Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Theory and Applications of Dependent Nonparametric Bayesian Models for Bounded and Unbounded Responses

ANDRÉS FELIPE BARRIENTOS

Advisor:
PROF. FERNANDO A. QUINTANA
Co–Advisor:
PROF. ALEJANDRO JARA

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor in Statistics.

Members of the Examination Commitee:
PROF. GUIDO DEL PINO – PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
PROF. ROLANDO DE LA CRUZ – PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
PROF. JORGE GONZÁLEZ – PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
PROF. TIMOTHY HANSON – UNIVERSITY OF SOUTH CAROLINA

August 2012

Acknowledgements

First and foremost, I would like to express my profound gratitude to Dr. Fernado Quintana and Dr. Alejandro Jara not only for their support, guidance and encouragement througthout my dissertation work, but also for their friendship and professionalism. They have been my mentors through all these years, helping me to grow as a researcher and motivating me towards new research problems.

I would like to gratefully acknowledge the financial support during my dissertation work from the *"Programa Becas Latinoamericanas de CONICY"*, which was awarded to me for the period 2008–2012, and was crucial to the successful completion of my doctorate.

Although the majority of my friends and my colleagues at PUC are not explicitly listed here, I would like to acknowledge them for their support and help during the few past years. I would like specially to thank Alejandro Muñoz for his support and pacience at the begining of this path. His presence enriched me in numerous ways and taught me valuable lessons in life. I would also like to thank María José, Juan Francisco and Jaime for all the support and friendship they have given me. Additionally, I wish to thank Soledad for her administrative support.

Outside Chile I knew I could always count on my family and on great friends. I could never thank my mother enough, for her unconditional support and for always believing in me. I thank my friend, Hector Toro, who has been as a brother for me in every sense of the word, despite being geographically so far. I also owe special thanks to all those people in Colombia who encouraged me to come this university and to make my dreams come true.

To my wonderful Martha: I will never find the right words to express my deepest gratitude to you. Thank you for your *love, sweet presence, encouragement, wisdom, support, help, faith in me, endless patience* and, above everything, for letting me walk at your side in the life way.

Contents

Abstract

As the complexity of many scientific problems grows, the modelling and analysis of data coming from these problems requires of increasingly sophisticated statistical models. The constant search of such models has been one of the major stimulus for research in Bayesian nonparametric (BNP) methods. This dissertation presents advances in BNP models for predictor–dependent distributions (or density regression) by studying one of their most important properties (support) and proposing a novel class of these models. In order to contextualize the dissertation, an initial chapter is included presenting a literature review and some basic concepts which are useful to understand the main reasons that motivated this work. Those reasons are also included in this chapter. Because this project is based on two different works, the dissertation has been divided in two pieces that are self–contained and included in two different chapters, 2 and 3.

In the fist part, Chapter 2, we study the support properties of Dirichlet process–based models for sets of predictor–dependent probability distributions. Exploiting the connection between copulas and stochastic processes, we provide an alternative definition of MacEachern's dependent Dirichlet processes. Based on this definition, we provide sufficient conditions for the full weak support of different versions of the process. In particular, we show that under mild conditions on the copula functions, the version where only the support points or the weights are dependent on predictors have full weak support. In addition, we also characterize the Hellinger

and Kullback–Leibler support of mixtures induced by the different versions of the dependent Dirichlet process. A generalization of the results for the general class of dependent stick–breaking processes is also provided.

In the second part, Chapter 3, we propose a novel probability model for sets of predictor–dependent probability distributions with bounded domain. The proposal corresponds to an extension of the Dirichlet–Bernstein prior by using dependent stick–breaking processes. Appealing theoretical properties such as full support, continuity, marginal distribution, correlation structure, and consistency of the posterior distribution are studied. Practicable special cases of the general model are discussed and illustrated using simulated and real–life data. The simulated data is also used to compare the proposed methodology to existing methods.

Finally, Chapter 4 summarizes the dissertation and discusses possible generalizations and future work.

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and literature review

1.1.1 The general context

The definition and study of theoretical properties of probability models defined on infinite - dimensional spaces have received increasing attention in the statistical literature because these models are the basis for the Bayesian nonparametric (BNP) generalization of finite-dimensional statistical models (see, e.g., Ghosh & Ramamoorthi, 2003; Müller & Quintana, 2004; Hjort et al., 2010). These generalizations allow the user to gain model flexibility and robustness against mis-specification of a parametric statistical model. BNP models are specified by defining a stochastic process whose trajectories belong to a functional space, $\mathcal{G}$, such as the space of all probability measures defined on a given measurable space. The law governing such a process is then used as a prior distribution for a functional parameter in the Bayesian framework.

The increase in applications of BNP methods in the statistical literature has been motivated largely by the availability of simple and efficient methods for posterior computation in Dirichlet process mixture (DPM) models (Ferguson, 1983; Lo, 1984). The DPM models incorporate Dirichlet process (DP) priors (Ferguson, 1973, 1974) for components in Bayesian hierarchical models, resulting in an extremely flexible class of models. Due to the flexibility and ease in implementation, DPM models are now routinely implemented in a wide variety of applications, ranging from machine learning to genomics (see, e.g. Hjort et al., 2010). Furthermore, a rich theoretical literature about support, posterior consistency and rates of convergence (Lo, 1984; Ghosal et al., 1999; Lijoi et al., 2005; Ghosal & Van der Vaart, 2007) justify the use of DPM models for inference in single density estimation problems.

Let $\mathcal{G}$ be the space of all probability measures, with density w.r.t. Lebesgue measure, defined on an appropriate measurable space $(S, \mathcal{B}(S))$, with $S \subseteq \mathbb{R}^q$, and where $\mathcal{B}(S)$ is the Borel σ–field. A DPM model for density estimation is a $\mathcal{G}$–valued stochastic process, G, defined on an appropriated probability space $(\Omega, \mathcal{A}, P)$, such that for almost every $\omega \in \Omega$, the density function of G is given by

$$g(y \mid F(\omega)) = \int_{\Theta} \psi(y, \theta) F(\omega)(d\theta), \;\; y \in S, \tag{1.1}$$

where $\psi(\cdot, \theta)$ is a continuous density function on $(S, \mathcal{S})$, for every $\theta \in \Theta$, $\Theta \subseteq \mathbb{R}^q$, and F is a DP, whose sample paths are probability measures defined on $(\Theta, \mathcal{B}(\Theta))$, with $\mathcal{B}(\Theta)$ being the Borel σ–field. If F is DP with parameters (M, F_0), where $M \in \mathbb{R}_0^+$ and F_0 is a probability measure on $(\Theta, \mathcal{B}(\Theta))$, written as $F \mid M, F_0 \sim DP(MF_0)$, then the trajectories of the process can be a.s. represented by the following stick-breaking representation (Sethuraman, 1994):

$$F(B) = \sum_{i=1}^{\infty} w_i \delta_{\theta_i}(B), \;\; B \in \mathcal{B}(\Theta), \tag{1.2}$$

where $\delta_\theta(\cdot)$ is the Dirac measure at θ, $w_i = V_i \prod_{j<i}(1 - V_j)$, with $V_i \mid M \overset{iid}{\sim} Beta(1, M)$, and $\theta_i \mid F_0 \overset{iid}{\sim} F_0$. Discussion of properties and applications of DP can be found, for instance, in Ferguson (1973, 1974), Korwar & Hollander (1973), Antoniak (1974), Blackwell & MacQueen (1973), Cifarelli & Regazzini (1990), Hanson et al. (2005), Hjort & Ongaro (2005), Hjort et al. (2010) and in references therein. Recent work in BNP models has concentrated on different generalizations of the problem, which are described in the next sections.

1.1.2 Alternatives to Dirichlet process mixing

Alternative discrete probability models to the DP have been considered. Some examples are members of the general class of species sampling models (SSM) introduced by Pitman (1996). The class of SSM includes as special cases the DP and the normalized random measures (Nieto-Barajas et al., 2004), among many others. Members of this class can be represented in the form $G(B) = \sum_{i=1}^{\infty} w_i \delta_{\theta_i}(B) + (1 - \sum_{i=1}^{\infty} w_i \delta_{\theta_i}(B))\, G_0(B)$, $B \in \mathcal{B}(\Theta)$, where, the atoms θ_i are iid random variables with common distribution G_0, $\theta_i \stackrel{iid}{\sim} G_0$, which are assumed independent of the non-negative random weights w_i. The weights w_i are constrained such that $\sum_{i=1}^{\infty} w_i \leq 1$ a.s. The name of the class is motivated by the interpretation of the parameters; the ith weight w_i is interpreted as the relative frequency of the ith species in a species' list present in a certain population, and θ_i is interpreted as the tag assigned to that species. If $\sum_{i=1}^{\infty} w_i = 1$ the SSM is called proper and the corresponding prior random probability measure G is a.s. discrete. Some examples of SSM are the Dirichlet-multinomial processes (Muliere & Secchi, 1995), the ϵ-DP (Muliere & Tardella, 1998), the normalized inverse Gaussian processes (Lijoi et al., 2005), and the beta two-parameter processes (Ishwaran & Zarepour, 2000).

Perhaps one of the best known examples of the SSM is the stick-breaking process (Ishwaran & James, 2001). A discrete random probability measure of the form (1.2) is called a stick-breaking process if $\omega_i = V_i \prod_{j<i}(1 - V_j)$, $V_i \mid M \stackrel{ind}{\sim} Beta(a_i, b_i)$, and $\theta_i \mid F_0 \stackrel{iid}{\sim} F_0$, where $\{a_i\}_{i\geq 1}$ and $\{b_i\}_{i\geq 1}$ are sequences of positive numbers. These random weights, ω_i, define a proper SSM if and only if $\sum_{i\geq 1} E[\log(1 - V_i)] = -\infty$. In particular, there are two specific stick-breaking process that have been well studied: the Dirichlet process (Ferguson, 1973, 1974) where $a_i = 1$ and $b_i = M$, and the two parameter Poisson-Dirichlet processes (Pitman, 1995, 1996; Pitman & Yor, 1997; Ishwaran & James, 2001) where $a_i = 1 - a$ and $b_i = M + ai$ (with $0 \leq a \leq 1,\ M > -a$).

1.1.3 Continuous and absolutely continuous random probability measures

Alternatives formulations of the problem have been considered by using BNP models which admit directly continuous and absolutely continuous distributions, thus avoiding the convolution with a continuous kernel to generate probability measures with density w.r.t. Lebesgue measure. Some examples are the general class of tail-free processes (Freedman, 1963; Fabius, 1964; Ferguson, 1974), Polya trees (Ferguson, 1974; Mauldin et al., 1992; Lavine, 1992, 1994), mixtures of Polya trees (Lavine, 1992; Hanson & Johnson, 2002; Hanson, 2006; Christensen et al., 2008; Jara et al., 2009), randomized Polya trees (Paddock, 1999, 2002; Paddock et al., 2003), Gaussian processes (O'Hagan, 1992; Angers & Delampady, 1992), Wavelets (Müller & Vidakovic, 1998), logistic Gaussian processes (Tokdar & Ghosh, 2007, see, e.g.), and quantile pyramids (Hjort & Walker, 2009).

1.1.4 Models for related probability measures

Generalizations of (1.1) and (1.2) have been proposed to accommodate dependence of the data on predictors. To date, most of the extensions have focused on constructions that generalize the DPM model by considering

$$g(y \mid x, F_x(\omega)) = \int_{\Theta} \psi(y,\theta) F_x(\omega)(d\theta), \;\; y \in S, \tag{1.3}$$

where $g(y \mid x, F_x)$ is a conditional density indexed by the value of a continuous predictor $x \in \mathcal{X} \subset \mathbb{R}$, and the dependence is introduced through the mixing probability measure F_x. In this case, the parametric space, $\mathcal{G}$, corresponds to the product space given by $\prod_{x\in\mathcal{X}} \mathcal{P}(S)$, where $\mathcal{P}(S)$ is the space of all probability measures defined on $(S, \mathcal{B}(S))$. Notice that the inferential problem is related to the modeling of the collection of predictor-dependent probability measures $\{F_x : x \in \mathcal{X}\}$.

Some of the earliest developments on dependent DP models appeared in Cifarelli & Regazzini (1978), who defined dependence across related random measures by introducing a regression for the baseline measure of marginally DP distributed random measures. A more flexible construction was proposed by MacEachern (1999, 2000), called the dependent Dirichlet process (DDP). The key idea behind the DDP is to create a set of marginally DP-distributed random measures and to introduce dependence by modifying the stick-breaking representation of each element in the set. Specifically, MacEachern (1999, 2000) generalized expression (1.2) by assuming

$$F_x(B) = \sum_{i=1}^{\infty} w_i(x)\delta_{\theta_i(x)}(B), \quad B \in \mathbb{B}, \tag{1.4}$$

where the point masses $\theta_i(x)$, $i = 1, \ldots$, are independent stochastic processes with index set $\mathcal{X}$, and the weights take the form $w_i(x) = V_i(x)\prod_{j<i}[1 - V_j(x)]$, with $V_i(x)$, $l = 1, \ldots$, being independent stochastic processes with index set $\mathcal{X}$ and $Beta(1, M)$ marginal distribution. MacEachern (2000) also studied a version of the process with predictor-independent weights, $F_x(B) = \sum_{i=1}^{\infty} w_i\delta_{\theta_i(x)}(B)$. Versions of the predictor-independent weights DDP have been successfully applied in a variety of applications (see, e.g. De Iorio et al., 2004; Gelfand et al., 2005; Jara et al., 2010). Other extensions of the DP for dealing with related probability distributions include the DPM mixture of normals model for the joint distribution of the response and predictors (Müller et al., 1996), the hierarchical mixture of DPM (Müller et al., 2004), the hierarchical DP (Teh et al., 2006), the order-based DDP model (Griffin & Steel, 2006), the nested DP (Rodriguez et al., 2008), the kernel-stick breaking (Dunson & Park, 2008), among many others. Based on a different formulation of the conditional density estimation problem, Tokdar et al. (2010) and Jara & Hanson (2011) proposed alternatives to convolutions of dependent stick-breaking approaches of the form (1.3) and (1.4), which yield conditional probability measures with density w.r.t. Lebesgue measure without the need of convolutions.

The development of any BNP model always has to keep in mind that there are some properties which are expected to be satisfied by the model. In the case of models for related probability measures, it is expected that the following properties are necessarily met: (i) the support of the prior distribution induced by the process $\{G_x : x \in \mathcal{X}\}$ should be large; (ii) a continuity property in the sense that G_x converges, at least in probability, to G_{x_0} , as $x \longrightarrow x_0$; (iii) increasing (and decreasing) dependence of G_x and G_{x_0} with decreasing (increasing) distance between x and x_0; and (iv) posterior computation can proceed efficiently through a straightforward MCMC algorithm. A commonly studied property, but not necessarily expected, is given by, (v) the marginal distribution of G_x ideally follows a familiar distribution at any given level of the predictor x. Since the parametric space corresponds to a product space, the study of the support property requires to consider generalizations of the standard topologies for spaces of singles probability measures. The next section discusses some aspects related to this property.

1.1.5 Support property in models for related probability measures

The support of a probability measure, also known as topological support, is defined as the smallest closed set with probability one. Here, the parametric space $\mathcal{G}$ must be endowed with a topology and the collection of all Borel sets on $\mathcal{G}$ forms the σ–algebra where the prior distribution is defined. It is said that a probability measure has full support, given a topology, if $\mathcal{G}$ is its topological support. In this context, a probability measure satisfies the large support property if its support contains a sufficiently large amount of elements of $\mathcal{G}$. This property can be considered a minimum requirement and almost a "necessary condition" for a nonparametric model to be considered "nonparametric", because it ensures that a nonparametric prior does not assign too much mass on small or restricted sets of probability measures. This property is also important because it is a typically required condition for frequentist consistency of the posterior distribution.

As was discussed above, the definition of the support of probability models depends on the choice of the topology. These topologies are usually defined through a base of open neighborhoods. In the context of models for related probability measures, the most natural topologies that could be considered are weak product, L_q product, ν-integrated L_q and L_∞, where $q = 1, \ldots, \infty$ and ν is a probability measure defined on the $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. A brief description of these topologies is given below:

Let $\mathcal{P}(S)^{\mathcal{X}} = \prod_{x \in \mathcal{X}} \mathcal{P}(S)$ be the product space formed by all the sets of predictor–dependent probability distributions of the form $\{H_x : x \in \mathcal{X}\}$, where for every $x \in \mathcal{X}$, H_x is a probability measure defined on $(S, \mathcal{B}(S))$. Product spaces are commonly equipped with a natural topology called the product topology, which is defined as the coarsest topology for which all the projections $\{H_x : x \in \mathcal{X}\} \mapsto H_{x_0}$, $x_0 \in \mathcal{X}$, are continuous. In this context, the notion of the continuity in the weak and L_q product topology is given by the weak convergence of probability measures and the L_q–distance between density functions, respectively.

Notice that for the L_q product topology the product space $\mathcal{P}(S)^{\mathcal{X}}$ must be restricted to sets of predictor–dependent probability distributions such that for every $x \in \mathcal{X}$, H_x has density w.r.t Lebesgue measure. Although the product topology can be a natural choice in product spaces, it is not the only one. Any product space can be identified with some space of functions, and any notion of distance between those functions can be used to induce a topology on this space. Here, the elements of $\mathcal{P}(S)^{\mathcal{X}}$ are identified with functions of the form $x \mapsto H_x$, and thus, the ν-integrated L_q and L_∞ topologies are induced by the metrics,

$$d_{\nu,L_q}(x \mapsto P_x, x \mapsto H_x) = \int_{\mathcal{X}} \int_S |p_x(y) - h_x(y)|^q \, dy\nu(dx),$$

and

$$d_{L_\infty}(x \mapsto P_x, x \mapsto H_x) = \sup_{x \in \mathcal{X}} \sup_{y \in S} |p_x(y) - h_x(y)|,$$

respectively, and where for every $x \in \mathcal{X}$, p_x and h_x is the density w.r.t. Lebesgue measure of P_x and H_x, respectively. Again, the space $\mathcal{P}(S)^{\mathcal{X}}$ must be restricted so that these metrics are well defined.

Notice that the topologies presented above can be defined through a base of open neighborhoods. Therefore, to show that under a specific topology, the set $\mathcal{M} \subseteq \mathcal{P}(S)^{\mathcal{X}}$ is the support of a particular prior distribution defined on $\mathcal{P}(S)^{\mathcal{X}}$, it is necessary to show that, for the topology being considered, $\mathcal{M}$ is a closed set, and for every open basic set, $\mathcal{N}$, the prior probability of the set $\mathcal{N} \bigcap \mathcal{M}$ is positive.

Although the formal definition of support depends on the considered topology, there exists an important kind of support that may not be interpreted in a topological sense, namely the Kullback–Leibler support. The topological sense is lost because the Kullback–Leibler divergence is asymmetric and not a metric. The importance of this support arises from the fact that it is required to show the consistency of the posterior distribution. The characterization of this support is done using a similar strategy to that used in topological spaces. Specifically, we say that $\{H_x : x \in \mathcal{X}\} \in \mathcal{P}(S)^{\mathcal{X}}$ belongs to the Kullback–Leibler support, if every neighbourhood of $\{H_x : x \in \mathcal{X}\}$ has positive prior probability, where such neighbourhoods have been defined using the Kullback–Leibler divergence.

1.2 Motivation

The motivation for the developments of this thesis is summarized in the following two sections.

1.2.1 Characterizing the support in models for related probability measures.

Although there exists a wide variety of probability models for related probability distributions, there is a scarcity of results characterizing the support of the proposed processes. Some recent results have been provided by Pati et al. (2011) and Norets & Pelenis (2011), in the context of dependent mixtures of Gaussians induced by probit stick–breaking processes (Chung & Dunson, 2009), and dependent mixtures of location–scale distributions induced by finite mixing distributions (Norets, 2010) and kernel stick–breaking processes (Dunson & Park, 2008), respectively. However, these results have been obtained for very specific BNP models and they could not be easily extended to a broader class of models. Another result was provided by MacEachern (1999, 2000), who partially characterized the support of the DDP, leaving a gap with regard to this property.

The lack of results that provide conditions, necessary or sufficient, to ensure the property of large support in models for related probability measures is an issue that deserves to be addressed. In particular, it is necessary to determine conditions that characterize the support of dependent processes that serve as mixing probability measures, for instance the DDP, and to study its relationship with the support of mixtures induced by these processes. Given the context above, one of the developments of this thesis was motivated by making a contribution addressed to study the support of some of the most popular models for related probability measures. In particular, we study the support of the DDP and dependent stick–breaking process, and its relationship with the support of mixture models induced by them. The weak product, L_1 product and Kullback-Leibler product support are considered in this study. The use of product topologies allows us to define the DDP or the stick–breaking process in terms of processes whose laws are entirely determined by their respective finite dimensional distributions. Moreover, given the connection between copulas and stochastic processes, we are also able to define the DDP and the stick–breaking process now using the structure provided by copulas.

1.2.2 A BNP model for bounded density regression.

In practice, it is becoming more common to deal with regression problems where the observed data have complex structures and where the use of standard statistical methods is limited. Depending on the complexity, we can find different alternative methods to deal with this kind of data, including the increasingly popular BNP methods. In this context, there is a rich variety of models for related probability distributions that have been proposed. The list is huge but the underlying challenge is the same. Some examples include the DDP MacEachern (1999, 2000), the DPM mixture of normals model for the joint distribution of the response and predictors (Müller et al., 1996), the hierarchical mixture of DPM (Müller et al., 2004), the order–based DDP model (Griffin & Steel, 2006), the predictor–dependent weighted mixture of DP (Dunson et al., 2007), the kernel–stick breaking (Dunson & Park, 2008), the probit–stick breaking processes (Chung & Dunson, 2009; Rodriguez & Dunson, 2009), the dependent skew DP model (Quintana, 2010), the Geometric stick-breaking processes for continuous-time (Mena et al., 2011), among many others.

To the best of our knowledge, all of the previous (and many others) approaches have considered models for densities on the real line and none have focused on densities defined on a known closed interval of the real line, $[l, u]$, $-\infty < l < u < \infty$. In principle, all these models could be applied if the data were suitably transformed. Since suitable transformations mean considering bijective functions, implying that the edges of the interval, l and u, are identified with $-\infty$ and ∞, it follows that the transformed densities would not be defined on the edges of the domain, i.e, at l and u. A natural solution to deal with this issue would consist of defining the values of the transformed densities evaluated at l and u, as the right-hand limit at l and left-hand limit at u, respectively. However, this solution has two important disadvantages: $(*)$ it is hard to compute these limits in practice; and $(**)$ the limits could be equal to infinity, implying an unbounded likelihood function if at least one of the observations is equal to l or u. The unboundedness of likelihood might be controlled by the prior distribution; but imposing restrictions on the prior is not a really easy task to do in this context. Therefore, models based on transformations should be used with some care, especially in cases where a part of the data associated to the response variable are concentrated on the edges of the interval.

In this context, the lack of BNP models for related probability distributions where the response variable belongs to a closed interval of the real line is the main motivation to propose a novel BNP model addressed to fill this gap. Here, we proposed an BNP model for predictor–dependent probability distributions well defined on a closed interval, which satisfies the desired properties (i)-(v), section 1.1.4.

1.3 Outline of this dissertation

The work in this thesis can be divided in two parts that have been developed inside a BNP framework, in particular, in the context of probability models for predictor–dependent probability distributions. Each part is presented in individual chapters, to be precise in chapters 2 and 3, which can be read independently because they are self-contained in terms of the notation and abbreviations. Both chapters include an introduction, development and a final section with the conclusions and future work. The outline of this dissertation is as follows.

In Chapter 2, we provide an alternative characterization of the weak support of the two versions of MacEachern's DDP, namely, a version where both weights and support points are functions of the predictors, and a version where only the support points are functions of the predictors. We also characterize the weak support of a version of the DDP model where only the weights depend on predictors. In addition, we also provide sufficient conditions for the full Hellinger support of mixture models induced by DDP priors, and characterize their Kulback–Leibler support. Our results are based on an alternative definition of MacEachern's DDP, which exploits the connection between stochastic processes and copulas. Specifically, families of copulas are used to define the finite dimensional distributions of stochastic processes with given marginal distributions. The alternative formulation of the DDP makes explicit the parameters of the process, and their role on the support properties.

In Chapter 3, we propose a novel probability model for sets of predictor–dependent probability distributions with bounded domain. Dependent stick-breaking processes are used to extend the Dirichlet–Bernstein prior proposed by Petrone (1999a,b). Theoretical properties associated with the proposal, such as continuity, correlation structure, support and consistency of the posterior distribution, are studied. In particular, the weak product, L_∞ product, L_∞, Kullback–Leibler product and, ν–integrated and L_∞ Kullback–Leibler support were characterized for the proposal model. Practicable special cases of the general model are discussed and illustrated using simulated and real data. The simulated data is also used to compare the proposed methodology to existing methods.

