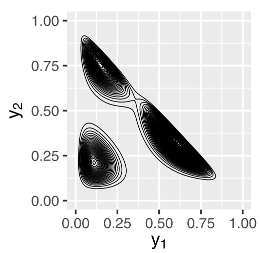
(a) True, $x = 0.25$



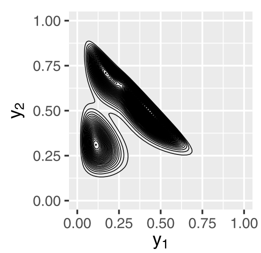
(b) True, $x = 0.50$



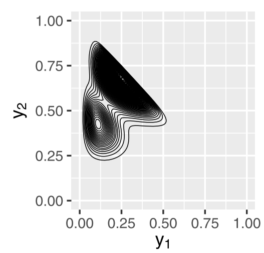
(c) True, $x = 0.75$



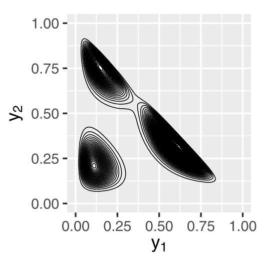
(d) $n = 250$, $x = 0.25$



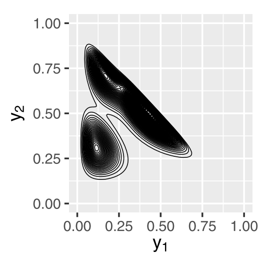
(e) $n = 250$, $x = 0.50$



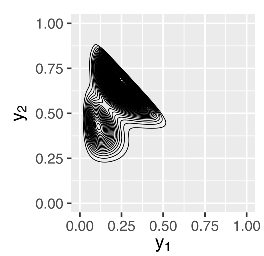
(f) $n = 250$, $x = 0.75$



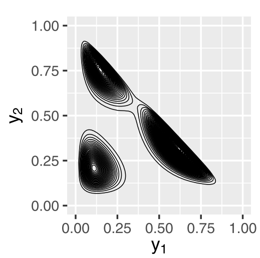
(g) $n = 500$, $x = 0.25$



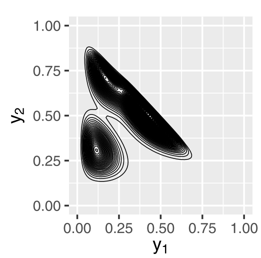
(h) $n = 500$, $x = 0.50$



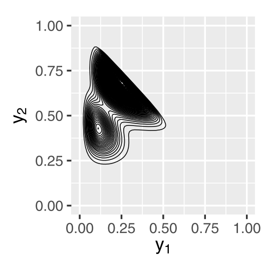
(i) $n = 500$, $x = 0.75$



(j) $n = 1000$, $x = 0.25$

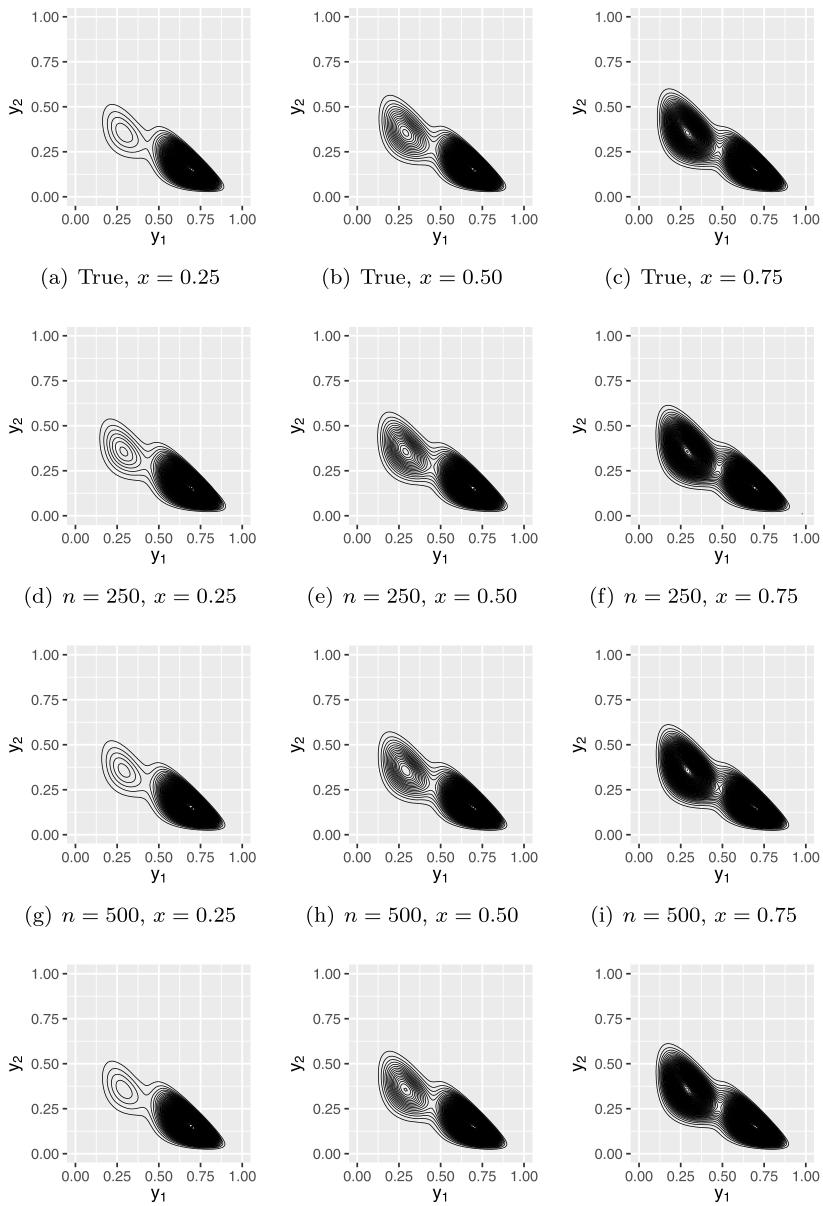


(k) $n = 100$, $x = 0.50$



(l) $n = 1000$, $x = 0.75$

FIG 2. *Simulation study - Scenario II: contour plots of the true density (first row) and mean across replicates of the posterior mean of the conditional density for* $n = 250$ *(second row),* $n = 500$ *(third row), and* $n = 1000$ *(fourth row). The results are shown under Prior I for* $(\gamma^{\eta}, \gamma^{\mathbf{z}})$*. Results are displayed for* $x = 0.25$ *(first column),* $x = 0.50$ *(second column), and* $x = 0.75$ *(third column).*



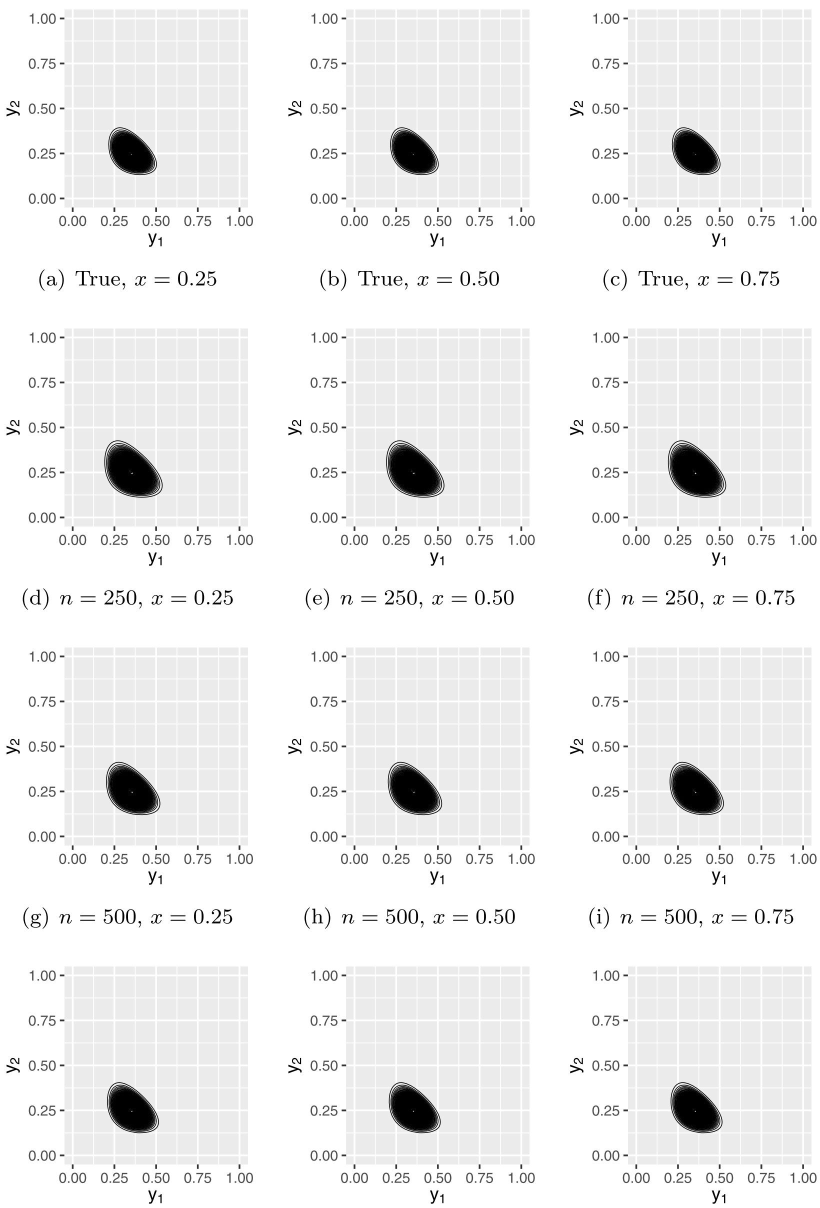
(a) True, $x = 0.25$ (b) True, $x = 0.50$ (c) True, $x = 0.75$

(d) $n = 250$, $x = 0.25$ (e) $n = 250$, $x = 0.50$ (f) $n = 250$, $x = 0.75$

(g) $n = 500$, $x = 0.25$ (h) $n = 500$, $x = 0.50$ (i) $n = 500$, $x = 0.75$

(j) $n = 1000$, $x = 0.25$ (k) $n = 100$, $x = 0.50$ (l) $n = 1000$, $x = 0.75$

FIG 3. *Simulation study - Scenario III: contour plots of the true density (first row) and mean across replicates of the posterior mean of the conditional density for $n = 250$ (second row), $n = 500$ (third row), and $n = 1000$ (fourth row). The results are shown under Prior I for $(\gamma^{\eta}, \gamma^{\mathbf{z}})$. Results are displayed for $x = 0.25$ (first column), $x = 0.50$ (second column), and $x = 0.75$ (third column).*



(a) True, $x = 0.25$ (b) True, $x = 0.50$ (c) True, $x = 0.75$

(d) $n = 250$, $x = 0.25$ (e) $n = 250$, $x = 0.50$ (f) $n = 250$, $x = 0.75$

(g) $n = 500$, $x = 0.25$ (h) $n = 500$, $x = 0.50$ (i) $n = 500$, $x = 0.75$

(j) $n = 1000$, $x = 0.25$ (k) $n = 100$, $x = 0.50$ (l) $n = 1000$, $x = 0.75$

FIG 4. *Simulation study - Scenario IV: contour plots of the true density (first row) and mean across replicates of the posterior mean of the conditional density for $n = 250$ (second row), $n = 500$ (third row), and $n = 1000$ (fourth row). The results are shown under Prior I for $(\gamma^{\eta}, \gamma^{\mathbf{z}})$. Results are displayed for $x = 0.25$ (first column), $x = 0.50$ (second column), and $x = 0.75$ (third column).*

estimate the per capita daily production and characterization of solid waste in the city. The data set contains information about 261 block sides, for which solid waste was separated in different kinds of materials, including food, hygienic, and others. Finally, the proportion of these materials were registered for each block side. Additionally, the socioeconomic level of the residents in the area was recorded including the categories "low–low", "low", "medium–low", "medium", "medium–high", and "high". See Klinger et al. (2009) for more details regarding this data set.

In this analysis, the proportion of food, hygienic solid, and other type of waste were considered as the response vector on the 2–dimensional simplex. The socioeconomic level was considered as a categorical covariate with a dummy variable representation leading to $p = 6$ predictors. Here, we consider a model specification similar to the one detailed in Section 5. Here we set $\sigma^2_\eta = \sigma^2_{\boldsymbol{z}} = 100$ and $\lambda = 25$. We refer the reader to Appendix H for a description regarding the selection of τ^η_l and $\tau^{\boldsymbol{z}}_l$.

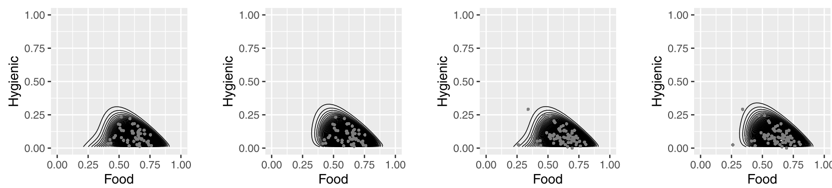
A single Markov chain with $300,000$ samples was generated. Posterior inference was based on a reduced chain with $10,000$ samples obtained after a $100,000$ burn-in period and keeping 1 every 20 samples. To assess the performance of the proposed model, we also fit a parametric Dirichlet regression (PDR) model to the data. Due to presence of zero-coordinate vectors in the data, we transform the response vectors, $\boldsymbol{y}$, using the transformation proposed by Smithson & Verkuilen (2006), given by $\boldsymbol{y}^* = [\boldsymbol{y}(n-1) + 1/(m+1)]/n$, where n is the size of the sample and m is the dimension of the simplex. Thus the parametric model was applied to the transformed responses $\boldsymbol{y}^*$, such that

$$\boldsymbol{y}^*_i \mid \boldsymbol{x}_i, \boldsymbol{\gamma} \sim \mathrm{dir}(\boldsymbol{\gamma}(\boldsymbol{x}_i)),$$

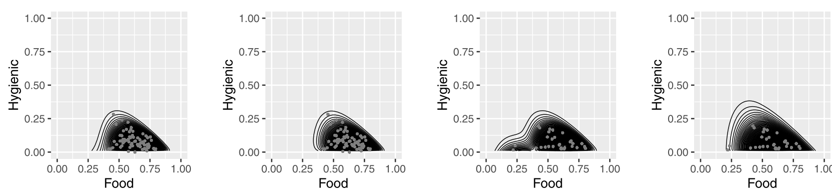
where $\boldsymbol{\gamma}(\boldsymbol{x}_i) = (\gamma_1(\boldsymbol{x}_i), \ldots, \gamma_m(\boldsymbol{x}_i))$, with $\log(\gamma_l(\boldsymbol{x})) = \boldsymbol{x}^t\boldsymbol{\beta}_l$, $l = 1, \ldots, m$. We complete the model specification by assuming $\boldsymbol{\beta}_l \sim N_{p+1}(\boldsymbol{m}, \boldsymbol{\Sigma})$, with $\boldsymbol{m} = \boldsymbol{0}$, $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_{p+1}$, $\sigma^2 = 100$, and $\boldsymbol{I}_p$ being a $p \times p$ identity matrix. The models were compared by means of their posterior predictive abilities, quantified by the log pseudo marginal likelihood (LPML) and the widely applicable information criterion (WAIC) (Watanabe & Opper, 2010). The LPML, developed by Geisser & Eddy (1979), is given by $\sum_{i=1}^n \log p_M(\boldsymbol{y}_i \mid \boldsymbol{Y}_{-i})$, where $p_M(\boldsymbol{y}_i \mid \boldsymbol{Y}_{-i})$ is the posterior predictive distribution for observation $\boldsymbol{y}_i$, based on the data $\boldsymbol{Y}_{-i}$, under model M, with $\boldsymbol{Y}_{-i}$ denoting the observed data matrix after removing the ith observation. The $p_M(\boldsymbol{y}_i \mid \boldsymbol{Y}_{-i})$ is also known as the conditional predictive ordinate of observation i under model M and the method of Gelfand & Dey (1994) was used in its computation. The WAIC is given by

$$\mathrm{WAIC} = -\frac{1}{n}\sum_{i=1}^n \log E_{post}\left[p_M(\boldsymbol{y}_i \mid \boldsymbol{\theta})\right] + \frac{1}{n}\sum_{i=1}^n Var_{post}\left[\log p_M(\boldsymbol{y}_i \mid \boldsymbol{\theta})\right], \quad (12)$$

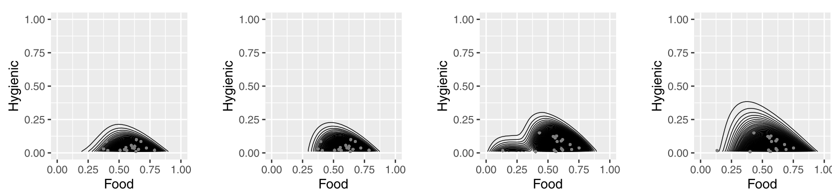
where $p_M(\boldsymbol{y}_i \mid \boldsymbol{\theta})$ is the density function for observation $\boldsymbol{y}_i$, given parameter $\boldsymbol{\theta}$, under model M, and E_{post} and Var_{post} denote the posterior mean and posterior variance, respectively. The second term on the right hand side of expression (12)



(a) DMBPPP: low-low (b) PDR: low-low (c) DMBPPP: low (d) PDR: low



(e) DMBPPP: medium-low (f) PDR: medium-low (g) DMBPPP: medium (h) PDR: medium



(i) DMBPPP: medium-high (j) PDR: medium-high (k) DMBPPP: high (l) PDR: high

Fig 5. *Solid waste data: Posterior mean of the conditional density. The results for the for the DMBPP model under spike-and-slab Prior I are presented in the first and third columns. The results for the PDR model are presented in the second and fourth columns. Panels (a) and (b), (c) and (d), (e) and (f), (i) and (j), and (k) and (l) present the results for socioeconomic level low-low, low, medium-low, medium, medium-high, and high, respectively. The x–axis and y–axis denote the proportion of food and hygienic waste, respectively.*

is a penalty for overfitting. In what follows, we compute WAIC as described by Gelman et al. (2013), page 173, and report $-nWAIC$. Models with greater values of LPML and $-nWAIC$ are to be preferred. For the parametric model, we consider the transformed data $\boldsymbol{y}_i = \boldsymbol{y}_i^*$ in the computation of the LPML and WAIC criteria.

Figure 5 displays the conditional density estimates for the DMBPP model and the PDR model, for each socioeconomic level. The results for the DMBPP model re shown under spike-and-slab Prior I. They suggest that different socioeconomic levels show different recycling behaviors and the advantages of the nonparametric model are evident. The DMBPP model shows to be more flexible in estimating the conditional densities than the parametric model for varying values of the predictor, specially when the socioeconomic levels are "medium" and "high". We highlight that the parametric fit was only possible due to a pre-transformation of the data.

The LPML for the DMBPP and PDR model was 778 and 649, respectively. The $-nWAIC$ the DMBPP and PDR model was 778 and 649 was 778 and 650, respectively. These goodness-of-fit criteria support and agree in that the DMBPP model provides a better fit for this data set than the PDR model. For DMBPP model, the posterior probability of $(\gamma^{\eta}, \gamma^{\mathbf{z}}) = (0, 0)$ was approximately equal to zero, which is a formal test for the hypothesis that the densities for solid waste are the same across the socioeconomic levels. A probability close to zero is interpreted as little evidence in favor of this hypothesis. Additional results for the DMBPP model under spike-and-slap Prior II can be found in Appendix I, which are robust regarding the prior specification.

7. Discussion

We have proposed a novel and general class of probability models for sets of predictor-dependent probability distributions supported on simplex spaces. The proposal corresponds to an extension of dependent univariate Bernstein polynomial processes proposed by Barrientos et al. (2017) and is based on the modified class of MBP proposed by Barrientos et al. (2015).

The proposed model class has appealing theoretical properties such as full support, well behaved correlation function, and consistent posterior distribution. The incorporation of a spike-and-slab prior for the predictor dependent stochastic processes involved in the model adapts well to the complexity of the underlying data-generating distribution. The approach also allows the user to formally test whether all predictors are simultaneously related to the compositional response. The study of the theoretical properties of the model selection component of our approach is the subject of ongoing research.

Appendix A: Dealing with compositional data and zero-valued entries

In this appendix, we explain how the proposed approach is able to handle compositional observations with zero-valued entries. To simplify the discussion, we limit ourselves to the scenario where no predictors are present, but the arguments extend naturally to the predictor-dependent case. In some circumstances, such as in the Colombian solid waste application, some compositional observations may have entries equal to zero. As a result, analysts must determine whether such zeros are systematic and, if so, whether they should be handled using a zero-inflated modeling approach. Analysts might use standard modeling techniques if there is no indication that the statistical model must be degenerated to account for zero-valued entries (as we assume throughout this paper). Analysts should be aware that using standard modeling techniques can cause potential problems. In what follows, we illustrate some of those potential problems in a specific, yet common, scenario.

Suppose we will model compositional data with zero-valued entries using either a single Dirichlet distribution or a standard mixture of Dirichlet densities.

Assume that, out of a sample of n observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$, only the first entry of $\boldsymbol{y}_1$ is exactly equal to zero, i.e., $y_{1,1} = 0$. For the remaining entries of $\boldsymbol{y}_1$ and all entries of $\boldsymbol{y}_2, \ldots, \boldsymbol{y}_n$, assume they take values on $(0, 1)$. If we use the Dirichlet distribution to model $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ directly, we need first to notice that the likelihood is equal to

$$y_{1,1}^{\gamma_1 - 1} \times \mathcal{L}_{-y_{1,1}}\left((\gamma_1, \ldots, \gamma_{m+1}); (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)\right),$$

where

$$\begin{aligned}\mathcal{L}_{-y_{1,1}}&\left((\gamma_1, \ldots, \gamma_{m+1}); (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)\right) \\ &= \left\{ \frac{\Gamma(\gamma_1 + \ldots + \gamma_{m+1})}{\Gamma(\gamma_1) \times \ldots \times \Gamma(\gamma_{m+1})} y_{1,2}^{\gamma_2 - 1} \times \ldots \times y_{1,m}^{\gamma_m - 1} \left(1 - \sum_{j=1}^{m} y_{i,j}\right)^{\gamma_{m+1} - 1} \right\} \\ &\qquad\qquad \times \prod_{i=2}^{n} \operatorname{dir}\left(\boldsymbol{y}_i \mid (\gamma_1, \ldots, \gamma_{m+1})\right),\end{aligned}$$

and

$$\begin{aligned}\operatorname{dir}&\left(\boldsymbol{y}_i \mid (\gamma_1, \ldots, \gamma_{m+1})\right) = \\ &\frac{\Gamma(\gamma_1 + \ldots + \gamma_{m+1})}{\Gamma(\gamma_1) \times \ldots \times \Gamma(\gamma_{m+1})} y_{i,1}^{\gamma_1 - 1} \times \ldots \times y_{i,m}^{\gamma_m - 1} (1 - \sum_{j=1}^{m} y_{i,j})^{\gamma_{m+1} - 1}.\end{aligned}$$

Since we are assuming that $y_{1,1} = 0$, we have to consider the following three cases:

- If $\gamma_1 < 1$, then $(\gamma_1, \ldots, \gamma_{m+1}) \mapsto y_{1,1}^{\gamma_1 - 1} \mathcal{L}_{-y_{1,1}}\left((\gamma_1, \ldots, \gamma_{m+1}); (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)\right) \equiv \infty$.
- If $\gamma_1 > 1$, then $(\gamma_1, \ldots, \gamma_{m+1}) \mapsto y_{1,1}^{\gamma_1 - 1} \mathcal{L}_{-y_{1,1}}\left((\gamma_1, \ldots, \gamma_{m+1}); (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)\right) \equiv 0$.
- If $\gamma_1 = 1$, then $(\gamma_1, \ldots, \gamma_{m+1}) \mapsto y_{1,1}^{\gamma_1 - 1} \mathcal{L}_{-y_{1,1}}\left((\gamma_1, \ldots, \gamma_{m+1}); (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)\right)$ is not constant and will take values on $(0, \infty)$. For this case, we adopt the convention that $0^0 = 1$.

Notice that, only in the case $\gamma_1 = 1$, we will be able to make inferences about

$$(\gamma_2, \ldots, \gamma_{m+1}).$$

Since the constraint $\gamma_1 = 1$ is entirely driven by the fact that the first entry of $\boldsymbol{y}_1$ is equal to zero, adopting this modeling approach is inadequate.

We might also consider using a mixture model for it is a more general and flexible approach. For example, consider a standard DP mixture of Dirichlet densities of the form

$$\sum_{j=1}^{\infty} w_j \operatorname{dir}\left(\cdot \mid \boldsymbol{\gamma}_j = (\gamma_{1,j}, \ldots, \gamma_{m+1,j})\right).$$

Using this mixture model will lead to issues similar to those described above if the prior distribution for the atoms, $\boldsymbol{\gamma}_j$, is absolutely continuous with respect to Lebesgue. We could overcome this issue if the prior distribution for $\boldsymbol{\gamma}_j$ is such that $P\{\gamma_{1,j} = 1\} > 0$, which will allow making inferences while remaining flexible. For a more general scenario where we assume that zeros might occur in multiple entries and observations, we will be able to make inferences as long as the assumption $P\{\gamma_{l,j} = 1 \,:\, j \in \mathscr{J}\} > 0$ is met for each $\mathscr{J} \subseteq \{1, \ldots, m\}$. This particular assumption is naturally met (under mild conditions) when using the class of mixture of Dirichlet densities derived from the MBP described in Section 2, that is $\sum_{j=1}^{\infty} w_j \mathrm{dir}\left(\cdot \mid \alpha(k, \lceil k\boldsymbol{\theta}_j \rceil)\right)$. The assumption is met when the prior for $\boldsymbol{\theta}_j$ has full support on Δ_m^0 or, equivalently, the prior for $\boldsymbol{\theta}_j$ has positive density (with respect to Lebesgue) on Δ_m^0.

Appendix B: Formal definition of special cases of the general model

Definition 2. *Let $\mathscr{V}$ and $\mathscr{H}$ be two sets of functions as defined before. Let $F = \{F_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$ be a $\mathscr{P}(\Delta_m)$-valued stochastic process such that:*

(i) $v_1, v_2, \ldots,$ *are independent* $[0, 1]$*-valued random variables with common distribution indexed by a finite-dimensional parameter* $\boldsymbol{\Psi}_v$.

(ii) $\boldsymbol{z}_j = \{\boldsymbol{z}_j(\boldsymbol{x}) : \boldsymbol{x} \in \mathscr{X}\}$, $j \geq 1$, *are independent and identically distributed real-valued stochastic processes with law indexed by a finite-dimensional parameter* $\boldsymbol{\Psi}_z$.

(iii) $k \in \mathbb{N}$ *is a discrete random variable with distribution indexed by a finite-dimensional parameter* λ.

(iv) For every $\boldsymbol{x} \in \mathscr{X}$, *the density function of* $F_{\boldsymbol{x}}$, *w.r.t. Lebesgue measure, is given by the following dependent mixture of Dirichlet densities,*

$$f_{\boldsymbol{x}}(\cdot) = \sum_{j=1}^{\infty} w_j \mathrm{dir}\left(\cdot \mid \alpha(k, \lceil k\boldsymbol{\theta}_j(\boldsymbol{x}) \rceil)\right), \tag{13}$$

where $\alpha(k, \boldsymbol{j}) = (\boldsymbol{j},\ k + m - \|\mathbf{j}\|_1)$, $\boldsymbol{\theta}_j(\boldsymbol{x})$ *and* $\lceil k\boldsymbol{\theta}_j(\boldsymbol{x}) \rceil$ *are defined as in Definition 1, and* $w_j = v_j \prod_{l<j} [1 - v_l]$.

The process $F = \{F_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$ will be referred to as single–weight dependent MBP process with parameters $(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_z, \mathscr{H})$, and denoted by $w\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_z, \mathscr{H})$ and $w\mathrm{DMBPP}$ for short.

Definition 3. *Let $\mathscr{V}$ and $\mathscr{H}$ be two sets of functions as defined before. Let $F = \{F(\boldsymbol{x}, \cdot\,) : \boldsymbol{x} \in \mathscr{X}\}$ be a $\mathscr{P}(\Delta_m)$-valued stochastic process such that:*

(i) $\eta_j = \{\eta_j(\boldsymbol{x}) : \boldsymbol{x} \in \mathscr{X}\}$, $j \geq 1$, *are independent and identically distributed real-valued stochastic processes with law indexed by a finite-dimensional parameter* $\boldsymbol{\Psi}_\eta$.

(ii) $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots,$ *are independent* Δ_m^0*-valued random vectors with common distribution indexed by a finite-dimensional parameter* $\boldsymbol{\Psi}_{\boldsymbol{\theta}}$.

(iii) $k \in \mathbb{N}$ *is a discrete random variable with distribution indexed by a finite-dimensional parameter* λ.

(iv) For every $\boldsymbol{x} \in \mathscr{X}$, the density function of $F_{\boldsymbol{x}}$, w.r.t. Lebesgue measure, is given by the following dependent mixture of Dirichlet densities,

$$f_{\boldsymbol{x}}(\cdot) = \sum_{j=1}^{\infty} w_j(\boldsymbol{x}) \mathrm{dir}\left(\cdot \mid \alpha(k, \lceil k\boldsymbol{\theta}_j \rceil)\right), \tag{14}$$

where $w_j(\boldsymbol{x})$ are defined as in Definition 1 and

$$\lceil k\boldsymbol{\theta}_j \rceil = (\lceil k\theta_{j1} \rceil, \ldots, \lceil k\theta_{jm} \rceil).$$

The process $F = \{F_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$ will be referred to as single–atoms dependent MBP process with parameters $(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$, and denoted by θDMBPP$(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$ and θDMBPP for short.

Appendix C: Topological bases and sub-bases

A sub-base for the weak product topology for the space

$$\mathscr{P}(\Delta_m)^{\mathscr{X}} = \prod_{\boldsymbol{x} \in \mathscr{X}} \mathscr{P}(\Delta_m),$$

is given by sets of the form $B^W_{f,\epsilon,\boldsymbol{x}_0}(\{Q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}) = \prod_{\boldsymbol{x} \in \mathscr{X}} \Delta^W_{f,\epsilon,\boldsymbol{x}_0}(Q_{\boldsymbol{x}})$, where

$$\Delta^W_{f,\epsilon,\boldsymbol{x}_0}(Q_{\boldsymbol{x}}) = \begin{cases} \mathscr{P}(\Delta_m), & \text{if } \boldsymbol{x} \neq \boldsymbol{x}_0, \\ \left\{ M_{\boldsymbol{x}} \in \mathscr{P}(\Delta_m) : \left| \int_{\Delta_m} f \, dM_{\boldsymbol{x}} - \int_{\Delta_m} f \, dQ_{\boldsymbol{x}} \right| < \epsilon \right\}, & \text{if } \boldsymbol{x} = \boldsymbol{x}_0, \end{cases}$$

for every $f : \Delta_m \longrightarrow \mathbb{R}$ bounded continuous function, $\epsilon > 0$, $\boldsymbol{x}_0 \in \mathscr{X}$ and $Q_{\boldsymbol{x}} \in \mathscr{P}(\Delta_m)$.

Let $\mathscr{D}(\Delta_m) \subset \mathscr{P}(\Delta_m)$ be the space of all probability measures defined on Δ_m that are absolutely continuous w.r.t. Lebesgue measure and with continuous density function. A sub-base for the L_∞ product topology for the space

$$\mathscr{D}(\Delta_m)^{\mathscr{X}} = \prod_{\boldsymbol{x} \in \mathscr{X}} \mathscr{D}(\Delta_m)$$

is given by sets of the form $B^{L_\infty}_{\epsilon,\boldsymbol{x}_0}(\{Q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}) = \prod_{\boldsymbol{x} \in \mathscr{X}} \Delta^{L_\infty}_{\epsilon,\boldsymbol{x}_0}(Q_{\boldsymbol{x}})$, where

$$\Delta^{L_\infty}_{\epsilon,\boldsymbol{x}_0}(Q_{\boldsymbol{x}}) = \begin{cases} \mathscr{D}(\Delta_m), & \text{if } \boldsymbol{x} \neq \boldsymbol{x}_0, \\ \left\{ M_{\boldsymbol{x}} \in \mathscr{D}(\Delta_m) : \sup_{\boldsymbol{y} \in \Delta_m} \mid m_{\boldsymbol{x}}(\boldsymbol{y}) - q_{\boldsymbol{x}}(\boldsymbol{y}) \mid < \epsilon \right\}, & \text{if } \boldsymbol{x} = \boldsymbol{x}_0, \end{cases}$$

for every $\epsilon > 0$, $\boldsymbol{x}_0 \in \mathscr{X}$ and $Q_{\boldsymbol{x}} \in \mathscr{D}(\Delta_m)$, where $m_{\boldsymbol{x}}$ and $q_{\boldsymbol{x}}$ denote the density function of $M_{\boldsymbol{x}}$ and $Q_{\boldsymbol{x}}$, respectively.

Now, assume that the predictor vector $\boldsymbol{x}$ contains only continuous predictors and that the predictor space $\mathscr{X}$ is compact. A base for the L_∞ topology for the space $\mathscr{D}(\Delta_m)^{\mathscr{X}} = \prod_{\boldsymbol{x} \in \mathscr{X}} \mathscr{D}(\Delta_m)$, is given by sets of the form

$$B^{L_\infty}_{\epsilon}(\{Q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}) = \left\{ \{M_{\boldsymbol{x}} : \boldsymbol{x} \Subset \mathscr{X}\} \in \mathscr{D}(\Delta_m)^{\mathscr{X}} : \sup_{\boldsymbol{x} \in \mathscr{X}} \sup_{\boldsymbol{y} \in \Delta_m} \mid m_{\boldsymbol{x}}(\boldsymbol{y}) - q_{\boldsymbol{x}}(\boldsymbol{y}) \mid < \epsilon \right\},$$

for every $\epsilon > 0$ and $Q_{\boldsymbol{x}} \in \mathscr{D}(\Delta_m)$.

Appendix D: Proof of theoretical results

D.1. Proof of Theorem 1

The proof of this theorem follows the same reasoning as the proof of Theorem 1 in Barrientos et al. (2015). First, we re-define the three versions of the model by means of a mapping S and the measurability of the process is then proved by showing that the mapping S is continuous. This is stated in Lemma 1 below, which is an extension of Lemma B.1.1 in the supplementary material of Barrientos et al. (2015).

The proofs of parts (i), (ii), and (iii) in Lemma 1 follow the same reasoning as the corresponding proofs for Lemma B.1.1. The main difference in the following proofs comes from the number of weights that the multivariate Bernstein polynomial on the $m-$dimensional symplex has. For completeness, the proof of Theorem 1 is provided below.

Let $\mathbf{T}$, $\mathbf{T}^{\theta}$ and $\mathbf{T}^{w}$ be dependent stick-breaking processes of the form:

- $\mathbf{T} = \{T_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$, where $T_{\boldsymbol{x}}(\omega, \cdot) = \sum_{j=1}^{\infty} w_j(\boldsymbol{x}, \omega)\delta_{\boldsymbol{\theta}_j(\boldsymbol{x},\omega)}(\cdot)$, where $w_j(\boldsymbol{x}, \omega)$ and $\boldsymbol{\theta}_j(\boldsymbol{x}, \omega)$ are defined as in Definition 1.
- $\mathbf{T}^{\theta} = \{T^{\theta}_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$, where $T^{\theta}_{\boldsymbol{x}}(\omega, \cdot) = \sum_{j=1}^{\infty} w_j(\boldsymbol{x}, \omega)\delta_{\boldsymbol{\theta}_j(\omega)}(\cdot)$, where $w_j(\boldsymbol{x}, \omega)$ and $\boldsymbol{\theta}_j(\omega)$ are defined as in Definition 2.
- $\mathbf{T}^{w} = \{T^{w}_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$, where $T^{w}_{\boldsymbol{x}}(\omega, \cdot) = \sum_{j=1}^{\infty} w_j(\omega)\delta_{\boldsymbol{\theta}_j(\boldsymbol{x},\omega)}(\cdot)$, where $w_j(\omega)$ and $\boldsymbol{\theta}_j(\boldsymbol{x}, \omega)$ are defined as in Definition 3.

Let S be a mapping defined on $\mathbb{N} \times \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}}$ of the form

$$S(k_0, \mathcal{Q}) := \{H(k_0, Q_{\boldsymbol{x}}) : \boldsymbol{x} \in \mathscr{X}\}, \tag{15}$$

where $k_0 \in \mathbb{N}$, $\mathcal{Q} = \{Q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}}$ and $H(k_0, Q_{\boldsymbol{x}})$ is the probability measure associated to the Bernstein polynomial of degree k_0 of the measure $Q_{\boldsymbol{x}}$. F can be expressed as $S(k, \mathbf{T})$, $S(k, \mathbf{T}^{\theta})$ or $S(k, \mathbf{T}^{w})$, when F corresponds to DMBPP$(\lambda, \boldsymbol{\Psi}_{\eta}, \boldsymbol{\Psi}_{\boldsymbol{z}}, \mathscr{V}, \mathscr{H})$, θDMBPP$(\lambda, \boldsymbol{\Psi}_{\eta}, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$ and wDMBPP$(\lambda, \boldsymbol{\Psi}_{v}, \boldsymbol{\Psi}_{\boldsymbol{z}}, \mathscr{H})$, respectively. Since $\mathbf{T}$, $\mathbf{T}^{\theta}$ and $\mathbf{T}^{w}$ are well-defined stochastic processes, to prove the measurability of F, it suffices to prove the measurability of S which is proven by showing that mapping S is continuous. For this, it is necessary to consider some topologies in the space where the mapping is valued and defined. This topologies and spaces are described below.

Let $\mathscr{T}_1$ be the weak product topology for the space $\mathscr{P}(\Delta_m)^{\mathscr{X}}$ and let $\mathscr{T}_2$ and $\mathscr{T}_3$ be the L_{∞} product topology and L_{∞} topology for the space $\mathscr{D}(\Delta_m)^{\mathscr{X}}$, respectively. A sub-base for the weak product topology, $\mathscr{T}_4$, for the space $\mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}} = \prod_{\boldsymbol{x}\in\mathscr{X}} \mathscr{P}(\tilde{\Delta}_m)$ is given by sets of the form $\tilde{B}^{W}_{f,\epsilon,\boldsymbol{x}_0}(\mathcal{Q}) = \prod_{\boldsymbol{x}\in\mathscr{X}} \tilde{\Delta}^{W}_{f,\epsilon,\boldsymbol{x}_0}(Q_{\boldsymbol{x}})$, where $\tilde{\Delta}^{W}_{f,\epsilon,\boldsymbol{x}_0}(Q_{\boldsymbol{x}}) = \Delta^{W}_{f,\epsilon,\boldsymbol{x}_0}(Q_{\boldsymbol{x}}) \bigcap \mathscr{P}(\tilde{\Delta}_m)$, with $\mathcal{Q} \in \mathscr{P}(\Delta_m)^{\mathscr{X}}$, $f : \Delta_m \longrightarrow \mathbb{R}$ a bounded continuous function, $\epsilon > 0$ and $\boldsymbol{x}_0 \in \mathscr{X}$. A sub-base for the product topology, $\mathscr{L}_1$, for the space $\mathbb{N} \times \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}}$ is given by sets of the form $B^{D\times W}_{f,\epsilon,\boldsymbol{x}_0}(\mathcal{Q}) = \prod_{\boldsymbol{x}\in\mathscr{X}} \left[\{k_0\} \times \tilde{\Delta}^{W}_{f,\epsilon,\boldsymbol{x}_0}(Q_{\boldsymbol{x}})\right]$. Finally, a sub-base for the product topology, $\mathscr{L}_2$, for the space $\mathbb{N} \times \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}}$ is given by sets of the form $B^{D\times L_\infty}_{\epsilon,N}(k_0, \mathcal{Q}) = \{k_0\} \times \tilde{\Delta}^{L_\infty}_{\epsilon,N}(\mathcal{Q})$, where $\tilde{\Delta}^{L_\infty}_{\epsilon,N}(\mathcal{Q})$ is given by

$$\left\{ \{M_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}} : \max_{\mathbf{j} \in \mathcal{H}^0_{N,m}} \sup_{\boldsymbol{x} \in \mathscr{X}} |M_{\boldsymbol{x}}(A_{\mathbf{j},N}) - Q_{\boldsymbol{x}}(A_{\mathbf{j},N})| < \epsilon \right\}, \quad (16)$$

where $k_0 \in \mathbb{N}$, $N \in \mathbb{N}$, $\epsilon > 0$, $A_{\mathbf{j},N} = \left(\frac{j_1-1}{N}, \frac{j_1}{N}\right] \times \ldots, \times \left(\frac{j_m-1}{N}, \frac{j_m}{N}\right]$ and $\mathcal{Q} \in \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}}$.

The following lemma states that mapping S defined by expression (15), is continuous under $\mathscr{T}_1$, $\mathscr{T}_2$ and $\mathscr{T}_3$ in the space where S is valued, thus ensuring that F is measurable under $\mathscr{B}_1$, $\mathscr{B}_2$ and $\mathscr{B}_3$, respectively.

Lemma 1. *Let S be a mapping defined as in Equation (15), then*

(i) $S : \left(\mathbb{N} \times \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}}, \mathscr{L}_1\right) \longrightarrow \left(\mathscr{P}(\Delta_m)^{\mathscr{X}}, \mathscr{B}_1\right)$,

(ii) $S : \left(\mathbb{N} \times \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}}, \mathscr{L}_1\right) \longrightarrow \left(\mathscr{D}(\Delta_m)^{\mathscr{X}}, \mathscr{B}_2\right)$,

(iii) $S : \left(\mathbb{N} \times \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}}, \mathscr{L}_2\right) \longrightarrow \left(\mathscr{D}(\Delta_m)^{\mathscr{X}}, \mathscr{B}_3\right)$,

are continuous.

The proof of each part of Lemma 1 is given below:

(i) Let $\mathcal{Q} \in \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}}$, $k_0 \in \mathbb{N}$ and

$$V\left(S(k_0, \mathcal{Q}); \epsilon\right) = \bigcap_{i=1}^{L} \bigcap_{j=1}^{K_i} B^W_{f_{ij},\epsilon,\boldsymbol{x}_i}\left(S(k_0, \mathcal{Q})\right),$$

where L, K_i, $i \in \{1, \ldots, L\}$, are positive integers, f_{ij}, $j = 1, \ldots, K_i$, $i = 1, \ldots, L$, are bounded continuous functions, $\epsilon > 0$ and $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L) \in \mathscr{X}^L$. The proof is based on finding and open set $U \in \mathscr{L}_1$ such that $(k_0, \mathcal{Q}) \in U$ and $S\left(U\right) \subseteq V\left(S(k_0, \mathcal{Q}); \epsilon\right)$.
Notice that for every $\mathcal{M} = \{M_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}}$,

$$\begin{aligned}
&\left| \int_{\Delta_m} f_{ij}\ dH(k_0, M_{\boldsymbol{x}_i}) - \int_{\Delta_m} f_{ij}\ dH(k_0, Q_{\boldsymbol{x}_i}) \right| \\
&\quad \le \int_{\Delta_m} |f_{ij}(\boldsymbol{y})| \sum_{\mathbf{j} \in \mathcal{H}^0_{k_0,m}} \left| M_{\boldsymbol{x}_i}\left(A_{\mathbf{j},k_0}\right) - Q_{\boldsymbol{x}_i}\left(A_{\mathbf{j},k_0}\right) \right| \mathrm{dir}\left(\ \boldsymbol{y} \mid \alpha(k_0, \mathbf{j})\right), \\
&\quad \le \frac{M_0(k_0 + m - 1)!}{m!(k_0 - 1)!}\ N_{k_0}(\mathcal{M}, \mathcal{Q}),
\end{aligned}$$

where $\alpha\left(k, \boldsymbol{j}\right) = (\boldsymbol{j},\ k + m - \|\mathbf{j}\|_1)$ and

$$N_{k_0}(\mathcal{M}, \mathcal{Q}) = \max_{i \in \{1,\ldots,L\}} \max_{\mathbf{j} \in \mathcal{H}^0_{k_0,m}} \left| M_{\boldsymbol{x}_i}\left(A_{\mathbf{j},k_0}\right) - Q_{\boldsymbol{x}_i}\left(A_{\mathbf{j},k_0}\right) \right|,$$

$M_0 = \max_{i \in \{1,\ldots,L\}} \max_{j \in \{1,\ldots,K_i\}} \sup_{\boldsymbol{y} \in \Delta_m} |f_{ij}(\boldsymbol{y})|$, $\|\quad \cdot\quad \|_1$ denotes the l_1–norm, and $A_{\mathbf{j},k_0} = \left(\frac{j_1-1}{k_0}, \frac{j_1}{k_0}\right] \times \ldots \times \left(\frac{j_m-1}{k_0}, \frac{j_m}{k_0}\right]$. From Lemma 1 in

Barrientos et al. (2012), there exists $\mathcal{Q}' = \{Q'_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}}$ such that for every $\boldsymbol{x} \in \mathscr{X}$, $Q'_{\boldsymbol{x}}$ is absolutely continuous w.r.t Lebesgue measure and such that,

$$N_{k_0}(\mathcal{Q}', \mathcal{Q}) \leq \frac{m!(k_0-1)!}{2M_0(k_0+m-1)!}\epsilon.$$

Since $Q'_{\boldsymbol{x}_i}$, $i = 1, \ldots, L$, is an absolutely continuous measure, w.r.t. Lebesgue measure, then $A_{\mathbf{j},k_0}$, $\mathbf{j} \in \mathcal{H}^0_{k_0,m}$, are sets of $Q'_{\boldsymbol{x}_i}$ continuity, i.e., the boundaries of $A_{\mathbf{j},k_0}$ have null $Q'_{\boldsymbol{x}_i}$ measure, for every $\mathbf{j} \in \mathcal{H}^0_{k_0,m}$ and every $i = 1, \ldots, L$. Thus, the set

$$\begin{aligned} U'(\mathcal{Q}';\tilde{\epsilon}) &= \bigcap_{i=1}^{L} \left\{ M_{\boldsymbol{x}_i} \in \mathscr{P}(\tilde{\Delta}_m) : \max_{\mathbf{j}\in\mathcal{H}^0_{k_0,m}} \left| M_{\boldsymbol{x}_i}(A_{\mathbf{j},k_0}) - Q'_{\boldsymbol{x}_i}(A_{\mathbf{j},k_0}) \right| \leq \tilde{\epsilon} \right\}, \\ &= \left\{ \mathcal{M} \in \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}} : N_{k_0}(\mathcal{M}, \mathcal{Q}') \leq \tilde{\epsilon} \right\}, \end{aligned} \tag{17}$$

belongs to $\mathscr{T}_4$. Notice that if $\tilde{\epsilon} = \frac{m!(k_0-1)!}{2M_0(k_0+m-1)!}\epsilon$, then

$$\left| \int_{\Delta_m} f_{ij}\ dH(k_0, M_{\boldsymbol{x}_i}) - \int_{\Delta_m} f_{ij}\ dH(k_0, Q_{\boldsymbol{x}_i}) \right| < \epsilon,$$

where $H(k_0, Q_{\boldsymbol{x}})$ is the probability measure associated to the multivariate Bernstein polynomial of measure $Q_{\boldsymbol{x}}$ of degree k_0. Therefore, if $U = \{k_0\} \times U'(\mathcal{Q}';\tilde{\epsilon})$, then $U \in \mathscr{L}_1$, $(k_0, \mathcal{Q}) \in U$ and $S(U) \subseteq V(S(k_0, \mathcal{Q}), \epsilon)$, which completes the proof of (i) in Lemma 1.