In Chapter 4, we provide a review of the results presented in this Dissertation, and discuss possible directions of future research. A final appendix contains proofs of theorems, technical results and figures showing some of the estimated densities.

Chapter 2

On the support of MacEachern's dependent Dirichlet processes and extensions

This chapter has been published as:

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.

2.1 Introduction

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

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

$$G(B) = \sum_{i=1}^{\infty} W_i \delta_{\theta_i}(B), \quad B \in \mathscr{S}, \tag{2.1}$$

where $\delta_\theta(\cdot)$ is the Dirac measure at θ, $W_i = V_i \prod_{j<i}(1 - V_j)$ for all $i \geq 1$, with $V_i \mid \alpha \overset{iid}{\sim}$ Beta$(1, \alpha)$, and $\theta_i \mid G_0 \overset{iid}{\sim} G_0$. MacEachern (1999, 2000) generalized expression (2.1) by considering

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

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

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

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

In this paper we provide an alternative characterization of the weak support of the two versions of MacEachern's DDP discussed above, namely, a version where both weights and support points are functions of the predictors, and a version where only the support points are functions of the predictors. We also characterize the weak support of a version of the DDP model where only the weights depend on predictors. Finally, we provide sufficient conditions for the full Hellinger support of mixture models induced by DDP priors, and characterize their Kulback–Leibler support. Our results are based on an alternative definition of MacEachern's DDP, which exploits the connection between stochastic processes and copulas. Specifically, families of copulas are used to define the finite dimensional distributions of stochastic processes with given marginal distributions. The alternative formulation of the DDP makes explicit the parameters of the process, and their role on the support properties. The rest of this paper is organized as follows. Section 2.2 provides the alternative definition of MacEachern's DDP. Section 2.3 contains the main results about the support of the various DDP versions, as well as extensions to more general stick–breaking constructions. A general discussion and possible future research lines are given in Section 2.4.

2.2 MacEachern's dependent Dirichlet processes

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

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

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

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

2.2.1 Copulas and stochastic processes

Copulas are functions that are useful for describing and understanding the dependence structure between random variables. The basic idea is the ability to express a multivariate distribution as a function of its marginal distributions. If H is a d–variate CDF with marginal CDF's given by $F_1, \ldots, F_d$, then by Sklar's theorem (Sklar, 1959), there exists a copula function

$C : [0,1]^d \longrightarrow [0,1]$ such that $H(t_1,\ldots,t_d) = C(F_1(t_1),\ldots,F_d(t_d))$, for all $t_1,\ldots,t_d \in \mathbb{R}$, and this representation is unique if the marginal distributions are absolutely continuous w.r.t. Lebesgue measure. Thus by the probability integral transform, a copula function is a d–variate CDF on $[0,1]^d$ with uniform marginals on $[0,1]$, which fully captures the dependence among the associated random variables, irrespective of the marginal distributions. Examples and properties of copulas can be found, for example, in Joe (1997).

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

Corollary 2.1. *Let $\mathscr{C}_{\mathscr{X}} = \{C_{x_1,\ldots,x_d} : x_1,\ldots,x_d \in \mathscr{X}, d > 1\}$ be a collection of copula functions and $\mathscr{D}_{\mathscr{X}} = \{F_x : x \in \mathscr{X}\}$ a collection of one–dimensional probability distributions defined on a common measurable space $(\mathscr{D}, \mathscr{B}(\mathscr{D}))$, where $\mathscr{D} \subseteq \mathbb{R}$. Assume that for every integer $d > 1$, $x_1,\ldots,x_d \in \mathscr{X}$, $u_i \in [0,1]$, $i = 1,\ldots,d$, $k \in \{1,\ldots,d\}$, and permutation $\pi = (\pi_1,\ldots,\pi_d)$ of $\{1,\ldots,d\}$, the elements in $\mathscr{C}_{\mathscr{X}}$ satisfy the following consistency conditions:*

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

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

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

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

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

such that

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

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

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

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

From a practical point of view, it is easy to specify a family of copulas satisfying conditions (i) and (ii) in Corollary 2.1. An obvious approach is to consider the family of copula functions arising from the finite–dimensional distributions of known and tractable stochastic processes. The family of copula functions associated with Gaussian or t–Student processes could be considered as natural candidates in many applications for which $\mathscr{X} \subseteq \mathbb{R}^p$. The finite–dimensional copula functions of Gaussian processes are given by

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

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

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

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

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

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

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

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

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

2.2.2 The alternative definition

Let $\mathscr{C}_{\mathscr{X}}^{V}$ and $\mathscr{C}_{\mathscr{X}}^{\theta}$ be two sets of copulas satisfying the consistency conditions of Corollary 2.1. As earlier, let $\alpha_{\mathscr{X}} = \{\alpha_x : x \in \mathscr{X}\}$ be a set such that, for every $x \in \mathscr{X}$, $\alpha_x \in \mathbb{R}_0^+$, and let $G_{\mathscr{X}}^0 = \{G_x^0 :\ x \in \mathscr{X}\}$ be a set of probability measures defined on a common measurable space $(S, \mathscr{S})$, where $S \subseteq \mathbb{R}^q$, $q \in \mathbb{N}$, and $\mathscr{S} = \mathscr{B}(S)$ is the Borel σ–field of S. Finally, let $\mathscr{P}(S)$ be the set of all Borel probability measures defined on $(S, \mathscr{S})$.

Definition 2.1. *Let $\{G_x : x \in \mathscr{X}\}$ be a $\mathscr{P}(S)$–valued stochastic process on an appropriate probability space $(\Omega, \mathscr{A}, P)$ such that:*

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

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

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

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

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

In what follows, two simplifications of the general definition of the process will be of interest. If the stochastic processes in (i) of Definition 2.1 are replaced by independent and identically distributed Beta$(1, \alpha)$ random variables, with $\alpha > 0$, the process will be referred to as "single weights" DDP. This is to emphasize the fact that the weights in the stick–breaking representation (iii) of Definition 2.1, are not indexed by predictors x.

Definition 2.2. *Let $\{G_x : x \in \mathscr{X}\}$ be a $\mathscr{P}(S)$–valued stochastic process on an appropriate probability space $(\Omega, \mathscr{A}, P)$ such that:*

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

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

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

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

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

The second simplification is when the stochastic processes in (ii) of Definition 2.1 are replaced by independent random vectors with common distribution G^0, where G^0 is supported on the measurable space $(S, \mathscr{S})$. In this case the process will be referred to as "single atoms" DDP, to emphasize the fact that the support points in the stick–breaking representation are not indexed by predictors x.

Definition 2.3. *Let $\{G_x : x \in \mathscr{X}\}$ be a $\mathscr{P}(S)$–valued stochastic process on an appropriate probability space $(\Omega, \mathscr{A}, P)$ such that:*

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

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

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

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

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

2.3 The main results

2.3.1 Preliminaries

As is widely known, the definition of the support of probability models on functional spaces depends on the choice of a "distance" defining the basic neighborhoods. The results presented here are based on three different notions of distance between probability measures. Theorems 2.1 through 2.3 below are based on neighborhoods created using any distance that metrizes the weak star topology, namely, any distance d_W such that, for two probability measures F and G_n defined on a common measurable space, $d_W(G_n,F)\longrightarrow 0$ if and only if G_n converges weakly to F as n goes to infinity. If F and G are probability measures absolutely continuous with respect to a common dominating measure, stronger notions of distance can be considered. The results summarized in Theorems 2.4 and 2.5 are based on neighborhoods created using the Hellinger distance, which is topologically equivalent to the L_1 distance, and the Kullback–Leibler divergence, respectively. If f and g are versions of the densities of F and G with respect to a dominating measure λ, respectively, the L_1 distance, Hellinger distance and the Kullback–Leibler divergence are defined by

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

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

and

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

respectively.

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

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

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

2.3.2 Weak support of dependent Dirichlet processes

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

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

Theorem 2.1. *Let $\{G_x : x \in \mathscr{X}\}$ be a* DDP $\left(\alpha_{\mathscr{X}}, \mathscr{C}_{\mathscr{X}}^{\theta}, \mathscr{C}_{\mathscr{X}}^{V}, G_{\mathscr{X}}^0\right)$. *If $\mathscr{C}_{\mathscr{X}}^{\theta}$ and $\mathscr{C}_{\mathscr{X}}^{V}$ are col-*

lections of copulas with positive density w.r.t. Lebesgue measure, on the appropriate unitary hyper–cubes, then $\mathscr{P}(\Theta)^{\mathscr{X}}$ is the weak support of the process, i.e., the DDP has full weak support.

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

Let $\mathscr{P}_n = \{P_x^n : x \in \mathscr{X}\} \in \mathscr{P}(\Theta)^{\mathscr{X}}$ be a collection of probability measures with support contained in Θ. Let $\{\mathscr{P}_n\}_{n\geq 1} \subset \mathscr{P}(\Theta)^{\mathscr{X}}$ be a sequence of such collections, satisfying the condition that for all $x \in \mathscr{X}$, $P_x^n \overset{weakly}{\longrightarrow} P_x$, when $n \longrightarrow \infty$, where P_x is a probability measure. Since S is closed and $P_x^n \overset{weakly}{\longrightarrow} P_x$, Portmanteau's theorem implies that $P_x(\Theta) \geq \limsup_n P_x^n(\Theta)$, for every $x \in \mathscr{X}$. It follows that $\mathscr{P}(\Theta)^{\mathscr{X}}$ is a closed set. Now, let $\Theta^{\mathscr{X}} = \prod_{x\in\mathscr{X}} \Theta$. Since Θ is the support of G_x^0, for every $x \in \mathscr{X}$, it follows that

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

i.e.,

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

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

$$\prod_{i=1}^{T}\left\{P_{x_i} \in \mathscr{P}(\Theta) : \left|\int f_{ij}\,dP_{x_i} - \int f_{ij}\,dP_{x_i}^0\right| < \epsilon_i,\ j = 1, \ldots, K_i\right\}, \tag{2.2}$$

where $x_1, \ldots, x_T \in \mathscr{X}$, T and K_i, $i = 1, \ldots, T$, are positive integers, f_{ij}, $i = 1, \ldots, T$, and $j = 1, \ldots, K_i$, are bounded continuous functions and ϵ_i, $i = 1, \ldots, T$, are positive constants. To show that neighborhoods of the form (2.2) have positive probability mass, it suffices to show they contain certain subset–neighborhoods with positive probability mass. In particular, we consider subset–neighborhoods of probability measures which are absolutely continuous w.r.t. the corresponding centering distributions and that adopt the form

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

where Q_{x_i} is a probability measure absolutely continuous w.r.t. $G^0_{x_i}$, $i=1,\dots,T$, $A_{i1},\dots,A_{im_i}\subseteq\Theta$ are measurable sets with Q_{x_i}–null boundary, and $\epsilon^*>0$. For discrete centering distributions $G^0_{x_1},\dots,G^0_{x_T}$, the existence of a subset–neighborhood $U\left(Q_{x_1},\dots,Q_{x_T},\{A_{ij}\},\epsilon^*\right)$ of the set (2.2) is immediately ensured. The case of centering distributions that are absolutely continuous w.r.t. Lebesgue measure follows after Lemma A.1 in Appendix A.

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

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

where A^1_{ij} is interpreted as A_{ij} and A^0_{ij} is interpreted as A^c_{ij}. Note that

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

is a measurable partition of Θ such that

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

It follows that sets of the form

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

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

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

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

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

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

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

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

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

Thus, to show that (2.2) has positive P–measure, it suffices to show that

$$P\{\omega \in \Omega : [G(x_i,\omega)(A_0),\ldots,G(x_i,\omega)(A_k)] \in B\left(\mathbf{s}_{x_i},\epsilon\right),\ i = 1,\ldots,T\} > 0. \tag{2.3}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

It then follows from the DDP definition that

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

where,

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

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

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

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

By the definition of the process,

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

and

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

It follows that

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

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

$$P\left\{\omega \in \Omega : \left[V_j\left(x_1, \omega\right), \ldots, V_j\left(x_T, \omega\right)\right] \in Q_j^{\omega}, j = 1, \ldots, k+1\right\} = \int_{Q_1} \int_{Q_2(\mathbf{v}_1)} \cdots \int_{Q_{k+1}(\mathbf{v}_1, \ldots, \mathbf{v}_k)} f_{x_1,\ldots,x_T}^{V}\left(\mathbf{v}_1\right) \cdots f_{x_1,\ldots,x_T}^{V}\left(\mathbf{v}_{k+1}\right) d\mathbf{v}_{k+1} \cdots d\mathbf{v}_2 d\mathbf{v}_1 > 0, \tag{2.4}$$

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

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

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

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

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

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

Theorem 2.2. *Let* $\{G_x : x \in \mathscr{X}\}$ *be a* $\mathrm{wDDP}\left(\alpha, \mathscr{C}_{\mathscr{X}}^{\theta}, G_{\mathscr{X}}^0\right)$. *If* $\mathscr{C}_{\mathscr{X}}^{\theta}$ *is a collection of copulas with positive density w.r.t. Lebesgue measure, on the appropriate unitary hyper–cubes, then* $\mathscr{P}\left(\Theta\right)^{\mathscr{X}}$ *is the weak support of the process.*

Proof: Using a similar reasoning as in the proof of Theorem 2.1, it suffices to prove (2.3), that is

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

As in the proof of Theorem 2.1, we consider constraints for the elements of the wDDP that imply the previous relation. Since the rational numbers are dense in $\mathbb{R}$, there exist $M_i, m_{ij}\in\mathbb{N}$ such that for $i=1,\ldots,T$, and $j=0,\ldots,k-1$,

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

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

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

Therefore, for any

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

that verifies

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

we have

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

and

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

where,

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

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

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

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

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

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

It follows that

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

The non–singularity of the Beta distribution implies that

$$P\left\{\omega\in\Omega:V_l\left(\omega\right)\in Q_l^{\omega},l=1,\ldots,N\right\}>0. \tag{2.5}$$

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

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

which completes the proof. □

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

Theorem 2.3. *Let $\{G_x:x\in\mathscr{X}\}$ be a $\theta\mathrm{DDP}\left(\alpha_{\mathscr{X}},\mathscr{C}_{\mathscr{X}}^{V},G^0\right)$, where the support of G_0 is Θ. If $\mathscr{C}_{\mathscr{X}}^{V}$ is a collection of copulas with positive density w.r.t. to Lebesgue measure, on the appropriate unitary hyper–cubes, then the support of the process is $\mathscr{P}\left(\Theta\right)^{\mathscr{X}}$.*

Proof: In analogy with the proofs of Theorems 2.1 and 2.2, it suffices to prove (2.3), that is

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

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

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

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

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

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

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

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

where,

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

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

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

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

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

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

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

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

where,

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

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

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

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

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

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

It follows that

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

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

$$\begin{aligned} & P\left\{\omega\in\Omega:[V_j(x_1,\omega),\cdots,V_j(x_T,\omega)]\in Q_j^{\omega}, j=1,\ldots,k+1\right\}= \\ & \int_{Q_1}\int_{Q_2(\mathbf{v}_1)}\cdots\int_{Q_{k+1}(\mathbf{v}_1,\cdots,\mathbf{v}_k)} f_{x_1,\ldots,x_T}^{V}(\mathbf{v}_1)\cdots f_{x_1,\ldots,x_T}^{V}(\mathbf{v}_{k+1})\,d\mathbf{v}_{k+1}\cdots d\mathbf{v}_2 d\mathbf{v}_1>0. \end{aligned} \tag{2.6}$$

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

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

which completes the proof of the theorem. □

2.3.3 The support of dependent Dirichlet process mixture models

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

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

Theorem 2.4. *Let ψ be a non–negative valued function defined on the product measurable space $(\mathscr{Y} \times \Theta, \mathscr{B}(\mathscr{Y}) \otimes \mathscr{B}(\Theta))$, where $\mathscr{Y} \subseteq \mathbb{R}^n$ is the sample space with corresponding Borel σ–field $\mathscr{B}(\mathscr{Y})$, $\Theta \subseteq \mathbb{R}^q$ is the parameter space with corresponding Borel σ–field $\mathscr{B}(\Theta)$ and $\mathscr{B}(\mathscr{Y}) \otimes \mathscr{B}(\Theta)$ denotes the product σ–field on $\mathscr{Y} \times \Theta$. Assume that ψ satisfies the following conditions:*

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

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

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

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

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

then

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

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

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

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

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

If $\{G_x : x \in \mathscr{X}\}$ *is a* DPP, *a* wDDP *or a* θDDP, *satisfying the conditions of Theorem 2.1, 2.2 or 2.3, respectively, then the Hellinger support of the process*

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

is

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

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

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

$$\begin{aligned}\prod_{i=1}^{T}\Bigg\{\int_\Theta \psi\left(\cdot,\theta\right) P_{x_i}\left(d\theta\right) :& \\ \int_{\mathscr{Y}}\left|\int_\Theta \psi\left(y,\theta\right) P_{x_i}\left(d\theta\right) - f_{x_i}^0\left(y\right)\right| \lambda\left(dy\right) < \epsilon,&\ P_{x_i} \in \mathscr{P}\left(\Theta\right)\Bigg\},\end{aligned} \tag{2.7}$$

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

To show that the DDP mixture model assigns positive probability mass to sets of the form (2.7), we construct a weak neighborhood around $\{P_x^0 : x \in \mathscr{X}\} \in \mathscr{P}\left(\Theta\right)^{\mathscr{X}}$ such that every element in it satisfies (2.7). This is done by appropriately defining the bounded and continuous functions that determine the weak neighborhood.

Let ν, ρ and η be positive constants. Fix a compact set $K_{x_i} \subset \mathscr{B}\left(\mathscr{Y}\right)$ such that

$$\int_{K_{x_i}^c} f_{x_i}^0\left(y\right)\lambda\left(dy\right) < \frac{\epsilon}{8},$$

and define

$$h_{i,1}^0\left(\theta\right) = \int_{K_{x_i}^c} \psi\left(y,\theta\right)\lambda\left(dy\right),$$

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

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

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

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

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

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

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

$$\prod_{i=1}^{T} \{P_{x_i} \in \mathscr{P}(\Theta) : \\ \left| \int h^l_{i,j_l} dP_{x_i} - \int h^l_{i,j_l} dP^0_{x_i} \right| < \nu,\ l = 0, 1,\ j_0 = 1, 2,\ 1 \leq j_1 \leq n_i \right\}, \tag{2.8}$$

for $\nu > 0$. We will show that for appropriate choices of η, ν, r, and ρ, every collection $\{P_{x_1}, \ldots, P_{x_T}\}$ in sets of the form (2.8), satisfies

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

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

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

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

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

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

$$\begin{aligned}\int_{K^c_{x_i}} \left| \int_{\Theta} \psi\left(y,\theta\right) P_{x_i}\left(d\theta\right) - f^0_{x_i}\left(y\right) \right| \lambda\left(dy\right) &\leq \int h^0_{i,1} dP_{x_i} + \int_{K^c_{x_i}} f^0_{x_i}\left(y\right) \lambda\left(dy\right) \\ &\leq \nu + \frac{\epsilon}{4}. \qquad (2.9)\end{aligned}$$

In addition, note that

$$\int_{K_{x_i}} \left| \int_{\Theta} \psi\left(y,\theta\right) P_{x_i}\left(d\theta\right) - f^0_{x_i}\left(y\right) \right| \lambda\left(dy\right) \leq B_1 + B_2 + B_3 \qquad (2.10)$$

where,

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

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

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

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

Now, since

$$P_{x_i}\left[\overline{B}\left(0, r\right)^C\right] \le \nu + \int h^0_{i,2} dP^0_{x_i} \le \nu + P^0_{x_i}\left[\overline{B}\left(0, r-\nu\right)^C\right] \le \nu + \rho,$$

it follows that

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

Finally, by (2.9) and (2.10), if

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

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

and

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

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

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

Theorem 2.5. *Assume that ψ belongs to a location–scale family, $\psi(\cdot, \theta) = \sigma^{-n} k\left(\frac{\cdot - \mu}{\sigma}\right)$, where $\mu = \left(\mu_1, \ldots, \mu_n\right)$ is an n–dimensional vector, and $\sigma \in \mathbb{R}^+$. Let k be a non–negative valued function defined on $(\mathscr{Y} \times \Theta, \mathscr{B}(\mathscr{Y}) \otimes \mathscr{B}(\Theta))$, where $\mathscr{Y} \subseteq \mathbb{R}^n$ is the sample space with corresponding Borel σ–field $\mathscr{B}(\mathscr{Y})$ and $\Theta \subseteq \mathbb{R}^n \times \mathbb{R}^+$ is the parameter space with corresponding Borel σ–field $\mathscr{B}(\Theta)$. Assume k satisfies the following conditions:*

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

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

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

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

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

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

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

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

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

If $\{G_x : x \in \mathscr{X}\}$ *is a* DPP*, a* wDDP *or a* θDDP*, where* $\mathbb{R}^n \times \mathbb{R}^+$ *is the support of the corresponding centering distributions, and satisfying the conditions of Theorem 2.1, 2.2 or 2.3, respectively, then*

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

for $\epsilon > 0$.

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

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

and

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

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

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

and from Theorems 2.1, 2.2 and 2.3, it follows that

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

which completes the proof. $\square$

Notice that the conditions of Theorem 2.5 are satisfied for most of the important location–scale kernels. In fact, Wu & Ghosal (2008) show that conditions (i) – (iv) are satisfied by the normal, skew–normal, double–exponential, logistic and t-Student kernels.

2.3.4 Extensions to more general dependent processes

Although the previous results about the support of models for collections of probability distributions are focused on MacEachern's DDP, similar results can be obtained for more general dependent processes. Natural candidates for the definition of dependent processes include the general class of stick–breaking (SB) processes, which includes the DP, the two–parameter Poisson–Dirichlet processes (Pitman & Yor, 1997), the beta two–parameter processes (Ishwaran & James, 2001) and the geometric stick–breaking processes (Mena et al., 2011), as important special cases. A SB probability measure is given by expression (2.1), but where the beta distribution associated with the SB construction of the weights can be replaced by any collection of distributions defined on the unit interval $[0, 1]$ such that the resulting weights add up to one almost surely. Specifically, the weights are given by $W_i = V_i \prod_{j<i}(1 - V_j)$, for every $i \geq 1$, where $V_i \mid H_i \overset{ind}{\sim} H_i$, with H_i being a probability measure on $[0, 1]$, for every $i \in \mathbb{N}$, and such that

$$\sum_{i=1}^{\infty} W_i \overset{a.s.}{=} 1. \tag{2.11}$$

Notice that, under an SB prior, it can be shown that a necessary and sufficient condition for expression (2.11) to hold is that $\sum_{i=1}^{\infty} \log\left(1 - E_{H_i}(V_i)\right) = -\infty$.

For every $i \in \mathbb{N}$, let $\mathscr{C}_{\mathscr{X}}^{V_i}$ be a set of copulas satisfying the consistency conditions of Corollary 2.1 and set $\mathscr{C}_{\mathscr{X},\mathbb{N}}^{V} = \left\{\mathscr{C}_{\mathscr{X}}^{V_i} : \ i \in \mathbb{N}\right\}$. For every $i \in \mathbb{N}$, let $\mathscr{V}_{\mathscr{X}}^{V_i} = \{H_{(i,x)} : x \in \mathscr{X}\}$ be a collection of probability distributions defined on $([0, 1], \mathscr{B}([0, 1]))$ and set $\mathscr{V}_{\mathscr{X},\mathbb{N}} = \left\{\mathscr{V}_{\mathscr{X}}^{V_i} : \ i \in \mathbb{N}\right\}$.