(ii) Let $\mathcal{Q} \in \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}}$, $k_0 \in \mathbb{N}$ and $V(S(k_0, \mathcal{Q}); \epsilon) = \bigcap_{i=1}^{L} B^{L_\infty}_{\epsilon,\boldsymbol{x}_i}(S(k_0, \mathcal{Q}))$, where L is a positive integer, $\epsilon > 0$ and $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L) \in \mathscr{X}^L$. The proof is based on finding and open set $U \in \mathscr{L}_1$ such that $(k_0, \mathcal{Q}) \in U$ and $S(U) \subseteq V(S(k_0, \mathcal{Q}); \epsilon)$.
Notice that for every $\mathcal{M} = \{M_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}}$,

$$\begin{aligned} &\sup_{\boldsymbol{y}\in\Delta_m} \left| b(\boldsymbol{y} \mid k_0, M_{\boldsymbol{x}_i}) - b(\boldsymbol{y} \mid k_0, Q_{\boldsymbol{x}_i}) \right| \\ &\qquad \leq \sup_{\boldsymbol{y}\in\Delta_m} \sum_{\mathbf{j}\in\mathcal{H}^0_{k_0,m}} \left| M_{\boldsymbol{x}_i}(A_{\mathbf{j},k_0}) - Q_{\boldsymbol{x}_i}(A_{\mathbf{j},k_0}) \right| \mathrm{dir}\left(\ \boldsymbol{y} \mid \alpha(k_0, \mathbf{j})\right), \\ &\qquad \leq \frac{M_0(k_0+m-1)!}{m!(k_0-1)!}\ N_{k_0}(\mathcal{M}, \mathcal{Q}), \end{aligned}$$

where $M_0 = \max_{\mathbf{j}\in\mathcal{H}^0_{k_0,m}} \sup_{\boldsymbol{y}\in\Delta_m} \mathrm{dir}\left(\ \boldsymbol{y} \mid \mathbf{j}, k_0 + m - \|\mathbf{j}\|_1\right)$, $b(\boldsymbol{y} \mid k_0, M_{\boldsymbol{x}_i})$ stands for the density function of the multivariate Bernstein polynomial of function $M_{\boldsymbol{x}_i}$ of degree k_0, and $N_{k_0}(\mathcal{M}, \mathcal{Q})$ and $A_{\mathbf{j},k_0}$ are defined as in part (i) of the proof. By the same arguments from part (i), it follows that if $U = \{k_0\} \times U'(\mathcal{Q}';\tilde{\epsilon})$, where $U'(\mathcal{Q}';\tilde{\epsilon})$ is defined as in (17), with $\tilde{\epsilon} = \frac{m!(k_0-1)!}{2M_0(k_0+m-1)!}\epsilon$, then $U \in \mathscr{L}_1$, $(k_0, \mathcal{Q}) \in U$ and $S(U) \subseteq V(S(k_0, \mathcal{Q}), \epsilon)$, which completes the proof of (ii) in Lemma 1.

(iii) Let $\mathcal{Q} \in \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}}$, $k_0 \in \mathbb{N}$ and $V\left(S(k_0, \mathcal{Q}); \epsilon\right) = B_\epsilon^{L_\infty}(S(k_0, \mathcal{Q}))$. The proof is based on finding and open set $U \in \mathscr{L}_2$ such that $(k_0, \mathcal{Q}) \in U$ and $S\left(U\right) \subseteq V\left(S(k_0, \mathcal{Q}); \epsilon\right)$.
Notice that for every $\mathcal{M} = \{M_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}}$,

$$\begin{aligned}\sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{\boldsymbol{y}\in\Delta_m} &|b(\boldsymbol{y} \mid k_0, M_x) - b(\boldsymbol{y} \mid k_0, Q_x)| \\ &\leq \frac{M_0(k_0+m-1)!}{m!(k_0-1)!} \sup_{\boldsymbol{x}\in\mathscr{X}} \max_{\mathbf{j}\in\mathcal{H}^0_{k_0,m}} \left|M_{\boldsymbol{x}}\left(A_{\mathbf{j},k_0}\right) - Q_{\boldsymbol{x}}\left(A_{\mathbf{j},k_0}\right)\right|,\end{aligned}$$

where $M_0 = \max_{\mathbf{j}\in\mathcal{H}^0_{k_0,m}} \sup_{\boldsymbol{y}\in\Delta_m}, \mathrm{dir}\left(\, \boldsymbol{y} \mid \mathbf{j}, k_0 + m - \|\mathbf{j}\|_1\right)$, and $A_{\mathbf{j},k_0}$ are defined as in the proof of (i). Then, if $U = \{k_0\} \times \tilde{\Delta}^{L_\infty}_{\tilde{\epsilon},k_0}(\mathcal{Q})$, where $\tilde{\Delta}^{L_\infty}_{\tilde{\epsilon},k_0}(\mathcal{Q})$ is defined as in (16), with $\tilde{\epsilon} = \frac{m!(k_0-1)!}{M_0(k_0+m-1)!}\epsilon$, then $U \in \mathscr{L}_2$, $(k_0, \mathcal{Q}) \in U$ and $S\left(U\right) \subseteq V(S(k_0, \mathcal{Q}), \epsilon)$, which completes the proof of (iii) in Lemma 1.

D.2. Proof of Theorem 2

Proving that $\mathscr{P}\left(\Delta_m\right)^{\mathscr{X}}$ and $\mathscr{D}\left(\Delta_m\right)^{\mathscr{X}}$ are the support of F under the weak product and L_∞ product topology, are direct extensions of the proofs of Theorems 2 and 3 in Barrientos et al. (2017), respectively. For completeness, the proof of this theorem is provided in what follows. First, we prove that $\mathscr{P}\left(\Delta_m\right)^{\mathscr{X}}$ is the support of F under the weak product topology. Then we prove that $\mathscr{D}\left(\Delta_m\right)^{\mathscr{X}}$ is the support of F under the L_∞ product topology. In each case all three versions of F are considered.

To prove that $\mathscr{P}\left(\Delta_m\right)^{\mathscr{X}}$ is the support of F under the weak product topology, it suffices to prove that any open set of the weak product topology has positive $P \circ F^{-1}$–measure. Let $\mathcal{Q} \in \mathscr{P}(\Delta_m)^{\mathscr{X}}$ and $V\left(\mathcal{Q}; \epsilon\right) = \bigcap_{i=1}^{L} \bigcap_{j=1}^{K_i} B^W_{f_{ij},\epsilon,\boldsymbol{x}_i}\left(\mathcal{Q}\right)$, where L, K_i, $i = 1, \ldots, L$, are positive integers, f_{ij}, $j = 1, \ldots, K_i$, $i = 1, \ldots, L$, are bounded continuous functions, $\epsilon > 0$ and $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L) \in \mathscr{X}^L$. From Lemma 1 in Barrientos et al. (2012), there exists $\mathcal{Q}' = \{Q'_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \mathscr{P}(\Delta_m)^{\mathscr{X}}$, such that for every $\boldsymbol{x} \in \mathscr{X}$, $Q'_{\boldsymbol{x}}$ is absolutely continuous w.r.t Lebesgue measure and such that $Q'_{\boldsymbol{x}} = Q_{\boldsymbol{x}}$ if $\boldsymbol{x} \neq \boldsymbol{x}_i$ and

$$\left|\int_{\Delta_m} f_{ij}\, Q_{\boldsymbol{x}_i} - \int_{\Delta_m} f_{ij} dQ'_{\boldsymbol{x}_i}\right| < \frac{\epsilon}{2},$$

if $\boldsymbol{x} = \boldsymbol{x}_i$, $i = 1, \ldots, L$. Then, $V\left(\mathcal{Q}'; \epsilon/2\right) \subset V\left(\mathcal{Q}; \epsilon\right)$. Since for every $\boldsymbol{x} \in \mathscr{X}$, $H(k, Q'_{\boldsymbol{x}})$ converges weakly to $Q'_{\boldsymbol{x}}$ as $k \to \infty$, for every $\epsilon > 0$, there exists large enough $k_0 \in \mathbb{N}$ such that

$$\left|\int_{\Delta_m} f_{ij} dH(k_0, Q'_{\boldsymbol{x}_i}) - \int_{\Delta_m} f_{ij} dQ'_{\boldsymbol{x}_i}\right| < \frac{\epsilon}{4},$$

then $V\left(S(k_0, \mathcal{Q}'); \epsilon/4\right) \subset V\left(\mathcal{Q}'; \epsilon/2\right)$. By Lemma 1 part (i), there exists $U = \{k_0\} \times U'(\mathcal{Q}'; \tilde{\epsilon}) \in \mathscr{L}_1$, with $\tilde{\epsilon} = \frac{m!(k_0-1)!}{4M_0(k_0+m-1)!}\epsilon$, $k_0 \in \mathbb{N}$ and $U'(\mathcal{Q}'; \tilde{\epsilon}) \in \mathscr{T}_4$, such that $S(U) \subset V\left(S(k_0, \mathcal{Q}'); \epsilon/4\right)$. Thus, to prove this theorem, it suffices to prove that $P \circ F^{-1}(V(\mathcal{Q};\epsilon)) \geq P\{\omega \in \Omega : (k(\omega), \overline{\mathbf{T}}) \in U\} > 0$, where $U = \{k_0\} \times U'(\mathcal{Q}'; \tilde{\epsilon})$, with $U'(\mathcal{Q}'; \tilde{\epsilon})$ defined as in (17) and $\overline{\mathbf{T}}$ is either $\mathbf{T}$, $\mathbf{T}^{\theta}$ or $\mathbf{T}^{w}$. Before considering each case, note that there are $N = \frac{(k_0+m-1)!}{m!(k_0-1)!}$ disjoint sets in $\mathcal{H}^0_{k_0,m}$. Each of these sets is denoted by $A_{[l],N}$, $l = 1, \ldots, N$.

When $\overline{\mathbf{T}}$ is $\mathbf{T}$, the assumption that the stochastic processes η_j and $\boldsymbol{z}_j$ are well defined and have full support implies that

$$
\begin{aligned}
P\{\omega \in \Omega : (k(\omega), \mathbf{T}) \in U\} \geq\ & P\{\omega \in \Omega : k(\omega) = k_0\} \\
& \times P\left\{\omega \in \Omega : \max_{i\in\{1,\ldots,L\}} \max_{\mathbf{j}\in\mathcal{H}^0_{k_0,m}} \left| M_{\boldsymbol{x}_i}(A_{\mathbf{j},k_0}) - Q'_{\boldsymbol{x}_i}(A_{\mathbf{j},k_0}) \right| \leq \tilde{\epsilon} \right\}, \\
\geq\ & P\{\omega \in \Omega : k(\omega) = k_0\} \\
& \times \prod_{l=1}^{N} P\left\{\omega \in \Omega : (\boldsymbol{\theta}_l(\boldsymbol{x}_1,\omega), \ldots, \boldsymbol{\theta}_l(\boldsymbol{x}_L,\omega)) \in A^L_{[l],N}\right\} \\
& \times P\left\{\omega \in \Omega : (V_l(\boldsymbol{x}_1,\omega), \ldots, V_l(\boldsymbol{x}_L,\omega)) \in B^L_l, l \in \{1,\ldots,N\}\right\} \\
& \times \prod_{l=N+1}^{\infty} P\left\{\omega \in \Omega : (\boldsymbol{\theta}_l(\boldsymbol{x}_1,\omega), \ldots, \boldsymbol{\theta}_l(\boldsymbol{x}_L,\omega)) \in \tilde{\Delta}^L_m\right\} \\
& \times \prod_{l=N+1}^{\infty} P\left\{\omega \in \Omega : (V_l(\boldsymbol{x}_1,\omega), \ldots, V_l(\boldsymbol{x}_L,\omega)) \in [0,1]^L\right\}, \\
>\ & 0,
\end{aligned}
$$

where

$$
\begin{aligned}
B^L_1 &= \bigotimes_{i=1}^{L} \left\{ Q'_{\boldsymbol{x}_i}(A_{[1],N}) - \frac{\tilde{\epsilon}}{4(N-1)} \ ;\ Q'_{\boldsymbol{x}_i}(A_{[1],N}) + \frac{\tilde{\epsilon}}{4(N-1)} \right\}, \\
B^L_l &= \bigotimes_{i=1}^{L} \left\{ \frac{Q'_{\boldsymbol{x}_i}(A_{[l],N}) - \frac{\tilde{\epsilon}}{4(N-1)}}{\prod_{l_1<l}[1 - V_{l_1}(\boldsymbol{x}_i,\omega)]} \ ;\ \frac{Q'_{\boldsymbol{x}_i}(A_{[l],N}) + \frac{\tilde{\epsilon}}{4(N-1)}}{\prod_{l_1<l}[1 - V_{l_1}(\boldsymbol{x}_i,\omega)]} \right\}, \quad l = 2,\ldots,N-1, \\
B^L_N &= \bigotimes_{i=1}^{L} \left\{ \frac{Q'_{\boldsymbol{x}_i}(A_{[N],N}) - \frac{\tilde{\epsilon}}{3}}{\prod_{l_1<N}[1 - V_{l_1}(\boldsymbol{x}_i,\omega)]} \ ;\ \frac{Q'_{\boldsymbol{x}_i}(A_{[N],N}) - \frac{\tilde{\epsilon}}{4}}{\prod_{l_1<N}[1 - V_{l_1}(\boldsymbol{x}_i,\omega)]} \right\},
\end{aligned}
$$

$A^L_{[l],N} = \bigotimes_{i=1}^{L} A_{[l],N}$, $l = 1,\ldots,N$, $\tilde{\Delta}^L_m = \bigotimes_{i=1}^{L} \tilde{\Delta}_d$ and $[0,1]^L = \bigotimes_{i=1}^{L}[0,1]$. This completes the proof that F considered as DMBPP$(\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_{\boldsymbol{z}}, \mathscr{V}, \mathscr{H})$ has weak product support.

When $\overline{\mathbf{T}}$ is $\mathbf{T}^{\theta}$, the assumption that the stochastic processes η_j and the random vectors $\boldsymbol{\theta}_j$ are well defined and have full support imply that

$$
\begin{aligned}
P\{\omega \in \Omega : (k(\omega), \mathbf{T}^{\theta}) \in U\} \geq\ & P\{\omega \in \Omega : k(\omega) = k_0\} \\
& \times \prod_{l=1}^{N} P\left\{\omega \in \Omega : (\boldsymbol{\theta}_l(\omega), \ldots, \boldsymbol{\theta}_l(\omega)) \in A^L_{[l],N}\right\}
\end{aligned}
$$

$$\times P\left\{\omega \in \Omega : (V_l(\boldsymbol{x}_1,\omega),\ldots,V_l(\boldsymbol{x}_L,\omega)) \in B_l^L, l \in \{1,\ldots,N\}\right\},$$
$$> 0,$$

where B_1^L, B_l^L, $l = 2,\ldots,N-1$, B_N^L and $A_{[l],N}^L$, $l = 1,\ldots,N$, are defined as above. This completes the proof that F considered as $\theta\text{DMBPP}(\lambda, \boldsymbol{\Psi}_{\boldsymbol{z}}, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$ has weak product support.

Finally, when $\overline{\mathbf{T}}$ is $\mathbf{T}^w$. Since Δ_m is a separable space and $\tilde{\Delta}_m$ is dense in Δ_m, then the space of measures whose support points are finite subsets of $\tilde{\Delta}_m$ is dense in $\mathscr{P}(\Delta_m)$ (Parthasarathy, 1967). Then, for each $\boldsymbol{x} \in \mathscr{X}$, there exists a probability measure $\tilde{Q}_{\boldsymbol{x}}(\cdot) = \sum_{j=1}^R \tilde{w}_j \delta_{\tilde{\boldsymbol{\theta}}_j(\boldsymbol{x})}(\cdot)$, defined on $\tilde{\Delta}_m$, where R is an integer, $\tilde{w}_j \in [0,1]$, $j = 1,\ldots,R$, $\sum_{j=1}^R \tilde{w}_j = 1$, and $\tilde{\boldsymbol{\theta}}_j(\boldsymbol{x}) \in \tilde{\Delta}_m$ are continuous functions of $\boldsymbol{x}$, $j = 1,\ldots,R$, such that, for every $\boldsymbol{x} \in \mathscr{X}$, $\mathbf{j} \in \mathcal{H}_{k_0,m}^0$,

$$\left|\tilde{Q}_{\boldsymbol{x}}(A_{\mathbf{j},k_0}) - Q'_{\boldsymbol{x}}(A_{\mathbf{j},k_0})\right| < \frac{\tilde{\epsilon}}{2}.$$

Then $U'(\tilde{Q};\tilde{\epsilon}/2) \subset U'\left(Q';\tilde{\epsilon}\right)$, where $U'(\mathcal{Q};\epsilon)$ is defined as in (17). Thus, it suffices to prove that

$$P\left\{\omega \in \Omega : (k(\omega), \mathbf{T}^w) \in \{k_0\} \times U'(\tilde{Q};\tilde{\epsilon}/2)\right\} > 0.$$

Consider $\left\{\tilde{A}_{[\tilde{l}],M}\right\}_{\tilde{l}=1}^M$, a finer partition of $\mathcal{H}_{k_0,m}^0$ than $\left\{A_{[l],N}\right\}_{l=1}^N$, such that $A_{[1],N} = \bigcup_{\tilde{l}=1}^{n_1} \tilde{A}_{[\tilde{l}],M}$ and $A_{[l],N} = \bigcup_{\tilde{l}=n_{l-1}+1}^{n_l} \tilde{A}_{\tilde{l},M}$, $l = 1,\ldots,N$, where $\sum_{l=1}^N n_l = M$. Then, the assumption that the stochastic processes $\boldsymbol{z}_j$ and the random variables v_j are well defined and have full support imply that

$$\begin{aligned}
&P\{\omega \in \Omega : (k(\omega), \mathbf{T}^w) \in U\} \geq P\{\omega \in \Omega : k(\omega) = k_0\} \\
&\times P\left\{\omega \in \Omega : \left(\lceil k_0 \boldsymbol{\theta}_1(\boldsymbol{x}_i,\omega)\rceil - \lceil k_0 \tilde{\boldsymbol{\theta}}_j(\boldsymbol{x}_i)\rceil\right) = \mathbf{0}, m = 1,\ldots,L, j = 1,\ldots,n_1\right\} \\
&\times \prod_{l=2}^N P\left\{\omega \in \Omega : \left(\lceil k_0 \boldsymbol{\theta}_l(\boldsymbol{x}_i,\omega)\rceil - \lceil k_0 \tilde{\boldsymbol{\theta}}_j(\boldsymbol{x}_i)\rceil\right) = \mathbf{0}, i = 1,\ldots,L, j = n_{l-1}+1,\ldots,n_l\right\} \\
&\times P\left\{\omega \in \Omega : v_l(\omega) \in B_l^L, l = 1,\ldots,N\right\}, \\
&> 0,
\end{aligned}$$

where

$$B_1^L = \left\{\sum_{j=1}^{n_1} \tilde{w}_j - \frac{\tilde{\epsilon}}{8(N-1)} \,;\, \sum_{j=1}^{n_1} \tilde{w}_j + \frac{\tilde{\epsilon}}{8(N-1)}\right\},$$

$$B_l^L = \left\{\frac{\displaystyle\sum_{j=n_{l-1}+1}^{n_l} \tilde{w}_j - \frac{\tilde{\epsilon}}{8(N-1)}}{\displaystyle\prod_{l_1<l}[1 - V_{l_1}(\omega)]} \,;\, \frac{\displaystyle\sum_{j=n_{l-1}+1}^{n_l} \tilde{w}_j + \frac{\tilde{\epsilon}}{8(N-1)}}{\displaystyle\prod_{l_1<l}[1 - V_{l_1}(\omega)]}\right\}, \quad l = 2,\ldots,N-1,$$

$$B_N^L = \left\{ \frac{1 - \sum_{j=1}^{n_{N-1}} \tilde{w}_j - \frac{\tilde{\epsilon}}{6}}{\prod_{l_1 < N} [1 - V_{l_1}(\omega)]} \; ; \; \frac{1 - \sum_{j=1}^{n_{N-1}} \tilde{w}_j - \frac{\tilde{\epsilon}}{8}}{\prod_{l_1 < N} [1 - V_{l_1}(\omega)]} \right\},$$

which completes the proof that F considered as wDMBPP$(\lambda, \mathbf{\Psi}_v, \mathbf{\Psi}_{\boldsymbol{z}}, \mathscr{H})$ has weak product support. Thus the proof that $\mathscr{P}(\Delta_m)^{\mathscr{X}}$ is the support of F under the weak product topology is completed.

Now we will prove that $\mathscr{D}(\Delta_m)^{\mathscr{X}}$ is the support of F under the L_∞ product toplogy. Note that it suffices to prove that any open set of the L_∞ product topology has positive $P \circ F^{-1}$–measure. Let $\mathcal{Q} \in \mathscr{D}(\Delta_m)^{\mathscr{X}}$ and $V(\mathcal{Q}; \epsilon) = \bigcap_{i=1}^{L} \Delta_{\epsilon, \boldsymbol{x}_i}^{L_\infty}(\mathcal{Q})$, where L is a positive integer, $\epsilon > 0$, and $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_L) \in \mathscr{X}^L$. Recall that for every $\boldsymbol{x} \in \mathscr{X}$, $Q_{\boldsymbol{x}} \in \mathscr{D}(\Delta_m)$ is an absolutely continuous measures, w.r.t. Lebesgue measure, with continuous density, $q_{\boldsymbol{x}}$. By Theorem 1 in Barrientos et al. (2015), for every $\epsilon > 0$, there exists large enough $k_0 \in \mathbb{N}$, such that for every $\boldsymbol{x} \in \mathscr{X}$,

$$\sup_{\boldsymbol{y} \in \Delta_m} |b(\boldsymbol{y} \mid k_0, Q_{\boldsymbol{x}}) - q_{\boldsymbol{x}}(\boldsymbol{y})| < \frac{\epsilon}{2},$$

where $b(\boldsymbol{y} \mid k, Q_{\boldsymbol{x}})$ stands for the density function of the multivariate Bernstein polynomial of degree k of function $Q_{\boldsymbol{x}}$. Then $V(S(k_0, \mathcal{Q}); \epsilon/2) \subset V(\mathcal{Q}; \epsilon)$. By Lemma 1 part (ii), there exists $U = \{k_0\} \times U'(\mathcal{Q}'; \tilde{\epsilon}) \in \mathscr{L}_1$, with $\tilde{\epsilon} = \frac{m!(k_0-1)!}{2M_0(k_0+m-1)!}\epsilon$, $k_0 \in \mathbb{N}$ and $U'(\mathcal{Q}'; \tilde{\epsilon}) \in \mathscr{T}_4$, such that $S(U) \subset V(S(k_0, \mathcal{Q}); \epsilon/2)$. In analogy with the weak product support proof, it suffices to prove that $P \circ F^{-1}(V(\mathcal{Q}; \epsilon)) \geq P\{\omega \in \Omega : (k(\omega), \overline{\mathbf{T}}) \in \{k_0\} \times U'(\mathcal{Q}; \tilde{\epsilon})\} > 0$, where $\overline{\mathbf{T}}$ is either $\mathbf{T}$, $\mathbf{T}^\theta$ or $\mathbf{T}^w$. By the same arguments used to prove the weak product support of F, it follows that $P \circ F^{-1}(V(\mathcal{Q}; \epsilon)) > 0$, when F is considered as DMBPP, θDMBPP, or wDMBPP. This completes the proof that $\mathscr{D}(\Delta_m)^{\mathscr{X}}$ is the support of F under the L_∞ product topology, and thus completes the proof of the theorem. $\square$

D.3. Proof of Theorem 3

The proof of Theorem 3 follows the same reasoning of the proof of Theorem 4 in Barrientos et al. (2017). First, we state and prove Lemma 2 below, which is used in the proof of this theorem and is an extension of Lemma B.4.1 in the supplementary material of Barrientos et al. (2017), and then we prove that $\tilde{\mathscr{D}}(\Delta_m)^{\mathscr{X}}$ is contained in the support of F under the L_∞ topology.

Lemma 2. *Let* $\{Q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \tilde{\mathscr{D}}(\Delta_m)^{\mathscr{X}}$ *be an absolutely continuous measure, w.r.t. Lebesgue measure, such that the mapping* $(\boldsymbol{x}, \boldsymbol{y}) \mapsto q_{\boldsymbol{x}}(\boldsymbol{y})$ *is continuous, and consider* $\mathscr{X}$ *a compact space on* $\mathbb{R}^p$. *Denote* $b_{k, Q_{\boldsymbol{x}}}(\boldsymbol{y})$*, the density function, w.r.t. Lebesgue measure, of the multivariate Bernstein polynomial of degree* k *of function* $Q_{\boldsymbol{x}}$. *Then for every* $\epsilon > 0$, *there exists* $k_0 \in \mathbb{N}$ *such that for every*

$k \geq k_0$,

$$\sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{\boldsymbol{y}\in\Delta_m} \mid b(\boldsymbol{y} \mid k, Q_{\boldsymbol{x}}) - q_{\boldsymbol{x}}(\boldsymbol{y}) \mid < \epsilon.$$

Proof of Lemma 2

Without loss of generality, consider $\mathscr{X} = [0,1]^p$, and a uniform marginal distribution for $\boldsymbol{X}$ on $\mathscr{X}$. Then, $q_{\boldsymbol{x}}(\boldsymbol{y})$ denotes a joint density function on $\Delta_m \times \mathscr{X}$. Note that $b(\boldsymbol{y} \mid k, Q_{\boldsymbol{x}})$ can be written as

$$b(\boldsymbol{y} \mid k, Q_{\boldsymbol{x}}) = \sum_{\boldsymbol{j}\in\mathcal{H}^0_{k,m}} \left[\int_{A_{\mathbf{j},k}} q_{\boldsymbol{x}}(\boldsymbol{y})d\boldsymbol{y}\right] \times \mathrm{dir}(\boldsymbol{y} \mid \alpha(k,\boldsymbol{j})).$$

Now, consider $r \in \mathbb{N}$, $\boldsymbol{l} = (l_1,\ldots,l_r)$, where $l_s \in \mathbb{N}$, $s = 1,\ldots,r$, are positive integers, and define

$$\begin{aligned} p_{k,\boldsymbol{l},Q_{\boldsymbol{x}}}(\boldsymbol{y}) \quad = \quad & \sum_{\boldsymbol{j}\in\mathcal{H}^0_{k,m}} \sum_{i_1=1}^{l_1} \cdots \sum_{i_r=1}^{l_r} \left[\int_{B_{i_1}} \cdots \int_{B_{i_r}} \int_{A_{\mathbf{j},k}} q_{\boldsymbol{x}}(\boldsymbol{y})d\boldsymbol{y}dx_r \ldots dx_1\right] \\ & \times \prod_{s=1}^{r} \beta(x_s \mid a_s, b_s)\mathrm{dir}(\boldsymbol{y} \mid \alpha(k,\boldsymbol{j})), \end{aligned}$$

where $B_{i_s} = \left(\frac{i_s-1}{l_s}, \frac{i_s}{l_s}\right]$, $a_s = i_s$, $b_s = l_s - i_s + 1$, $s = 1,\ldots,r$, and $\beta(\cdot \mid a, b)$ stands for a beta density with parameters a and b. Since $(\boldsymbol{x},\boldsymbol{y}) \mapsto q_{\boldsymbol{x}}(\boldsymbol{y})$ is a continuous mapping, it is easy to show that $p_{k,\boldsymbol{l},Q_{\boldsymbol{x}}}(\boldsymbol{y})$ can uniformly approximate any continuous density function defined on $\Delta_m \times \mathscr{X}$. Thus, every $k > k_0$, $l_s > l_{s,0}$, $s = 1,\ldots,r$, it follows that

$$\sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{\boldsymbol{y}\in\Delta_m} |p_{k,\boldsymbol{l},Q_{\boldsymbol{x}}}(\boldsymbol{y}) - q_{\boldsymbol{x}}(\boldsymbol{y})| < \epsilon/2.$$

Now, noting that

$$\sum_{i_1=1}^{l_1} \cdots \sum_{i_r=1}^{l_r} \left[\int_{B_{i_1}} \cdots \int_{B_{i_r}} \int_{A_{\mathbf{j},k}} q_{\boldsymbol{x}}(\boldsymbol{y})d\boldsymbol{y}\; dx_r \ldots dx_1\right] \times \prod_{s=1}^{r} \beta(x_s \mid a_s, b_s), \quad (18)$$

is the density function of the multivariate Bernstein polynomial of degree $l_1, \ldots$, l_r, of the mapping

$$\boldsymbol{x} \mapsto \int_{A_{\mathbf{j},k}} q_{\boldsymbol{x}}(\boldsymbol{y})d\;\boldsymbol{y},$$

defined on $\mathscr{X}$, it follows that (18) converges uniformly to $\int_{A_{\mathbf{j},k}} q_{\boldsymbol{x}}(\boldsymbol{y})d\boldsymbol{y}$, as $(l_1,\ldots,l_r) \to \infty$, component-wise. Therefore, for every $l_s > l_{s,1}$, $s = 1,\ldots,r$,

$$\sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{\boldsymbol{y}\in\Delta_m} |b(\boldsymbol{y} \mid k, Q_{\boldsymbol{x}}) - p_{k,\boldsymbol{l},Q_{\boldsymbol{x}}}(\boldsymbol{y})| < \sum_{\boldsymbol{j}\in\mathcal{H}^0_{k,m}} \frac{\tilde{\epsilon}}{2}\; \mathrm{dir}(\boldsymbol{y} \mid \alpha(k,\boldsymbol{j})) < \frac{\epsilon}{2},$$

where $\tilde{\epsilon} = \frac{m!(k-1)!}{M_0(k+m-1)!}\epsilon$, with $M_0 = \max_{\mathbf{j}\in\mathcal{H}^0_{k,m}} \sup_{\boldsymbol{y}\in\Delta_m} \mathrm{dir}(\boldsymbol{y} \mid \alpha(k, \boldsymbol{j}))$. Finally, for $k > k_0$, $l_s > \max\{l_{s,0}, l_{s,1}\}$, $s = 1, \ldots, r$, and an application of the triangle inequality, it follows that

$$\sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{\boldsymbol{y}\in\Delta_m} |b(\boldsymbol{y} \mid k, Q_{\boldsymbol{x}}) - q_{\boldsymbol{x}}(\boldsymbol{y})| < \epsilon,$$

which completes to proof of the lemma. $\square$

Now, note that to prove that $\tilde{\mathscr{D}}(\Delta_m)^{\mathscr{X}}$ is contained in the support of F under the L_∞ topology, it suffices to prove that any open set of the L_∞ topology has positive $P \circ F^{-1}$–measure. Let $\mathcal{Q} \in \tilde{\mathscr{D}}(\Delta_m)^{\mathscr{X}}$ and $V(\mathcal{Q}; \epsilon) = B^{L_\infty}_{\epsilon}(\mathcal{Q})$, $\epsilon > 0$. Recall that $\mathscr{X}$ is compact, and $Q_{\boldsymbol{x}} \in \tilde{\mathscr{D}}(\Delta_m)$ is an absolutely continuous measures, w.r.t. Lebesgue measure, with continuous density, $q_{\boldsymbol{x}}$, sucth that $(\boldsymbol{x}, \boldsymbol{y}) \mapsto q_{\boldsymbol{x}}(\boldsymbol{y})$ is continuous. From Lemma 2, there exists large enough k_0, such that,

$$\sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{\boldsymbol{y}\in\Delta_m} |b(\boldsymbol{y} \mid k, Q_{\boldsymbol{x}}) - q_{\boldsymbol{x}}(\boldsymbol{y})| < \frac{\epsilon}{2},$$

where $\mathrm{bp}(\boldsymbol{y} \mid k, Q_{\boldsymbol{x}})$ stands for the density function of the multivariate Bernstein polynomial of function $Q_{\boldsymbol{x}}$ of degree k. Then, $V(S(k_0, \mathcal{Q}); \epsilon/2) \subset V(\mathcal{Q}; \epsilon)$. By Lemma 1 part (iii), there exists $U = \{k_0\} \times \tilde{\Delta}^{L_\infty}_{\tilde{\epsilon},k_0}(\mathcal{Q}) \in \mathscr{L}_2$, with $\tilde{\epsilon} = \frac{m!(k_0-1)!}{2M_0(k_0+m-1)!}\epsilon$, $k_0 \in \mathbb{N}$, such that $S(U) \subset V(S(k_0, \mathcal{Q}); \epsilon/2)$. Thus, to prove this theorem, it suffices to prove that $P \circ F^{-1}(V(\mathcal{Q}; \epsilon)) \geq P\{\omega \in \Omega : (k(\omega), \overline{\mathbf{T}}) \in \{k_0\} \times U^\star(\mathcal{Q}; \tilde{\epsilon})\} > 0$, where

$$U^\star(\mathcal{Q}; \tilde{\epsilon}) = \left\{ \mathcal{M} \in \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}} : \sup_{\boldsymbol{x}\in\mathscr{X}} \max_{\mathbf{j}\in\mathcal{H}^0_{k_0,m}} |M_{\boldsymbol{x}}(A_{\mathbf{j},k_0}) - Q_{\boldsymbol{x}}(A_{\mathbf{j},k_0})| \leq \tilde{\epsilon} \right\}, \tag{19}$$

and $\overline{\mathbf{T}}$ is either $\mathbf{T}$, $\mathbf{T}^\theta$ or $\mathbf{T}^w$.

First we assume that $\overline{\mathbf{T}}$ is $\mathbf{T}$ and following a similar reasoning as in the proof of Theorem 2. Since the stochastic processes η_j and $\boldsymbol{z}_j$ are well defined and have full support, $A_{[l],N} \in \mathscr{B}(\Delta_m)$ and the mappings

$$\begin{aligned} &\boldsymbol{x} \mapsto Q_{\boldsymbol{x}}(A_{[l],N}), \\ &\boldsymbol{x} \mapsto \frac{Q_{\boldsymbol{x}}(A_{[l],N})/2}{\prod_{l_1<l} [1 - V_{l_1}(\boldsymbol{x}, \omega)]}, \end{aligned}$$

are continuous, it follows that,

$$\begin{aligned} &P \circ F^{-1}(V(\mathcal{Q}; \epsilon)) \\ &\quad \geq P\{\omega \in \Omega : k(\omega) = k_0\} \\ &\qquad \times \prod_{l=1}^{N} P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} |\boldsymbol{\theta}_l(\boldsymbol{x}, \omega)| \in A_{[l],N}\right\} \end{aligned}$$

$$\times P\left\{\omega\in\Omega:\sup_{\boldsymbol{x}\in\mathscr{X}}\left|V_1(\boldsymbol{x},\omega)-Q_{\boldsymbol{x}}(A_{[l],N})/2\right|<\frac{\tilde{\epsilon}}{2N}\right\}$$
$$\times P\left\{\omega\in\Omega:\sup_{\boldsymbol{x}\in\mathscr{X}}\left|V_l(\boldsymbol{x},\omega)-\frac{Q_{\boldsymbol{x}}(A_{[l],N})/2}{\prod_{l_1<l}[1-V_{l_1}(\boldsymbol{x},\omega)]}\right|<\frac{\tilde{\epsilon}}{2N},l=2,\dots,N\right\},$$
$$>0,$$

which completes the proof that when F is a DMBPP$(\lambda,\boldsymbol{\Psi}_\eta,\boldsymbol{\Psi}_{\boldsymbol{z}},\mathscr{V},\mathscr{H})$ it has full L_∞ support.

Now assume that $\overline{\mathbf{T}}$ is $\mathbf{T}^\theta$. Given the above proof, it is straightforward to prove that F considered as θDMBPP$(\lambda,\boldsymbol{\Psi}_\eta,\mathscr{V},\boldsymbol{\Psi}_{\boldsymbol{\theta}})$ has L_∞ support.

Finally, assume that $\overline{\mathbf{T}}$ is $\mathbf{T}^w$. Consider the partition $\left\{A_{[l],N}\right\}_{l=1}^N$ of Δ_m and for each $l=1,\dots,N$, consider $\{\mathscr{X}_{l,j}\}_{j=1}^{N_l}$, a partition of space $\mathscr{X}$, $N_l\in\mathbb{N}$, $N_l>N$. Since $Q_{\boldsymbol{x}}\in\tilde{\mathscr{D}}(\Delta_m)$ are such that $(\boldsymbol{y},\boldsymbol{x})\mapsto q_{\boldsymbol{x}}(\boldsymbol{y})$ are continuous, then $(\boldsymbol{y},\boldsymbol{x})\mapsto Q_{\boldsymbol{x}}(\boldsymbol{y})$ are continuous and can be approximated by functions of the form,

$$\overline{Q}_{\boldsymbol{x}}(\boldsymbol{y})=\sum_{l=1}^N\sum_{j=1}^{N_l}a_{l,j}\ \mathbb{I}(\boldsymbol{x},\boldsymbol{y})_{\left\{\mathscr{X}_{l,j}\times A_{[l],N}\right\}},$$

where $\{a_{l,j}\}_{j=1}^{N_l}$, $l=1,\dots,N$ are positive constants, $\mathbb{I}_A$ denotes the indicator function of set A, $\boldsymbol{x}\in\mathscr{X}$ and $\boldsymbol{y}\in\Delta_m$. Now, for each $l=1,\dots,N$, consider the mapping $(a_{l,1},\dots,a_{l,N_l})\mapsto\tilde{w}_{l,j}=a_{l,j}/\sum_{j=1}^{N_l}a_{l,j}$ and the continuous mappings $\boldsymbol{x}\mapsto\tilde{\boldsymbol{\theta}}_{l,j}(\boldsymbol{x})$, where $\tilde{w}_{l,j}\in[0,1]$, $\sum_{j=1}^{N_l}\tilde{w}_{l,j}=1$, $\tilde{\boldsymbol{\theta}}_{l,j}(\boldsymbol{x})\in\tilde{\Delta}_m$ and $\tilde{\boldsymbol{\theta}}(\mathscr{X}_{l,1},\dots,\mathscr{X}_{l,N_l})=\left\{\tilde{A}_{[l,j],N_l}\right\}_{j=1}^{N_l}$ is a finer partition of $\mathcal{H}^0_{k_0,m}$ than $\left\{A_{[l],N}\right\}_{l=1}^N$, such that $A_{[l],N}=\bigcup_{j=1}^{n_l}\tilde{A}_{[l,j],N_l}$, $n_l<N_l$. Thus, for each $l=1,\dots,N$, $\overline{Q}_{\boldsymbol{x}}(A_{[l],N})$ can be written as a measure of the form

$$\tilde{Q}_{\boldsymbol{x}}(A_{[l],N})=\sum_{j=1}^{N_l}\tilde{w}_{l,j}\ \mathbb{I}\left\{\tilde{\boldsymbol{\theta}}_{l,j}(\boldsymbol{x})\right\}_{\left\{\tilde{A}_{[l,j],N_l}\right\}},$$

such that, for every $l=1,\dots,N$,

$$\sup_{\boldsymbol{x}\in\mathscr{X}}\left|\tilde{Q}_{\boldsymbol{x}}(A_{[l],N})-Q_{\boldsymbol{x}}(A_{[l],N})\right|<\frac{\tilde{\epsilon}}{2}.$$

Then $U^\star(\tilde{\mathcal{Q}};\tilde{\epsilon}/2)\subset U^\star(\mathcal{Q};\tilde{\epsilon})$, where $U^\star(\mathcal{Q};\epsilon)$ is defined as (19). Thus, in analogy with the previous proofs, it suffices to prove that

$$P\left\{\omega\in\Omega:(k(\omega),\mathbf{T}^w)\in\{k_0\}\times U^\star(\tilde{\mathcal{Q}};\tilde{\epsilon}/2)\right\}>0.$$

Following a similar reasoning as in the proof of Theorem 2 when $\overline{\mathbf{T}}$ was considered as $\mathbf{T}^w$ and by the assumption that the stochastic processes η_j and $\boldsymbol{z}_j$ are well defined and have full support, $A_{[l],N}\in\mathscr{B}(\Delta_m)$, and the mappings

$$\boldsymbol{x}\mapsto k_0\tilde{\boldsymbol{\theta}}_{l,j}(\boldsymbol{x}),\quad j=1,\dots,n_l,\quad l=1,\dots,N,$$

are continuous, it follows that,

$$
\begin{aligned}
P & \circ F^{-1}(V(\mathcal{Q};\epsilon)) \\
& \geq P\{\omega \in \Omega : k(\omega) = k_0\} \\
& \quad \times \prod_{l=1}^{N} P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} \left(\lceil k_0 \boldsymbol{\theta}_l(\boldsymbol{x},\omega)\rceil - \lceil k_0 \tilde{\boldsymbol{\theta}}_{l,j}(\boldsymbol{x},\omega)\rceil\right) = \boldsymbol{0}, j = 1,\ldots,n_l\right\} \\
& \quad \times P\left\{\omega \in \Omega : \left| v_1(\omega) - \sum_{j=1}^{n_1} \tilde{w}_{1,j}(\omega)/2 \right| < \frac{\tilde{\epsilon}}{2N}\right\} \\
& \quad \times P\left\{\omega \in \Omega : \left| v_l(\omega) - \frac{\sum_{j=1}^{n_l} \tilde{w}_{l,j}(\omega)/2}{\prod_{l_2<l}[1 - v_{l_2}(\omega)]} \right| < \frac{\tilde{\epsilon}}{2N}, l = 2,\ldots,N\right\}, \\
& > 0,
\end{aligned}
$$

which completes the proof that when F is a $w\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_{\boldsymbol{z}}, \mathscr{H})$ it has full L_∞ support. □

D.4. Proof of Theorem 4

The proof of this Theorem is an extension of the proof of Corollary 1 in Barrientos et al. (2017). For completeness, the proof is given below.

Let $\{Q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \tilde{\mathscr{D}}(\Delta_m)^{\mathscr{X}}$ with continuous density function $\{q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$. Here we will prove that, for every $\delta > 0$, any Kullback-Leibler neighborhood of $\{Q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$ has positive $P \circ F^{-1}$–measure. This is,

$$
P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} KL(q_{\boldsymbol{x}}, f(\boldsymbol{x},\omega)) < \delta\right\} > 0,
$$

where $KL(q,p) = \int_{\Delta_m} q(\boldsymbol{y}) \log\left(\frac{q(\boldsymbol{y})}{p(\boldsymbol{y})}\right) d\boldsymbol{y}$. Since $\mathscr{X}$ and Δ_m are compact sets and $(\boldsymbol{x},\boldsymbol{y}) \mapsto q_{\boldsymbol{x}}(\boldsymbol{y})$ is a continuous mapping, it follows that $\inf_{\boldsymbol{x}\in\mathscr{X}} \inf_{\boldsymbol{y}\in\Delta_m} q_{\boldsymbol{x}}(\boldsymbol{y})$ exists and is bounded.

First, suppose that $\inf_{\boldsymbol{x}\in\mathscr{X}} \inf_{\boldsymbol{y}\in\Delta_m} q_{\boldsymbol{x}}(\boldsymbol{y}) > 0$. If for every $\epsilon > 0$, $\sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{\boldsymbol{y}\in\Delta_m} |f(\boldsymbol{x},\omega)(\boldsymbol{y}) - q_{\boldsymbol{x}}(\boldsymbol{y})| < \epsilon$, then $\inf_{\boldsymbol{x}\in\mathscr{X}} \inf_{\boldsymbol{y}\in\Delta_m} f(\boldsymbol{x},\omega)(\boldsymbol{y}) > 0$ and for every $\epsilon' > 0$, there exists $\epsilon > 0$ such that for every $\boldsymbol{x} \in \mathscr{X}$ and every $\boldsymbol{y} \in \Delta_m$,

$$
\log\left(\frac{q_{\boldsymbol{x}}(\boldsymbol{y})}{f(\boldsymbol{x},\omega)(\boldsymbol{y})}\right) < \epsilon'.
$$

This in turn implies that $\sup_{\boldsymbol{x}\in\mathscr{X}} KL(q_{\boldsymbol{x}}, f(\boldsymbol{x},\omega)) < \epsilon'$. From Theorem 3, it follows that

$$
P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} KL(q_{\boldsymbol{x}}, f(\boldsymbol{x},\omega)) < \epsilon'\right\} > P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{\boldsymbol{y}\in\Delta_m} |f(\boldsymbol{x},\omega) - q_{\boldsymbol{x}}| < \epsilon\right\} > 0.
$$

Now, suppose that $\inf_{\boldsymbol{x}\in\mathscr{X}} \inf_{\boldsymbol{y}\in\Delta_d} q_{\boldsymbol{x}}(\boldsymbol{y}) \approx 0$. Here we use a similar reasoning as in the proof of Theorem 2 of Petrone & Wasserman (2002). Consider $a > 0$ and

$$q^1_{\boldsymbol{x}}(\boldsymbol{y}) = \frac{q_{\boldsymbol{x}}(\boldsymbol{y}) \vee a}{\int_{\Delta_m} q_{\boldsymbol{x}}(\boldsymbol{y}) \vee a \; d\boldsymbol{y}},$$

where $a \vee b$ stands for the maximum between a and b. Clearly $q^1_{\boldsymbol{x}}(\boldsymbol{y})$ is a density function such that $q_{\boldsymbol{x}}(\boldsymbol{y}) \leq C q^1_{\boldsymbol{x}}(\boldsymbol{y})$, with $C = \int_{\Delta_m} q_{\boldsymbol{x}}(\boldsymbol{y}) \vee a \; d\boldsymbol{y}$, and is greater than zero. Hence $\sup_{\boldsymbol{x}\in\mathscr{X}} KL(q^1_{\boldsymbol{x}}, f(\boldsymbol{x},\omega)) < \epsilon'$. Considering a and ϵ' sufficiently small, it follows that there exists $\tilde{\epsilon} > 0$, sucht that

$$KL(q_{\boldsymbol{x}}, f(\boldsymbol{x},\omega)) \leq (C+1)\log(C) + C\left\{KL(q^1_{\boldsymbol{x}}, f(\boldsymbol{x},\omega)) + \sqrt{KL(q^1_{\boldsymbol{x}}, f(\boldsymbol{x},\omega))}\right\} < \tilde{\epsilon}.$$

Thus, from the first part of this proof, it follows that

$$P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} KL(q_{\boldsymbol{x}}, f(\boldsymbol{x},\omega)) < \tilde{\epsilon}\right\} \geq P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} KL(q^1_{\boldsymbol{x}}, f(\boldsymbol{x},\omega)) < \epsilon'\right\} > 0,$$

which completes the proof of the theorem. □

D.5. Proof of Theorem 5

The following Lemma is used in the proofs of continuity and association structure of the processes. This Lemma states that equicontinuous families of functions preserve a.s. continuity and convergence in distribution of stochastic processes.