Definition 2.4. *Let $\{G_x : x \in \mathscr{X}\}$ be a $\mathscr{P}(S)$–valued stochastic process on an appropriate probability space $(\Omega, \mathscr{A}, P)$ such that:*

(i) $V_1, V_2, \ldots$ *are independent stochastic processes of the form* $V_i : \mathscr{X} \times \Omega \to [0, 1]$, $i \geq 1$,

with finite dimensional distributions determined by the set of copulas $\mathscr{C}_{\mathscr{X}}^{V_i}$ and the set of marginal distributions $\mathscr{V}_{\mathscr{X}}^{V_i}$, such that, for every $x \in \mathscr{X}$,

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

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

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

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

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

As in the DDP case, two simplifications of the general definition of the DSBP can be considered. If the stochastic processes in (i) of Definition 2.4 are replaced by independent random variables with label–specific distribution H_i, then the process will be referred to as "single weights" DSBP, to emphasize the fact that the weights in the stick–breaking representation (iii) of Definition 2.4, are not indexed by predictors x. In this case, the process is parameterized by $\left(\mathscr{C}_{\mathscr{X}}^{\theta}, \mathscr{V}_{\mathbb{N}}^{V}, G_{\mathscr{X}}^{0}\right)$, and denoted by wDSBP$\left(\mathscr{C}_{\mathscr{X}}^{\theta}, \mathscr{V}_{\mathbb{N}}^{V}, G_{\mathscr{X}}^{0}\right)$, where $\mathscr{V}_{\mathbb{N}}^{V} = \{H_i : \ i \in \mathbb{N}\}$ is a collection of probability distributions on $[0,1]$, such that condition (2.11) holds. If the stochastic processes in (ii) of Definition 2.4 are replaced by independent random vectors with common distribution G^0, where G^0 is supported on the measurable space $(S, \mathscr{S})$, then the process will be referred to as "single atoms" DSBP, to emphasize the fact that the support points in the stick–breaking representation are not indexed by predictors x. This version of the process is parameterized by $\left(\mathscr{C}_{\mathscr{X},\mathbb{N}}^{V}, \mathscr{V}_{\mathscr{X},\mathbb{N}}^{V}, G^0\right)$, and denoted by θDSBP$\left(\mathscr{C}_{\mathscr{X},\mathbb{N}}^{V}, \mathscr{V}_{\mathscr{X},\mathbb{N}}^{V}, G^0\right)$.

Theorem 2.6. *Let* $\{G_x : x \in \mathscr{X}\}$ *be a* DSBP$\big(\mathscr{C}^V_{\mathscr{X},\mathbb{N}}, \mathscr{C}^\theta_{\mathscr{X}}, \mathscr{V}^V_{\mathscr{X},\mathbb{N}}, G^0_{\mathscr{X}}\big)$*. If* $\Theta \subseteq S$ *is the support of* G^0_x*, for every* $x \in \mathscr{X}$*,* $\mathscr{C}^V_{\mathscr{X},\mathbb{N}}$ *and* $\mathscr{C}^\theta_{\mathscr{X}}$ *are collections of copulas with positive density w.r.t. Lebesgue measure, on the appropriate unitary hyper–cubes, and, for every* $i \in \mathbb{N}$*, the elements in* $\mathscr{V}^{V_i}_{\mathscr{X}}$ *have positive density on* $[0,1]$*, then* $\mathscr{P}(\Theta)^{\mathscr{X}}$ *is the weak support of the process, i.e., the* DSBP *has full weak support.*

Proof: The proof follows similar arguments to the ones of Theorem 2.1. Specifically, it is only needed to replace

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

in expression (2.4) by

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

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

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

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

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

□

Theorem 2.7. *Let* $\{G_x : x \in \mathscr{X}\}$ *be a* wDSBP$\big(\mathscr{C}^\theta_{\mathscr{X}}, \mathscr{V}^V_{\mathbb{N}}, G^0_{\mathscr{X}}\big)$*. If* $\Theta \subseteq S$ *is the support of* G^0_x*, for every* $x \in \mathscr{X}$*,* $\mathscr{C}^\theta_{\mathscr{X}}$ *is a collection of copulas with positive density w.r.t. Lebesgue measure, on the appropriate unitary hyper–cubes, and, for every* $i \in \mathbb{N}$*,* H_i *has positive density on* $[0,1]$*, then* $\mathscr{P}(\Theta)^{\mathscr{X}}$ *is the weak support of the process, i.e., the* wDSBP *has full weak support.*

Proof: The non–singularity of the $H_{(i)}$'s implies that condition (2.5) holds, for every $i \in \mathbb{N}$. The rest of the proof remains the same as for Theorem 2.2. □

Theorem 2.8. *Let $\{G_x : x \in \mathscr{X}\}$ be a θDSBP$\left(\mathscr{C}^V_{\mathscr{X},\mathbb{N}}, \mathscr{V}^V_{\mathscr{X},\mathbb{N}}, G^0\right)$, where Θ is the support of G_0. If $\mathscr{C}^V_{\mathscr{X},\mathbb{N}}$ is a collection of copulas with positive density w.r.t. to Lebesgue measure, on the appropriate unitary hyper–cubes, and, for every $i \in \mathbb{N}$, the elements in $\mathscr{V}^{V_i}_{\mathscr{X}}$ have positive density on $[0,1]$, then the support of the process is $\mathscr{P}\left(\Theta\right)^{\mathscr{X}}$.*

Proof: The proof follows similar arguments to the ones of Theorem 2.3. It is only needed to replace

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

in expression (2.6) by

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

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

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

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

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

□

Since the proofs of Theorems 2.4 and 2.5 depend on the dependent mixing distributions through their weak support only, the results are also valid for the different versions of the DSBP. Thus, the following theorems are stated without any proof.

Theorem 2.9. *Let ψ be a non–negative valued function defined on the product measurable space $(\mathscr{Y} \times \Theta, \mathscr{B}(\mathscr{Y}) \otimes \mathscr{B}(\Theta))$, where $\mathscr{Y} \subseteq \mathbb{R}^n$ is the sample space with corresponding Borel σ–field $\mathscr{B}(\mathscr{Y})$ and $\Theta \subseteq \mathbb{R}^q$ is the parameter space with corresponding Borel σ–field $\mathscr{B}(\Theta)$. Assume that ψ satisfies conditions (i) – (iii) of Theorem 2.4. If $\{G_x : x \in \mathscr{X}\}$ is a DSBP, a wDSBP or a θDSBP, satisfying the conditions of Theorem 2.6, 2.7 or 2.8, respectively, then the Hellinger support of the process $\left\{\int_\Theta \psi\left(\cdot,\theta\right) G_x\left(d\theta\right) : x \in \mathscr{X}\right\}$ is*

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

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

Theorem 2.10. *Assume that ψ belongs to a location–scale family, $\psi(\cdot,\theta) = \sigma^{-n} k\left(\frac{\cdot-\mu}{\sigma}\right)$, where $\mu = (\mu_1, \ldots, \mu_n)$ is an n–dimensional vector, and $\sigma \in \mathbb{R}^+$. Let k be a non–negative valued function defined on $(\mathscr{Y} \times \Theta, \mathscr{B}(\mathscr{Y}) \otimes \mathscr{B}(\Theta))$, where $\mathscr{Y} \subseteq \mathbb{R}^n$ is the sample space with corresponding Borel σ–field $\mathscr{B}(\mathscr{Y})$ and $\Theta \subseteq \mathbb{R}^n \times \mathbb{R}^+$ is the parameter space with corresponding Borel σ–field $\mathscr{B}(\Theta)$. Assume k satisfies conditions (i) – (iv) of Theorem 2.5 and that the elements in $\{f^0_{x_i} : i = 1, \ldots, T\}$ satisfy conditions (v) – (viii) of Theorem 2.5. If $\{G_x : x \in \mathscr{X}\}$ is a DSBP, a wDSBP or a θDSBP, where $\mathbb{R}^n \times \mathbb{R}^+$ is the support of the corresponding centering distributions, and satisfying the conditions of Theorem 2.6, 2.7 or 2.8, respectively, then*

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

for $\epsilon > 0$.

2.4 Concluding remarks and future research

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

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

Chapter 3

Fully nonparametric regression for bounded data using dependent Bernstein polynomials

This chapter has been submitted for publication as:

BARRIENTOS, A. F., JARA, A. & QUINTANA, F. A. (2012). Fully nonparametric regression for bounded data using dependent Bernstein polynomials. *Submitted*.

3.1 Introduction

This paper deals with the problem of defining a fully nonparametric regression model for a continuous response variable with bounded support $y \in [l, u]$, $-\infty < l < u < +\infty$, based on a set of predictors $\boldsymbol{x} \in \mathscr{X} \subseteq \mathbb{R}^p$. The nonparametric regression model is induced by assuming $y \mid G_{\boldsymbol{x}} \overset{ind.}{\sim} G_{\boldsymbol{x}}$, where $G_{\boldsymbol{x}}$ is a probability measure defined on $([l, u], \mathscr{B}([l, u]))$, and by defining a probability model for the set of predictor-dependent continuous probability distributions $\mathscr{G} = \{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$, allowing the complete shape of the elements of $\mathscr{G}$ to change flexibly with the values of $\boldsymbol{x}$.

The problem of defining priors over related random probability distributions has received increasing attention over the past few years. To date, much effort has focused on constructions that generalize the widely used class of Dirichlet process priors (Ferguson, 1973, 1974). Some exceptions are Tokdar et al. (2010), Karabatsos & Walker (2011), Trippa et al. (2011) and Jara & Hanson (2011), who proposed models based on logistic Gaussian processes, on infinite ordered-category probit regressions, on dependent beta processes and tailfree processes, respectively. MacEachern (1999, 2000) proposed the dependent Dirichlet process (DDP) to define a full joint model on the set $\mathscr{G}$, where marginally every $G_{\boldsymbol{x}} \in \mathscr{G}$ is a Dirichlet process. The key idea behind the DDP is to introduce dependence by modifying the stick-breaking representation of each element in the set. Specifically, MacEachern (1999, 2000) proposed to consider discrete random measures of the form

$$G_{\boldsymbol{x}}(B) = \sum_{i=1}^{\infty} w_i(\boldsymbol{x}) \delta_{\theta_i(\boldsymbol{x})}(B),$$

where B is a measurable set in an appropriate space, the point masses $\theta_i(\boldsymbol{x})$, $i = 1, \ldots$, are independent stochastic processes with index set $\mathscr{X}$, and the weights take the form $w_i(\boldsymbol{x}) = V_i(\boldsymbol{x}) \prod_{j<i}[1 - V_j(\boldsymbol{x})]$, with $V_i(\boldsymbol{x})$, $i = 1, \ldots$, being independent stochastic processes with index set $\mathscr{X}$ and $\mathrm{beta}(1, M)$ marginal distribution. MacEachern (2000) also studied a version of the process with predictor-independent weights, $G_{\boldsymbol{x}}(B) = \sum_{i=1}^{\infty} w_i \delta_{\theta_i(\boldsymbol{x})}(B)$, and showed that this version of the model has full support when flexible point mass processes are consid-

ered. Versions of the predictor-independent weights DDP have been successfully applied to ANOVA (De Iorio et al., 2004), survival (De Iorio et al., 2009; Jara et al., 2010), spatial modeling (Gelfand et al., 2005), functional data (Dunson & Herring, 2006), time series (Caron et al., 2006), discriminant analysis (De la Cruz et al., 2007), and longitudinal data analysis (Müller et al., 2005).

Other extensions of the DP for dealing with related probability distributions include the DPM mixture of normals model for the joint distribution of the response and predictors (Müller et al., 1996), the hierarchical mixture of DPM (Müller et al., 2004), the hierarchical DP (Teh et al., 2006), the order-based DDP model (Griffin & Steel, 2006), the nested DP (Rodriguez et al., 2008), the predictor-dependent weighted mixture of DP (Dunson et al., 2007), the kernel-stick breaking process (Dunson & Park, 2008), the matrix-stick breaking process (Dunson et al., 2008), the local DP (Chung & Dunson, 2011), the logit-stick breaking processes (Ren et al., 2011), the probit-stick breaking processes (Chung & Dunson, 2009; Rodríguez & Dunson, 2011), the cluster-X model (Müller & Quintana, 2010), the PPMx model (Müller et al., 2011), and the dependent skew DP model (Quintana, 2010), among many others. Dependent neutral to the right processes and correlated two-parameter Poisson-Dirichlet processes have been proposed by Epifani & Lijoi (2010) and Leisen & Lijoi (2011), respectively, by considering suitable Lévy copulas. The general class of dependent normalized completely random measures has been discussed, for instance, by Nipoti (2011) and Lijoi et al. (2012).

To the best of our knowledge, all of the existing approaches have focussed on densities on the real line, considering dependent mixtures of Gaussian models. While the normal kernel is a sensible choice in such settings, its usefulness is rather limited when considering densities on a known bounded interval. Even though an appropriate transformation could be applied to the data for the sake of the analysis using standard procedures, the estimates based on a normal kernel suffers from boundary effects at l and u. Since appropriate transformations mean considering bijective functions, implying that the edges of the interval, l and u, are identified with $-\infty$ and ∞, it follows that the transformed densities would not be defined on the edges of the domain, i.e, at l and u. Therefore, models based on transformations should be used with some care, especially in cases where a part of the data associated to the response variable are concentrated on the edges of the interval. In contrast, the class of models considered here is not restricted to any particular type of boundary behavior, and is thus more appropriated for data sets which are concentrated at the edges of the response support.

In the context of single density estimation problems, Petrone (1999a,b) and Kottas (2006) proposed models for probability distributions supported on $[0, 1]$ and $[0, T]$, respectively. In related work, Robert & Rousseau (2003) developed a goodness of fit method using beta mixtures with unknown number of components, and Mallick & Gelfand (1994) and Gelfand & Mallick (1995) considered mixtures of beta distribution functions to model random monotonic functions. We extend the class of Dirichlet-Bernstein priors of Petrone (1999a,b), to deal with sets of predictor-dependents probability distributions with bounded support.

The rest of the paper is organized as follows. Random Bernstein polynomials are briefly described in Section 3.2, so as to make the discussion self contained. Section 3.3 introduces the general version of the proposed model and its main theoretical properties are established. Proofs of these results are provided in an accompanying supplementary material file. Simplifications of the general model class are discussed in Section 3.4. The models are illustrated and compared to the existing methods using simulated data in Section 3.5, which also contains the results of a real-life data analysis. A final discussion section concludes the article.

3.2 Random Bernstein polynomials

Bernstein polynomials were introduced by Bernstein (1912) to give a proof of Weiertrass' approximation theorem. If $G : [0, 1] \longrightarrow \mathbb{R}$, the associated Bernstein polynomial of degree k is given by

$$\mathrm{BP}(y|k, G) = \sum_{j=0}^{k} G(j/k) \left(\begin{array}{c} k \\ j \end{array} \right) y^j (1-y)^{k-j}, y \in [0, 1]. \tag{3.1}$$

If G is the CDF of a probability measure defined on the unit interval, then (3.1) is also a CDF on $[0, 1]$ and represents a mixture of beta distributions. If $G(0) = 0$, its density function is given by

$$\mathrm{bp}(y \mid k, G) = \sum_{j=1}^{k} w_{j,k} \beta(y \mid j, k-j+1), \tag{3.2}$$

where $w_{j,k} = G(j/k) - G((j-1)/k)$, and $\beta(\cdot|a, b)$ stands for a beta density with parameters a and b. For a single-density estimation problem, Petrone (1999a,b) proposed a hierarchical prior, called the Bernstein polynomial prior (BPP). This consists of a random density given by (3.2), where k has probability mass function ρ, and given k, $\boldsymbol{w}_k = (w_{1,k}, \ldots, w_{k,k})$ has distribution H_k on the simplex

$$\Delta_{k-1} = \left\{ (w_1, \ldots, w_k) \in \mathbb{R}^k : 0 \leq w_j \leq 1, j = 1, \ldots, k, \sum_{j=1}^{k} w_j = 1 \right\}.$$

Petrone (1999a,b) referred to (3.2) as the Bernstein polynomial density with parameters k and $\boldsymbol{w}_k$, and showed that if ρ assigns positive mass to all naturals, and the density of H_k is positive for any point in Δ_k, then the weak support of the BPP is the space of all probability measures on $([0, 1], \mathscr{B}([0, 1]))$. Letting $\zeta_{j,k} = M(G_0(j/k) - G_0((j-1)/k))$, $j = 1, \ldots, k$, G_0 being a probability distribution on $(0, 1]$ and M being a positive constant, Petrone (1999a,b) used the fact that assuming

$$\boldsymbol{w}_k = (w_{1,k}, \ldots, w_{k,k}) \sim \text{Dirichlet}(\zeta_{1,k}, \ldots, \zeta_{k,k}),$$

is equivalent to assume that G follows a Dirichlet process (DP) prior, $G \mid M, G_0 \sim DP(MG_0)$. Petrone (1999a,b) refers to the later model as the Bernstein-Dirichlet prior (BDP), and discussed a Markov chain Monte Carlo (MCMC) algorithm to scan its posterior distribution. Petrone & Wasserman (2002) studied the consistency of the posterior distribution for BPP. They showed that under the same conditions that guarantee the full weak support of the prior, the posterior distribution is weakly consistent at any bounded continuous density on $[0, 1]$. Furthermore, they showed that under tail conditions on ρ the posterior is consistent with respect to the Hellinger metric.

3.3 The general model

3.3.1 The definition

Suppose that we observe regression data $\{(\boldsymbol{x}_i, y_i) : i = 1, \ldots, n\}$, where $\boldsymbol{x}_i \in \mathscr{X} \subseteq \mathbb{R}^p$ is a p-dimensional vector of predictors and y_i is a continuous $[l, u]$-valued outcome. Since the bounded support of the response variable can be rescaled to the unit interval, we will assume that $l = 0$ and $u = 1$ without loss of generality. To introduce dependence in the random probability measures with bounded support, we replace the DP mixing distribution in the definition of the BDP prior by a dependent stick-breaking process, which is defined by using transformed stochastic processes indexed by predictors $\boldsymbol{x} \in \mathscr{X}$. 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 $h_{\boldsymbol{x}} : \mathbb{R} \longrightarrow (0, 1]$, and such that for every $a \in \mathbb{R}$, $v_{\boldsymbol{x}}(a)$ and $h_{\boldsymbol{x}}(a)$ are continuous functions of $\boldsymbol{x}$. Let $\mathscr{P}([0, 1])$ be the set of all probability measures defined on $([0, 1], \mathscr{B}([0, 1]))$.

Definition 3.1. *Let $\mathscr{V}$ and $\mathscr{H}$ be two set of functions as before. Let $\mathscr{G} = \{G(\boldsymbol{x}, \omega) : \boldsymbol{x} \in \mathscr{X}\}$ be a $\mathscr{P}([0, 1])$-valued stochastic process on an appropriate probability space $(\Omega, \mathscr{A}, P)$ such that:*

(i) $\eta_1, \eta_2, \ldots$*, are independent and identically distributed real-valued stochastic processes of the form* $\eta_i : \mathscr{X} \times \Omega \longrightarrow \mathbb{R}$, $i \geq 1$, *with law indexed by a finite-dimensional parameter* $\boldsymbol{\Psi}_1$ *and marginal distributions* $\{F_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$.

(ii) $z_1, z_2, \ldots$*, are independent and identically distributed real-valued stochastic processes of the form* $z_i : \mathscr{X} \times \Omega \longrightarrow \mathbb{R}$, $i \geq 1$, *with law indexed by a finite-dimensional parameter* $\boldsymbol{\Psi}_2$ *and marginal distributions* $\{H_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$.

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

(iv) For every $\boldsymbol{x} \in \mathscr{X}$ *and almost every* $\omega \in \Omega$, *the density function of* $G(\boldsymbol{x}, \omega)$, *w.r.t.*

Lebesgue measure, is given by the following dependent mixture of beta densities:

$$g(\boldsymbol{x},\omega)(\cdot) = \sum_{j=1}^{\infty} w_j(\boldsymbol{x},\omega)\beta\left(\cdot \mid \lceil k(\omega)\theta_j(\boldsymbol{x},\omega)\rceil, k(\omega) - \lceil k(\omega)\theta_j(\boldsymbol{x},\omega)\rceil + 1\right), \quad (3.3)$$

where $\lceil\cdot\rceil$ *denotes the ceiling function,* $\theta_j(\boldsymbol{x},\omega) = h_{\boldsymbol{x}}(z_j(\boldsymbol{x},\omega))$*, and*

$$w_j(\boldsymbol{x},\omega) = v_{\boldsymbol{x}}\left\{\eta_j(\boldsymbol{x},\omega)\right\}\prod_{i<j}\left[1 - v_{\boldsymbol{x}}\left\{\eta_i(\boldsymbol{x},\omega)\right\}\right].$$

The process, $\mathscr{G} = \{G_{\boldsymbol{x}} \doteq G(\boldsymbol{x},\omega) : \boldsymbol{x} \in \mathscr{X}\}$*, will be referred to as dependent Bernstein polynomial process with parameters* $(\lambda, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \mathscr{V}, \mathscr{H})$*, and denoted by* $\mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \mathscr{V}, \mathscr{H})$.

Notice that, for every $\omega \in \Omega$, expression (3.3) is indeed a density w.r.t. Lebesgue measure since, for every $\boldsymbol{x} \in \mathscr{X}$,

$$\sum_{i=1}^{\infty} \log\left[1 - E_{F_{\boldsymbol{x}}}\left(v_{\boldsymbol{x}}\left\{\eta_i(\boldsymbol{x},\cdot)\right\}\right)\right] = -\infty,$$

which is a sufficient and necessary condition for the weights to add up to one with probability one. In addition, it follows immediately from Definition 3.1 that the trajectories of the process are sets of Bernstein polynomial densities. In fact, (3.3) is equivalent to

$$g(\boldsymbol{x},\omega)(\cdot) = \sum_{j=1}^{k(\omega)} W_j(\boldsymbol{x},\omega)\beta\left(\cdot \mid j, k(\omega) - j + 1\right),$$

where $W_j(\boldsymbol{x},\omega) = \sum_{i=1}^{\infty} w_i(\boldsymbol{x},\omega) I\{\theta_i(\boldsymbol{x},\omega)\}_{\{\lceil k(\omega)\theta_i(\boldsymbol{x},\omega)\rceil = j\}}$, with $I\{\cdot\}_A$ being the indicator function for the set A.

The choice of the transformation functions $\mathscr{V}$ and $\mathscr{H}$ induce interesting properties of the DBPP. For instance, it is easy to show that if, for every $\boldsymbol{x} \in \mathscr{X}$, the elements in $\mathscr{V}$ are such that $v_{\boldsymbol{x}}(\cdot) = B^{-1}(F_{\boldsymbol{x}}(\cdot) \mid 1, M_{\boldsymbol{x}})$ and the elements in $\mathscr{H}$ are such that $h_{\boldsymbol{x}}(\cdot) = G_{0,\boldsymbol{x}}^{-1}(H_{\boldsymbol{x}}(\cdot))$, with $B^{-1}(\cdot \mid a, b)$ being the inverse CDF of a beta distribution with parameters (a, b), then marginally

$G_{\boldsymbol{x}}$ follows a Bernstein-Dirichlet prior with parameters $(\lambda, M_{\boldsymbol{x}}, G_{0,\boldsymbol{x}})$, for every $\boldsymbol{x} \in \mathscr{X}$, that is

$$G_{\boldsymbol{x}} \mid \lambda, M_{\boldsymbol{x}}, G_{0\boldsymbol{x}} \sim \mathrm{BDP}(\lambda, M_{\boldsymbol{x}}, G_{0\boldsymbol{x}}),$$

where $F_{\boldsymbol{x}}(\cdot)$ stands for the CDF of the marginal distribution of $\eta_i(\boldsymbol{x}, \cdot)$, for every $i \in \mathbb{N}$, $M_{\boldsymbol{x}} \in \mathbb{R}_0^+ = [0, +\infty)$, $G_{0\boldsymbol{x}}^{-1}$ is the inverse CDF of a probability measure defined on $(0, 1]$ and $H_{\boldsymbol{x}}(\cdot)$ stands for the CDF of the marginal distribution of $z_i(\boldsymbol{x}, \cdot)$, for every $i \in \mathbb{N}$.

Under the same assumptions, it also follows that, for any given $k \in \mathbb{N}$,

$$E\left\{G_{\boldsymbol{x}}(B_y) \mid k\right\} = \sum_{j=0}^{k} G_{0\boldsymbol{x}}(j/k) \frac{k!}{(k-j)!j!} y^j (1-y)^{k-j}, \tag{3.4}$$

and

$$Var\left\{G_{\boldsymbol{x}}(B_y) \mid k\right\} = \frac{1}{M_{\boldsymbol{x}}} \left\{ \sum_{j=0}^{k} c(j,k,y)^2 \varepsilon(j,k,\boldsymbol{x}) - \left(\sum_{j=0}^{k} c(j,k,y) \varepsilon(j,k,\boldsymbol{x}) \right)^2 \right\}, \tag{3.5}$$

where $B_y = [0, y]$, $c(j,k,y) = \sum_{l=j}^{k} \frac{k!}{(k-l)!l!} y^l (1-y)^{k-l}$, $\varepsilon(0,k,\boldsymbol{x}) = 0$, and $\varepsilon(j,k,\boldsymbol{x}) = G_{0\boldsymbol{x}}(j/k) - G_{0\boldsymbol{x}}((j-1)/k)$.

3.3.2 The association structure and continuity of the process

The characteristics of the stochastic processes used in Definition 3.1 determine important properties of the resulting DBPP. Natural choices for longitudinal or spatial modeling are appropriate Gaussian processes. Regardless of the specific choice, the use of almost surely (a.s.) continuous stochastic processes ensures that the DBPP is a.s. continuous from the left and has a limit from the right. The following theorem is proved in the Section B.1 (Appendix B).

Theorem 3.1. *Let* $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \sim \mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \mathscr{V}, \mathscr{H})$. *If for every* $j \in \mathbb{N}$, *the stochastic processes* η_j *and* z_j *are* P*-a.s. continuous, then for every* $\{\boldsymbol{x}_j\}_1^\infty \subset \mathscr{X}$, *such that*

$\lim_{j\to+\infty} \boldsymbol{x}_j \longrightarrow \boldsymbol{x}_0 \in \mathscr{X}$ *and* $x_{jl} \leq x_{0l}$, $l = 1, \ldots, p$,

$$\lim_{j\to+\infty} \sup_{B\in\mathscr{B}([0,1])} |G_{\boldsymbol{x}_j}(B) - G_{\boldsymbol{x}_0}(B)| = 0, \ P\text{-a.s.},$$

that is, $G_{\boldsymbol{x}_j}$ *converges* P*-a.s. in total variation norm to* $G_{\boldsymbol{x}_0}$*, when* $\boldsymbol{x}_j \longrightarrow \boldsymbol{x}_0^-$*. In addition, for every* $\{\boldsymbol{x}_j\}_1^\infty \subset \mathscr{X}$*, such that* $\lim_{j\to+\infty} \boldsymbol{x}_j \longrightarrow \boldsymbol{x}_0 \in \mathscr{X}$ *and* $x_{jl} \geq x_{0l}$, $l = 1, \ldots, p$*, there exists a random probability measure on* $([0,1], \mathscr{B}([0,1]))$, $\tilde{G}_{\boldsymbol{x}_0}$*, such that*

$$\lim_{j\to+\infty} \sup_{B\in\mathscr{B}([0,1])} |G_{\boldsymbol{x}_j}(B) - \tilde{G}_{\boldsymbol{x}_0}(B)| = 0, \ P\text{-a.s.},$$

that is, $G_{\boldsymbol{x}_j}$ *converges* P*-a.s. in total variation norm to* $\tilde{G}_{\boldsymbol{x}_0}$*, when* $\boldsymbol{x}_j \longrightarrow \boldsymbol{x}_0^+$*.*

The association structure of DBPP is completely determined by the dependence structure of the stochastic processes used in Definition 3.1. General analytical expressions for the correlation function are not possible to derive because they depend on the specific laws of the associated stochastic processes. However, we show that, under mild conditions on the stochastic processes defining the DBPP, the correlation between the corresponding random measures approaches to one as the predictor values get closer. The following theorem is proved in the Section B.1 (Appendix B).

Theorem 3.2. *Let* $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \sim \mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \mathscr{V}, \mathscr{H})$*. If for every* $\{\boldsymbol{x}_j\}_1^\infty$*, with* $\boldsymbol{x}_j \in \mathscr{X}$*, such that* $\lim_{j\to+\infty} \boldsymbol{x}_j \longrightarrow \boldsymbol{x}_0 \in \mathscr{X}$*, we have* $z_i(\boldsymbol{x}_j, \cdot) \xrightarrow{\mathscr{L}} z_i(\boldsymbol{x}_0, \cdot)$ *and* $\eta_i(\boldsymbol{x}_j, \cdot) \xrightarrow{\mathscr{L}} \eta_i(\boldsymbol{x}_0, \cdot)$*, as* $j \longrightarrow +\infty$*, then, for all* $y \in (0,1)$*,*

$$\lim_{j\to+\infty} \rho\left[G(\boldsymbol{x}_j, \cdot)(B_y), G(\boldsymbol{x}_0, \cdot)(B_y)\right] = 1,$$

where $\rho(A, B)$ *denotes the Pearson correlation between* A *and* B*, and* $B_y = [0, y]$*.*

If the stochastic processes defining the DBPP 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 on the centering distributions of the DBPP the correlation between the corresponding random measures can approach to zero. The following theorem, proved in the Section B.1 (Appendix B), shows that under the assumptions previously discussed for the DBPP, 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 Bernstein polynomial.

Theorem 3.3. *Let* $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \sim \mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \mathscr{V}, \mathscr{H})$. *Asumme 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* $Cov\left[\mathbb{I}_{\{z_i(\boldsymbol{x}_1,\cdot)\in A_1\}}, \mathbb{I}_{\{z_i(\boldsymbol{x}_2,\cdot)\in A_2\}}\right] = 0$ *and* $Cov\left[\mathbb{I}_{\{\eta_i(\boldsymbol{x}_1,\cdot)\in A_3\}}, \mathbb{I}_{\{\eta_i(\boldsymbol{x}_2,\cdot)\in A_4\}}\right] = 0$*, for every* $A_1, A_2, A_3, A_4 \in \mathscr{B}(\mathbb{R})$*. Assume also that for every* $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathscr{X}$ *such that* $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| > \gamma$*, and for every sequence* $\{(\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j})\}_1^\infty \subset \mathscr{X}^2$*, such that* $\lim_{j\to+\infty}(\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j}) = (\boldsymbol{x}_1, \boldsymbol{x}_2)$*, we have* $(z_i(\boldsymbol{x}_{1j}, \cdot), z_i(\boldsymbol{x}_{2j}, \cdot)) \xrightarrow{\mathscr{L}} (z_i(\boldsymbol{x}_1, \cdot), z_i(\boldsymbol{x}_2, \cdot))$ *and* $(\eta_i(\boldsymbol{x}_{1j}, \cdot), \eta_i(\boldsymbol{x}_{2j}, \cdot)) \xrightarrow{\mathscr{L}} (\eta_i(\boldsymbol{x}_1, \cdot), \eta_i(\boldsymbol{x}_2, \cdot))$*, as* $j \longrightarrow +\infty$*. Then, for every* $y \in [0,1]$*,*

$$\begin{aligned}\lim_{j\to+\infty} Cov\left[G(\boldsymbol{x}_{1j}, \cdot)(B_y), G(\boldsymbol{x}_{2j}, \cdot)(B_y)\right] &= \\ Cov\left[\sum_{l=1}^{k(\cdot)} G^*_{0\boldsymbol{x}_1}\left(A_{l,k(\cdot)}\right) \mathrm{BIN}(l \mid k(\cdot), y),\ \sum_{l=1}^{k(\cdot)} G^*_{0\boldsymbol{x}_2}\left(A_{l,k(\cdot)}\right) \mathrm{BIN}(l \mid k(\cdot), y)\right]\end{aligned}$$

where $B_y = [0, y]$, $A_{l,k} = [0, l/k]$, $G^*_{0\boldsymbol{x}}$ *stands for the marginal probability measure of* $\theta_i(\boldsymbol{x}, \cdot)$ *and* $\mathrm{BIN}(\ \cdot \mid k, y)$ *stands for the probability mass function of the binomial distribution with parameters* (k, y).

Remark 3.1. *It is easy to see that if the* DBPP *is specified such that the marginal distribution of* k *is degenerated, then the correlation between the corresponding random measures goes to zero, since* $\lim_{j\to+\infty} Cov\left[G(\boldsymbol{x}_{1j}, \cdot)(B_y), G(\boldsymbol{x}_{2j}, \cdot)(B_y)\right] = 0$.

Remark 3.2. *If the* DBPP *is specified such that* $G^*_{0\boldsymbol{x}_1} = G^*_{0\boldsymbol{x}_2}$*, then*

$$\lim_{j\to+\infty} Cov\left[G(\boldsymbol{x}_{1j}, \cdot)(B_y), G(\boldsymbol{x}_{2j}, \cdot)(B_y)\right] \geq 0,$$

since

$$\lim_{j\to+\infty} Cov\left[G(\boldsymbol{x}_{1j},\cdot)(B_y), G(\boldsymbol{x}_{2j},\cdot)(B_y)\right] = Var\left[\sum_{l=1}^{k(\cdot)} G^*_{0\boldsymbol{x}_1}\left(A_{l,k(\cdot)}\right)\mathrm{BIN}(l \mid k(\cdot), y)\right].$$

Remark 3.3. *If $G^*_{0\boldsymbol{x}_1}$ or $G^*_{0\boldsymbol{x}_2}$ is the $U(0,1)$ distribution, then*

$$\lim_{j\to+\infty} Cov\left[G(\boldsymbol{x}_{1j},\cdot)(B_y), G(\boldsymbol{x}_{2j},\cdot)(B_y)\right] = 0,$$

since

$$\sum_{l=1}^{k(\cdot)} G^*_{0\boldsymbol{x}_t}\left(A_{l,k(\cdot)}\right)\mathrm{BIN}(l \mid k(\cdot), y) = \sum_{j=1}^{k(\cdot)} \frac{j}{k(\cdot)}\begin{pmatrix} k(\cdot) \\ j \end{pmatrix} y^j (1-y)^{k(\cdot)-j} = y,$$

which is constant as a function of k for $t = 1$ or 2, and every $y \in [0,1]$.

Although the trajectories of the DBPP are a.s. continuous from the left only, its autocorrelation function is continuous under mild conditions on the stochastic processes defining the DBPP. The following theorem is proved in the Section B.1 (Appendix B).

Theorem 3.4. *Let $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \sim \mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \mathscr{V}, \mathscr{H})$. Assume that for every $\{(\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j})\}_1^\infty \subset \mathscr{X}^2$, such that $\lim_{j\to+\infty}(\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j}) = (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathscr{X}^2$, we have that $(z_i(\boldsymbol{x}_{1j},\cdot), z_i(\boldsymbol{x}_{2j},\cdot)) \xrightarrow{\mathscr{L}} (z_i(\boldsymbol{x}_1,\cdot), z_i(\boldsymbol{x}_2,\cdot))$ and $(\eta_i(\boldsymbol{x}_{1j},\cdot), \eta_i(\boldsymbol{x}_{2j},\cdot)) \xrightarrow{\mathscr{L}} (\eta_i(\boldsymbol{x}_1,\cdot), \eta_i(\boldsymbol{x}_2,\cdot))$, as $j \longrightarrow +\infty$. Then, for every $y \in [0,1]$,*

$$\lim_{j\to\infty} \rho\left[G(\boldsymbol{x}_{1j},\cdot)(B_y), G(\boldsymbol{x}_{2j},\cdot)(B_y)\right] = \rho\left[G(\boldsymbol{x}_1,\cdot)(B_y), G(\boldsymbol{x}_2,\cdot)(B_y)\right],$$

where $B_y = [0, y]$.

3.3.3 The support of the process

Large support is an important and basic property that any Bayesian nonparametric model should ideally possess. In fact, assigning positive mass to neighborhoods of any collection of probability distributions $\{F_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$ is a minimum requirement (and almost a "necessary" property) for a model to be considered "nonparametric". This property is also important because it is typically a required condition for frequentist consistency of the posterior distribution. As is widely known, the definition of the support of probability models on functional spaces strongly depends on the choice of a "distance" defining the basic neighborhoods. Therefore, it is first necessary to make explicit the topology under consideration. The results presented here are based on generalizations of standard topologies for spaces of single probability measures. Specifically, we consider the weak product topology, L_∞ product topology and L_∞ topology.

A sub-base of the weak product topology for the space $\mathscr{P}\left([0,1]\right)^{\mathscr{X}} = \prod_{\boldsymbol{x}\in\mathscr{X}} \mathscr{P}\left([0,1]\right)$, is given by sets of the form $B^{\mathrm{W}}_{f,\epsilon,\boldsymbol{x}_0}\left(\{F_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}\right) = \prod_{\boldsymbol{x}\in\mathscr{X}} \Delta^{\mathrm{W}}_{f,\epsilon,\boldsymbol{x}_0}\left(F_{\boldsymbol{x}}\right)$, where

$$\Delta^{\mathrm{W}}_{f,\epsilon,\boldsymbol{x}_0}\left(F_{\boldsymbol{x}}\right) = \begin{cases} \mathscr{P}([0,1]), & \text{if } \boldsymbol{x} \in \mathscr{X}, \boldsymbol{x} \neq \boldsymbol{x}_0, \\ \left\{P_{\boldsymbol{x}} \in \mathscr{P}([0,1]) : \left|\int_{[0,1]} f dP - \int_{[0,1]} f dF_{\boldsymbol{x}}\right| < \epsilon\right\}, & \text{if } \boldsymbol{x} \in \mathscr{X}, \boldsymbol{x} = \boldsymbol{x}_0, \end{cases}$$

for every continuous and bounded function f, $\boldsymbol{x}_0 \in \mathscr{X}$ and $\epsilon > 0$. The following theorem provides sufficient conditions for $\mathscr{P}\left([0,1]\right)^{\mathscr{X}}$ to be the support of the DBPP under the weak product topology, that is, it provides sufficient conditions under which $\mathscr{P}\left([0,1]\right)^{\mathscr{X}}$ is the smallest closet set of $P \circ \mathscr{G}^{-1}$-measure one under the weak product topology. The proof of the theorem is given in the Section B.1 (Appendix B).

Theorem 3.5. *Let* $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \sim \mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \mathscr{V}, \mathscr{H})$*. If for every* $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d) \in \mathscr{X}^d$*,* $d \geq 1$*, the joint distribution of* $(\eta_i(\boldsymbol{x}_1, \cdot), \ldots, \eta_i(\boldsymbol{x}_d, \cdot))$ *and* $(z_i(\boldsymbol{x}_1, \cdot), \ldots, z_i(\boldsymbol{x}_d, \cdot))$ *have full support on* $\mathbb{R}^d$*, and* $k(\cdot)$ *has full support on* $\mathbb{N}$*, then* $\mathscr{P}\left([0,1]\right)^{\mathscr{X}}$ *is the support of* $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$ *under the weak product topology.*

Let $\mathscr{D}\left([0,1]\right) \subset \mathscr{P}\left([0,1]\right)$ be the set of all probability measures defined on $\left([0,1], \mathscr{B}\left([0,1]\right)\right)$ that are absolutely continuous w.r.t. Lebesgue measure and with continuous density function on $[0,1]$. A sub-base of the L_∞ product topology for the space $\mathscr{D}\left([0,1]\right)^{\mathscr{X}} = \prod_{\boldsymbol{x}\in\mathscr{X}} \mathscr{D}\left([0,1]\right)$, is given by sets of the form $B^{L_\infty}_{\epsilon,\boldsymbol{x}_0}(\{F_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}) = \prod_{\boldsymbol{x}\in\mathscr{X}} \Delta^{L_\infty}_{\epsilon,\boldsymbol{x}_0}(F_{\boldsymbol{x}})$, where

$$\Delta^{L_\infty}_{\epsilon,\boldsymbol{x}_0}(F_{\boldsymbol{x}}) = \begin{cases} \mathscr{D}([0,1]), & \text{if } \boldsymbol{x} \in \mathscr{X}, \boldsymbol{x} \neq \boldsymbol{x}_0, \\ \left\{P_{\boldsymbol{x}} \in \mathscr{D}([0,1]) : \sup_{y\in[0,1]} |p_{\boldsymbol{x}}(y) - f_{\boldsymbol{x}}(y)| < \epsilon\right\}, & \text{if } \boldsymbol{x} \in \mathscr{X}, \boldsymbol{x} = \boldsymbol{x}_0, \end{cases}$$

where $p_{\boldsymbol{x}}$ and $f_{\boldsymbol{x}}$ denote the densities of $P_{\boldsymbol{x}}$ and $F_{\boldsymbol{x}}$ w.r.t. Lebesgue measure, respectively. The following theorem shows that, under the same assumptions of Theorem 3.5, $\mathscr{D}([0,1])^{\mathscr{X}}$ is the support of the DBPP under the L_∞ product support. The proof of the following theorem is provided in the Section B.1 (Appendix B).

Theorem 3.6. *Let $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \sim \mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \mathscr{V}, \mathscr{H})$. If for every $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d \in \mathscr{X}^d$, $d \geq 1$, the joint distributions of $(\eta_i(\boldsymbol{x}_1, \cdot), \ldots, \eta_i(\boldsymbol{x}_d, \cdot))$ and $(z_i(\boldsymbol{x}_1, \cdot), \ldots, z_i(\boldsymbol{x}_d, \cdot))$ have full support on $\mathbb{R}^d$, and $k(\cdot)$ has full support on $\mathbb{N}$, then $\mathscr{D}([0,1])^{\mathscr{X}}$ is the support of $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$ under the L_∞ product topology.*

If stronger assumptions on the predictor space $\mathscr{X}$ and the parameter space are imposed, a stronger support property (L_∞) can be obtained. Specifically, assume that the predictor space $\mathscr{X}$ is a compact set and consider the sub-space $\tilde{\mathscr{D}}([0,1])^{\mathscr{X}} \subset \mathscr{D}([0,1])^{\mathscr{X}}$, where

$$\tilde{\mathscr{D}}([0,1])^{\mathscr{X}} = \left\{\{F_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \mathscr{D}([0,1])^{\mathscr{X}} : (y, \boldsymbol{x}) \longrightarrow f_{\boldsymbol{x}}(y) \text{ is continuous}\right\},$$

with $f_{\boldsymbol{x}}$ denoting the density of $F_{\boldsymbol{x}} \in \mathscr{D}([0,1])$ w.r.t. Lebesgue measure. A base of the L_∞ topology for the space $\mathscr{D}([0,1])^{\mathscr{X}} = \prod_{\boldsymbol{x}\in\mathscr{X}} \mathscr{D}([0,1])$, is given by sets of the form

$$B^{L_\infty}_{\epsilon}(\{F_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}) = \left\{\{P_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \mathscr{D}([0,1])^{\mathscr{X}} : \sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{y\in[0,1]} |p_{\boldsymbol{x}}(y) - f_{\boldsymbol{x}}(y)| < \epsilon\right\},$$

where $\epsilon > 0$ and, for every $\boldsymbol{x} \in \mathscr{X}$, $p_{\boldsymbol{x}}$ and $f_{\boldsymbol{x}}$ denote the densities of $P_{\boldsymbol{x}}$ and $F_{\boldsymbol{x}}$ w.r.t. Lebesgue measure, respectively. The following theorem, proved in the Section B.1 (Appendix B), provides sufficient conditions for $\widetilde{\mathscr{D}}([0,1])^{\mathscr{X}}$ to be the support of the DBPP under the L_∞ topology.

Theorem 3.7. *Let $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \sim \mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \mathscr{V}, \mathscr{H})$. If $\mathscr{X}$ is a compact set, $k(\cdot)$*

has full support on $\mathbb{N}$, and the processes used in the definition of the DBPP are such that, for any $[0,1]$-valued continuous function defined on $\mathscr{X}$, f, and $\epsilon > 0$, we have that

$$P\left\{\omega \in \Omega : \sup_{\boldsymbol{x} \in \mathscr{X}} |v_{\boldsymbol{x}}(\eta_i(\boldsymbol{x}, \omega)) - f(\boldsymbol{x})| < \epsilon\right\} > 0,$$

and

$$P\left\{\omega \in \Omega : \sup_{\boldsymbol{x} \in \mathscr{X}} |h_{\boldsymbol{x}}(z_i(\boldsymbol{x}, \omega)) - f(\boldsymbol{x})| < \epsilon\right\} > 0,$$

then $\widetilde{\mathscr{D}}\left([0,1]\right)^{\mathscr{X}}$ is the support of $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$ under the L_∞ topology.

An important consequence of the previous theorem is that the DBPP can assign positive mass to arbitrarily small neighborhoods of any collection of probability measures $\{Q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \tilde{\mathscr{D}}\left([0,1]\right)^{\mathscr{X}}$, based on the supremum over the predictor space of Kullback-Leibler (KL) divergences between the predictor-dependent probability measures. The following corollary is proved in the Section B.1 (Appendix B).

Corollary 3.1. *Let $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \sim \mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \mathscr{V}, \mathscr{H})$. Assume that $\mathscr{X}$ is a compact set, $k(\cdot)$ has full support on $\mathbb{N}$, and that the processes used in the definition of the DBPP are such that, for any $\epsilon > 0$ and $[0,1]$-valued continuous function f defined on $\mathscr{X}$, we have*

$$P\left\{\omega \in \Omega : \sup_{\boldsymbol{x} \in \mathscr{X}} |v_{\boldsymbol{x}}(\eta_i(\boldsymbol{x}, \omega)) - f(\boldsymbol{x})| < \epsilon\right\} > 0,$$

and

$$P\left\{\omega \in \Omega : \sup_{\boldsymbol{x} \in \mathscr{X}} |h_{\boldsymbol{x}}(z_i(\boldsymbol{x}, \omega)) - f(\boldsymbol{x})| < \epsilon\right\} > 0.$$

Then,

$$P\left\{\omega \in \Omega : \sup_{\boldsymbol{x} \in \mathscr{X}} \int_0^1 q_{\boldsymbol{x}}(y) \log\left(\frac{q_{\boldsymbol{x}}(y)}{g(\boldsymbol{x}, \omega)(y)}\right) dy < \epsilon\right\} > 0,$$

for every $\epsilon > 0$ and every $\{Q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \tilde{\mathscr{D}}([0,1])$, with density functions $\{q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$.

3.3.4 The asymptotic behavior of the posterior distribution

Let Q be the true probability measure generating the predictors, with density w.r.t. a corresponding σ-additive measure denoted by q. Suppose that the response variable and predictors are drawn independently from a probability distribution of the form $m_0(y, \boldsymbol{x}) = q(\boldsymbol{x})q_0(y \mid \boldsymbol{x})$, where $q_0(y \mid \boldsymbol{x})$ denotes a fixed conditional density on $[0,1]$, $\boldsymbol{x} \in \mathscr{X}$. Let $m^{(\cdot)}(y, \boldsymbol{x}) = q(\boldsymbol{x})g(\boldsymbol{x}, \cdot)(y)$ be the random joint distribution for the response and predictors arising when $g(\boldsymbol{x}, \cdot)(y)$ is given by (3.3). Since the KL divergence between m_0 and a realization $m^{(\omega)}$ of the implied joint distribution under the DBPP can be bounded by the supremum over the predictor space of KL divergences between the predictor-dependent probability measures,

$$
\begin{aligned}
\mathrm{KL}(m_0, m^{(\omega)}) &= \int_{\mathscr{X}} \int_{[0,1]} m_0(y, \boldsymbol{x}) \log\left(\frac{m_0(y, \boldsymbol{x})}{m^{(\omega)}(y, \boldsymbol{x})}\right) dy d\boldsymbol{x}, \\
&= \int_{\mathscr{X}} q(\boldsymbol{x}) \int_{[0,1]} q_0(y \mid \boldsymbol{x}) \log\left(\frac{q_0(y \mid \boldsymbol{x})}{g(\boldsymbol{x}, \omega)(y)}\right) dy d\boldsymbol{x}, \\
&\leq \sup_{\boldsymbol{x} \in \mathscr{X}} \int_{[0,1]} q_0(y \mid \boldsymbol{x}) \log\left(\frac{q_0(y \mid \boldsymbol{x})}{g(\boldsymbol{x}, \omega)(y)}\right) dy,
\end{aligned}
$$

it follows that, for every $\delta > 0$,

$$
\begin{aligned}
P\left\{\omega \in \Omega : \sup_{\boldsymbol{x} \in \mathscr{X}} \int_{[0,1]} q_0(y \mid \boldsymbol{x}) \log\left(\frac{q_0(y \mid \boldsymbol{x})}{g(\boldsymbol{x}, \omega)(y)}\right) dy < \delta\right\}& \\
\geq P\left\{\omega \in \Omega : \mathrm{KL}(m_0, m^{(\omega)}) < \delta\right\},& \\
> 0,&
\end{aligned}
$$

under the assumptions of Theorem 3.7 and Corollary 3.1. Thus, by Schwartz's theorem (Schwartz, 1965) it follows that the posterior distribution associated with the random joint distribution induced by the DBPP model is weakly consistent, that is, the posterior measure of any weak neighborhood, of any joint distribution of the form $m_0(y, \boldsymbol{x}) = q(\boldsymbol{x})q_0(y \mid \boldsymbol{x})$, converges to 1 as the sample size goes to infinity. This result is summarized in the following theorem.