Lemma 3. *Let $\mathscr{F} = \{f_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$ be a set of known bijective continuous functions such that for every $\boldsymbol{x} \in \mathscr{X}$, $f_{\boldsymbol{x}} : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ is such that for every $\boldsymbol{a} \in \mathbb{R}^m$, $f_{\boldsymbol{x}}(\boldsymbol{a})$ is a continuous functions of $\boldsymbol{x}$. In addition asume that $\mathscr{F}$ is an equicontinuous family of functions of $\boldsymbol{a}$ or $\{\boldsymbol{x} \mapsto f_{\boldsymbol{x}}(\boldsymbol{a}) : \boldsymbol{a} \in \mathbb{R}^m, f_x \in \mathscr{F}\}$ is an equicontinuous family of functions of $\boldsymbol{x}$. Let $g_i : \mathscr{X} \times \Omega \longrightarrow \mathbb{R}^m$, $i \geq 1$, be stochastic processes defined on an appropiate probability space $(\Omega, \mathscr{A}, P)$.*

- *(i) If for every $i \in \mathbb{N}$, the stochastic process g_i is P-a.s. continuous, then $\boldsymbol{x} \mapsto f_{\boldsymbol{x}}\{g_i(\boldsymbol{x},\cdot)\}$, $i \in \mathbb{N}$ is P-a.s. continuous.*
- *(ii) Consider $\{\boldsymbol{x}_j\}_{j=1}^{\infty} \subset \mathscr{X}$, such that $\lim_{j\to\infty} \boldsymbol{x}_j = \boldsymbol{x}_0 \in \mathscr{X}$. If $g_i(\boldsymbol{x}_j,\cdot) \xrightarrow{\mathscr{L}} g_i(\boldsymbol{x}_0,\cdot)$, as $j \longrightarrow \infty$, then $f_{\boldsymbol{x}_j}\{g_i(\boldsymbol{x}_j,\cdot)\} \xrightarrow{\mathscr{L}} f_{\boldsymbol{x}_0}\{g_i(\boldsymbol{x}_0,\cdot)\}$, as $j \longrightarrow \infty$.*

Proof of Lemma 3

(i) First consider, for every $\boldsymbol{x} \in \mathscr{X}$, $f_{\boldsymbol{x}}$ an equicontinuous of function of $\boldsymbol{a}$. Consider $\boldsymbol{x}_0 \in \mathscr{X}$. Since $f_{\boldsymbol{x}}(g_i(\boldsymbol{x}_0,\omega))$ is a continuous function of $\boldsymbol{x}$, there exists $\delta_1 > 0$ such that for all $\boldsymbol{x} \in B(\boldsymbol{x}_0, \delta_1)$, $|f_{\boldsymbol{x}}(g_i(\boldsymbol{x}_0,\omega)) -$

$f_{\boldsymbol{x}_0}(g_i(\boldsymbol{x}_0,\omega))| < \epsilon/2$. By assumption, g_i, being a P-a.s. continuous stochastic process, implies that, for almost every $\omega \in \Omega$, and for every $\epsilon_2 > 0$, there exists $\delta_2 > 0$ such that for all $\boldsymbol{x} \in B(\boldsymbol{x}_0, \delta_2)$, $|g_i(\boldsymbol{x},\omega) - g_i(\boldsymbol{x}_0,\omega)| < \epsilon_2$. Hence, by the equicontinuity of $f_{\boldsymbol{x}}$, for almost every $\omega \in \Omega$, and every $g_i(\boldsymbol{x},\omega) \in B(g_i(\boldsymbol{x}_0,\omega),\epsilon_2)$, $|f_{\boldsymbol{x}}(g_i(\boldsymbol{x},\omega)) - f_{\boldsymbol{x}}(g_i(\boldsymbol{x}_0,\omega))| < \epsilon/2$. Finally, considering $\delta = \min\{\delta_1,\delta_2\}$ which does not depend on $f_{\boldsymbol{x}}$, by the triangle inequality, it follows that for every $\omega \in \Omega$ and every $\boldsymbol{x} \in B(\boldsymbol{x}_0,\delta)$, $|f_{\boldsymbol{x}}\{g_i(\boldsymbol{x},\omega)\} - f_{\boldsymbol{x}_0}\{g_i(\boldsymbol{x}_0,\omega)\}| < \epsilon$, which completes this part of the proof. Now consider $\{\boldsymbol{x} \mapsto f_{\boldsymbol{x}}(\boldsymbol{a}) : \boldsymbol{a} \in \mathbb{R}^m, f_x \in \mathscr{F}\}$ an equicontinuous family of functions of $\boldsymbol{x}$. The proof is similar to the previous. By the equicontinuity consideration, there exists $\delta_1 > 0$ such that for every $\boldsymbol{x} \in B(\boldsymbol{x}_0,\delta_1)$, $|f_{\boldsymbol{x}}(g_i(\boldsymbol{x},\omega)) - f_{\boldsymbol{x}_0}(g_i(\boldsymbol{x},\omega))| < \epsilon/2$. Since g_i is a P-a.s. continuous stochastic process, for almost every $\omega \in \Omega$, and for every $\epsilon_2 > 0$, there exists $\delta_2 > 0$ such that for every $\boldsymbol{x} \in B(\boldsymbol{x}_0,\delta_2)$, $|g_i(\boldsymbol{x},\omega) - g_i(\boldsymbol{x}_0,\omega)| < \epsilon_2$. Due to continuity of $f_{\boldsymbol{x}}$ as a function for $\boldsymbol{a}$, it follows that for almost every $\omega \in \Omega$, and every $g_i(\boldsymbol{x},\omega) \in B(g_i(\boldsymbol{x}_0,\omega),\epsilon_2)$, $|f_{\boldsymbol{x}_0}(g_i(\boldsymbol{x},\omega)) - f_{\boldsymbol{x}_0}(g_i(\boldsymbol{x}_0,\omega))| < \epsilon/2$. Finally, considering $\delta = \min\{\delta_1,\delta_2\}$ which does not depend on $f_{\boldsymbol{x}}$, by the triangle inequality, it follows that for every $\omega \in \Omega$ and every $\boldsymbol{x} \in B(\boldsymbol{x}_0,\delta)$, $|f_{\boldsymbol{x}}\{g_i(\boldsymbol{x},\omega)\} - f_{\boldsymbol{x}_0}\{g_i(\boldsymbol{x}_0,\omega)\}| < \epsilon$, which completes the proof of the first part of the lemma.

(ii) Consider $\mathscr{F}$ an equicontinuous family of functions of $\boldsymbol{a}$ or $\{\boldsymbol{x} \mapsto f_{\boldsymbol{x}}(\boldsymbol{a}) : \boldsymbol{a} \in \mathbb{R}^m, f_x \in \mathscr{F}\}$ an equicontinuous family of functions of $\boldsymbol{x}$. If $g_i(\boldsymbol{x}_j,\cdot) \xrightarrow{\mathscr{L}} g_i(\boldsymbol{x}_0,\cdot)$, as $j \longrightarrow \infty$, then by baby Skorohod's theorem (Resnick, 2019), there exist random variables $\{\tilde{g}_i(\boldsymbol{x}_j,\cdot)\}_{j\geq 0}$ defined on the Lebesgue probability space $([0,1],\mathscr{B}([0,1]),\lambda)$, where λ is the Lebesgue measure, such that for each fixed $j \geq 0$, $g_i(\boldsymbol{x}_j,\cdot) \overset{d}{=} \tilde{g}_i(\boldsymbol{x}_j,\cdot)$, and $\tilde{g}_i(\boldsymbol{x}_j,\cdot) \longrightarrow \tilde{g}_i(\boldsymbol{x}_0,\cdot)$ λ-a.s. as $j \longrightarrow \infty$. Since $f_{\boldsymbol{x}}(\boldsymbol{a})$ is a continuous function of $\boldsymbol{a}$, it follows that for $\boldsymbol{x} \in \mathscr{X}$, $f_{\boldsymbol{x}}\{g_i(\boldsymbol{x}_j,\cdot)\} \overset{d}{=} f_{\boldsymbol{x}}\{\tilde{g}_i(\boldsymbol{x}_j,\cdot)\}$. In particular, $f_{\boldsymbol{x}_j}\{g_i(\boldsymbol{x}_j,\cdot)\} \overset{d}{=} f_{\boldsymbol{x}_j}\{\tilde{g}_i(\boldsymbol{x}_j,\cdot)\}$ and $f_{\boldsymbol{x}_0}\{g_i(\boldsymbol{x}_j,\cdot)\} \overset{d}{=} f_{\boldsymbol{x}_0}\{\tilde{g}_i(\boldsymbol{x}_j,\cdot)\}$. Since $\tilde{g}_i(\boldsymbol{x}_j,\cdot) \longrightarrow \tilde{g}_i(\boldsymbol{x}_0,\cdot)$ λ-a.s. as $j \longrightarrow \infty$ then $\tilde{g}_i$ is λ-a.s continuous. Therefore, by Lemma 3 part *(i)*, $f_{\boldsymbol{x}_j}\{\tilde{g}_i(\boldsymbol{x}_j,\cdot)\} \longrightarrow f_{\boldsymbol{x}_0}\{\tilde{g}_i(\boldsymbol{x}_0,\cdot)\}$ λ-a.s. as $j \longrightarrow \infty$ which implies that $f_{\boldsymbol{x}_j}\{\tilde{g}_i(\boldsymbol{x}_j,\cdot)\} \xrightarrow{\mathscr{L}} f_{\boldsymbol{x}_0}\{\tilde{g}_i(\boldsymbol{x}_0,\cdot)\}$. Thus

$$f_{\boldsymbol{x}_j}\{g_i(\boldsymbol{x}_j,\cdot)\} \overset{d}{=} f_{\boldsymbol{x}_j}\{\tilde{g}_i(\boldsymbol{x}_j,\cdot)\} \xrightarrow{\mathscr{L}} f_{\boldsymbol{x}_0}\{\tilde{g}_i(\boldsymbol{x}_0,\cdot)\} \overset{d}{=} f_{\boldsymbol{x}_0}\{g_i(\boldsymbol{x}_0,\cdot)\},$$

as $j \longrightarrow \infty$, which completes proof of the lemma. □

Now we provide the proof of the theorem. Firstly, assume that F is a DMBPP $(\lambda,\boldsymbol{\Psi}_\eta,\boldsymbol{\Psi}_z,\mathscr{V},\mathscr{H})$. Since the elements of $\mathscr{V}$ and $\mathscr{H}$ are equicontinuous functions of $\boldsymbol{x}$, and for every $i \geq 1$, the stochastic processes η_i and $\boldsymbol{z}_i$ are P–a.s. continuous, by Lemma 3 and continuous mapping theorem, it follows that $\boldsymbol{x} \mapsto v_{\boldsymbol{x}}(\eta_i(\boldsymbol{x},\cdot))$, $\boldsymbol{x} \mapsto w_i(\boldsymbol{x},\cdot)$, and $\boldsymbol{x} \mapsto \boldsymbol{\theta}_i(\boldsymbol{x},\cdot\,)$ are P–a.s. continuous mappings. Now, the ceiling function being continuous from the left and having a limit from the right implies that, for $i \geq 1$, and almost every $\omega \in \Omega$, $\lceil k(\omega)\boldsymbol{\theta}_i(\boldsymbol{x}_l,\omega)\rceil$ has a limit, as $l \longrightarrow \infty$. Note that there exists $M > 0$ such that, for every $\boldsymbol{y} \in \Delta_m$, $i \geq 1$, $\boldsymbol{x} \in \mathscr{X}$ and

$\omega \in \Omega$, $\mathrm{d}(\boldsymbol{y} \mid \alpha(k(\omega), \lceil k(\omega)\boldsymbol{\theta}_i(\boldsymbol{x},\omega)\rceil)) \leq M$, where $\alpha(k, \mathbf{j}) = (\mathbf{j},\ k+m-\sum_{l=1}^{m} j_l)$, and that for every $\boldsymbol{x} \in \mathscr{X}$ and $\omega \in \Omega$, $\sum_{i=1}^{\infty} w_i(\boldsymbol{x},\omega) = 1$. Then by dominated convergence theorem for series, the density, w.r.t. Lebesgue measure, of $F(\boldsymbol{x},\cdot)$, $f(\boldsymbol{x}_l,\omega)$, has *a.s.* a limit, say $\tilde{f}(\boldsymbol{x}_0,\cdot)$. This is, for every $\boldsymbol{y} \in \Delta_m$,

$$\Pr\left\{\omega \in \Omega : \lim_{l\to\infty} f(\boldsymbol{x}_l,\omega)(\boldsymbol{y}) = \tilde{f}(\boldsymbol{x}_0,\omega)(\boldsymbol{y}),\right\} = 1,$$

Let $\tilde{F}(\boldsymbol{x},\omega)$ be a probability measure with density function $\tilde{f}(\boldsymbol{x},\omega)$. A direct application of Scheffe's theorem implies that $F(\boldsymbol{x}_l,\cdot)$ converges in total variation norm to $\tilde{F}(\boldsymbol{x}_0,\cdot)$ as $l \longrightarrow \infty$, a.s., this is,

$$P\left\{\omega \in \Omega : \lim_{l\to\infty} \sup_{B\in\mathscr{B}(\Delta_m)} |F(\boldsymbol{x}_l,\omega)(B) - \tilde{F}(\boldsymbol{x}_0,\omega)(B)| = 0,\right\} = 1,$$

which completes the proof of the theorem when F is a DMBPP.

Now, assume that F is wDMBPP$(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_z, \mathscr{H})$. This proof follows the same arguments as in the previous part, but the arguments related to the weights of the process are not needed. Thus, there exists a probability measure $\tilde{F}(\boldsymbol{x},\omega)$, such that

$$P\left\{\omega \in \Omega : \lim_{l\to\infty} \sup_{B\in\mathscr{B}(\Delta_m)} |F(\boldsymbol{x}_l,\omega)(B) - \tilde{F}(\boldsymbol{x}_0,\omega)(B)| = 0,\right\} = 1,$$

which completes the proof of the theorem. □

D.6. Proof of Theorem 6

The proof of this theorem follows the reasoning of the proof of Theorem 6 in Barrientos et al. (2017). For completeness the proof is stated below. Here, we prove that for every $\{\boldsymbol{x}_l\}_{l=0}^{\infty}$, with $\boldsymbol{x}_l \in \mathscr{X}$, such that $\lim_{l\to\infty} \boldsymbol{x}_l = \boldsymbol{x}_0$,

$$P\left\{\omega \in \Omega : \lim_{l\to\infty} \sup_{B\in\mathscr{B}(\Delta_m)} |F(\boldsymbol{x}_l,\omega)(B) - F(\boldsymbol{x}_0,\omega)(B)| = 0,\right\} = 1.$$

By assumption, for every $i \geq 1$, the stochastic processes η_i are *a.s.* continuous, i.e., for every $i \geq 1$, $\boldsymbol{x} \longmapsto \eta_i(\boldsymbol{x},\cdot)$ is an *a.s.* continuous function. By Lemma 3, the equicontinuity assumption of $\mathscr{V}$ as a function of $\boldsymbol{x}$, and continuous mapping theorem, it follows that for every $i \geq 1$, $\boldsymbol{x} \longmapsto w_i(\boldsymbol{x},\cdot)$ is an *a.s.* continuous function. Therefore for every $i \geq 1$ and every $\{\boldsymbol{x}_l\}_{l=0}^{\infty}$, such that $\lim_{l\to\infty} \boldsymbol{x}_l = \boldsymbol{x}_0$, we have that $\lim_{l\to\infty} w_i(\boldsymbol{x}_l,\cdot) = w_i(\boldsymbol{x}_0,\cdot)$, *a.s.*. Noting that there exists $M > 0$ such that, for every $\boldsymbol{y} \in \Delta_m$, $i \geq 1$, and $\omega \in \Omega$, $\mathrm{d}(\boldsymbol{y} \mid \alpha(k(\omega), \lceil k(\omega)\boldsymbol{\theta}_i(\omega)\rceil)) \leq M$, and that for every $\boldsymbol{x} \in \mathscr{X}$ and $\omega \in \Omega$, $\sum_{i=1}^{\infty} w_i(\boldsymbol{x},\omega) = 1$, dominated convergence theorem for series implies that the density, w.r.t. Lebesgue measure, of $F(\boldsymbol{x},\cdot)$ is *a.s.* continuos, i.e., for every $\boldsymbol{y} \in \Delta_m$,

$$\Pr\left\{\omega \in \Omega : \lim_{l\to\infty} f(\boldsymbol{x}_l,\omega)(\boldsymbol{y}) = f(\boldsymbol{x}_0,\omega)(\boldsymbol{y}),\right\} = 1.$$

Finally, a direct application of Scheffe's theorem implies that $F(\boldsymbol{x}_j,\cdot)$ converges in total variation norm to $F(\boldsymbol{x}_0,\cdot)$ as $j \longrightarrow \infty$, a.s., which completes the proof of the theorem. □

D.7. Proof of Theorem 7

Here we prove that for every $\boldsymbol{y} \in \tilde{\Delta}_m$, every $\{\boldsymbol{x}_l\}_{l=0}^{\infty}$, with $\boldsymbol{x}_l \in \mathscr{X}$, such that $\lim_{l\to\infty} \boldsymbol{x}_l = \boldsymbol{x}_0$,

$$\frac{E\{F(\boldsymbol{x}_l,\cdot)(B_{\boldsymbol{y}})F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}})\} - E\{F(\boldsymbol{x}_l,\cdot)(B_{\boldsymbol{y}})\}E\{F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}})\}}{\sqrt{Var\{F(\boldsymbol{x}_l,\cdot)(B_{\boldsymbol{y}})\}Var\{F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}})\}}} \longrightarrow 1,$$

as $l \longrightarrow \infty$, where $B_{\boldsymbol{y}} = [0,y_1]\times\ldots\times[0,y_m]$, and expectations are obtained by the law of total expectation conditioning on the degree of the polynomial. We state the complete proof for the general definition of F. The proof for the simplified versions of F are straightforward. In order to reduce the notation, $k(\omega)$ is denoted by k when necessary.

First, assume that F is a DMBPP$(\lambda,\boldsymbol{\Psi}_\eta,\boldsymbol{\Psi}_z,\mathscr{V},\mathscr{H})$. Notice that for every $\boldsymbol{y}\in\Delta_m$ and every $\boldsymbol{x}\in\mathscr{X}$,

$$E\{F(\boldsymbol{x},\cdot)(B_{\boldsymbol{y}})|k=k_0\} = \sum_{\mathbf{j}\in\mathcal{H}_{k_0,m}} E\left\{F^*(\boldsymbol{x},\cdot)\left(\frac{\mathbf{j}}{k}\right)|k=k_0\right\}\mathrm{Mult}(\mathbf{j}\mid\beta(k_0,\boldsymbol{y})),$$

where $\mathcal{H}_{k,m} = \{(j_1,\ldots,j_m)\in\{0,\ldots,k\}^m : \sum_{l=1}^m j_l \le k+m-1\}$, $\beta(k_0,\boldsymbol{y}) = (k_0+m-1,\boldsymbol{y})$, $(\mathbf{j}/k) = (j_1/k,\ldots,j_m/k)$, $\mathrm{Mult}(\cdot\mid k,\boldsymbol{y})$ denotes the probability mass function of a multinomial distribution with parameters $(k,\boldsymbol{y})$, and

$$F^*(\boldsymbol{x},\cdot)\left(\frac{\mathbf{j}}{k}\right) = \sum_{i=1}^{\infty} w_i(\boldsymbol{x},\cdot)\mathbb{I}\{\boldsymbol{\theta}_i(\boldsymbol{x},\cdot)\}_{\{A_{\mathbf{j},k}\}},$$

where $A_{\mathbf{j},k} = [0,j_1/k]\times\ldots\times[0,j_m/k]$. Since the stochastic processes $\{\eta_i\}_{i\ge1}$ and $\{\boldsymbol{z}_i\}_{i\ge1}$ are independent and identically distributed, it follows that,

$$\begin{aligned} E\left\{F^*(\boldsymbol{x},\cdot)\left(\frac{\mathbf{j}}{k}\right)\middle| k=k_0\right\} &= \sum_{i=1}^{\infty} E\{w_i(\boldsymbol{x},\cdot)\}E\left\{\mathbb{I}\{\boldsymbol{\theta}_1(\boldsymbol{x},\cdot)\}_{\{A_{\mathbf{j},k_0}\}}\right\}, \\ &= G_{0,\boldsymbol{x}}(A_{\mathbf{j},k_0}), \end{aligned}$$

where $G_{0,\boldsymbol{x}}(A) = G_0(\boldsymbol{x},\cdot)(A)$ denotes the distribution function of $\boldsymbol{\theta}_1(\boldsymbol{x},\cdot)$ defined on $\tilde{\Delta}_m$. Thus

$$E\{F(\boldsymbol{x},\cdot)(B_{\boldsymbol{y}})\mid k=k_0\} = \sum_{\mathbf{j}\in\mathcal{H}_{k_0,m}} G_{0,\boldsymbol{x}}(A_{\mathbf{j},k_0})\,\mathrm{Mult}(\mathbf{j}\mid\beta(k_0,\boldsymbol{y})).$$

Noting that for every $\boldsymbol{x},\boldsymbol{x}_0\in\mathscr{X}$ and every $\boldsymbol{y}\in\Delta_m$,

$$\begin{aligned} &E\{F(\boldsymbol{x},\cdot)(B_{\boldsymbol{y}})F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}})|\,k=k_0\} \\ &\quad = \sum_{\substack{\mathbf{j}_1\in\mathcal{H}_{k,m},\\ \mathbf{j}_2\in\mathcal{H}_{k,m}}} E\left\{F^*(\boldsymbol{x},\boldsymbol{x}_0,\cdot)\left(\frac{\mathbf{j}_1}{k},\frac{\mathbf{j}_2}{k}\right)\middle| k=k_0\right\}\times\bar{\mathrm{M}}(\mathbf{j}_1,\mathbf{j}_2\mid\beta(k_0,\boldsymbol{y})), \end{aligned}$$

where $\bar{M}(\mathbf{j}_1, \mathbf{j}_2 \mid \beta(k, \boldsymbol{y})) = \mathrm{Mult}(\mathbf{j}_1 \mid \beta(k, \boldsymbol{y})) \times \mathrm{Mult}(\mathbf{j}_2 \mid \beta(k, \boldsymbol{y}))$, and

$$F^*(\boldsymbol{x}, \boldsymbol{x}_0, \cdot)\left(\frac{\mathbf{j}_1}{k}, \frac{\mathbf{j}_2}{k}\right) = \sum_{i=1}^{\infty} w_i(\boldsymbol{x}, \cdot) w_i(\boldsymbol{x}_0, \cdot) \mathbb{I}\left\{\boldsymbol{\theta}_i(\boldsymbol{x}, \cdot)\right\}_{\left\{A_{\mathbf{j}_1,k}\right\}} \mathbb{I}\left\{\boldsymbol{\theta}_i(\boldsymbol{x}_0, \cdot)\right\}_{\left\{A_{\mathbf{j}_2,k}\right\}},$$
$$+ \sum_{\substack{i, i_1 = 1, \\ i \neq i_1}}^{\infty} w_i(\boldsymbol{x}, \cdot) w_{i_1}(\boldsymbol{x}_0, \cdot) \mathbb{I}\left\{\boldsymbol{\theta}_i(\boldsymbol{x}, \cdot)\right\}_{\left\{A_{\mathbf{j}_1,k}\right\}} \mathbb{I}\left\{\boldsymbol{\theta}_{i_1}(\boldsymbol{x}_0, \cdot)\right\}_{\left\{A_{\mathbf{j}_2,k}\right\}}.$$