Theorem 3.8. *Let* $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \sim \mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \mathscr{V}, \mathscr{H})$. *Assume that* $\mathscr{X}$ *is a compact set,* $k(\cdot)$ *has full support on* $\mathbb{N}$, *and that the processes used in the definition of the DBPP are* *such that, for any $\epsilon > 0$ and $[0,1]$-valued continuous function f defined on $\mathscr{X}$, we have*

$$P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} |v_{\boldsymbol{x}}(\eta_i(\boldsymbol{x},\omega)) - f(\boldsymbol{x})| < \epsilon\right\} > 0,$$

and

$$P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} |h_{\boldsymbol{x}}(z_i(\boldsymbol{x},\omega)) - f(\boldsymbol{x})| < \epsilon\right\} > 0.$$

Then the posterior distribution associated with the random joint distribution induced by the DBPP model, $m^{(\cdot)}(y,\boldsymbol{x}) = q(\boldsymbol{x})g(\boldsymbol{x},\cdot)(y)$, where q is the density generating the predictors, is weakly consistent at any joint distribution of the form $m_0(y,\boldsymbol{x}) = q(\boldsymbol{x})q_0(y \mid \boldsymbol{x})$, where $\{q_0(\cdot \mid \boldsymbol{x}) : \boldsymbol{x} \in \mathscr{X}\} \in \widetilde{\mathscr{D}}\left([0,1]\right)^{\mathscr{X}}$.

3.4 Simplified versions of the general model

In the search of parsimonious models, it is of interest to know whether simplified versions of the general model class proposed in the previous section retain most of its appealing properties. In this section we study two simplifications of the general model class, by considering dependent-stick breaking processes where only the support points or only the weights are indexed by the predictors.

3.4.1 The $w\mathrm{DBPP}$

We first consider the case where the dependence in the probability measures with bounded support involves a dependent stick-breaking process with common weights across probability measures, and support points given by stochastic processes indexed by predictors $\boldsymbol{x} \in \mathscr{X}$. The resulting process is referred to as 'single weights' DBPP and denoted by $w\mathrm{DBPP}$.

Definition 3.2. *Let $\mathscr{H}$ be a set of functions as before. Let $\mathscr{G} = \{G(\boldsymbol{x},\omega) : \boldsymbol{x} \in \mathscr{X}\}$ be a $\mathscr{P}\left([0,1]\right)$-valued stochastic process on an appropriate probability space $(\Omega, \mathscr{A}, P)$ such that:*

(i) $v_1, v_2, \ldots$ are independent random variables of the form $v_i : \Omega \longrightarrow [0,1]$, $i \geq 1$, and with common distribution indexed by a finite-dimensional parameter α.

(ii) $z_1, z_2, \ldots$, *are independent and identically distributed real-valued stochastic processes of the form* $z_i : \mathscr{X} \times \Omega \longrightarrow \mathbb{R}$, $i \geq 1$, *with law indexed by a finite-dimensional parameter* $\mathbf{\Psi}_2$ *and marginal distributions* $\{H_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$.

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

(iv) For every $\boldsymbol{x} \in \mathscr{X}$ *and almost every* $\omega \in \Omega$, *the density function of* $G(\boldsymbol{x}, \omega)$, *w.r.t. Lebesgue measure, is given by a common-weights dependent mixture of beta densities,*

$$g(\boldsymbol{x}, \omega)(\cdot) = \sum_{j=1}^{\infty} w_j(\omega) \beta\left(\cdot \mid \lceil k(\omega)\theta_j(\boldsymbol{x}, \omega)\rceil, k(\omega) - \lceil k(\omega)\theta_j(\boldsymbol{x}, \omega)\rceil + 1\right),$$

where $\lceil \cdot \rceil$ *denotes the ceiling function,* $\theta_j(\boldsymbol{x}, \omega) = h_{\boldsymbol{x}}(z_j(\boldsymbol{x}, \omega))$, *and*

$$w_j(\omega) = v_j(\omega) \prod_{i<j} \left[1 - v_i(\omega)\right].$$

The process $\mathscr{G} = \{G_{\boldsymbol{x}} \doteq G(\boldsymbol{x}, \omega) : \boldsymbol{x} \in \mathscr{X}\}$ *will be referred to as 'single-weights' dependent Bernstein polynomial process with parameters* $(\alpha, \lambda, \mathbf{\Psi}_2, \mathscr{H})$, *and denoted by* $w\mathrm{DBPP}(\alpha, \lambda, \mathbf{\Psi}_2, \mathscr{H})$.

As shown in the Section B.2 (Appendix B), under equivalent assumptions on the parameters defining the process, the 'single weights' DBPP retains all of the properties shown for the general version of the model,

3.4.2 The $\theta\mathrm{DBPP}$

We now consider the case where the dependence in the probability measures is introduced via the use of dependent stick-breaking processes with common support points across probability measures, and weights corresponding to stochastic processes indexed by predictors $\boldsymbol{x} \in \mathscr{X}$. The resulting process is referred to as 'single atoms' DBPP and denoted by $\theta\mathrm{DBPP}$.

Definition 3.3. *Let $\mathscr{V}$ be a set of functions as before. Let $\mathscr{G} = \{G(\boldsymbol{x},\omega) : \boldsymbol{x} \in \mathscr{X}\}$ be a $\mathscr{P}([0,1])$-valued stochastic process on an appropriate probability space $(\Omega, \mathscr{A}, P)$ such that:*

(i) $\eta_1, \eta_2, \ldots$*, are independent and identically distributed real-valued stochastic processes of the form* $\eta_i : \mathscr{X} \times \Omega \longrightarrow \mathbb{R}$*,* $i \geq 1$*, with law indexed by a finite-dimensional parameter* $\boldsymbol{\Psi}_1$ *and marginal distributions* $\{F_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$*.*

(ii) $\theta_1, \theta_2, \ldots$*, are independent random variables of the form* $\theta_i : \Omega \longrightarrow [0,1]$*,* $i \geq 1$*, and with common distribution* G_0*.*

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

(iv) For every $\boldsymbol{x} \in \mathscr{X}$ *and almost every* $\omega \in \Omega$*, the density function of* $G(\boldsymbol{x},\omega)$*, w.r.t. Lebesgue measure, is given by a dependent mixture of beta densities,*

$$g(\boldsymbol{x},\omega)(\cdot) = \sum_{j=1}^{\infty} w_j(\boldsymbol{x},\omega)\beta\left(\cdot \mid \lceil k(\omega)\theta_j(\omega)\rceil, k(\omega) - \lceil k(\omega)\theta_j(\omega)\rceil + 1\right), \tag{3.6}$$

where $\lceil \cdot \rceil$ *denotes the ceiling function and*

$$w_j(\boldsymbol{x},\omega) = v_{\boldsymbol{x}}\left\{\eta_j(\boldsymbol{x},\omega)\right\} \prod_{i<j} \left[1 - v_{\boldsymbol{x}}\left\{\eta_i(\boldsymbol{x},\omega)\right\}\right].$$

The process $\mathscr{G} = \{G_{\boldsymbol{x}} \doteq G(\boldsymbol{x},\omega) : \boldsymbol{x} \in \mathscr{X}\}$ *will be referred to as 'single-atoms' dependent Bernstein polynomial process with parameters* $(\lambda, \boldsymbol{\Psi}_1, \mathscr{V}, G_0)$*, and denoted by* $\theta\mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_1, \mathscr{V}, G_0)$*.*

As shown in the Section B.3 (Appendix B), the properties of the 'single atoms' DBPP have some interesting differences with the general model class. On the one hand, the $\theta\mathrm{DBPP}$ has full support under the three topologies we considered, and its posterior distribution is also weakly consistent. In addition, the correlation of corresponding random measures has identical behavior when the predictor values get close, and the correlation function is also continuous as a function of the predictors. On the other hand, however, the correlation between the associated random measures when the predictor values are far apart reaches a different limit, and it is difficult to establish conditions on the prior specification ensuring that this limit is zero. Another interesting property of the $\theta\mathrm{DBPP}$ compared to the general model class is that the use of a.s. continuous stochastic processes in the weights guarantees a.s. continuity of the 'single atoms' DBPP (and not only from the left). The following theorem is proved in the Section B.4 (Appendix B).

Theorem 3.9. *Let $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \sim \theta\mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_1, \mathscr{V}, G_0)$. Assume that for every $j \in \mathbb{N}$, the stochastic process η_j is P-a.s. continuous. Then, for every $\{\boldsymbol{x}_j\}_1^\infty \subset \mathscr{X}$, such that $\lim_{j\rightarrow+\infty} \boldsymbol{x}_j \longrightarrow \boldsymbol{x}_0 \in \mathscr{X}$,*

$$\lim_{j\rightarrow+\infty} \sup_{B\in\mathscr{B}([0,1])} |G_{\boldsymbol{x}_j}(B) - G_{\boldsymbol{x}_0}(B)| = 0,\ P\text{-a.s.},$$

for every $\boldsymbol{x}_0 \in \mathscr{X}$, that is, $G_{\boldsymbol{x}_j}$ converges P-a.s. in total variation norm to $G_{\boldsymbol{x}_0}$, as $\boldsymbol{x}_j \longrightarrow \boldsymbol{x}_0$.

3.5 Illustrations

We illustrate the behavior of the models with simulated and real-life data. In these illustrations we consider special cases of the general models, where the stochastic processes used in the definition of the DBPP correspond to Gaussian processes arising from linear (in the coefficients) regression models, with random and normally distributed coefficients. The computational implementation of the models is based on MCMC methods. The MCMC algorithms can be based on a finite dimensional approximation of the dependent sitck-breaking process, or on the use of the slice sampler (Walker, 2007) or the retrospective sampler algorithm (Papaspiliopoulos & Roberts, 2008). A full description of the MCMC implementation used here is given in Section B.5 (Appendix B). User-friendly functions implementing these methods were written in compiled language and incorporated into the R library DPpackage (Jara, 2007; Jara et al., 2011).

3.5.1 Simulated data

To illustrate the performance of the proposed models and to compare them to the existing methods, nine simulated data sets were generated; one for each of three different scenarios and three sample sizes ($n = 250$, $n = 500$ and $n = 1,000$). In all cases, a single continuous covariate x was considered, with values generated from the $U(0,1)$ distribution. The three different scenarios are given in Table 3.1. They represent varying degrees of complexity and shapes as x varies in the predictor space. All models exhibit a multi-modal behavior. The conditional distributions for Scenario I have a bi-modal behavior for low values of the predictor, and the modes merge as the predictor value increases. The conditional distributions for Scenario II have positive density at 1, while $f(y \mid x) \to 0$ as $y \to 0$, for every $x \in (0,1)$. Finally, the conditional distributions for Scenario III have positive density at 0 and 1, for every $x \in (0,1)$. Additionally, a central mode is also present, and the density value at the mode increases as the value of the predictor increases.

Table 3.1: Simulated data: True models.

Scenario	Conditional density
I	$f(y \mid x) = 0.5 \times \text{Beta}(y \mid 20, 1.1 + 20x) + 0.5 \times \text{Beta}(y \mid 1.1 + 5x, 5)$.
II	$f(y \mid x) = 0.5 \times \text{Beta}(y \mid 20, 1.1 + 20(x + 0.27)) + 0.5 \times \text{Beta}(y \mid 1.1 + 5(x + 0.27), 1)$.
III	$f(y \mid x) = 0.3 \times \text{Beta}(y \mid 1, 10) + 0.5 \times \text{Beta}(y \mid 1.1 + 20x, 8) + 0.2 \times \text{Beta}(y \mid 10, 1)$.

Particular cases of the general models were considered by assuming

$$v_x(\cdot) = h_x(\cdot) = \exp\{\cdot\}/\left(1 + \exp\{\cdot\}\right),$$

for every $x \in (0,1)$. Furthermore, we considered Gaussian processes (GP) in the definition of the models, by exploiting the connection between GP and linear models. Specifically, we assume that $\eta_i(x,\omega) = \mathbf{d}_\eta(x)^T \boldsymbol{\gamma}_i^\eta(\omega)$ and $\boldsymbol{\gamma}_i^\eta(\cdot) \mid \boldsymbol{\mu}^\eta, \mathbf{S}^\eta \overset{iid}{\sim} N_{r_1}\left(\boldsymbol{\mu}^\eta, \mathbf{S}^\eta\right)$, and that $z_i(x,\omega) = \mathbf{d}_z(x)^T \boldsymbol{\gamma}_i^z(\omega)$ and $\boldsymbol{\gamma}_i^z(\cdot) \mid \boldsymbol{\mu}^z, \mathbf{S}^z \overset{iid}{\sim} N_{r_2}\left(\boldsymbol{\mu}^z, \mathbf{S}^z\right)$, $i = 1, 2, \ldots$, where $\mathbf{d}_\eta(x)$ and $\mathbf{d}_z(x)$ are r_1- and r_2-dimensional design vectors, respectively, including linear and/or non-linear functions of the predictor x. The corresponding versions of the DBPP using this specification are referred to as linear DBPP (LDBPP), linear wDBPP (wLDBPP) and linear θDBPP (θLDBPP). The model specification was completed by assuming

$$k \mid \lambda \sim \text{Poisson}(\lambda)\mathbb{I}_{\{k>1\}},$$

$$\boldsymbol{\mu}^{\eta} \mid \boldsymbol{m}_0^{\eta}, \mathbf{S}_0^{\eta} \sim N_{r_1}\left(\mathbf{m}_0^{\eta}, \mathbf{S}_0^{\eta}\right), \quad \mathbf{S}^{\eta} \mid \nu^{\eta}, \boldsymbol{\Psi}^{\eta} \sim IW_{r_1}\left(\nu^{\eta}, \boldsymbol{\Psi}^{\eta}\right),$$

$$\boldsymbol{\mu}^{z} \mid \mathbf{m}_0^{z}, \mathbf{S}_0^{z} \sim N_{r_2}\left(\mathbf{m}_0^{z}, \mathbf{S}_0^{z}\right), \quad \mathbf{S}^{z} \mid \nu^{z}, \boldsymbol{\Psi}^{z} \sim IW_{r_2}\left(\nu^{z}, \boldsymbol{\Psi}^{z}\right),$$

$$v_j \mid \alpha \overset{iid}{\sim} \text{Beta}(1, \alpha), \quad \theta_j \mid a, b \overset{iid}{\sim} \text{Beta}(a, b),$$

where $IW_r(\nu, \mathbf{A})$ denotes the r-dimensional inverted-Wishart distribution with degrees of freedom ν and scale matrix $\mathbf{A}$.

Two versions of each linear DBPP model were considered. In version 1, we set $\mathbf{d}_{\eta}(x) = (1, x)^T$ and/or $\mathbf{d}_z(x) = (1, x)^T$. In version 2, random B-splines regression models (see, e.g. Eilers & Marx, 1996; Lang & Brezger, 2004) were considered. In this case, $\mathbf{d}_{\eta}(x) = (1, \psi_1(x), \ldots, \psi_6(x))^T$ and/or $\mathbf{d}_z(x) = (1, \psi_1(x), \ldots, \psi_6(x))^T$, where $\psi_j(x)$ corresponds to the jth B-spline basis function evaluated at x. The models were fit by assuming $\lambda = 25$, $\mathbf{m}_0^{\eta} = \mathbf{0}_{r_1}$, $\mathbf{m}_0^{z} = \mathbf{0}_{r_2}$, $\mathbf{S}_0^{\eta} = 2.25 \times \mathbf{I}_{r_1}$, $\mathbf{S}_0^{z} = 2.25 \times \mathbf{I}_{r_2}$, $\nu^{\eta} = r_1 + 2$, $\nu^{z} = r_2 + 2$ and $\alpha = a = b = 1$. For each simulated dataset, one Markov chain was generated completing a conservative total number of 110,000 scans of the Markov chain cycle described in Appendix E of the supplementary material. Standard tests (not shown), as implemented in the BOA R library (Smith, 2007), suggested convergence of the chains. Because of storage limitations, the full chain was subsampled every 10 iterations, after a burn-in period of 10,000 samples, to give a reduced chain of length 10,000.

For comparison purposes, we considered the linear dependent Dirichlet process (LDDP) of De Iorio et al. (2004, 2009) and the weight dependent Dirichlet process (WDDP) of Müller et al. (1996). For the approach of Müller et al. (1996), we consider the multivariate extension of the univariate Dirichlet process mixture of normals model of Escobar & West (1995) to fit the complete transformed data $\boldsymbol{w}_i = (\log(y_i/(1-y_i)), x_i)^T$, and focus on the conditional densities $f(y \mid x)$ arising from the model. The Dirichlet process mixture model is given by

$$\boldsymbol{w}_i \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i \overset{ind.}{\sim} N_2\left(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right), \quad (\boldsymbol{\mu}, \boldsymbol{\Sigma}_i) \mid Q_1 \overset{iid}{\sim} Q_1, \quad Q_1 \mid M_1, Q_{01} \sim DP\left(M_1, Q_{01}\right),$$

where the baseline distribution Q_{01} is the conjugate normal-inverted-Wishart (IW) distribution $Q_{01} \equiv N_2\left(\boldsymbol{\mu} \mid \mathbf{m}_1, \kappa_0^{-1}\boldsymbol{\Sigma}\right) IW_2\left(\boldsymbol{\Sigma} \mid \nu_1, \boldsymbol{\Psi}_1\right)$. To complete the model specification, the following hyper-priors were assumed: $M_1 \mid a_{01}, b_{01} \sim \Gamma\left(a_{01}, b_{01}\right)$, $\mathbf{m}_1 \mid \mathbf{m}_2, \mathbf{S}_2 \sim N_2\left(\mathbf{m}_2, \mathbf{S}_2\right)$, $\kappa_0 \mid \tau_1, \tau_2 \sim \Gamma\left(\tau_1/2, \tau_2/2\right)$, and $\boldsymbol{\Psi}_1 \mid \nu_2, \boldsymbol{\Psi}_2 \sim IW_2\left(\nu_2, \boldsymbol{\Psi}_2\right)$. The LDDP, on the other hand, can be represented as Dirichlet process mixture of linear (in the coefficients) regression models

$$\log(y_i/(1-y_i)) \mid \boldsymbol{\gamma}_i, \sigma_i^2 \overset{ind.}{\sim} N\left(\mathbf{d}(x_i)^T\boldsymbol{\gamma}_i, \sigma_i^2\right),$$

$$\left(\boldsymbol{\gamma}_i, \sigma_i^2\right) \mid Q_2 \overset{iid}{\sim} Q_2, \quad Q_2 \mid M_2, Q_{02} \sim DP\left(M_2, Q_{02}\right),$$

where $\mathbf{d}(x)$ is a r_3-dimensional design vector, respectively, including linear and/or non-linear functions of the predictor x, and $Q_{02} \equiv N_{r_3}\left(\boldsymbol{\gamma} \mid \boldsymbol{\mu}_{\boldsymbol{\gamma}}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\right) \Gamma\left(\sigma^{-2} \mid s_1/2, s_2/2\right)$. The LDDP model specification is completed with the following hyper-priors: $M_2 \mid a_{02}, b_{02} \sim \Gamma\left(a_{02}, b_{02}\right)$, $s_2 \mid \tau_{s_1}, \tau_{s_2} \sim \Gamma(\tau_{s_1}/2, \tau_{s_2}/2)$, $\boldsymbol{\mu}_{\boldsymbol{\gamma}} \mid \mathbf{a}, \mathbf{A} \sim N_{r_3}\left(\mathbf{a}, \mathbf{A}\right)$, and $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} \mid \nu_{\boldsymbol{\gamma}}, \boldsymbol{\Psi}_{\boldsymbol{\gamma}} \sim IW_{r_3}\left(\nu_{\boldsymbol{\gamma}}, \boldsymbol{\Psi}_{\boldsymbol{\gamma}}\right)$. Marginalized versions of Dirichlet process-based models were fit, where Q_1 and Q_2 are integrated out, using standard algorithms to fit Dirichlet process mixture models. Credible intervals for the conditional densities in this case were obtained from MCMC samples using the $\epsilon-$DP approach proposed by Muliere & Tardella (1998), with $\epsilon = 0.01$. Two versions of the LDDP were considered. Model LDDP1 corresponds to a mixture of linear regression models, that is, $\mathbf{d}(x) = (1, x)^T$. Model LDDP2 corresponds to a mixture of B-splines regression models, where $\mathbf{d}(x) = (1, \psi_1(x), \ldots, \psi_6(x))^T$. The MCMC specification was similar to the DBPP models and the prior specification was as in Jara et al. (2011).

The discrepancy between estimated, $\widehat{f}\left(\cdot \mid x\right)$, and true model, $f\left(\cdot \mid x\right)$, was measured using an estimate to the L_∞ $\left(\widehat{L}_\infty\right)$ distance,

$$\widehat{L}_\infty = \max_{l} \max_{m} \left| \widehat{f}\left(y_m \mid x_l\right) - f\left(y_m \mid x_l\right) \right|,$$

which is based on grid of equally-spaced values of the response $\{y_m\}_1^M$ and of the predictor $\{x_l\}_1^L$. In addition, we also considered the estimate to the integrated-L_1 distance $\left(\widehat{IL}_1\right)$, given by

$$\widehat{IL}_1 = \frac{1}{M}\frac{1}{L}\sum_{l=1}^{L}\sum_{m=1}^{M} \left| \widehat{f}\left(y_m \mid x_l\right) - f\left(y_m \mid x_l\right) \right|,$$

Table 3.2 shows the values for $\widehat{L}_\infty$ and $\widehat{IL}_1$ for each model, scenario and sample size. The results indicate that the best version of our model outperformed the competitors for every scenario and sample size, using both the $\widehat{L}_\infty$ and $\widehat{IL}_1$ criteria. As expected, behavior of the models was similar under Scenario I, the least problematic for the competitors of our proposed model, because there is no boundary problem. However, the number of versions of the proposed model outperforming the competitors tends to increase with the sample size; for $n = 1,000$, three out of six versions of the proposed model outperformed the competitors under the $\widehat{L}_\infty$ and $\widehat{IL}_1$ criteria. When the boundary problem was present (Scenarios II and III), 5 or 6 (out of 6) of the versions of the proposed model outperformed the competitors using the most demanding criteria; the $\widehat{L}_\infty$ value for the worst competitor was as high as 72 times the corresponding value for the best version of the proposed model.

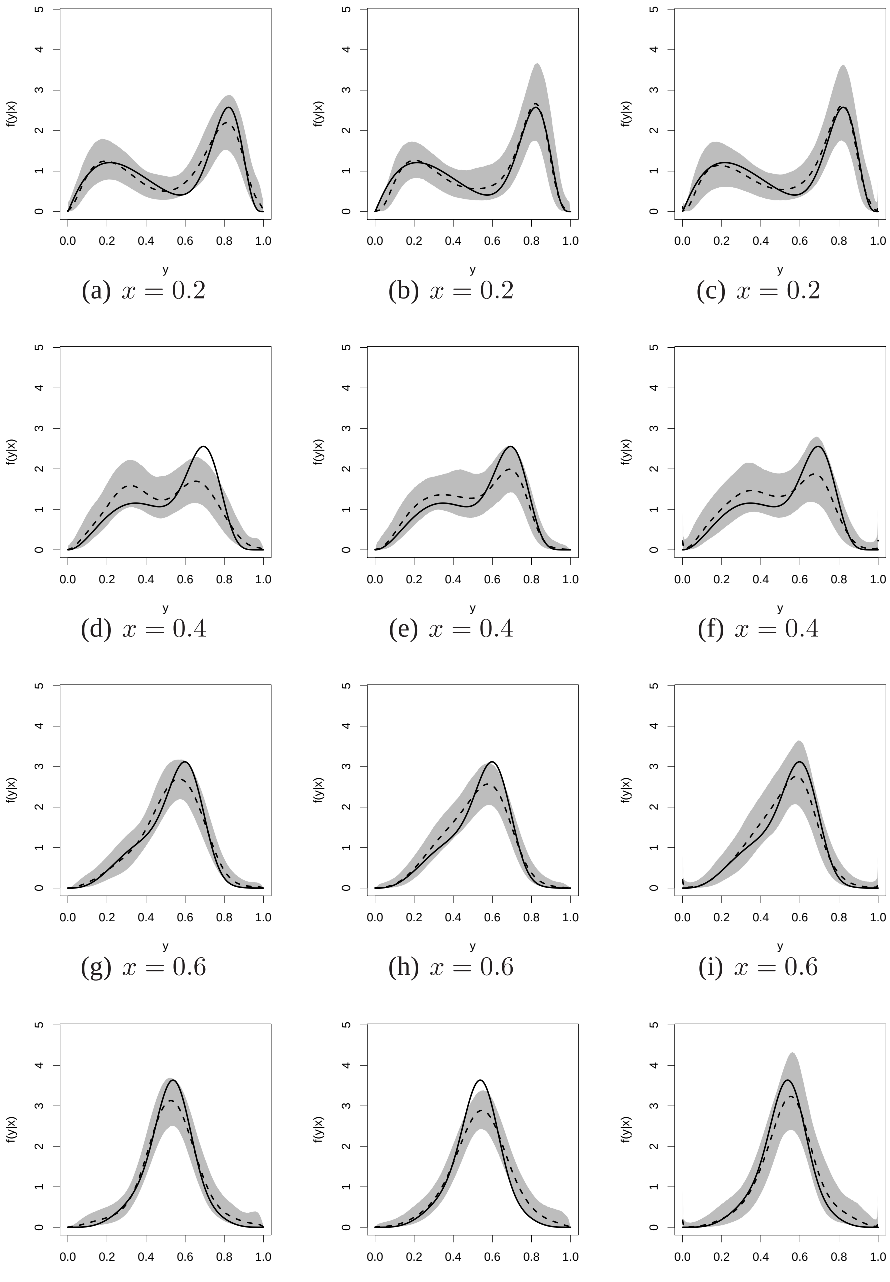
The posterior inferences for the conditional densities showed that for each scenario, sample size and version of the proposed model, the estimates correspond approximately to the true densities. In most of the cases, the true model was completely covered by 95% point-wise highest probability density (HPD) bands, and the quality of the estimation improved as the sample size increases. Under Scenarios II and III, poor results were obtained using the LDDP and WDDP models. Indeed, the density estimates diverged substantially from the true densities at the extremes of the support, confirming that these models are not suitable for this type of behavior. Figures 3.1, 3.2 and 3.3 illustrate these findings. They show, for the sample size $n = 500$, the predictive density, evaluated in a grid of size 200 at four values of the predictor for

Table 3.2: Simulated data: Estimated L_∞ (integrated L_1) for each model, under the different simulation scenarios and sample sizes.