Applying a similar reasoning as before, it follows that, for every $\boldsymbol{x}, \boldsymbol{x}_0 \in \mathscr{X}$ and every $\boldsymbol{y} \in \Delta_m$,

$$\begin{aligned} &E\left\{F(\boldsymbol{x}, \cdot)(B_{\boldsymbol{y}}) F(\boldsymbol{x}_0, \cdot)(B_{\boldsymbol{y}}) \mid k = k_0\right\} \\ &\quad = \sum_{\substack{\mathbf{j}_1 \in \mathcal{H}_{k_0,m} \\ ,\mathbf{j}_2 \in \mathcal{H}_{k_0,m}}} \left\{ \sum_{i=1}^{\infty} E\left\{w_i(\boldsymbol{x}, \cdot) w_i(\boldsymbol{x}_0, \cdot)\right\} G_{0,\boldsymbol{x},\boldsymbol{x}_0}\left(A_{\mathbf{j}_1,k_0} \times A_{\mathbf{j}_2,k_0}\right), \right. \\ &\qquad \left. + \sum_{\substack{i, i_1 = 1, \\ i \neq i_1}}^{\infty} E\left\{w_i(\boldsymbol{x}, \cdot) w_{i_1}(\boldsymbol{x}_0, \cdot)\right\} G_{0,\boldsymbol{x}}\left(A_{\mathbf{j}_1,k_0}\right) G_{0,\boldsymbol{x}_0}\left(A_{\mathbf{j}_2,k_0}\right) \right\}, \\ &\qquad \times \bar{M}(\mathbf{j}_1, \mathbf{j}_2 \mid \beta(k_0, \boldsymbol{y})), \end{aligned}$$

where $G_{0,\boldsymbol{x},\boldsymbol{x}_0}(A) = G_0((\boldsymbol{x}, \boldsymbol{x}_0), \cdot)(A)$ denotes the joint distribution function of $(\boldsymbol{\theta}_i(\boldsymbol{x}, \cdot), \boldsymbol{\theta}_i(\boldsymbol{x}_0, \cdot))$ defined on $\tilde{\Delta}_m^2$. In particular, for $\boldsymbol{x} = \boldsymbol{x}_0$,

$$\begin{aligned} E\left\{F(\boldsymbol{x}, \cdot)(B_{\boldsymbol{y}})^2 \,\middle|\, k = k_0\right\} = &\sum_{\substack{\mathbf{j}_1 \in \mathcal{H}_{k_0,m} \\ ,\mathbf{j}_2 \in \mathcal{H}_{k_0,m}}} \left\{ \sum_{i=1}^{\infty} E\left\{w_i(\boldsymbol{x}, \cdot)^2\right\} G_{0,\boldsymbol{x}}\left(A_{\min\{\mathbf{j}_1,\mathbf{j}_2\},\mathrm{k}_0}\right), \right. \\ &\left. + \sum_{\substack{i, i_1 = 1, \\ i \neq i_1}}^{\infty} E\left\{w_i(\boldsymbol{x}, \cdot) w_{i_1}(\boldsymbol{x}, \cdot)\right\} G_{0,\boldsymbol{x}}\left(A_{\mathbf{j}_1,k_0}\right) G_{0,\boldsymbol{x}}\left(A_{\mathbf{j}_2,k_0}\right) \right\}, \\ &\times \bar{M}(\mathbf{j}_1, \mathbf{j}_2 \mid \beta(k_0, \boldsymbol{y}), \end{aligned}$$

where $A_{\min\{\mathbf{j}_1,\mathbf{j}_2\},k} = [0, \min\{j_{11}, j_{21}\}/k] \times \ldots \times [0, \min\{j_{1m}, j_{2m}\}/k]$. By assumption, for every $i \geq 1$, and every $\{\boldsymbol{x}_l\}_{l=0}^{\infty}$, with $\boldsymbol{x}_l \in \mathscr{X}$, such that $\lim_{l \to \infty} \boldsymbol{x}_l = \boldsymbol{x}_0$, the processes $\eta_i(\boldsymbol{x}_l, \cdot)$ and $\boldsymbol{z}_i(\boldsymbol{x}_l, \cdot)$ converge in distribution to $\eta_i(\boldsymbol{x}_0, \cdot)$ and $\boldsymbol{z}_i(\boldsymbol{x}_0, \cdot)$, respectively, as $l \to \infty$. Since $\mathscr{V}$ and $\mathscr{H}$ are sets of equicontinuous functions of $\boldsymbol{x}$, by Lemma 3, and continuous mapping theorem, it follows that $w_i(\boldsymbol{x}_l, \cdot)$ converges in distribution to $w_i(\boldsymbol{x}_0, \cdot)$ and $\boldsymbol{\theta}_i(\boldsymbol{x}_l, \cdot)$ converges in distribution to $\boldsymbol{\theta}_i(\boldsymbol{x}_0, \cdot)$, as $l \to \infty$. Thus, for every $a \in \tilde{\Delta}_m$, $\lim_{l \to \infty} G_{0,\boldsymbol{x}_l}(a) = G_{0,\boldsymbol{x}_0}(a)$. Noting that $w_i(\boldsymbol{x}, \cdot)$ are bounded variables, Portmanteau's theorem implies that the mappings $\boldsymbol{x} \mapsto E\{w_i(\boldsymbol{x}, \cdot)\}$, $\boldsymbol{x} \mapsto E\{w_i(\boldsymbol{x}, \cdot)^2\}$ and $\boldsymbol{x} \mapsto E\{w_i(\boldsymbol{x}, \cdot) w_i(\boldsymbol{x}_0, \cdot)\}$, are continuous. Now, considering $\boldsymbol{y} \in \tilde{\Delta}_m$, the above expressions and few applications of dominated convergence theorem for series, it follows that,

$$\begin{aligned}\lim_{j\to\infty} E\left\{F(\boldsymbol{x}_j,\cdot)(B_{\boldsymbol{y}})\right\} &= \sum_{k_0=1}^{\infty}\lim_{l\to\infty} E\left\{F(\boldsymbol{x}_l,\cdot)(B_{\boldsymbol{y}}) \mid k_0\right\} P\{\omega\in\Omega : k(\omega)=k_0\},\\ &= E\left\{F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}})\right\},\end{aligned}$$

$$\begin{aligned}\lim_{l\to\infty} E\left\{F(\boldsymbol{x}_l,\cdot)(B_{\boldsymbol{y}})^2\right\} &= \sum_{k_0=1}^{\infty}\lim_{l\to\infty} E\left\{F(\boldsymbol{x}_l,\cdot)(B_{\boldsymbol{y}})^2 \;\middle|\; k_0\right\} P\{\omega\in\Omega : k(\omega)=k_0\},\\ &= E\left\{F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}})^2\right\},\end{aligned}$$

and

$$\begin{aligned}\lim_{j\to\infty} E\left\{F(\boldsymbol{x}_j,\cdot)(B_{\boldsymbol{y}})F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}})\right\} &= \sum_{k_0=1}^{\infty}\lim_{j\to\infty} E\left\{F(\boldsymbol{x}_j,\cdot)(B_{\boldsymbol{y}})F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}}) \mid k_0\right\},\\ &\quad\times P\{\omega\in\Omega : k(\omega)=k_0\},\\ &= E\left\{F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}})^2\right\}.\end{aligned}$$

Thus the proof is completed when F is a DMBPP$(\lambda,\boldsymbol{\Psi}_\eta,\boldsymbol{\Psi}_{\boldsymbol{z}},\mathscr{V},\mathscr{H})$. □

D.8. Proof of Theorem 8

The proof of this theorem is a straightforward extension of the proof of Theorem 8 in Barrientos et al. (2017). For completeness we state the proof in what follows. Here we use the law of total covariance conditioning on the degree of the polynomial.

Assume that F is a DMBPP$(\lambda,\boldsymbol{\Psi}_\eta,\boldsymbol{\Psi}_z,\mathscr{V},\mathscr{H})$. By assumption, for every $i\geq 1$, and every $\{(\boldsymbol{x}_{1l},\boldsymbol{x}_{2l})\}_{l=1}^{\infty}$ with $(\boldsymbol{x}_{1l},\boldsymbol{x}_{2l})\in\mathscr{X}^2$ and $(\boldsymbol{x}_1,\boldsymbol{x}_2)\in\mathscr{X}^2$, such that $(\boldsymbol{x}_{1l},\boldsymbol{x}_{2l})\longrightarrow(\boldsymbol{x}_1,\boldsymbol{x}_2)$, as $l\to\infty$, the joint processes $(\eta_i(\boldsymbol{x}_{1l},\cdot),\eta_i(\boldsymbol{x}_{2l},\cdot))$ and $(\boldsymbol{z}_i(\boldsymbol{x}_{1l},\cdot),\boldsymbol{z}_i(\boldsymbol{x}_{2l},\cdot))$ converge in distribution to the processes $(\eta_i(\boldsymbol{x}_1,\cdot),\eta_i(\boldsymbol{x}_2,\cdot))$, and $(\boldsymbol{z}_i(\boldsymbol{x}_1,\cdot),\boldsymbol{z}_i(\boldsymbol{x}_2,\cdot))$, as $l\to\infty$, respectively. Since $\mathscr{V}$ and $\mathscr{H}$ are sets of equicontinuous functions of $\boldsymbol{x}$, by Lemma 3 and continuous mapping theorem, it follows that for every $i\geq 1$, $(w_i(\boldsymbol{x}_{1l},\cdot),w_i(\boldsymbol{x}_{2l},\cdot))$ and $(\boldsymbol{\theta}_i(\boldsymbol{x}_{1l},\cdot),\boldsymbol{\theta}_i(\boldsymbol{x}_{2l},\cdot))$ converge in distribution to $(w_i(\boldsymbol{x}_1,\cdot),w_i(\boldsymbol{x}_2,\cdot))$ and $(\boldsymbol{\theta}_i(\boldsymbol{x}_1,\cdot),\boldsymbol{\theta}_i(\boldsymbol{x}_2,\cdot))$, as $l\to\infty$, respectively. Thus, for every $\boldsymbol{a}_1\in\tilde{\Delta}_m$ and $\boldsymbol{a}_2\in\tilde{\Delta}_m$, $\lim_{l\to\infty}G_{0,\boldsymbol{x}_{1l},\boldsymbol{x}_{2l}}(\boldsymbol{a}_1,\boldsymbol{a}_2)=G_{0,\boldsymbol{x}_1,\boldsymbol{x}_2}(\boldsymbol{a}_1,\boldsymbol{a}_2)$, where $G_{0,\boldsymbol{x}_1,\boldsymbol{x}_2}$ denotes the joint distribution function of$(\boldsymbol{\theta}_i(\boldsymbol{x}_1,\cdot),\boldsymbol{\theta}_i(\boldsymbol{x}_2,\cdot))$. Noting that for every $\boldsymbol{x}$, $w_i(\boldsymbol{x},\cdot)$ are bounded variables, Portmanteau's theorem implies that mappings $\boldsymbol{x}\mapsto E\left\{w_i(\boldsymbol{x},\cdot)\right\}$ and $(\boldsymbol{x}_1,\boldsymbol{x}_2)\mapsto E\left\{w_i(\boldsymbol{x}_1,\cdot)w_{i_1}(\boldsymbol{x}_2,\cdot)\right\}$, $i,i_1\in\mathbb{N}$, are continuous. In addition, for $(\boldsymbol{x}_1,\boldsymbol{x}_2)\in\mathscr{X}^2$ such that $\|\boldsymbol{x}_1-\boldsymbol{x}_2\|>\gamma$, the assumption

$$Cov\left[\mathbb{I}_{\{A_1\}}\left\{\boldsymbol{z}_i(\boldsymbol{x}_1,\cdot)\right\},\mathbb{I}_{\{A_2\}}\left\{\boldsymbol{z}_i(\boldsymbol{x}_2,\cdot)\right\}\right]=0,$$

and

$$Cov\left[\mathbb{I}\left\{\eta_i(\boldsymbol{x}_1,\cdot)\right\}_{\{A_1\}},\mathbb{I}\left\{\eta_i(\boldsymbol{x}_2,\cdot)\right\}_{\{A_2\}}\right]=0,$$

imply that

$$E\left\{\mathbb{I}\left\{\boldsymbol{z}_i(\boldsymbol{x}_1,\cdot)\right\}_{\{A_1\}}\mathbb{I}\left\{\boldsymbol{z}_i(\boldsymbol{x}_2,\cdot)\right\}_{\{A_2\}}\right\}=E\left\{\mathbb{I}\left\{\boldsymbol{z}_i(\boldsymbol{x}_1,\cdot)\right\}_{\{A_1\}}\right\}E\left\{\mathbb{I}\left\{\boldsymbol{z}_i(\boldsymbol{x}_2,\cdot)\right\}_{\{A_2\}}\right\},$$
$$E\left\{\mathbb{I}\left\{\eta_i(\boldsymbol{x}_1,\cdot)\right\}_{\{A_1\}}\mathbb{I}\left\{\eta_i(\boldsymbol{x}_2,\cdot)\right\}_{\{A_2\}}\right\}=E\left\{\mathbb{I}\left\{\eta_i(\boldsymbol{x}_1,\cdot)\right\}_{\{A_1\}}\right\}E\left\{\mathbb{I}\left\{\eta_i(\boldsymbol{x}_2,\cdot)\right\}_{\{A_2\}}\right\}.$$

Therefore, for every $(\boldsymbol{x}_1,\boldsymbol{x}_2)\in\mathscr{X}^2$ such that $\|\boldsymbol{x}_1-\boldsymbol{x}_2\|>\gamma$, it follows that $G_{0,\boldsymbol{x}_1,\boldsymbol{x}_2}(\boldsymbol{a}_1,\boldsymbol{a}_2)=G_{0,\boldsymbol{x}_1}(\boldsymbol{a}_1)G_{0,\boldsymbol{x}_2}(\boldsymbol{a}_2)$, and that for $i,i_1\in\mathbb{N}$,

$$E\left\{w_i(\boldsymbol{x}_1,\cdot)w_{i_1}(\boldsymbol{x}_2,\cdot)\right\}=E\left\{w_i(\boldsymbol{x}_1,\cdot)\right\}E\left\{w_{i_1}(\boldsymbol{x}_2,\cdot)\right\}.$$

Now, considering the expressions from the proof of Theorem 7, for every $\boldsymbol{y}\in\Delta_m$, $(\boldsymbol{x}_1,\boldsymbol{x}_2)\in\mathscr{X}^2$ such that $\|\boldsymbol{x}_1-\boldsymbol{x}_2\|>\gamma$, and an application of dominated convergence theorem, it follows that

$$\begin{aligned}\lim_{l\to\infty}&E\left\{F(\boldsymbol{x}_{1l},\cdot)(B_{\boldsymbol{y}})F(\boldsymbol{x}_{2l},\cdot)(B_{\boldsymbol{y}})\mid k=k_0\right\}\\&=\sum_{\substack{\mathbf{j}_1\in\mathcal{H}_{k_0,m}\\ \mathbf{j}_2\in\mathcal{H}_{k_0,m}}}\sum_{\substack{i=1,\\ i_1=1}}^{\infty}E\left\{w_i(\boldsymbol{x}_1,\cdot)\right\}E\left\{w_{i_1}(\boldsymbol{x}_2,\cdot)\right\}G_{0,\boldsymbol{x}_1}\left(A_{\mathbf{j}_1,k_0}\right)G_{0,\boldsymbol{x}_2}\left(A_{\mathbf{j}_2,k_0}\right),\\&\quad\times\bar{M}(\mathbf{j}_1,\mathbf{j}_2\mid\beta(k_0,\boldsymbol{y})),\\&=\lim_{l\to\infty}E\left\{F(\boldsymbol{x}_{1l},\cdot)(B_{\boldsymbol{y}})\mid k=k_0\right\}E\left\{F(\boldsymbol{x}_{2l},\cdot)(B_{\boldsymbol{y}})\mid k=k_0\right\}.\end{aligned}$$

Thus,

$$\lim_{l\to\infty}Cov\left\{F(\boldsymbol{x}_{1l},\cdot)(B_{\boldsymbol{y}}),F(\boldsymbol{x}_{2l},\cdot)(B_{\boldsymbol{y}})\mid k=k_0\right\}=0.$$

Finally, by dominated convergence theorem for series, it follows that

$$\begin{aligned}\lim_{l\to\infty}&Cov\left[F(\boldsymbol{x}_{1l},\cdot)(B_{\boldsymbol{y}}),F(\boldsymbol{x}_{2l},\cdot)(B_{\boldsymbol{y}})\right]\\&=E\left\{\lim_{l\to\infty}Cov\left[F(\boldsymbol{x}_{1l},\cdot)(B_{\boldsymbol{y}}),F(\boldsymbol{x}_{2l},\cdot)(B_{\boldsymbol{y}})\mid k_0\right]\right\},\\&+Cov\left[\lim_{l\to\infty}E\left\{F(\boldsymbol{x}_{1l},\cdot)(B_{\boldsymbol{y}})\mid k_0\right\},\lim_{l\to\infty}E\left\{F(\boldsymbol{x}_{2l},\cdot)(B_{\boldsymbol{y}})\mid k(\cdot)\right\}\right],\\&=Cov\left[E\left\{F(\boldsymbol{x}_1,\cdot)(B_{\boldsymbol{y}})\mid k_0\right\},E\left\{F(\boldsymbol{x}_2,\cdot)(B_{\boldsymbol{y}})\mid k(\cdot)\right\}\right],\end{aligned}$$

where for every $\boldsymbol{x}\in\mathscr{X}$,

$$E\left\{F(\boldsymbol{x},\cdot)(B_{\boldsymbol{y}})\mid k=k_0\right\}=\sum_{\mathbf{j}\in\mathcal{H}_{k_0,m}}G_{0,\boldsymbol{x}}\left(A_{\mathbf{j},k_0}\right)\mathrm{Mult}(\mathbf{j}\mid k_0+m-1,\boldsymbol{y}).$$

which completes this part of the proof.

Assume now that F is a wDMBPP$(\lambda,\boldsymbol{\Psi}_v,\boldsymbol{\Psi}_z,\mathscr{H})$. By the same arguments as when F is the general model, for every $\boldsymbol{y}\in\Delta_m$ and $(\boldsymbol{x}_1,\boldsymbol{x}_2)\in\mathscr{X}^2$ and an application of dominated convergence theorem, it follows that

$$\lim_{l\to\infty}E\left\{F(\boldsymbol{x}_{1l},\cdot)(B_{\boldsymbol{y}})F(\boldsymbol{x}_{2l},\cdot)(B_{\boldsymbol{y}})\mid k=k_0\right\}$$

$$= \sum_{\substack{\mathbf{j}_1 \in \mathcal{H}_{k_0,m} \\ \mathbf{j}_2 \in \mathcal{H}_{k_0,m}}} \sum_{\substack{i=1, \\ i_1=1}}^{\infty} E\left\{w_i(\cdot) w_{i_1}(\cdot)\right\} G_{0,\boldsymbol{x}_1}\left(A_{\mathbf{j}_1,k_0}\right) G_{0,\boldsymbol{x}_2}\left(A_{\mathbf{j}_2,k_0}\right) \bar{M}(\mathbf{j}_1, \mathbf{j}_2 \mid \beta(k_0, \boldsymbol{y})),$$

and

$$\lim_{l\to\infty} E\left\{F(\boldsymbol{x}_{1l}, \cdot)(B_{\boldsymbol{y}}) \mid k = k_0\right\} E\left\{F(\boldsymbol{x}_{2l}, \cdot)(B_{\boldsymbol{y}}) \mid k = k_0\right\}$$
$$= \sum_{\substack{\mathbf{j}_1 \in \mathcal{H}_{k_0,m} \\ \mathbf{j}_2 \in \mathcal{H}_{k_0,m}}} \sum_{\substack{i=1, \\ i_1=1}}^{\infty} E\left\{w_i(\cdot)\right\} E\left\{w_{i_1}(\cdot)\right\} G_{0,\boldsymbol{x}_1}\left(A_{\mathbf{j}_1,k_0}\right) G_{0,\boldsymbol{x}_2}\left(A_{\mathbf{j}_2,k_0}\right)$$
$$\times \bar{M}(\mathbf{j}_1, \mathbf{j}_2 \mid \beta(k_0, \boldsymbol{y}))$$

Since $Cov\left[\sum_{i=1}^{\infty} w_i(\omega), \sum_{i_1=1}^{\infty} w_{i_1}(\omega)\right] = 0$, it follows that

$$\lim_{l\to\infty} Cov\left\{F(\boldsymbol{x}_{1l}, \cdot), (B_{\boldsymbol{y}}) F(\boldsymbol{x}_{2l}, \cdot)(B_{\boldsymbol{y}}) \mid k = k_0\right\} = 0.$$

Finally, the proof is completed using the same arguments as in the first part. □

D.9. Proof of Theorem 9

The proof of this theorem is a straightforward extension of the proof of Theorem 9 in Barrientos et al. (2017). For completeness we state the proof in what follows. We use the law of total covariance conditioning on the degree of the polynomial. Assume that F is a θDMBPP$(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$. By the same arguments as in the proof of the first part of Theorem 8, for every $\boldsymbol{y} \in \Delta_m$ and $(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathscr{X}^2$ such that $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| > \gamma$, and few applications of dominated convergence theorem, it follows that

$$\lim_{l\to\infty} E\left\{F(\boldsymbol{x}_{1l}, \cdot)(B_{\boldsymbol{y}}) F(\boldsymbol{x}_{2l}, \cdot)(B_{\boldsymbol{y}}) \mid k = k_0\right\}$$
$$= \sum_{\substack{\mathbf{j}_1 \in \mathcal{H}_{k_0,m} \\ \mathbf{j}_2 \in \mathcal{H}_{k_0,m}}} \sum_{\substack{i=1, \\ i_1=1}}^{\infty} E\left\{w_i(\boldsymbol{x}_1, \cdot)\right\} E\left\{w_{i_1}(\boldsymbol{x}_2, \cdot)\right\}$$
$$\times E\left\{\mathbb{I}\left\{\boldsymbol{\theta}_i(\cdot)\right\}_{\{A_{\mathbf{j}_1,k_0}\}} \mathbb{I}\left\{\boldsymbol{\theta}_{i1}(\cdot)\right\}_{\{A_{\mathbf{j}_2,k_0}\}}\right\} \times \bar{M}(\mathbf{j}_1, \mathbf{j}_2 \mid \beta(k_0, \boldsymbol{y})),$$

and

$$\lim_{l\to\infty} E\left\{F(\boldsymbol{x}_{1l}, \cdot)(B_{\boldsymbol{y}}) \mid k = k_0\right\} E\left\{F(\boldsymbol{x}_{2l}, \cdot)(B_{\boldsymbol{y}}) \mid k = k_0\right\}$$
$$= \sum_{\substack{\mathbf{j}_1 \in \mathcal{H}_{k_0,m} \\ \mathbf{j}_2 \in \mathcal{H}_{k_0,m}}} \sum_{\substack{i=1, \\ i_1=1}}^{\infty} E\left\{w_i(\boldsymbol{x}_1, \cdot)\right\} E\left\{w_{i_1}(\boldsymbol{x}_2, \cdot)\right\},$$
$$\times E\left\{\mathbb{I}\left\{\boldsymbol{\theta}_i(\cdot)\right\}_{\{A_{\mathbf{j}_1,k_0}\}}\right\} E\left\{\mathbb{I}\left\{\boldsymbol{\theta}_{i1}(\cdot)\right\}_{\{A_{\mathbf{j}_2,k_0}\}}\right\} \bar{M}(\mathbf{j}_1, \mathbf{j}_2 \mid \beta(k_0, \boldsymbol{y})).$$

Since $\{\boldsymbol{\theta}_i\}_{i\geq 1}$ are independent, then

$$\lim_{l\to\infty} Cov\left[F(\boldsymbol{x}_{1l}, \cdot)(B_{\boldsymbol{y}}), F(\boldsymbol{x}_{2l}, \cdot)(B_{\boldsymbol{y}}) \mid k_0\right]$$

$$= \sum_{\substack{\mathbf{j}_1 \in \mathcal{H}_{k_0,m} \\ \mathbf{j}_2 \in \mathcal{H}_{k_0,m}}} \sum_{i=1}^{\infty} E\left\{w_i(\boldsymbol{x}_1,\cdot)\right\} E\left\{w_i(\boldsymbol{x}_2,\cdot)\right\} Cov\left\{\mathbb{I}\left\{\boldsymbol{\theta}_i(\cdot)\right\}_{\left\{A_{\mathbf{j}_1,k_0}\right\}}, \mathbb{I}\left\{\boldsymbol{\theta}_i(\cdot)\right\}_{\left\{A_{\mathbf{j}_2,k_0}\right\}}\right\},$$
$$\times \bar{M}(\mathbf{j}_1,\mathbf{j}_2 \mid k_0+m-1,\boldsymbol{y}).$$