	Simulation Scenario								
	I			II			III		
Model	$n = 250$	$n = 500$	$n = 1000$	$n = 250$	$n = 500$	$n = 1000$	$n = 250$	$n = 500$	$n = 1000$
LDBPP1	6.19	6.57	6.92	1.09	2.83	0.89	2.78	3.25	2.76
	(0.20)	(0.16)	(0.13)	(0.16)	(0.17)	(0.11)	(0.24)	(0.16)	(0.13)
LDBPP2	4.59	6.44	3.80	2.76	3.59	3.43	5.38	2.18	2.73
	(0.27)	(0.19)	(0.15)	(0.24)	(0.20)	(0.15)	(0.35)	(0.18)	(0.17)
wLDBPP1	6.30	7.50	7.57	1.01	1.39	1.91	2.44	2.41	2.98
	(0.24)	(0.23)	(0.21)	(0.16)	(0.16)	(0.12)	(0.25)	(0.22)	(0.17)
wLDBPP2	7.25	8.33	4.48	4.09	1.89	2.45	7.40	1.86	2.44
	(0.27)	(0.23)	(0.17)	(0.23)	(0.18)	(0.12)	(0.42)	(0.18)	(0.15)
θLDBPP1	6.28	6.26	6.34	1.48	1.33	1.06	2.95	2.48	2.73
	(0.25)	(0.19)	(0.16)	(0.27)	(0.20)	(0.17)	(0.31)	(0.23)	(0.17)
θLDBPP2	3.91	6.10	6.77	1.93	3.16	2.63	3.44	4.27	1.88
	(0.28)	(0.18)	(0.19)	(0.27)	(0.21)	(0.14)	(0.33)	(0.22)	(0.19)
LDDP1	6.66	7.53	7.43	19.27	4.62	2.89	26.14	3.48	3.74
	(0.24)	(0.22)	(0.20)	(0.20)	(0.17)	(0.12)	(0.38)	(0.24)	(0.19)
LDDP2	34.62	6.38	22.65	59.48	15.22	26.12	138.29	103.31	60.54
	(0.32)	(0.20)	(0.18)	(0.28)	(0.25)	(0.14)	(0.58)	(0.24)	(0.20)
WDDP	4.52	6.86	6.68	28.79	9.15	3.05	170.06	8.01	4.75
	(0.26)	(0.17)	(0.17)	(0.34)	(0.27)	(0.13)	(0.50)	(0.24)	(0.20)

the best version of the proposed model and LDDP model, according to the $\widehat{L}_\infty$ criteria, and the WDDP model. The results for the remaining sample sizes are given in Section B.6 (Appendix B).

We note that these results are for one random sample from particular models, and conclusions should be drawn carefully. Nonetheless, these examples do show that the class of DBPP models is highly flexible and that misleading results can be obtained by using transformations of the data along with flexible models for data defined on the real line.



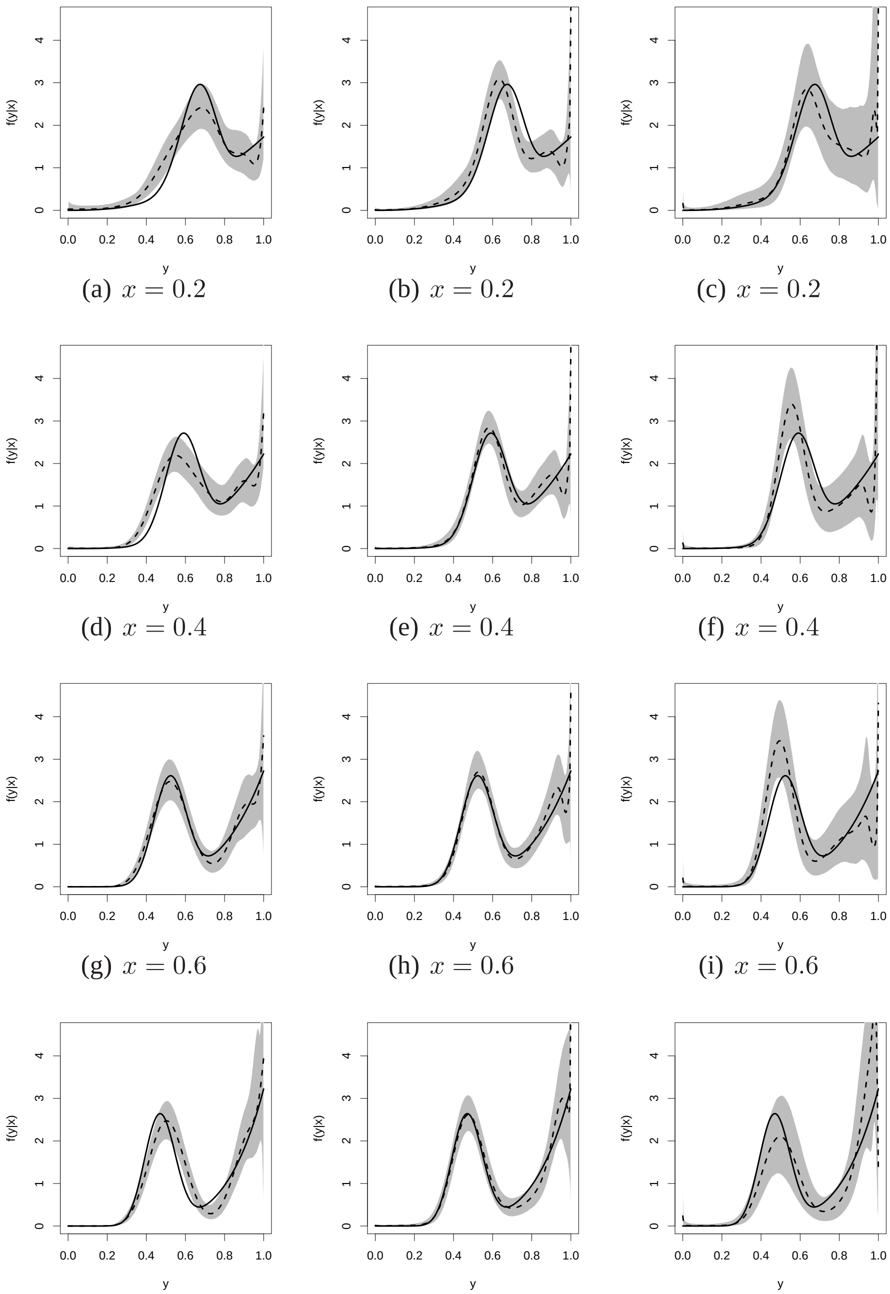
(a) $x = 0.2$ (b) $x = 0.2$ (c) $x = 0.2$

(d) $x = 0.4$ (e) $x = 0.4$ (f) $x = 0.4$

(g) $x = 0.6$ (h) $x = 0.6$ (i) $x = 0.6$

(j) $x = 0.8$ (k) $x = 0.8$ (l) $x = 0.8$

Figure 3.1: Simulated data - Scenario I ($n = 500$): True (continuous line) and posterior mean (dotted line) for the conditional density. A band constructed using the 95%point-wise HPD intervals is presented in gray. Panels (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l) display the results for the best DBPP model (θDBPP2), the best LDDP model (LDDP2), both regarding the estimated L_∞ distance, and the weight dependent DP for four values of the predictor, respectively.



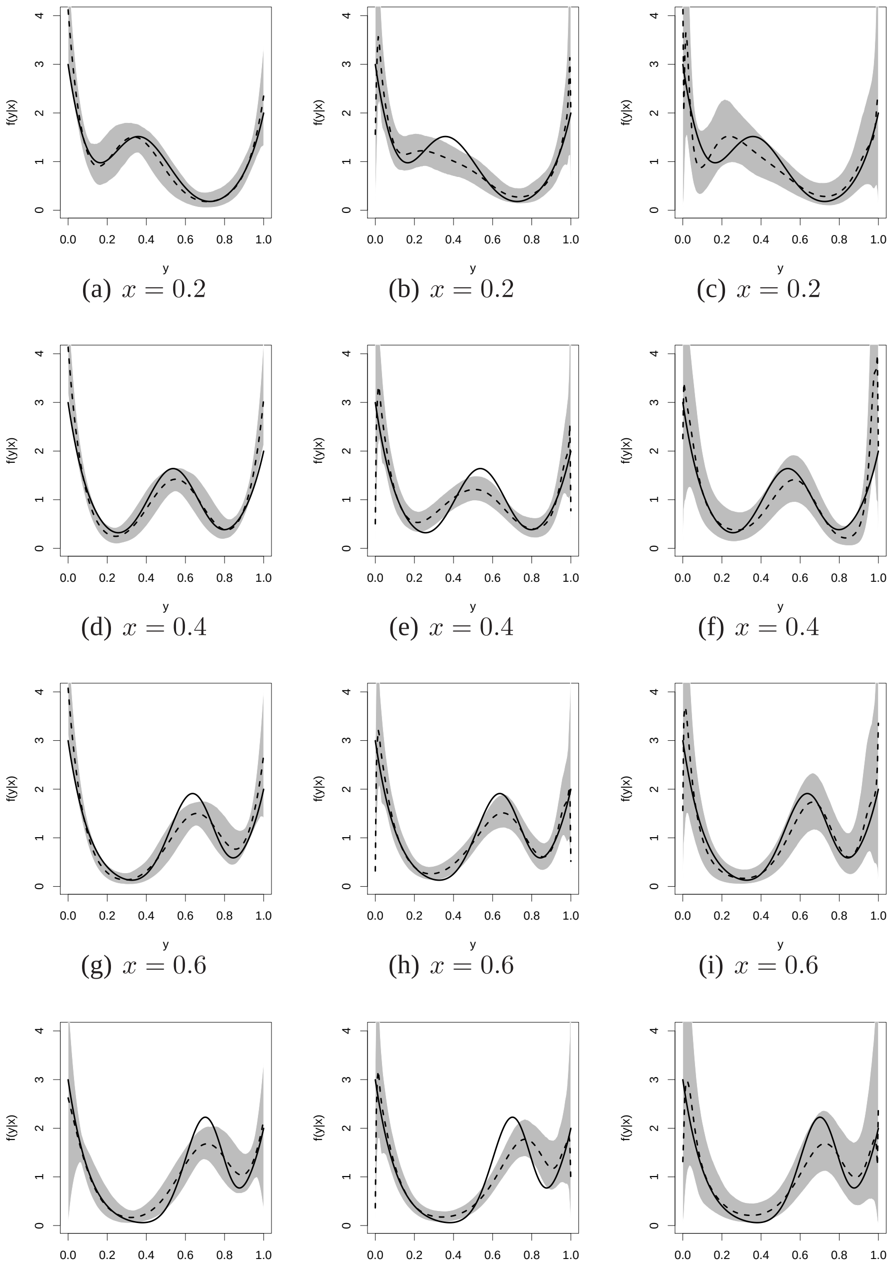
(a) $x = 0.2$ (b) $x = 0.2$ (c) $x = 0.2$

(d) $x = 0.4$ (e) $x = 0.4$ (f) $x = 0.4$

(g) $x = 0.6$ (h) $x = 0.6$ (i) $x = 0.6$

(j) $x = 0.8$ (k) $x = 0.8$ (l) $x = 0.8$

Figure 3.2: Simulated data - Scenario II ($n = 500$): True (continuous line) and posterior mean (dotted line) for the conditional density. A band constructed using the 95%point-wise HPD intervals is presented in gray. Panels (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l) display the results for the best DBPP model (θDBPP1), the best LDDP model (LDDP1), both regarding the estimated L_∞ distance, and the weight dependent DP for four values of the predictor, respectively.



(a) $x = 0.2$ (b) $x = 0.2$ (c) $x = 0.2$

(d) $x = 0.4$ (e) $x = 0.4$ (f) $x = 0.4$

(g) $x = 0.6$ (h) $x = 0.6$ (i) $x = 0.6$

(j) $x = 0.8$ (k) $x = 0.8$ (l) $x = 0.8$

Figure 3.3: Simulated data - Scenario III ($n = 500$): True (continuous line) and posterior mean (dotted line) for the conditional density. A band constructed using the 95%point-wise HPD intervals is presented in gray. Panels (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l) display the results for the best DBPP model (wDBPP2), the best LDDP model (LDDP1), both regarding the estimated L_∞ distance, and the weight dependent DP for four values of the predictor, respectively.

3.5.2 Solid waste data

We consider data about residentially generated solid waste in the city of Santiago de Cali, Colombia. The dataset contains information about 258 block sides and was collected to estimate the per capita daily production and characterization of solid waste in the city. The solid waste in each of the 258 block sides was separated in different kinds of materials, including food, hygienic waste, glass, metal and plastic. The proportions of these materials were registered for each block side. In addition, the socio-economic level of the houses associated to each block side was registered. The socioeconomic status was grouped in an ordinal scale of six levels: low-low, low, medium-low, medium, medium-high and high. We refer the reader to Klinger et al. (2009) for more details about these data.

The proportion of food and hygienic waste were considered as response variables. In both cases, the socio-economic level was used as a discrete predictor. As in the previous section, linear approximations to the general models were fit to the data, by assuming $\lambda = 25$, $\mathbf{m}_0^\eta = \mathbf{m}_0^z = \mathbf{0}_6$, $\mathbf{S}_0^\eta = \mathbf{S}_0^z = 2.25 \times \mathbf{I}_6$, $\nu^\eta = \nu^z = 8$ and $\alpha = a = b = 1$. For each model, one Markov chain was generated completing a conservative total number of 110,000 scans of the Markov chain cycle described in Appendix E of the supplementary material. Standard tests (not shown), as implemented in the BOA R library (Smith, 2007), suggested convergence of the chains. Because of storage limitations, the full chain was subsampled every 10 iterations, after a burn-in period of 10,000 samples, to give a reduced chain of length 10,000.

For comparison purposes, the parametric beta regression model, originally proposed by Ferrari & Cribari-Neto (2004) and later extended by Simas et al. (2010), was also fit to the data. The beta regression model proposed by Simas et al. (2010) is given by

$$y_i \mid \boldsymbol{x}_i, \boldsymbol{\gamma}^\mu, \boldsymbol{\gamma}^\phi \overset{ind}{\sim} \text{Beta}\left\{\mu\left(\boldsymbol{x}_i^T\boldsymbol{\gamma}^\mu\right)\phi\left(\boldsymbol{x}_i^T\boldsymbol{\gamma}^\phi\right), \left[1-\mu\left(\boldsymbol{x}_i^T\boldsymbol{\gamma}^\mu\right)\right]\phi\left(\boldsymbol{x}_i^T\boldsymbol{\gamma}^\phi\right)\right\},$$

where $\mu\left(\boldsymbol{x}_i^T\boldsymbol{\gamma}^\mu\right) = \left[1+\exp\left(-\boldsymbol{x}_i^T\boldsymbol{\gamma}^\mu\right)\right]^{-1}$ and $\phi\left(\boldsymbol{x}_i^T\boldsymbol{\gamma}^\phi\right) = \exp\left(\boldsymbol{x}_i^T\boldsymbol{\gamma}^\phi\right)$. The model specification was completed by assuming

$$\boldsymbol{\gamma}^\mu \mid m^\mu, \tau^\mu \sim N_6\left(m^\mu \times \mathbf{1}_6, \tau^\mu \times \mathbf{I}_6\right),$$

$$\boldsymbol{\gamma}^{\phi} \mid m^{\phi}, \tau^{\phi} \sim N_6 \left(m^{\phi} \times \mathbf{1}_6, \tau^{\phi} \times \mathbf{I}_6 \right),$$

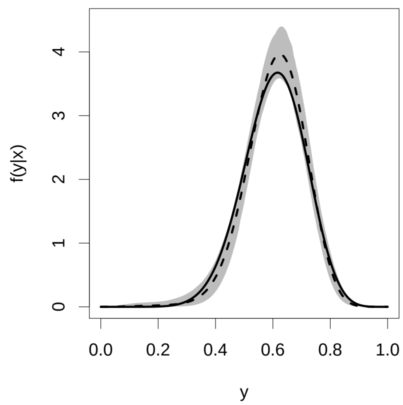
$$m^{\mu} \mid m_0^{\mu}, s_0^{\mu} \sim N(m_0^{\mu}, s_0^{\mu}), \;\; \tau^{\mu} \mid \tau_0^{\mu} \sim U(0, \tau_0^{\mu}),$$

$$m^{\phi} \mid m_0^{\phi}, s_0^{\phi} \sim N(m_0^{\phi}, s_0^{\phi}), \;\; \tau^{\phi} \mid \tau_0^{\phi} \sim U(0, \tau_0^{\phi}),$$

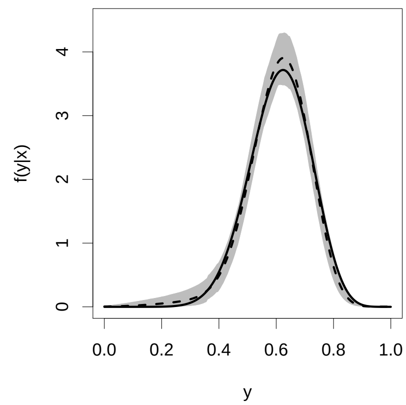
where $m_0^{\mu} = m_0^{\phi} = 0$, $s^{\mu} = s^{\phi} = 2.25$ and $\tau_0^{\mu} = \tau_0^{\phi} = 10$. Model comparison was performed using the log pseudo marginal likelihood (LPML), developed by Geisser & Eddy (1979) and further considered by Gelfand & Dey (1994). The log pseudo marginal likelihood for model M is defined as $\text{LPML}_M = \sum_{i=1}^{n} \log p_M \left(y_i \mid \mathbf{y}^{[-i]} \right)$, where $p_M \left(y_i \mid \mathbf{y}^{[-i]} \right)$ is the posterior predictive distribution for observation y_i, based on the data $\mathbf{y}^{[-i]}$, under model M, with $\mathbf{y}^{[-i]}$ being the observed data vector after removing the ith observation. Models with larger LPML values are to be preferred. The individual cross-validation predictive densities, known as conditional predictive ordinates (CPO), were also used. The CPOs measure the influence of individual observations and are often used as predictive model checking tools. The method suggested by Gelfand & Dey (1994) was used to obtain estimates of CPO statistics from the MCMC output.

For the proportion of food the three versions of the DBPP model behaved in a similar manner and outperformed the parametric beta regression model using the LPML criteria. The LPML values were 213.09, 212.12 and 215.26 for the LDBPP, wLDBPP and θLDBPP models, respectively. The LPML for the parametric model was 205.3. The conditional density estimates were similar across DBPP models, in agreement with what we previously found using the LPML criterion. More importantly, substantial differences between the DBPP models and the parametric beta regression model were observed, and the disagreement increases with the socioeconomic level in the corresponding ordinal scale. Figure 3.4 displays the results for the θLDBPP model for the six socioeconomic level. The results for the remaining DBPP models are given in Section B.7 (Appendix B).

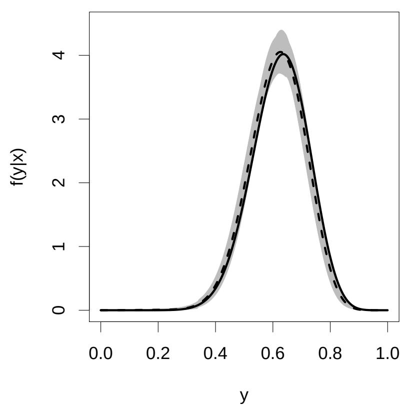
For the proportion of hygienic waste data, the DBPP models showed again a similar behavior regarding both, LPML and the posterior inference on conditional densities. The LPML



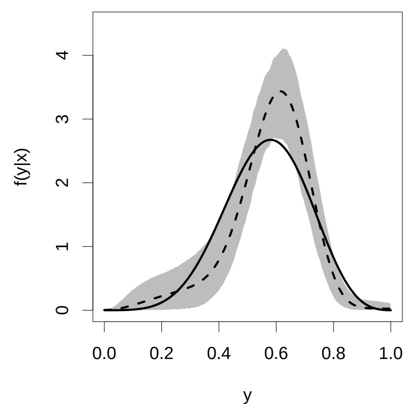
(a) Socioeconomic level: low-low



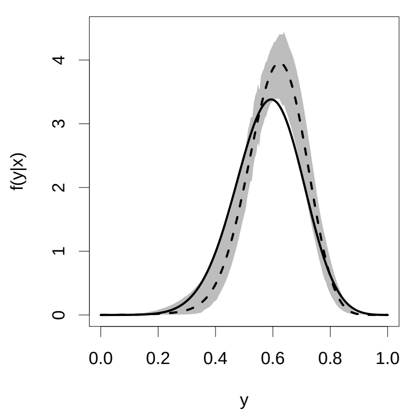
(b) Socioeconomic level: low



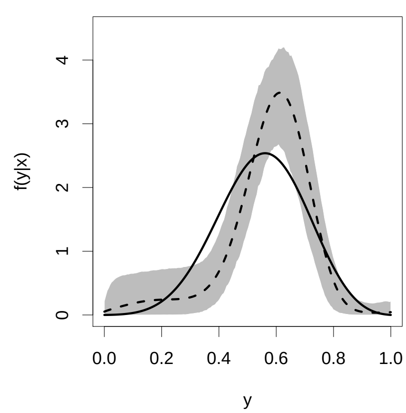
(c) Socioeconomic level: medium-low



(d) Socioeconomic level: medium



(e) Socioeconomic level: medium-high



(f) Socioeconomic level: high

Figure 3.4: Proportion of food - θLDBPP model. Panels (a), (b), (c), (d), (e) and (f) display the posterior mean (dashed line) and a 95% point-wise HPD band (grey area) for the conditional density at socioeconomic level low-low, low, medium-low, medium, medium-high and high, respectively, under the θLDBPP model. The posterior mean under the parametric beta regression model is given as a solid line for comparison purposes.

values were 420.78, 422.72 and 421.92 for LDBPP, wLDBPP and θLDBPP models, respectively. Figure 3.5 show the posterior inferences for the conditional densities at the different socioeconomic levels, under the θLDBPP model. The results for the remaining DBPP models are given in in Section B.8 (Appendix B). The results clearly show an important departure from the beta assumption. Specifically, the positive density at zero and the existence of a central mode observed for socioeconomic levels low-low, low, medium and medium-low cannot be obtained from a beta model.

The positive density observed at zero for the proportion of hygienic waste can be explained by the existence of zero values in the dataset. In fact, because of that, we were not able to fit the beta regression model to these data; the beta distribution is not always well defined at zero or one. A possible solution would be to consider a constrained parameter space for the model, such as

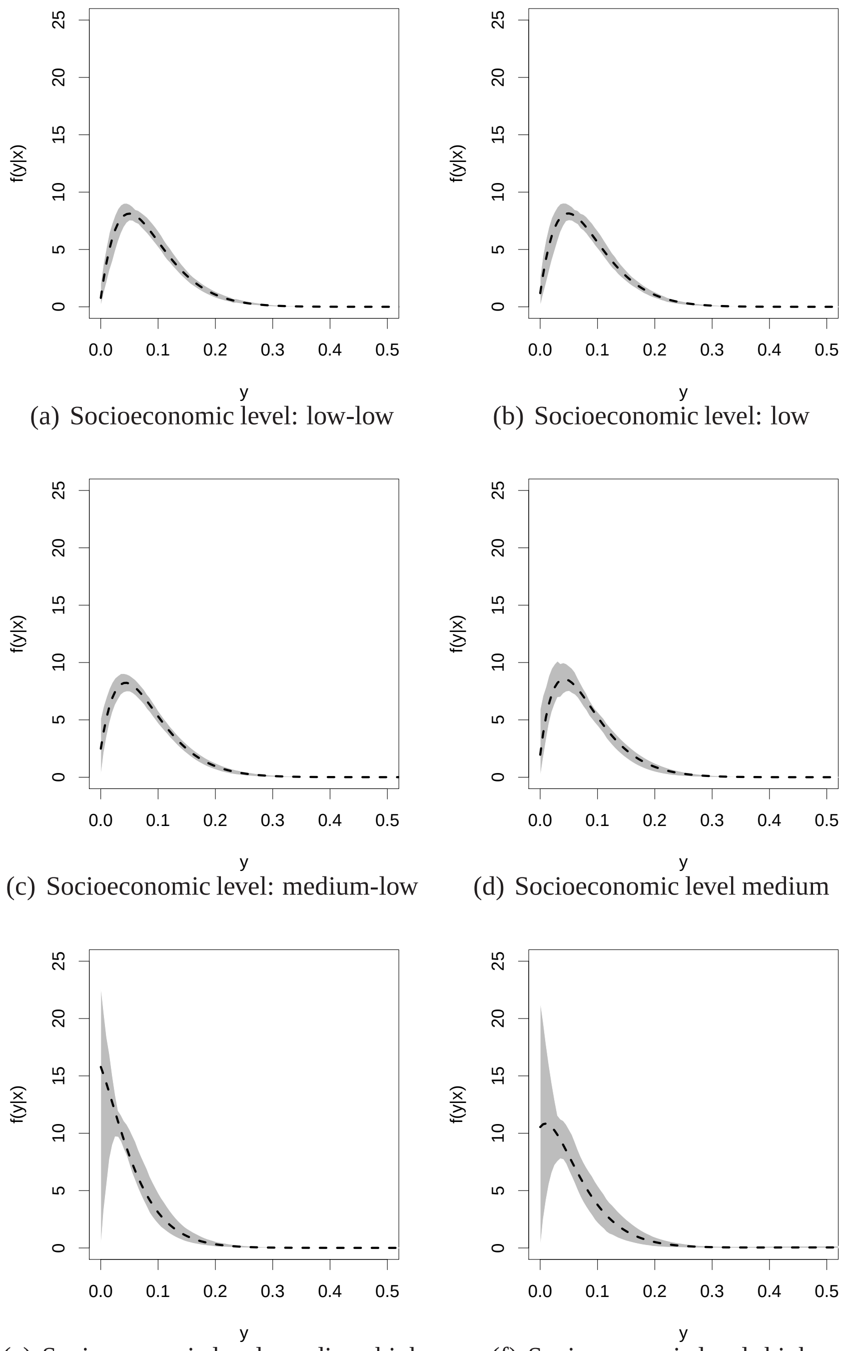
$$\left\{ \left(\boldsymbol{\gamma}^{\mu}, \boldsymbol{\gamma}^{\phi}\right) \in \mathbb{R}^{12} : \mu\left(\boldsymbol{x}^{T}\boldsymbol{\gamma}^{\mu}\right)\phi\left(\boldsymbol{x}^{T}\boldsymbol{\gamma}^{\phi}\right),\ [1-\mu\left(\boldsymbol{x}^{T}\boldsymbol{\gamma}^{\mu}\right)]\phi\left(\boldsymbol{x}^{T}\boldsymbol{\gamma}^{\phi}\right) \geq 1, \forall\, \boldsymbol{x} \in \mathscr{X} \right\}.$$

However, this solution would imply that for every $\boldsymbol{x} \in \mathscr{X}$, the conditional density would be a.s. equal to zero on the extreme values of the domain, which is clearly not supported by the data and we did not pursue that option here. This illustrates another advantage of the proposed class of models, namely that by construction, they are always well defined at every value of the unitary interval.

3.6 Concluding Remarks

We have proposed a novel class of probability models for sets of predictor-dependent probability distributions with bounded domain. The proposal corresponds to an extension of the Dirichlet-Bernstein prior by using dependent stick-breaking processes. The proposed class of models has appealing theoretical properties such as full support, continuity, known marginal distribution, well behaved correlation function, and its posterior distribution is consistent.

By using practicable special cases, the main advantages of the proposed class of models were illustrated using simulated and real-life data. The results suggest that the proposed models



(a) Socioeconomic level: low-low

(b) Socioeconomic level: low

(c) Socioeconomic level: medium-low

(d) Socioeconomic level medium

(e) Socioeconomic level: medium-high

(f) Socioeconomic level: high

Figure 3.5: Proportion of hygienic waste - θLDBPP model. Panels (a), (b), (c), (d), (e) and (f) display the posterior mean (dashed line) and a 95% point-wise HPD band (grey area) for the conditional density at socioeconomic level low-low, low, medium-low, medium, medium-high and high, respectively, under the θLDBPP model.

can outperform Bayesian nonparametric models for responses defined on the real line and the use of transformations, even when the boundary problem is not present. The results also suggest a clear advantage of the proposed class of models over parametric alternatives.

The extension of the class of models for dealing with multivariate bounded responses and/or mixed bounded and unbounded responses is the subject of ongoing research. The extension for response vectors defined on a corresponding simplex is also the subject of ongoing research.

Chapter 4

Conclusions and future work

In this dissertation, we have addressed two different topics in the context of Bayesian nonparametric (BNP) models for predictor–dependent probability measures. In Chapter 2, we studied the property of large support of MacEachern's dependent Dirichlet processes and extensions, using an alternative definition based on copulas. In Chapter 3, we proposed a novel probability model for sets of predictor–dependent probability distributions with bounded domain. This Chapter summarizes the main conclusions of this dissertation and gives some directions of future work.

4.1 Conclusions

Two main focuses have been developed in this dissertation, the main conclusions for each one are described below. In the first part, we focused on the study of the support properties of dependent Dirichlet process (DDP) and DDP mixture models, as well as those of more general dependent stick–breaking processes. The connection between copulas and stochastic processes served to provide sufficient conditions for weak, Hellinger and Kullback–Leibler support of models based on DDP's and general dependent stick–breaking processes. Those conditions were related to the support of the finite–dimensional distributions of the stochastic processes and the kernel used to define the mixture models. We also studied the support of simplified versions of the DDP, in particular, versions where only the weights or only the support points were indexed by the predictors. The results we obtained showed that any of the considered versions of the DDP maintains the large support property. In other words, the use of more parsimonious models does not necessarily imply a reduction of the support. This is an important conclusion since in practice it is more common to use dependent processes where only the weights or only the support points are indexed by the predictors.

The second focus of this dissertation was to propose a novel class of probability models for sets of predictor–dependent probability distributions whose domain is a closed interval. The use of dependent stick–breaking processes allowed to define a new class of dependent processes which extend the Dirichlet–Bernstein prior proposed by Petrone (1999a,b). The proposed process was called dependent Bernstein polynomial (DBPP). We showed that the DBPP satisfies the properties of full support, continuity, known marginal distribution, well behaved correlation function, and consistency of the posterior distribution. An important feature of the DBPP is that its trajectories are collections of densities well-defined on a closed interval. This feature allows the DBPP to be used in applications where the observations belong to a closed and bounded interval, including the case where some of these observations are concentrated in at least one of the edges of the interval. We also considered two simplified versions of the DBPP where only the weights or only the support points were indexed by the predictors. These versions satisfied the same properties as the general case.

Additionally, we showed the advantages of our proposal by applying the DBPP to simulated and real–life data and comparing the results to those obtained with other approaches. The approaches included beta regression models and BNP models for related probability measures defined on the real line. These BNP models were applied by using an appropriate transformation of the data. From the comparisons, one concluded that the proposed models can outperform those BNP and parametric approaches. Although the performance of the proposed model, compared to other approaches, was the best in all the considered scenarios, it was clearly much better in scenarios where some of the observations were concentrated in at least one of the edges of the interval. The DBPP model and its simplified versions turned out to be an attractive non-parametric alternative in the context of regression analysis for bounded data. This model can be easily used since user-friendly functions implementing these methods were written in compiled language and incorporated into the R library DPpackage (Jara, 2007; Jara et al., 2011).

4.2 Future work

The results presented in this dissertation can be applied to different contexts and extended in several directions. Some of the future works derived of this dissertation are described below.

As a extension of Chapter 2, we plan to consider stronger topologies and to study the support of general dependent processes under such topologies. In addition, we also plan to study the support of real–valued proccess which are defined as linear combinations of some design vector of the predictors and where the coefficients of the combinations are assumed random. The motivation here is given by the fact that those kind of processes are commonly used in practice to induce dependence.

The future work derived from Chapter 3 is focused in two different directions. The first one is motivated by educational data. Here, assuming that $T-1$ tests have been previously and sequentially applied to a group of students, the aim is to predict for each student the proportion of correctly answered questions of the T-th test. In order to propose a novel BNP model for such aim, our future plan is to include an autoregressive component in the DBPP following a similar approach to that used by Di Lucca et al. (2012). The second focus is to extend the DBPP model by replacing the predictor–dependent mixtures of beta distributions by predictor–dependent mixtures of Dirichlet distributions. The idea is to develop a BNP model for related probability measures whose density functions, w.r.t. Lebesgue measure, are defined on the k–dimensional simplex spaces, $k \in \mathbb{N}$. This topic is subject of current research.

Appendix A

Supplementary Material for Chapter 2

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

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

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

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

$\theta_1, \ldots, \theta_N \in \Theta$, such that

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

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

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

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

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

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

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

and

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

which implies that

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

Thus,

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

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

Appendix B

Supplementary Material for Chapter 3

B.1 Proofs of theoretical results associated with the DBPP

Proof of Theorem 3.1

Since the elements of $\mathscr{V}$ are continuous functions of $\boldsymbol{x}$ and, for every $j \in \mathbb{N}$, η_j is a P–a.s. continuous stochastic process, it follows that $\boldsymbol{x} \mapsto v_{\boldsymbol{x}}(\eta_j(\boldsymbol{x}, \cdot))$ and $\boldsymbol{x} \mapsto w_j(\boldsymbol{x}, \cdot)$, $j \in \mathbb{N}$, are P–a.s. continuous functions. Similarly, since the elements of $\mathscr{H}$ are continuous functions of $\boldsymbol{x}$ and, for every $j \in \mathbb{N}$, z_j is a P–as continuous stochastic process, it follows that $\boldsymbol{x} \mapsto h_{\boldsymbol{x}}(z_j(\boldsymbol{x}, \cdot))$, $j \in \mathbb{N}$, is a P–a.s. continuous function.

Now, since the ceiling function is continuous from the left and it has a limit from the right, it follows that, for almost every $\omega \in \Omega$ and every $\left\{\boldsymbol{x}_j^{(l)}\right\}_{j=1}^{\infty}$, with $\boldsymbol{x}_j^{(l)} \in \mathscr{X}$, such that $\lim_{j \to +\infty} \boldsymbol{x}_j^{(l)} = \boldsymbol{x}_0 \in \mathscr{X}$ and $x_{jm}^l \leq x_{0m}$, $m = 1, \ldots, p$,

$$\lim_{j \to +\infty} \left\lceil k(\omega)\theta_i\left(\boldsymbol{x}_j^{(l)}, \omega\right)\right\rceil = \left\lceil k(\omega)\theta_i\left(\boldsymbol{x}_0, \omega\right)\right\rceil.$$

Furthermore, for almost every $\omega \in \Omega$ and every $\left\{\boldsymbol{x}_j^{(r)}\right\}_{j=1}^{\infty}$, with $\boldsymbol{x}_j^{(r)} \in \mathscr{X}$, such that $\lim_{j\to+\infty} \boldsymbol{x}_j^{(r)} = \boldsymbol{x}_0 \in \mathscr{X}$ and $x_{jm}^r \geq x_{0m}$, for some $m = 1, \ldots, p$, it follows that

$$\lim_{j\to+\infty} \lceil k(\omega)\theta_i(\boldsymbol{x}_j^{(r)}, \omega)\rceil = \lceil k(\omega)\theta_i(\boldsymbol{x}_0, \omega)\rceil^{(r)} := \begin{cases} j & \text{if} \quad k(\omega)\theta_i(\boldsymbol{x}_0, \omega) \in (j-1, j) \\ j+1 & \text{if} \quad k(\omega)\theta_i(\boldsymbol{x}_0, \omega) = j \end{cases} .$$

Therefore, by the Lebesgue's dominated convergence theorem, it follows that the density w.r.t. Lebesgue measure of $G_{\boldsymbol{x}}$, is P–a.s. continuous from the left and it has a limit from the right, i.e., for every $y \in [0, 1]$,

$$P\left\{\omega \in \Omega : \lim_{j\to+\infty} g(\boldsymbol{x}_j^{(l)}, \omega)(y) = g(\boldsymbol{x}_0, \omega)(y), \ \lim_{j\to+\infty} g(\boldsymbol{x}_j^{(r)}, \omega)(y) = g^{(r)}(\boldsymbol{x}_0, \omega)(y)\right\} = 1,$$

where

$$g^{(r)}(\boldsymbol{x}_0, \omega)(y) = \sum_{i=1}^{\infty} w_j(\boldsymbol{x}_0, \omega)\beta(y|\lceil k(\omega)\theta_i(\boldsymbol{x}_0, \omega)\rceil^{(r)}, k(\omega) - \lceil k(\omega)\theta_i(\boldsymbol{x}_0, \omega)\rceil^{(r)} + 1).$$

Finally, let $G^{(r)}(\boldsymbol{x}_0, \omega)$ be a probability measure with density function $g^{(r)}(\boldsymbol{x}_0, \omega)$. A direct application of Scheffe's theorem implies that

$$P\left\{\omega \in \Omega : \lim_{j\to+\infty} \sup_{B\in\mathscr{B}([0,1])} |G(\boldsymbol{x}_j^{(l)}, \omega)(B) - G(\boldsymbol{x}_0, \omega)(B)| = 0, \right.$$
$$\left. \lim_{j\to+\infty} \sup_{B\in\mathscr{B}([0,1])} |G(\boldsymbol{x}_j^{(r)}, \omega)(B) - G^{(r)}(\boldsymbol{x}_0, \omega)(B)| = 0 \right\} = 1,$$

which completes the proof of the theorem. □

Proof of Theorem 3.2

Notice that for every $y \in [0,1]$ and every $\boldsymbol{x} \in \mathscr{X}$,

$$\begin{aligned} E\left\{G(\boldsymbol{x},\cdot)(B_y) \mid k\right\} &= E\left\{\sum_{l=1}^{k} \left[F^*(\boldsymbol{x},\cdot)(l)\right] \mathrm{BIN}(l \mid k,y) \,\middle|\, k\right\}, \\ &= \sum_{l=1}^{k} E\left\{F^*(\boldsymbol{x},\cdot)\,(l) \;\mid k\right\} \mathrm{BIN}\,(l \mid k,y)\,, \end{aligned}$$

where, $\mathrm{BIN}(\cdot \mid k,y)$ stands for the probability mass function of a binomial distribution with parameters (k,y), and

$$F^*(\boldsymbol{x},\cdot)\,(l) = \sum_{i=1}^{\infty} w_i(\boldsymbol{x},\cdot) I\{\theta_i(\boldsymbol{x},\cdot)\}_{\{\lceil k\theta_i(\boldsymbol{x},\cdot)\rceil \le l\}}.$$

Now, notice that the independence of the stochastic processes and the i.i.d. property of the corresponding elements, imply that

$$\begin{aligned} E\left\{F^*(\boldsymbol{x},\cdot)\,(l) \mid k\right\} &= E\left\{\sum_{i=1}^{\infty} w_i(\boldsymbol{x},\cdot) I\{\theta_i(\boldsymbol{x},\cdot)\}_{\{\lceil k\theta_i(\boldsymbol{x},\cdot)\rceil \le l\}} \,\middle|\, k\right\}, \\ &= \sum_{i=1}^{\infty} E\left\{w_i(\boldsymbol{x},\cdot)\right\} E\left\{I\{\theta_i(\boldsymbol{x},\cdot)\}_{\{\lceil k\theta_i(\boldsymbol{x},\cdot)\rceil \le l\}} \;\middle|\; k\right\}, \\ &= E\left\{I\{\theta_1(\boldsymbol{x},\cdot)\}_{\{\lceil k\theta_1(\boldsymbol{x},\cdot)\rceil \le l\}} \;\middle|\; k\right\}, \\ &= G^*_{0\boldsymbol{x}}\left(A_{l,k}\right), \end{aligned}$$

where, $A_{l,k} = [0, l/k]$ and $G^*_{0\boldsymbol{x}}$ stands for the marginal probability measure of $\theta_i(\boldsymbol{x},\cdot)$, for every $i \in \mathbb{N}$. It follows that

$$E\left\{G_{\boldsymbol{x}}(B_y) \mid k\right\} = \sum_{l=1}^{k} G^*_{0\boldsymbol{x}}\left(A_{j,k}\right) \mathrm{BIN}(l \mid k,y).$$

Applying a similar reasoning, it follows that, for every $\boldsymbol{x}, \boldsymbol{x}_0 \in \mathscr{X}$ and every $y \in [0,1]$,

$$
\begin{aligned}
&E\left\{G_{\boldsymbol{x}}(B_y)G_{\boldsymbol{x}_0}(B_y) \mid k\right\} = \\
&\qquad\sum_{l=1,l_1=1}^{k}\left[\sum_{i=1}^{\infty} E\left\{w_i(\boldsymbol{x},\cdot)w_i(\boldsymbol{x}_0,\cdot)\right\} G^*_{0,\boldsymbol{x},\boldsymbol{x}_0}(A_{l,k}\times A_{l_1,k})\right]\bar{B}(l,l_1 \mid k,y) + \\
&\qquad\sum_{l=1,l_1=1}^{k}\left[\sum_{i=1,i_1\neq i}^{\infty} E\left\{w_i(\boldsymbol{x},\cdot)w_{i_1}(\boldsymbol{x}_0,\cdot)\right\} G^*_{0\boldsymbol{x}}(A_{l,k})G^*_{0\boldsymbol{x}_0}(A_{l_1,k})\right]\bar{B}(l,l_1 \mid k,y),
\end{aligned}
$$

where, $\bar{B}(l,l_1 \mid k,y) = \mathrm{BIN}(l \mid k,y)\times\mathrm{BIN}(l_1 \mid k,y)$ and $G^*_{0,\boldsymbol{x},\boldsymbol{x}_0}$ corresponds to the marginal distribution of $(\theta_i(\boldsymbol{x},\cdot),\theta_i(\boldsymbol{x}_0,\cdot))$. In particular, for $\boldsymbol{x}=\boldsymbol{x}_0$,

$$
\begin{aligned}
&E\left\{G_{\boldsymbol{x}}(B_y)^2 \mid k\right\} = \\
&\qquad\sum_{l=1}^{k}\left[\sum_{i=1}^{\infty} E\left\{w_i(\boldsymbol{x},\cdot)^2\right\} G_{0\boldsymbol{x}}(A_{l,k})\right]\mathrm{BIN}(l \mid k,y)^2 + \\
&\qquad\sum_{l=1,l_1=1}^{k}\left[\sum_{i=1,i_1\neq i}^{\infty} E\left\{w_i(\boldsymbol{x},\cdot)w_{i_1}(\boldsymbol{x},\cdot)\right\} G_{0\boldsymbol{x}}(A_{l,k})G_{0\boldsymbol{x}}(A_{l_1,k})\right]\bar{B}(l,l_1 \mid k,y).
\end{aligned}
$$

Now, since the elements of $\mathscr{V}$ are continuous functions of $\boldsymbol{x}$ and, for every $i \in \mathbb{N}$ and every $\{\boldsymbol{x}_j\}_1^{\infty}$, with $\boldsymbol{x}_j \in \mathscr{X}$, such that $\lim_{j\to+\infty}\boldsymbol{x}_j \longrightarrow \boldsymbol{x}_0 \in \mathscr{X}$, $\eta_i(\boldsymbol{x},\cdot)$ converges in distribution to $\eta_i(\boldsymbol{x}_0,\cdot)$ as $j \longrightarrow +\infty$, it follows that $w_i(\boldsymbol{x}_j,\cdot)$ converges in distribution to $w_i(\boldsymbol{x}_0,\cdot)$, as $j \longrightarrow +\infty$, and that $\boldsymbol{x} \mapsto E\left\{w_i(\boldsymbol{x},\cdot)\right\}$, $\boldsymbol{x} \mapsto E\left\{w_i(\boldsymbol{x},\cdot)^2\right\}$ and $\boldsymbol{x} \mapsto E\left\{w_i(\boldsymbol{x},\cdot)w_{i_1}(\boldsymbol{x}_0,\cdot)\right\}$ are continuous functions. On the other hand, since the elements of $\mathscr{H}$ are continuous functions of $\boldsymbol{x}$ and, for every $i \in \mathbb{N}$ and every $\{\boldsymbol{x}_j\}_1^{\infty}$, with $\boldsymbol{x}_j \in \mathscr{X}$, such that $\lim_{j\to+\infty}\boldsymbol{x}_j \longrightarrow \boldsymbol{x}_0 \in \mathscr{X}$, $z_i(\boldsymbol{x},\cdot)$ converges in distribution to $z_i(\boldsymbol{x}_0,\cdot)$, as $j \longrightarrow +\infty$, it follows that $\theta_i(\boldsymbol{x}_j,\cdot)$ converges in distribution to $\theta_i(\boldsymbol{x}_0,\cdot)$, as $j \longrightarrow +\infty$. Now a few applications of Lebesgue dominated convergence theorem imply that

$$\begin{aligned}
\lim_{j\longrightarrow+\infty} E\left\{G_{\boldsymbol{x}_j}(B_y)^2\right\} &= \sum_{l=1}^{\infty} \lim_{j\longrightarrow+\infty} E\left\{G_{\boldsymbol{x}_j}(B_y)^2 \mid l\right\} P\{\omega\in\Omega : k(\omega)=l\}, \\
&= \sum_{l=1}^{\infty} E\left\{G_{\boldsymbol{x}_0}(B_y)^2 \mid l\right\} P\{\omega\in\Omega : k(\omega)=l\}, \\
&= E\left\{G_{\boldsymbol{x}_0}(B_y)^2\right\},
\end{aligned}$$

$$\begin{aligned}
\lim_{j\longrightarrow+\infty} E\left\{G_{\boldsymbol{x}_j}(B_y)G_{\boldsymbol{x}_0}(B_y)\right\} &= \sum_{l=1}^{\infty} \lim_{j\longrightarrow+\infty} E\left\{G_{\boldsymbol{x}_j}(B_y)G_{\boldsymbol{x}_0}(B_y) \mid l\right\} P\{\omega\in\Omega : k(\omega)=l\}, \\
&= \sum_{l=1}^{\infty} E\left\{G_{\boldsymbol{x}_0}(B_y)^2 \mid l\right\} P\{\omega\in\Omega : k(\omega)=l\}, \\
&= E\left\{G_{\boldsymbol{x}_0}(B_y)^2\right\},
\end{aligned}$$

and

$$\lim_{j\longrightarrow+\infty} E\left\{G_{\boldsymbol{x}_j}(B_y)\right\} = E\left\{G_{\boldsymbol{x}_0}(B_y)\right\},$$

which completes the proof of the theorem. □

Proof of Theorem 3.3

Notice that for every function $f_i : \mathscr{X}\times\mathscr{X} \longrightarrow [0,1]$ and $g_{ij} : \mathscr{X} \longrightarrow [0,1]$, $i=1,2$, $j=1,2$, it follows that

$$\begin{aligned}
&\left|f_1(\boldsymbol{x}_1,\boldsymbol{x}_2)f_2(\boldsymbol{x}_1,\boldsymbol{x}_2) - g_{11}(\boldsymbol{x}_1)g_{12}(\boldsymbol{x}_2)g_{21}(\boldsymbol{x}_1)g_{22}(\boldsymbol{x}_2)\right| \\
&= \left|f_1(\boldsymbol{x}_1,\boldsymbol{x}_2)f_2(\boldsymbol{x}_1,\boldsymbol{x}_2) \pm f_1(\boldsymbol{x}_1,\boldsymbol{x}_2)g_{21}(\boldsymbol{x}_1)g_{22}(\boldsymbol{x}_2) - g_{11}(\boldsymbol{x}_1)g_{12}(\boldsymbol{x}_2)g_{21}(\boldsymbol{x}_1)g_{22}(\boldsymbol{x}_2)\right|, \\
&\le f_1(\boldsymbol{x}_1,\boldsymbol{x}_2)\left|f_2(\boldsymbol{x}_1,\boldsymbol{x}_2) - g_{21}(\boldsymbol{x}_1)g_{22}(\boldsymbol{x}_2)\right| + g_{21}(\boldsymbol{x}_1)g_{22}(\boldsymbol{x}_2)\left|f_1(\boldsymbol{x}_1,\boldsymbol{x}_2) - g_{11}(\boldsymbol{x}_1)g_{12}(\boldsymbol{x}_2)\right|, \\
&\le \left|f_2(\boldsymbol{x}_1,\boldsymbol{x}_2) - g_{21}(\boldsymbol{x}_1)g_{22}(\boldsymbol{x}_2)\right| + \left|f_1(\boldsymbol{x}_1,\boldsymbol{x}_2) - g_{11}(\boldsymbol{x}_1)g_{12}(\boldsymbol{x}_2)\right|.
\end{aligned}$$