Finally, by dominated convergence theorem, it follows that, for every $\boldsymbol{y} \in \tilde{\Delta}_m$,

$$\begin{aligned}
&\lim_{l\to\infty} Cov\left[F(\boldsymbol{x}_{1l},\cdot)(B_{\boldsymbol{y}}), F(\boldsymbol{x}_{2l},\cdot)(B_{\boldsymbol{y}})\right] \\
&\quad = E\left\{\lim_{l\to\infty} Cov\left[F(\boldsymbol{x}_{1l},\cdot)(B_{\boldsymbol{y}}), F(\boldsymbol{x}_{2l},\cdot)(B_{\boldsymbol{y}}) \mid k_0\right]\right\}, \\
&\qquad + Cov\left[\lim_{l\to\infty} E\left\{F(\boldsymbol{x}_{1l},\cdot)(B_{\boldsymbol{y}}) \mid k_0\right\}, \lim_{l\to\infty} E\left\{F(\boldsymbol{x}_{2l},\cdot)(B_{\boldsymbol{y}}) \mid k(\cdot)\right\}\right], \\
&\quad = \sum_{k_0=1}^{\infty} \lim_{l\to\infty} Cov\left[F(\boldsymbol{x}_{1l},\cdot)(B_{\boldsymbol{y}}), F(\boldsymbol{x}_{2l},\cdot)(B_{\boldsymbol{y}}) \mid k_0\right] P(\{\omega\in\Omega : k(\omega)=k_0\}), \\
&\qquad + Cov\left[E\left\{F(\boldsymbol{x}_1,\cdot)(B_{\boldsymbol{y}}) \mid k_0\right\}, E\left\{F(\boldsymbol{x}_2,\cdot)(B_{\boldsymbol{y}}) \mid k(\cdot)\right\}\right],
\end{aligned}$$

which completes the proof of the theorem. □

D.10. Proof of Theorem 10

The proof of this theorem is an extension of the proof of Theorem 10 in Barrientos et al. (2017). For completeness we state the proof in what follows. We prove this theorem using the definition of correlation. Expectations are obtained by the law of total expectation, conditioning on the degree of the polynomial. Assume that F is a DMBPP($\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_z, \mathscr{V}, \mathscr{H}$), a θDMBPP($\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}}$) or a wDMBPP($\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_z, \mathscr{H}$). By assumption, for every $i \geq 1$, and every$\{(\boldsymbol{x}_{1l}, \boldsymbol{x}_{2l})\}_{l=1}^{\infty}$, with $(\boldsymbol{x}_{1l}, \boldsymbol{x}_{2l}) \in \mathscr{X}^2$, $(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathscr{X}^2$, such that

$$\lim_{l\to\infty} (\boldsymbol{x}_{1l}, \boldsymbol{x}_{2l}) = (\boldsymbol{x}_1, \boldsymbol{x}_2),$$

the joint processes $(\eta_i(\boldsymbol{x}_{1l},\cdot), \eta_i(\boldsymbol{x}_{2l},\cdot))$ and $(\boldsymbol{z}_i(\boldsymbol{x}_{1l},\cdot), \boldsymbol{z}_i(\boldsymbol{x}_{2l},\cdot))$ converge in distribution to $(\eta_i(\boldsymbol{x}_1,\cdot), \eta_i(\boldsymbol{x}_2,\cdot))$, and $(\boldsymbol{z}_i(\boldsymbol{x}_1,\cdot), \boldsymbol{z}_i(\boldsymbol{x}_2,\cdot))$, as $l \to \infty$, respectively. By the same arguments used in the proof of the first part of Theorem 8, it follows that for every $\boldsymbol{a}_1 \in \tilde{\Delta}_m$ and $\boldsymbol{a}_2 \in \tilde{\Delta}_m$, $\lim_{l\to\infty} G_{0,\boldsymbol{x}_{1l},\boldsymbol{x}_{2l}}(\boldsymbol{a}_1,\boldsymbol{a}_2) = G_{0,\boldsymbol{x}_1,\boldsymbol{x}_2}(\boldsymbol{a}_1,\boldsymbol{a}_2)$, where $G_{0,\boldsymbol{x}_1,\boldsymbol{x}_2}$ denotes the joint distribution function of$(\boldsymbol{\theta}_i(\boldsymbol{x}_1,\cdot), \boldsymbol{\theta}_i(\boldsymbol{x}_2,\cdot))$, and mappings $\boldsymbol{x} \mapsto E\left\{w_i(\boldsymbol{x},\cdot)\right\}$ and

$$(\boldsymbol{x}_1,\boldsymbol{x}_2) \mapsto E\left\{w_i(\boldsymbol{x}_1,\cdot) w_{i_1}(\boldsymbol{x}_2,\cdot)\right\},$$

$i, i_1 \in \mathbb{N}$, are continuous. By few applications of dominated convergence theorem, it follows that for $m = 1, 2$,

$$\begin{aligned}
\lim_{l\to\infty} E\left\{F(\boldsymbol{x}_{ml},\cdot)(B_{\boldsymbol{y}})\right\} &= E\left\{F(\boldsymbol{x}_m,\cdot)(B_{\boldsymbol{y}})\right\}, \\
\lim_{l\to\infty} E\left\{F(\boldsymbol{x}_{ml},\cdot)(B_{\boldsymbol{y}})^2\right\} &= E\left\{F(\boldsymbol{x}_m,\cdot)(B_{\boldsymbol{y}})^2\right\},
\end{aligned}$$

and

$$\lim_{l\to\infty} E\left\{F(\boldsymbol{x}_{1l},\cdot)(B_{\boldsymbol{y}})F(\boldsymbol{x}_{2l},\cdot)(B_{\boldsymbol{y}})\right\} = E\left\{F(\boldsymbol{x}_{1},\cdot)(B_{\boldsymbol{y}})F(\boldsymbol{x}_{2},\cdot)(B_{\boldsymbol{y}})\right\}.$$

Finally, for $\boldsymbol{y} \in \tilde{\Delta}_m$ and by the definition of correlation, the proof of the theorem is completed. □

D.11. Proof of Theorem 11

The proof of this theorem is an extension of the proof of Theorem 11 in Barrientos et al. (2017). For completeness we state the proof in what follows. Let $m(\boldsymbol{y},\boldsymbol{x}) = q(\boldsymbol{x})g_{\boldsymbol{x}}(\boldsymbol{y})$ be the random joint distribution for the response and predictors arising when $\{g_{\boldsymbol{x}}(\boldsymbol{y}) : \boldsymbol{x} \in \mathscr{X}\}$ is a DMBPP, wDMBPP or θDMBPP. Since the KL divergence between m_0 and the implied joint distribution m can be bounded by the supremum over the predictor space of KL divergences between the predictor-dependent probability measures,

$$\begin{aligned}
\mathrm{KL}(m_0, m) &= \int_{\mathscr{X}}\int_{\Delta_m} m_0(\boldsymbol{y},\boldsymbol{x}) \log\left(\frac{m_0(\boldsymbol{y},\boldsymbol{x})}{m(\boldsymbol{y},\boldsymbol{x})}\right) d\boldsymbol{y}d\boldsymbol{x},\\
&= \int_{\mathscr{X}} q(\boldsymbol{x})\int_{\Delta_m} q_0(\boldsymbol{y}\mid\boldsymbol{x}) \log\left(\frac{q_0(\boldsymbol{y}\mid\boldsymbol{x})}{g_{\boldsymbol{x}}(\boldsymbol{y})}\right) d\boldsymbol{y}d\boldsymbol{x},\\
&\leq \sup_{\boldsymbol{x}\in\mathscr{X}}\int_{\Delta_m} q_0(\boldsymbol{y}\mid\boldsymbol{x}) \log\left(\frac{q_0(\boldsymbol{y}\mid\boldsymbol{x})}{g_{\boldsymbol{x}}(\boldsymbol{y})}\right) d\boldsymbol{y},
\end{aligned}$$

when $\boldsymbol{x}$ contains only continuous predictors, it follows that, for every $\delta > 0$,

$$\Pr\left\{\mathrm{KL}(m_0, m) < \delta\right\} \geq \Pr\left\{\sup_{\boldsymbol{x}\in\mathscr{X}}\int_{\Delta_m} q_0(\boldsymbol{y}\mid\boldsymbol{x}) \log\left(\frac{q_0(\boldsymbol{y}\mid\boldsymbol{x})}{g_{\boldsymbol{x}}(\boldsymbol{y})}\right) d\boldsymbol{y} < \delta\right\} > 0,$$

under the assumptions of Theorem 4. Thus, by Schwartz's theorem (Schwartz, 1965), 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 $m_0(\boldsymbol{y},\boldsymbol{x}) = q(\boldsymbol{x})q_0(\boldsymbol{y}\mid\boldsymbol{x})$, converges to one as the sample size goes to infinity. □

D.12. Proof of Theorem 12

The proof of this theorem is a direct extension and follows the same arguments of the proof of Theorems 12, in Barrientos et al. (2017). Here the sequence of sieves is given by

$$\mathscr{F}_n = \left\{\tilde{m} : \tilde{m}(\boldsymbol{y},\boldsymbol{x}) = q(\boldsymbol{x})\sum_{j=1}^{\infty} w_j(\boldsymbol{x})\tilde{\mathrm{d}}_j(\boldsymbol{y}), \quad \{\tilde{\mathrm{d}}_j\}_{j=1}^{m_n} \in \tilde{\mathscr{B}}_{k_n}^{m_n},\right.$$

$$\left. \eta_j \in \mathcal{B}_{j,n}, j = 1, \ldots, m_n, \quad \sup_{\boldsymbol{x} \in \mathscr{X}} \sum_{j=m_n+1}^{\infty} w_j(\boldsymbol{x}) < \epsilon \right\},$$

where

$$\tilde{\mathcal{B}}_{k_n} = \left\{ \mathrm{dir}(\cdot \mid \mathbf{j}, \overline{k} + m - \|\mathbf{j}\|_1) : \mathbf{j} \in \mathcal{H}^0_{\overline{k},m}, \overline{k} = 1, \ldots, k_n \right\}.$$

Finally, we only need to note that the cardinality of $\tilde{\mathcal{B}}_{k_n}$ is $\sum_{k=1}^{k_n} \frac{(k+m-1)!}{m!(k-1)!} = \frac{k_n(k_n+m)!}{k_n!(m+1)!}$. □

Appendix E: The MCMC sampling scheme

In this section we provide further details on the Markov chain Monte Carlo (MCMC) sampling scheme of the posterior distributions involved in the model. We used the multivariate slice sampler, proposed by Neal (2003), to scan the posterior distribution of the parameters $\boldsymbol{\beta}^{\eta}_j$ and $\boldsymbol{\beta}^{z}_{j,l}$. Convergence of the posterior samples (not shown here) were evaluated by standard test as implemented in the CODA R library (Plummer et al., 2006) and by looking at the trace plots.

As described in the main document, the model considers a truncated version of the stick breaking representation of the predictor dependent mixing measures to a level N. As usual, for every $\boldsymbol{x} \in \mathscr{X}$, we set $v_N(\boldsymbol{x}) = 1$ to ensure the weights to add up to one. In what follows we provide expressions for the joint posterior distribution of $\{\boldsymbol{\beta}^{\eta}_j\}_{j=1}^{N-1}$, $\{\boldsymbol{\beta}^{\boldsymbol{z}}_{jl}\}_{j=1,l=1}^{N,m}$, β^{η}_{0j}, and $\beta^{\boldsymbol{z}}_{0jl}$, and describe the steps of the MCMC updating scheme.

Given data set $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)^t$, the likelihood can be written as $L(\boldsymbol{Y} \mid \ldots)$, with

$$L(\boldsymbol{Y} \mid \ldots) = \prod_{i=1}^{n} \left\{ \sum_{j=1}^{N} w_j(\boldsymbol{x}_i) \mathrm{dir}\left(\boldsymbol{y}_i \quad \middle| \lceil k\boldsymbol{\theta}_j(\boldsymbol{x}_i) \rceil, k + m - \|\lceil k\boldsymbol{\theta}_j(\boldsymbol{x}_i) \rceil\|_1 \right) \right\}.$$

The full conditional distributions of $\boldsymbol{\beta}^{\eta}_j$ and $\boldsymbol{\beta}^{z}_{jl}$ are given by

$$\pi\left(\boldsymbol{\beta}^{\eta}_j \mid \ldots\right) \propto L(\boldsymbol{Y} \mid \ldots) \times \left[N_p(\boldsymbol{\beta}^{\eta}_j \mid \mathbf{0}, \Sigma^{\eta}_1)^{1-\gamma^{\eta}} \times N_p(\boldsymbol{\beta}^{\eta}_j \mid \mathbf{0}, \Sigma^{\eta}_2)^{\gamma^{\eta}} \right],$$

$$\pi\left(\boldsymbol{\beta}^{z}_{j,l} \mid \ldots\right) \propto L(\boldsymbol{Y} \mid \ldots) \times \left[N_p(\boldsymbol{\beta}^{z}_{j,l} \mid \mathbf{0}, \Sigma^{z}_1)^{1-\gamma^{z}} \times N_p(\boldsymbol{\beta}^{z}_{j,l} \mid \mathbf{0}, \Sigma^{z}_2)^{\gamma^{z}} \right],$$

The full conditional distributions of the degree of the polynomial, k, is given by

$$\pi\left(k \mid \ldots\right) \propto L(\boldsymbol{Y} \mid \ldots) \times Poisson(k \mid \lambda) \mathbb{I}_{\{k \geq 1\}}.$$

Parameters γ^{η} and $\gamma^{\boldsymbol{z}}$ can be sampled from its conjugate posterior distribution. Note that $(\gamma^{\eta}, \gamma^{\boldsymbol{z}}) \mid \ldots \sim Discrete(w_1, w_2, w_3, w_4)$, where

$$w_l = \tilde{w}_l / \sum_{j=1}^{4} \tilde{w}_j,$$

$l = 1, \dots, 4$, with

$$\tilde{w}_1 \propto \prod_{j=1}^{N-1} N_p(\boldsymbol{\beta}_j^{\eta} \mid \mathbf{0}, \Sigma_1^{\eta}) \times \prod_{j=1}^{N} \prod_{l=1}^{d} N_p(\boldsymbol{\beta}_{j,l}^{z} \mid \mathbf{0}, \Sigma_1^{z}) \times \pi_1,$$
$$\tilde{w}_2 \propto \prod_{j=1}^{N-1} N_p(\boldsymbol{\beta}_j^{\eta} \mid \mathbf{0}, \Sigma_2^{\eta}) \times \prod_{j=1}^{N} \prod_{l=1}^{d} N_p(\boldsymbol{\beta}_{j,l}^{z} \mid \mathbf{0}, \Sigma_1^{z}) \times \pi_2,$$
$$\tilde{w}_3 \propto \prod_{j=1}^{N-1} N_p(\boldsymbol{\beta}_j^{\eta} \mid \mathbf{0}, \Sigma_1^{\eta}) \times \prod_{j=1}^{N} \prod_{l=1}^{d} N_p(\boldsymbol{\beta}_{j,l}^{z} \mid \mathbf{0}, \Sigma_2^{z}) \times \pi_3,$$
$$\tilde{w}_4 \propto \prod_{j=1}^{N-1} N_p(\boldsymbol{\beta}_j^{\eta} \mid \mathbf{0}, \Sigma_2^{\eta}) \times \prod_{j=1}^{N} \prod_{l=1}^{d} N_p(\boldsymbol{\beta}_{j,l}^{z} \mid \mathbf{0}, \Sigma_2^{z}) \times \pi_4,$$

where π_1, π_2, π_3, and π_4 denote the a priori probability of the binary pairs $(1,1)$, $(0,1)$, $(1,0)$, and $(0,0)$.

Appendix F: Model specification for the simulation study

As mentioned in the main document, the selection of the parameters τ_l^{η} and $\tau_l^{\boldsymbol{z}}$, for $l = 1, 2$, play a key role when selecting the version of the model that best fits the data. Recall that

$$\boldsymbol{\beta}_j^{\eta} \mid \gamma^{\eta} \overset{iid}{\sim} N_p(\mathbf{0}, \tau_1^{\eta}(\mathbb{X}^t\mathbb{X})^{-1})^{1-\gamma^{\eta}} \times N_p(\mathbf{0}, \tau_2^{\eta}(\mathbb{X}^t\mathbb{X})^{-1})^{\gamma^{\eta}},$$
$$\boldsymbol{\beta}_{jl}^{\boldsymbol{z}} \mid \gamma^{\boldsymbol{z}} \overset{iid}{\sim} N_p(\mathbf{0}, \tau_1^{\boldsymbol{z}}(\mathbb{X}^t\mathbb{X})^{-1})^{1-\gamma^{\boldsymbol{z}}} \times \gamma^{\boldsymbol{z}} N_p(\mathbf{0}, \tau_2^{\boldsymbol{z}}(\mathbb{X}^t\mathbb{X})^{-1})^{\gamma^{\boldsymbol{z}}},$$

for $j \geq 1$ and $l = 1, \dots, m$. The prior distributions on $\boldsymbol{\beta}_j^{\eta}$ and $\boldsymbol{\beta}_{jl}^{\boldsymbol{z}}$ induce prior distributions for the bounded stochastic processes defining the predictor dependent weights, $e^{\boldsymbol{\beta}_{0j}^{\eta} + \boldsymbol{x}^t \boldsymbol{\beta}_j^{\eta}} / (1 + e^{\boldsymbol{\beta}_{0j}^{\eta} + \boldsymbol{x}^t \boldsymbol{\beta}_j^{\eta}})$, and the predictor dependent atoms, $\left(e^{\boldsymbol{\beta}_{0j1}^{\boldsymbol{z}} + \boldsymbol{x}^t \boldsymbol{\beta}_{j1}^{\boldsymbol{z}}}, \dots, e^{\boldsymbol{\beta}_{0jm}^{\boldsymbol{z}} + \boldsymbol{x}^t \boldsymbol{\beta}_{jm}^{\boldsymbol{z}}}\right) / \left(1 + \sum_{l=1}^{m} e^{\boldsymbol{\beta}_{0jl}^{\boldsymbol{z}} + \boldsymbol{x}^t \boldsymbol{\beta}_{jl}^{\boldsymbol{z}}}\right)$. We aim to choose the parameters τ_1^{η} and $\tau_1^{\boldsymbol{z}}$ such that with high probability $\boldsymbol{x}^t \boldsymbol{\beta}_j^{\eta}$ and $\boldsymbol{x}^t \beta_{jl}^{\boldsymbol{z}}$ are close to zero and τ_2^{η} and $\tau_2^{\boldsymbol{z}}$ such that with high probability $\boldsymbol{x}^t \boldsymbol{\beta}_j^{\eta}$ and $\boldsymbol{x}^t \beta_{jl}^{\boldsymbol{z}}$ are away from zero.

Let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_G$ be a grid of values for the $[0,1]$–valued predictor that was considered in the simulation study. For each $\boldsymbol{x}_g$, $g = 1, \dots, G$, we define a sequence of values $\tilde{\tau}_1, \dots, \tilde{\tau}_L$ such that with high probability $\boldsymbol{x}_g^t \boldsymbol{\beta}_j^{\eta}$ lies between -4 and 4. We choose τ_{1g}^{η} as the largest $\tilde{\tau}_l$ such that with high probability $\boldsymbol{x}_g^t \boldsymbol{\beta}_j^{\eta}$ is between -0.2 and 0.2 and we choose τ_{2g}^{η} as the smallest $\tilde{\tau}_l > \tau_{1g}^{\eta}$ such that with high probability $\boldsymbol{x}_g^t \boldsymbol{\beta}_j^{\eta}$ is between -2.2 and 2.2. Finally, we set $\tau_1^{\eta} = \min\{\tau_{11}^{\eta}, \dots, \tau_{1G}^{\eta}\}$ and $\tau_2^{\eta} = \min\{\tau_{21}^{\eta}, \dots, \tau_{2G}^{\eta}\}$. We follow a very similar procedure to choose $\tau_1^{\boldsymbol{z}}$ and $\tau_2^{\boldsymbol{z}}$.

Appendix G: Additional results for the simulation study

In this section we provide additional results for the simulation study. More specifically, we provide the results measuring the performance of the model based on the estimate to the L_∞ distance and the contour plots of the density estimates obtained under Prior II for the binary latent variables.

Table 4 shows the mean of the L_∞ estimates across replicates for each Scenario, sample size and prior distribution for the binary latent variables. As expected, the integrated L_∞ distances between the true density and estimates decrease as the sample size increases. Regarding this criteria, the model shows the best density estimation performance for Scenario III, the single–atom true model, while the worst performance is observed for Scenario I, the fully predictor dependent true model, for every sample size and for both prior distributions for the binary latent variables.

TABLE 4
Mean (across Monte Carlo replicates) of the L_∞ distances between the truth and random measure estimates for each scenario, prior for the binary latent variables, and sample size.

	Prior I			Prior II		
Scenario	$n = 250$	$n = 500$	$n = 1000$	$n = 250$	$n = 500$	$n = 1000$
I	32.638	28.078	26.892	32.709	28.153	27.047
II	19.863	21.504	24.936	19.481	22.041	23.610
III	11.912	9.876	9.203	11.900	9.919	9.148
IV	26.837	17.814	11.267	26.825	17.838	11.089

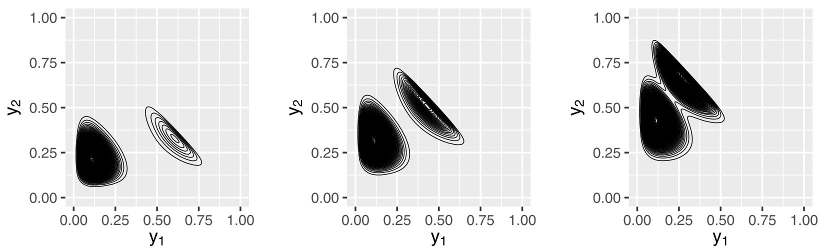
Figures 6 to 9 display the contour plot of the conditional density estimates mean (across replicates) for each sample size, selected values of the predictor, and Prior II for $(\gamma^\eta, \gamma^{\boldsymbol{z}})$, for Scenarios I to IV, respectively.

Appendix H: Model specification for the application to solid waste data

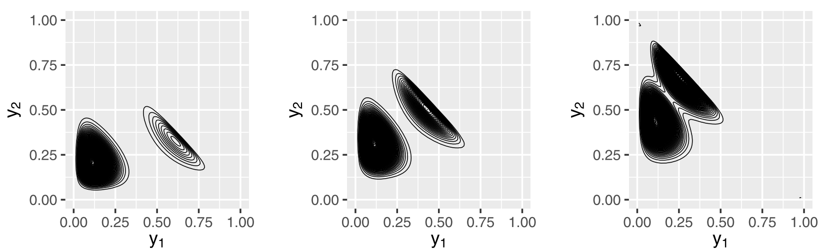
The selection of the parameters τ_l^η and $\tau_l^{\boldsymbol{z}}$, for $l = 1, 2$ follow the same reasoning as for the simulation study. Here, the $[0, 1]$–valued grid for the predictor is replaced by the six possible values that the predictor, in its dummy representation, can take.

Appendix I: Additional results for application to solid waste data

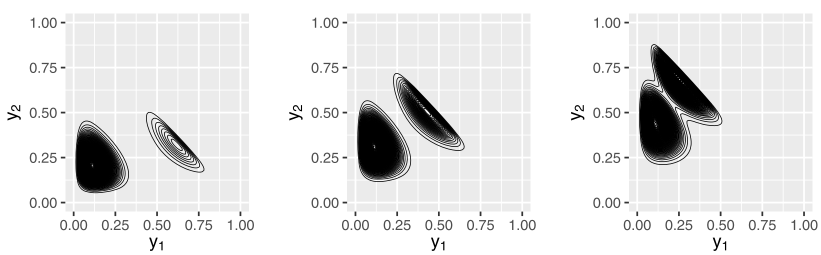
Figure 10 displays the conditional density estimates, as the posterior predictive mean, for the DMBPP model for each value of the categorical predictor “low–low”, “low”, “medium–low”, “medium”, “medium–high”, and “high”, under Prior II for the binary latent variables. The LPML and $-nWAIC$ values for the DMBPP model were 778.043 and 778.413, respectively, under Prior II for $(\gamma^\eta, \gamma^{\boldsymbol{z}})$.



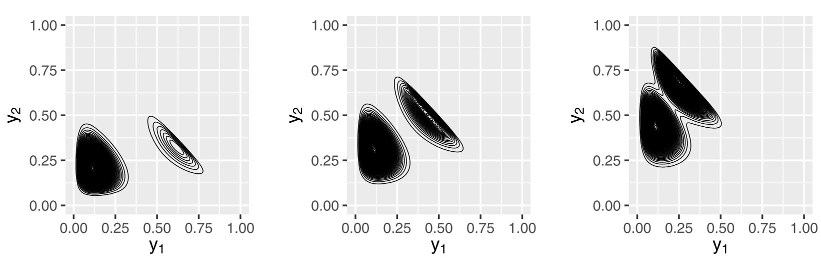
(a) True, $x = 0.25$ (b) True, $x = 0.50$ (c) True, $x = 0.75$



(d) $n = 250$, $x = 0.25$ (e) $n = 250$, $x = 0.50$ (f) $n = 250$, $x = 0.75$

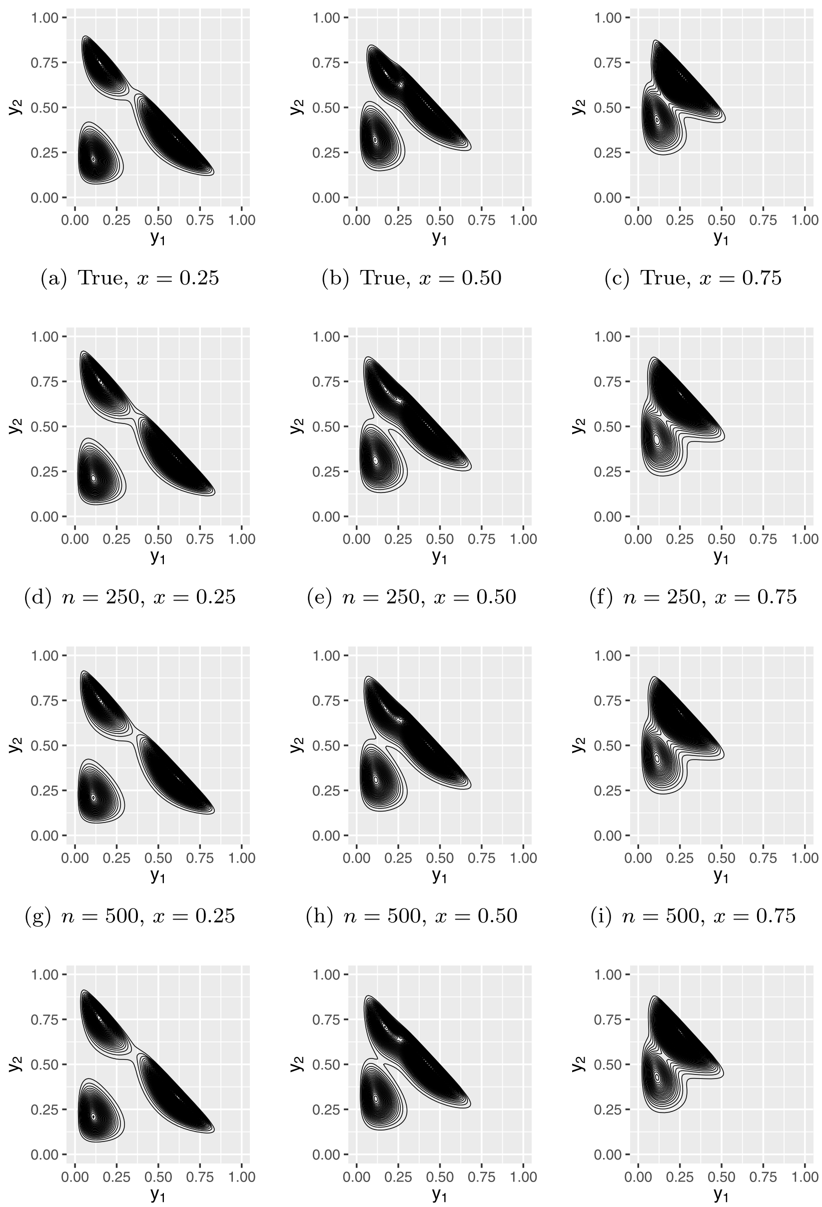


(g) $n = 500$, $x = 0.25$ (h) $n = 500$, $x = 0.50$ (i) $n = 500$, $x = 0.75$



(j) $n = 1000$, $x = 0.25$ (k) $n = 100$, $x = 0.50$ (l) $n = 1000$, $x = 0.75$

FIG 6. *Simulation study - Scenario I - Prior 2: contour plots of the true density (first row) and mean across replicates of the posterior mean of the conditional density for* $n = 250$ *(second row),* $n = 500$ *(third row), and* $n = 1000$ *(fourth row). The results are shown under Prior I for* $(\gamma^{\eta}, \gamma^{\mathbf{z}})$*. Results are displayed for* $x = 0.25$ *(first column),* $x = 0.50$ *(second column), and* $x = 0.75$ *(third column).*



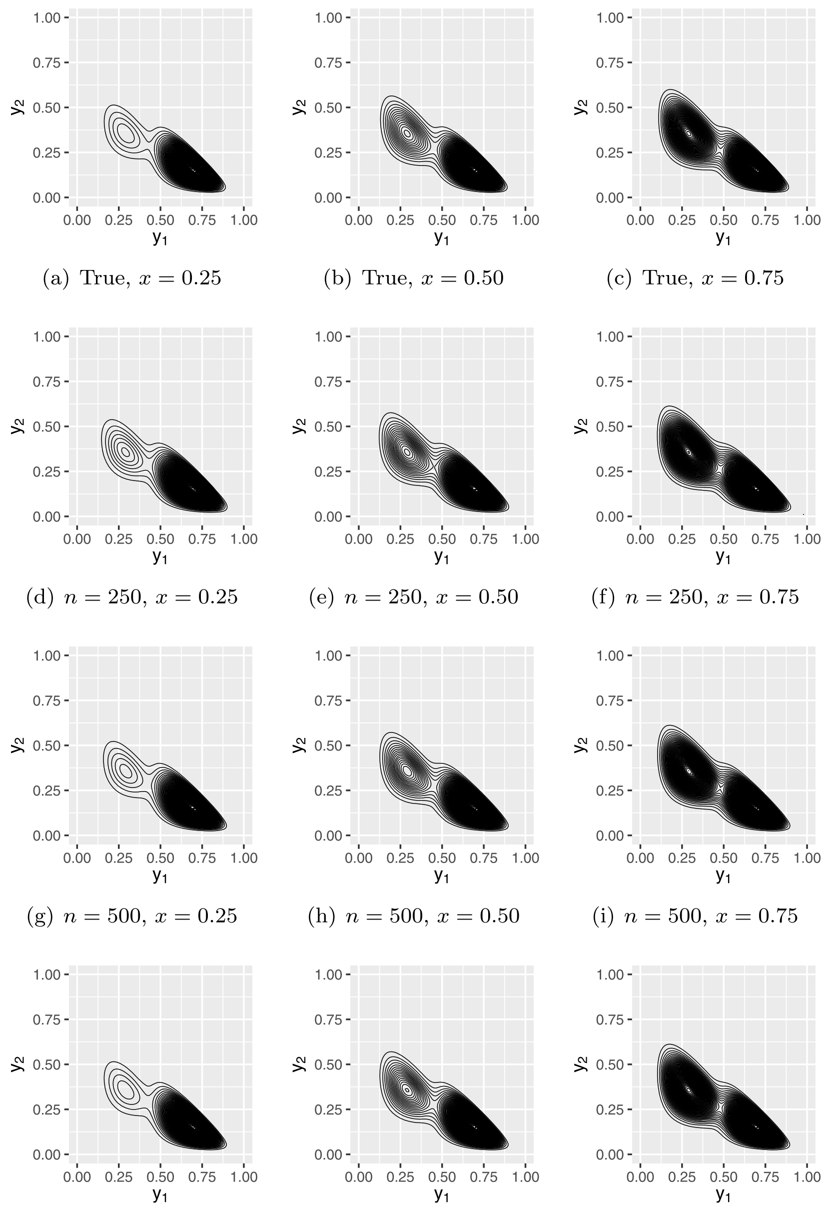
(a) True, $x = 0.25$ (b) True, $x = 0.50$ (c) True, $x = 0.75$

(d) $n = 250$, $x = 0.25$ (e) $n = 250$, $x = 0.50$ (f) $n = 250$, $x = 0.75$

(g) $n = 500$, $x = 0.25$ (h) $n = 500$, $x = 0.50$ (i) $n = 500$, $x = 0.75$

(j) $n = 1000$, $x = 0.25$ (k) $n = 100$, $x = 0.50$ (l) $n = 1000$, $x = 0.75$

FIG 7. *Simulation study - Scenario II - Prior II: Contour plots of the true density (first row) and mean across replicates of the posterior mean of the conditional density for* n = 250 *(second row),* n = 500 *(third row), and* n = 1000 *(fourth row). The results are shown under Prior I for* $(\gamma^{\eta}, \gamma^{\mathbf{z}})$. *Results are displayed for* $x = 0.25$ *(first column),* $x = 0.50$ *(second column), and* $x = 0.75$ *(third column).*



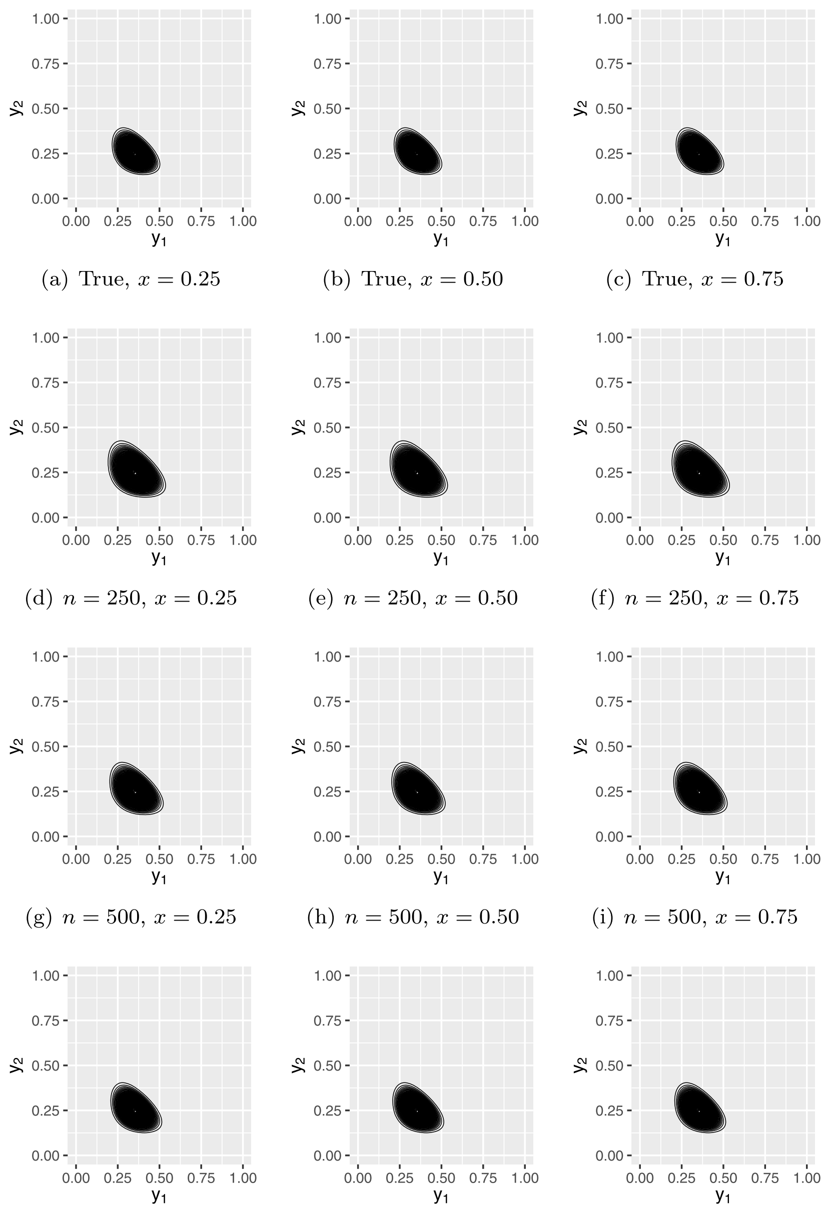
(a) True, $x = 0.25$ (b) True, $x = 0.50$ (c) True, $x = 0.75$

(d) $n = 250$, $x = 0.25$ (e) $n = 250$, $x = 0.50$ (f) $n = 250$, $x = 0.75$

(g) $n = 500$, $x = 0.25$ (h) $n = 500$, $x = 0.50$ (i) $n = 500$, $x = 0.75$

(j) $n = 1000$, $x = 0.25$ (k) $n = 100$, $x = 0.50$ (l) $n = 1000$, $x = 0.75$

FIG 8. *Simulation study - Scenario III - Prior II: Contour plots of the true density (first row) and mean across replicates of the posterior mean of the conditional density for* $n = 250$ *(second row),* $n = 500$ *(third row), and* $n = 1000$ *(fourth row). The results are shown under Prior I for* $(\gamma^{\eta}, \gamma^{\mathbf{z}})$*. Results are displayed for* $x = 0.25$ *(first column),* $x = 0.50$ *(second column), and* $x = 0.75$ *(third column).*



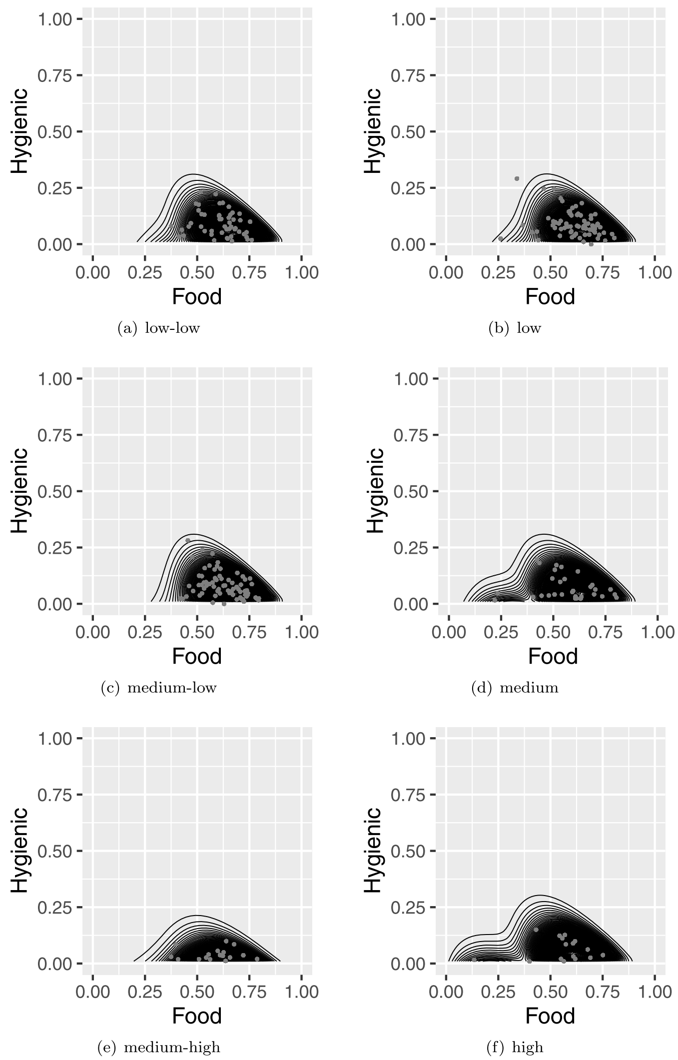
(a) True, $x = 0.25$ (b) True, $x = 0.50$ (c) True, $x = 0.75$

(d) $n = 250$, $x = 0.25$ (e) $n = 250$, $x = 0.50$ (f) $n = 250$, $x = 0.75$

(g) $n = 500$, $x = 0.25$ (h) $n = 500$, $x = 0.50$ (i) $n = 500$, $x = 0.75$

(j) $n = 1000$, $x = 0.25$ (k) $n = 100$, $x = 0.50$ (l) $n = 1000$, $x = 0.75$

FIG 9. *Simulation study - Scenario IV - Prior II: Contour plots of the true density (first row) and mean across replicates of the posterior mean of the conditional density for* $n = 250$ *(second row),* $n = 500$ *(third row), and* $n = 1000$ *(fourth row). The results are shown under Prior I for* $(\gamma^{\eta}, \gamma^{\mathbf{z}})$*. Results are displayed for* $x = 0.25$ *(first column),* $x = 0.50$ *(second column), and* $x = 0.75$ *(third column).*



(a) low-low (b) low

(c) medium-low (d) medium

(e) medium-high (f) high

Fig 10. *Application to Solid Waste Data - Prior II: Contour plot of conditional density estimates and data points for colorredDMBPP, under Prior II for* $(\gamma^{\eta}, \gamma^{z})$*, for each value of the discrete socioeconomic level predictor, low-low (panel (a)), low (panel (b)), medium-low (panel (c)), medium (panel (d)), medium-high (panel (e)), and high (panel (f)) . The x–axis and y–axis denote the proportion of food and hygienic waste, respectively.*

References

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society, Series B* 44 139–160. MR0676206

Atchison, J. & Shen, S. M. (1980). Logistic-normal distributions: Some properties and uses. *Biometrika* 67 261–272. MR0581723

Babu, G. J. & Chaubey, Y. P. (2006). Smooth estimation of a distribution and density function on a hypercube using Bernstein polynomials for dependent random vectors. *Statistics & Probability Letters* 76 959–969. MR2270097

Banerjee, S., Carlin, B. P. & Gelfand, A. E. (2003). *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC. MR3362184

Barndorff-Nielsen, O. (1973). On M-ancillarity. *Biometrika* 60 447–455. MR0345255

Barndorff-Nielsen, O. (1978). *Information and exponential families in statistical theory*. John Wiley & Sons, Ltd., Chichester. Wiley Series in Probability and Mathematical Statistics. MR0489333

Barrientos, A. F., Jara, A. & Quintana, F. A. (2012). On the support of MacEachern's dependent Drichlet processes and extensions. *Bayesian Analysis* 7 277– 310. MR2934952

Barrientos, A. F., Jara, A. & Quintana, F. A. (2015). Bayesian density estimation for compositional data using random Bernstein polynomials. *Journal of Statistical Planning and Inference* 166 116–125. MR3390138

Barrientos, A. F., Jara, A. & Quintana, F. A. (2017). Fully nonparametric regression for bounded data using dependent bernstein polynomials. *Journal of the American Statistical Association* 112 806–825. MR3671772

Di Marzio, M., Panzera, A. & Venieri, C. (2015). Non-parametric regression for compositional data. *Statistical Modelling* 15 113–133. MR3325749

Epifani, I. & Lijoi, A. (2010). Nonparametric priors for vectors of survival functions. *Statistica Sinica* 20 1455–1484. MR2777332

Florens, J.-P., Mouchart, M. & Rolin, J.-M. (1990). *Elements of Bayesian statistics*, vol. 134 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker, Inc., New York. MR1051656

Geisser, S. & Eddy, W. (1979). A predictive approach to model selection. *Journal of the American Statistical Association* 74 153–160. MR0529531

Gelfand, A. E. & Dey, D. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society, Series B* 56 501–514. MR1278223

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2013). *Bayesian data analysis*. CRC press. MR3235677

George, E. I. & McCulloch, R. E. (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association* 88 881–889.

Gueorguieva, R., Rosenheck, R. & Zelterman, D. (2008). Dirichlet component regression and its applications to psychiatric data. *Computational Statistics & Data Analysis* 52 5344–5355. MR2526600

Hijazi, R. H. (2003). *Analysis of compositional data using Dirichlet covariate models*. Ph.D. thesis, American University. MR2704518

Hijazi, R. H. & Jernigan, R. W. (2009). Modelling compositional data using Dirichlet regression models. *Journal of Applied Probability & Statistics* 4 77–91. MR2668780

Ishwaran, H. & James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* 96 161–173. MR1952729

Jara, A. & Hanson, T. (2011). A class of mixtures of dependent tail-free processes. *Biometrika* 98 553–566. MR2836406

Karabatsos, G. & Walker, S. G. (2012). Adaptive-modal Bayesian nonparametric regression. *Electronic Journal of Statistics* 6 2038–2068. MR3020256

Klinger, R., Olaya, J., Marmolejo, L. & Madera, C. (2009). A sampling plan for residentially generated solid waste quantification at urban zones of middle sized cities. *Revista Facultad de Ingeniería Universidad de Antioquia* 48 76–86.

Leisen, F. & Lijoi, A. (2011). Vectors of two–parameter Poisson–Dirichlet processes. *Journal of Multivariate Analysis* 102 482–495. MR2755010

Lijoi, A., Nipoti, B. & Prünster, I. (2014). Bayesian inference with dependent normalized completely random measures. *Bernoulli* 20 1260–1291. MR3217444

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

MacEachern, S. N. (2000). Dependent Dirichlet processes. Tech. rep., Department of Statistics, The Ohio State University.

Müller, P., Quintana, F. A., Jara, A. & E, H. T. (2015). *Bayesian Nonparametric Data Analysis*. New York, USA: Springer. MR3309338

Neal, R. (2003). Slice sampling. *The Annals of Statistics* 31 705–767. MR1994729

Ouimet, F. (2021). Asymptotic properties of bernstein estimators on the simplex. *Journal of Multivariate Analysis* 185 104784. MR4287788

Parthasarathy, K. R. (1967). *Probability Measures in Metric Spaces*. Providence, USA: AMS Chelsea Publishing. MR2169627

Pati, D., Dunson, D. B. & Tokdar, S. T. (2013). Posterior consistency in conditional distribution estimation. *Journal of Multivariate Analysis* 116 456–472. MR3049916

Petrone, S. (1999a). Bayesian density estimation using Bernstein polynomials. *The Canadian Journal of Statistics* 27 105–126. MR1703623

Petrone, S. (1999b). Random Bernstein polynomials. *Scandinavian Journal of Statistics* 26 373–393. MR1712051

Petrone, S. & Wasserman, L. (2002). Consistency of Bernstein polynomial posterior. *Journal of the Royal Statistical Society, Series B* 64 79–100. MR1881846

Plummer, M., Best, N., Cowles, K. & Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R news* 6 7–11.

Quintana, F., Müller, P., Jara, A. & MacEachern, S. (2022). The dependent Dirichlet process and related models. *Statistical Science* 37 24–41.

MR4371095

Resnick, S. (2019). *A probability path*. Springer. MR3135152

Schwartz, L. (1965). On Bayes procedures. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 4 10–26. MR0184378

Shimizu, T. K., Louzada, F., Suzuki, A. K. & Ehlers, R. S. (2021). Modeling compositional regression with uncorrelated and correlated errors: A Bayesian approach. *Journal of Data Science* 16 221–250.

Smithson, M. & Verkuilen, J. (2006). A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables. *Psychological methods* 11 54.

Tenbusch, A. (1994). Two-dimensional Bernstein polynomial density estimators. *Metrika* 41 233—-253. MR1293514

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics* 22 1701–1762. MR1329166

Trippa, L., Müller, P. & Johnson, W. (2011). The multivariate beta process and an extension of the Polya tree model. *Biometrika* 98 17–34. MR2804207

Tsagris, M., Alenazi, A. & Stewart, C. (2020). Non-parametric regression models for compositional data. *arXiv preprint arXiv:2002.05137*. MR3789428

Van der Merwe, S. (2019). A method for Bayesian regression modelling of composition data. *South African Statistical Journal* 53 55–64. MR3966355

Wang, H., Meng, J. & Tenenhaus, M. (2010). Regression modelling analysis on compositional data. In *Handbook of Partial Least Squares*. Springer, 381–406. MR2762258

Watanabe, S. & Opper, M. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* 11. MR2756194

Zellner, A. (1983). Applications of Bayesian analysis in econometrics. *Journal of the Royal Statistical Society, Series D* 32 23–34. MR0926622

Zheng, Y., Zhu, J. & Roy, A. (2010). Nonparametric bayesian inference for the spectral density function of a random field. *Biometrika* 97 238–245. MR2594432