The previous result implies that, for every $\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j} \in \mathscr{X}$,

$$
\begin{aligned}
&|E[G(\boldsymbol{x}_{1j},\cdot)(B_y), G(\boldsymbol{x}_{2j},\cdot)(B_y) \mid k] - E[G(\boldsymbol{x}_{1j},\cdot)(B_y) \mid k]\, E\{G(\boldsymbol{x}_{2j},\cdot)(B_y) \mid k]| \\
&= \Bigg| \sum_{l=1,l_1=1}^{k} \sum_{i=1,i_1=1}^{\infty} E[w_i(\boldsymbol{x}_{1j},\cdot) w_{i_1}(\boldsymbol{x}_{2j},\cdot)] \left[\mathbb{I}_{\{i \neq i_1\}} E\left[\mathbb{I}_{\left\{\theta_i(\boldsymbol{x}_{1j},\cdot) \in A_{l,k}\right\}}\right] E\left[\mathbb{I}_{\left\{\theta_{i_1}(\boldsymbol{x}_{2j},\cdot) \in A_{l_1,k}\right\}}\right] \right. \\
&\qquad \left. + \mathbb{I}_{\{i = i_1\}} E\left[\mathbb{I}_{\left\{(\theta_i(\boldsymbol{x}_{1j},\cdot), \theta_{i_1}(\boldsymbol{x}_{2j},\cdot)) \in A_{l,k} \times A_{l_1,k}\right\}}\right]\right] \bar{B}(l, l_1 \mid k, y) \\
&\quad - \sum_{l=1,l_1=1}^{k} \sum_{i=1,i_1=1}^{\infty} E[w_i(\boldsymbol{x}_{1j},\cdot)]\, E[w_{i_1}(\boldsymbol{x}_{2j},\cdot)] \\
&\qquad E\left[\mathbb{I}_{\left\{\theta_i(\boldsymbol{x}_{1j},\cdot) \in A_{l,k}\right\}}\right] E\left[\mathbb{I}_{\left\{\theta_{i_1}(\boldsymbol{x}_{2j},\cdot) \in A_{l_1,k}\right\}}\right] \bar{B}(l, l_1 \mid k, y) \Bigg|, \\
&\leq \sum_{l=1,l_1=1}^{k} \sum_{i=1,i_1=1}^{\infty} |E[w_i(\boldsymbol{x}_{1j},\cdot) w_{i_1}(\boldsymbol{x}_{2j},\cdot)] - E[w_i(\boldsymbol{x}_{1j},\cdot)]\, E[w_{i_1}(\boldsymbol{x}_{2j},\cdot)]| \\
&\quad + \sum_{l=1,l_1=1}^{k} \sum_{i=1}^{\infty} \left| E\left[\mathbb{I}_{\left\{(\theta_i(\boldsymbol{x}_{1j},\cdot), \theta_i(\boldsymbol{x}_{2j},\cdot)) \in A_{l,k} \times A_{l_1,k}\right\}}\right] - E\left[\mathbb{I}_{\left\{\theta_i(\boldsymbol{x}_{1j},\cdot) \in A_{l,k}\right\}}\right] E\left[\mathbb{I}_{\left\{\theta_i(\boldsymbol{x}_{2j},\cdot) \in A_{l_1,k}\right\}}\right] \right|,
\end{aligned}
$$

where, $\bar{B}(l, l_1 \mid k, y) = \mathrm{BIN}(l \mid k, y) \times \mathrm{BIN}(l_1 \mid k, y)$. Now, since the elements of $\mathscr{V}$ are continuous functions of $\boldsymbol{x}$ and, for every $i, i_1 \in \mathbb{N}$, $w_i(\boldsymbol{x}_{1j},\cdot) w_{i_1}(\boldsymbol{x}_{2j},\cdot)$ is a continuous function of $\{(\eta_i(\boldsymbol{x}_{1j},\cdot), \eta_i(\boldsymbol{x}_{2j},\cdot))\}_1^l$, $l = \max\{i, i_1\}$, it follows that

$$
\lim_{j \to \infty} |E[w_i(\boldsymbol{x}_{1j},\cdot) w_{i_1}(\boldsymbol{x}_{2j},\cdot)] - E[w_i(\boldsymbol{x}_{1j},\cdot)]\, E[w_{i_1}(\boldsymbol{x}_{2j},\cdot)]| = 0.
$$

On the other hand, since the elements of $\mathscr{H}$ are continuous functions of $\boldsymbol{x}$ and, for every $i \in \mathbb{N}$, it follows that

$$
\lim_{j \to \infty} \left| E\left[\mathbb{I}_{\left\{(\theta_i(\boldsymbol{x}_{1j},\cdot), \theta_i(\boldsymbol{x}_{2j},\cdot)) \in A_{l,k} \times A_{l_1,k}\right\}}\right] - E\left[\mathbb{I}_{\left\{\theta_i(\boldsymbol{x}_{1j},\cdot) \in A_{l,k}\right\}}\right] E\left[\mathbb{I}_{\left\{\theta_i(\boldsymbol{x}_{2j},\cdot) \in A_{l_1,k}\right\}}\right] \right| = 0.
$$

Thus, by Lebesgue's dominated convergence theorem, it follows that

$$\begin{aligned}
&\lim_{j\to\infty} \left|Cov\left[G(\boldsymbol{x}_{1j},\cdot)(B_y), G(\boldsymbol{x}_{2j},\cdot)(B_y) \mid k\right]\right| \\
&\quad = \lim_{j\to\infty} \left|E\left[G(\boldsymbol{x}_{1j},\cdot)(B_y)G(\boldsymbol{x}_{2j},\cdot)(B_y) \mid k\right] - E\left[G(\boldsymbol{x}_{1j},\cdot)(B_y) \mid k\right] E\left\{G(\boldsymbol{x}_{2j},\cdot)(B_y) \mid k\right]\right|, \\
&\quad = 0,
\end{aligned}$$

for every $k \in \mathbb{N}$, and, therefore,

$$\begin{aligned}
\lim_{j\to\infty} Cov\left[G(\boldsymbol{x}_{1j},\cdot)(B_y), G(\boldsymbol{x}_{2j},\cdot)(B_y)\right]& \\
= &\lim_{j\to\infty} E\left[Cov\left[G(\boldsymbol{x}_{1j},\cdot)(B_y), G(\boldsymbol{x}_{2j},\cdot)(B_y) \mid k\right]\right] \\
&+ \lim_{j\to\infty} Cov\left[E\left[G(\boldsymbol{x}_{1j},\cdot)(B_y) \mid k\right], E\left[G(\boldsymbol{x}_{2j},\cdot)(B_y) \mid k\right]\right], \\
= &E\left[\lim_{j\to\infty} Cov\left[G(\boldsymbol{x}_{1j},\cdot)(B_y), G(\boldsymbol{x}_{2j},\cdot)(B_y) \mid k\right]\right] \\
&+ Cov\left[\lim_{j\to\infty} E\left[G(\boldsymbol{x}_{1j},\cdot)(B_y) \mid k\right], \lim_{j\to\infty} E\left[G(\boldsymbol{x}_{2j},\cdot)(B_y) \mid k\right]\right], \\
= &Cov\left[E\left[G(\boldsymbol{x}_{1},\cdot)(B_y) \mid k\right], E\left[G(\boldsymbol{x}_{2},\cdot)(B_y) \mid k\right]\right],
\end{aligned}$$

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

$$\begin{aligned}
E\left[G(\boldsymbol{x},\cdot)(B_y) \mid k\right] &= \sum_{l=1}^{k} G_{0\boldsymbol{x}}^{*}\left(A_{l,k}\right) \begin{pmatrix} k \\ l \end{pmatrix} y^{l}(1-y)^{k-l}, \\
&\equiv \sum_{l=1}^{k} G_{0\boldsymbol{x}}^{*}\left(A_{l,k}\right) \mathrm{BIN}(l \mid k, y),
\end{aligned}$$

which completes the proof of the theorem. □

Proof of Theorem 3.4

Since the elements of $\mathscr{V}$ are continuous functions of $\boldsymbol{x}$ and, for every $i \in \mathbb{N}$ and every $\{(\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j})\}_1^{\infty}$, with $\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j} \in \mathscr{X}$, such that $\lim_{j\to+\infty}(\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j}) = (\boldsymbol{x}_1, \boldsymbol{x}_2)$, with $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathscr{X}$,

it follows that

$$(\eta_i(\boldsymbol{x}_{1j}, \cdot), \eta_i(\boldsymbol{x}_{2j}, \cdot)),$$

and $\eta_i(\boldsymbol{x}_{lj}, \cdot)$ converges in distribution to

$$(\eta_i(\boldsymbol{x}_1, \cdot), z_i(\boldsymbol{x}_2, \cdot)),$$

and $\eta_i(\boldsymbol{x}_l, \cdot)$, respectively, as $j \longrightarrow +\infty$, for $l = 1, 2$. It also follows that $(\boldsymbol{x}_1, \boldsymbol{x}_2) \mapsto E\left\{w_i(\boldsymbol{x}_l, \cdot)\right\}$, $(\boldsymbol{x}_1, \boldsymbol{x}_2) \mapsto E\left\{w_i(\boldsymbol{x}_l, \cdot)^2\right\}$, $l = 1, 2$, and $(\boldsymbol{x}_1, \boldsymbol{x}_2) \mapsto E\left\{w_i(\boldsymbol{x}_1, \cdot)w_{i_1}(\boldsymbol{x}_2, \cdot)\right\}$ are continuous functions, for every $i, i_1 \in \mathbb{N}$.

On the other hand, since the elements of $\mathscr{H}$ are continuous functions of $\boldsymbol{x}$ and, for every $i \in \mathbb{N}$ and every $\{(\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j})\}_1^\infty$, with $\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j} \in \mathscr{X}$, such that $\lim_{j \to +\infty}(\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j}) = (\boldsymbol{x}_1, \boldsymbol{x}_2)$, with $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathscr{X}$, it follows that

$$(z_i(\boldsymbol{x}_{1j}, \cdot), z_i(\boldsymbol{x}_{2j}, \cdot)),$$

and $z_i(\boldsymbol{x}_{lj}, \cdot)$ converges in distribution to

$$(z_i(\boldsymbol{x}_1, \cdot), z_i(\boldsymbol{x}_2, \cdot)),$$

and $z_i(\boldsymbol{x}_l, \cdot)$, respectively, as $j \longrightarrow +\infty$, for $l = 1, 2$. Finally, since the correlation, ρ, is a continuous function of $E\left\{w_i(\boldsymbol{x}_l, \cdot)\right\}$, $E\left\{w_i(\boldsymbol{x}_l, \cdot)^2\right\}$, $E\left\{w_i(\boldsymbol{x}_1, \cdot)w_{i_1}(\boldsymbol{x}_2, \cdot)\right\}$, $G^*_{0\boldsymbol{x}_l}(A_{j,k})$ and $G^*_{0,\boldsymbol{x}_1,\boldsymbol{x}_2}(A_{j,k} \times A_{j_1,k})$, $i, i_1, k \in \mathbb{N}$, $j, j_1 \in \{1, \ldots, k\}$ and $l = 1, 2$, then $(\boldsymbol{x}_1, \boldsymbol{x}_2) \mapsto \rho\left[G(\boldsymbol{x}_1, \cdot)(B_y), G(\boldsymbol{x}_2, \cdot)(B_y)\right]$ is also a continuous function. $\square$

Proof of Theorem 3.5

To prove the theorem it is sufficient to show that any set of the base for the product topology of weak convergence has positive $P \circ \mathscr{G}^{-1}$–measure. Let $U = \prod_{\boldsymbol{x} \in \mathscr{X}} U_{\boldsymbol{x}}$ be a set of the base, where $U_{\boldsymbol{x}}$ is a basic open set of the weak topology for $\mathscr{P}([0,1])$ and $U_{\boldsymbol{x}} = \mathscr{P}([0,1])$ for all but finitely many $\boldsymbol{x}$ in $\mathscr{X}$. It is easy to see that the measure of a basic open set for $\{Q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in$

$\mathscr{P}\left([0,1]\right)^{\mathscr{X}}$ is equal to the measure of a set of the form

$$\prod_{i=1}^{T}\left\{P_{\boldsymbol{x}_i}\in\mathscr{P}\left([0,1]\right):\left|\int f_{ij}dP_{\boldsymbol{x}_i}-\int f_{ij}dQ_{\boldsymbol{x}_i}\right|<\epsilon_i,\ j=1,\ldots,K_i\right\},$$

where $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_T\in\mathscr{X}$, T and K_i, $i=1,\ldots,T$, are positive integers, f_{ij}, $i=1,\ldots,T$, $j=1,\ldots,K_i$, are bounded continuous functions and ϵ_i, $i=1,\ldots,T$, are positive constants. Now notice that from Lemma 1 in Barrientos et al. (2012), it follows that for every $Q_{\boldsymbol{x}_i}\in\mathscr{P}([0,1])$, there exists $Q'_{\boldsymbol{x}_i}\in\mathscr{P}([0,1])$, absolutely continuous w.r.t. Lebesgue measure, such that

$$\left|\int f_{ij}dQ_{\boldsymbol{x}_i}-\int f_{ij}dQ'_{\boldsymbol{x}_i}\right|\leq\epsilon_i/2.$$

Thus, for almost every $\omega\in\Omega$,

$$\left|\int f_{ij}dG(\boldsymbol{x}_i,\omega)-\int f_{ij}dQ_{\boldsymbol{x}_i}\right|\leq\left|\int f_{ij}dG(\boldsymbol{x}_i,\omega)-\int f_{ij}dQ'_{\boldsymbol{x}_i}\right|+\epsilon_i/2.$$

Set $d_{ij}\left(\omega\right)=\left|\int f_{ij}dG(\boldsymbol{x}_i,\omega)-\int f_{ij}dQ'_{\boldsymbol{x}_i}\right|$. Now, borrowing the trick in Petrone (1999a), it follows that

$$\begin{aligned}d_{ij}\left(\omega\right)\leq&\left|\int f_{ij}dG(\boldsymbol{x}_i,\omega)-\int f_{ij}dH\left(Q'_{\boldsymbol{x}_i},\ k(\omega)\right)\right|+\left|\int f_{ij}dH\left(Q'_{\boldsymbol{x}_i},\ k(\omega)\right)-\int f_{ij}dQ'_{\boldsymbol{x}_i}\right|,\\ \equiv&\ d_{ij}^{(1)}\left(\omega\right)+d_{ij}^{(2)}\left(k(\omega)\right),\end{aligned}$$

where $H\left(Q'_{\boldsymbol{x}_i},\ k(\omega)\right)$ is the measure associated with the Bernstein polynomial of degree $k(\omega)$ of the measure $Q'_{\boldsymbol{x}_i}$. Since $H\left(Q'_{\boldsymbol{x}_i},\ k'\right)$ converges weakly to $Q'_{\boldsymbol{x}_i}$, as $k'\longrightarrow+\infty$, it follows that there exists $k_0\in\mathbb{N}$ such that, for almost every $\omega\in\Omega$ such that $k(\omega)>k_0$, $d_{ij}^{(2)}\left(k(\omega)\right)\leq\frac{\epsilon_i}{4}$.

On the other hand, it is easy to see that, for every $k_1 \in \mathbb{N}$,

$$\begin{aligned} d_{ij}^{(1)}(\omega) &= \left| \int_0^1 f_{ij}(y) \sum_{l=1}^{k_1} W_l(\boldsymbol{x}_i, \omega, k_1) \beta\left(y \mid l, k_1 - l + 1\right) dy \right. \\ &\quad \left. - \int_0^1 f_{ij}(y) \sum_{l=1}^{k_1} Q'_{\boldsymbol{x}_i}\left(\left(\frac{l-1}{k_1}, \frac{l}{k_1}\right]\right) \beta\left(y | l, k_1 - l + 1\right) dy \right|, \\ &\leq \int_0^1 \sum_{l=1}^{k_1} \left| W_l(\boldsymbol{x}_i, \omega, k_1) - Q'_{\boldsymbol{x}_i}\left(\left(\frac{l-1}{k_1}, \frac{l}{k_1}\right]\right) \right| |f_{i,j}(y)| \beta\left(y \mid l, k_1 - l + 1\right) dy, \\ &\leq M_{ij} \int_0^1 \sum_{l=1}^{k_1} \left| W_l(\boldsymbol{x}_i, \omega, k_1) - Q'_{\boldsymbol{x}_i}\left(\left(\frac{l-1}{k_1}, \frac{l}{k_1}\right]\right) \right| \beta\left(y \mid l, k_1 - l + 1\right) dy, \\ &\leq M_{ij} k_1 N_{ik_1}(\omega), \end{aligned}$$

where, $W_l(\boldsymbol{x}, \omega, k) = \sum_{i=1}^{\infty} w_i(\boldsymbol{x}, \omega) I\{\theta_i(\boldsymbol{x}, \omega)\}_{\{\lceil k\theta_i(\boldsymbol{x},\omega)\rceil = l\}}$, $M_{ij} = \sup_{y \in [0,1]} |f_{ij}(y)|$, and

$$N_{ik_1}(\omega) = \max_{l \in \{1,\ldots,k_1\}} \left| W_l(\boldsymbol{x}_i, \omega, k_1) - Q_{\boldsymbol{x}_i}\left(\left(\frac{l-1}{k_1}, \frac{l}{k_1}\right]\right) \right|.$$

Thus, if $N_{ik_1}(\omega) \leq \frac{\epsilon_i}{4M_{ij}k_1}$, then $d_{ij}^{(1)}(\omega) \leq \frac{\epsilon_i}{4}$, $i = 1, \ldots, T$, $j = 1, \ldots, K_i$. It follows that,

$$\begin{aligned} &P\left\{\omega \in \Omega : \left| \int f_{ij} dG(\boldsymbol{x}_i, \omega) - \int f_{ij} dQ_{x_i} \right| < \epsilon_i,\ i = 1, \ldots, T,\ j = 1, \ldots, K_i \right\} \geq \\ &\qquad \sum_{k_1 > k_0} P\left\{\omega \in \Omega : N_{ik_1}(\omega) \leq \frac{\epsilon_i}{4M_{ij}k_1}, i = 1, \ldots, T,\ j = 1, \ldots, K_i, k(\omega) = k_1 \right\}. \end{aligned}$$

Now, since by assumption the joint distribution of $(\eta_i(\boldsymbol{x}_1, \cdot), \ldots, \eta_i(\boldsymbol{x}_d, \cdot))$ and $(z_i(\boldsymbol{x}_1, \cdot), \ldots, z_i(\boldsymbol{x}_d, \cdot))$ have full support on $\mathbb{R}^d$, for every $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d \in \mathscr{X}^d$, $d \geq 1$, and $k(\cdot)$ has full support on $\mathbb{N}$, then Theorem 6 in Barrientos et al. (2012) ensures that

$$P\left\{\omega \in \Omega : N_{ik_1}(\omega) \leq \frac{\epsilon_i}{2M_{ij}k_1}, i = 1, \ldots, T,\ j = 1, \ldots, K_i,\ k(\omega) = k_1 \right\} > 0,$$

which completes the proof of the theorem. $\square$

Proof of Theorem 3.6

For every $T \geq 1$ and $i = 1, \ldots, T$, let $Q_{\boldsymbol{x}_i} \in \mathscr{D}([0,1])$. Now, a simple application of the triangle inequality implies that, for $i = 1, \ldots, T$, and almost every $\omega \in \Omega$,

$$\begin{aligned} \sup_{y \in [0,1]} |g(\boldsymbol{x}_i, \omega)(y) - q_{\boldsymbol{x}_i}(y)| \quad \leq \quad & \sup_{y \in [0,1]} |g(\boldsymbol{x}_i, \omega)(y) - \mathrm{bp}(y \mid k(\omega), Q_{\boldsymbol{x}_i})| + \\ & \sup_{y \in [0,1]} |\mathrm{bp}(y \mid k(\omega), Q_{\boldsymbol{x}_i}) - q_{\boldsymbol{x}_i}(y)| \,. \end{aligned}$$

Now, the continuity of $q_{\boldsymbol{x}_i}$ implies that it can be uniformly approximated by the density of the Bernstein polynomial of $Q_{\boldsymbol{x}_i}$ (see, Petrone & Wasserman, 2002, Theorem 2). Thus, for $i = 1, \ldots, K$, and $\epsilon > 0$, there exists $k_0 \in \mathbb{N}$, such that

$$\sup_{y \in [0,1]} |\mathrm{bp}(y \mid k_0, Q_{\boldsymbol{x}_i}) - q_{\boldsymbol{x}_i}(y)| < \epsilon/2.$$

On the other hand, it is easy to see that, for $i = 1, \ldots, K$, and every $k_1 \in \mathbb{N}$,

$$\begin{aligned} \sup_{y \in [0,1]} |g(\boldsymbol{x}_i, \omega)(y) - \mathrm{bp}(y \mid k_1, Q_{\boldsymbol{x}_i})| \quad = \quad & \sup_{y \in [0,1]} \left| \sum_{l=1}^{k_1} W_l(\boldsymbol{x}_i, \omega, k_1) \beta\left(y \mid l, k_1 - l + 1\right) \right. \\ & \left. - \sum_{l=1}^{k_1} Q'_{\boldsymbol{x}_i} \left(\left(\frac{l-1}{k_1}, \frac{l}{k_1} \right] \right) \beta\left(y | l, k_1 - l + 1\right) \right|, \\ \leq \quad & M k_1 N_{ik_1}(\omega), \end{aligned}$$

where, $W_l(\boldsymbol{x}, \omega, k) = \sum_{i=1}^{\infty} w_i(\boldsymbol{x}, \omega) I\{\theta_i(\boldsymbol{x}, \omega)\}_{\{\lceil k\theta_i(\boldsymbol{x}, \omega)\rceil = l\}}$,

$$M = \sup_{l \in \{1, \ldots, k_1\}} \sup_{y \in [0,1]} \beta\left(y \mid l, k_1 - l + 1\right),$$

and

$$N_{ik_1}(\omega) = \max_{l \in \{1, \ldots, k_1\}} \left| W_l(\boldsymbol{x}_i, \omega, k_1) - Q_{\boldsymbol{x}_i} \left(\left(\frac{l-1}{k_1}, \frac{l}{k_1} \right] \right) \right|.$$

It follows that if $N_{ik_1}(\omega) < \frac{\epsilon}{2Mk_1}$ and $k(\omega) = k_1 \geq k_0$ then,

$$\sup_{y\in[0,1]} |g(\boldsymbol{x}_i, \omega)(y) - \text{bp}(y \mid k(\omega), Q_{\boldsymbol{x}_i})| \leq \epsilon/2,$$

and

$$\sup_{y\in[0,1]} |g(\boldsymbol{x}_i, \omega)(y) - q_{\boldsymbol{x}_i}(y)| \leq \epsilon,$$

for $i = 1, \ldots, T$. Now, it is easy to show that

$$P\left\{\omega \in \Omega : \sup_{y\in[0,1]} |g(\boldsymbol{x}_i, \omega)(y) - q_{\boldsymbol{x}_i}(y)| < \epsilon,\ i = 1, \ldots, T\right\} \geq \sum_{k_1\geq k_0} P\left\{\omega \in \Omega : N_{ik_1}(\omega) \leq \frac{\epsilon}{4Mk_1}, i = 1, \ldots, T, k(\omega) = k_1\right\}.$$

Finally, since by assumption the joint distribution of $(\eta_i(\boldsymbol{x}_1, \cdot), \ldots, \eta_i(\boldsymbol{x}_d, \cdot))$ and $(z_i(\boldsymbol{x}_1, \cdot), \ldots, z_i(\boldsymbol{x}_d, \cdot))$ have full support on $\mathbb{R}^d$, for every $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d \in \mathscr{X}^d$, $d \geq 1$, and $k(\cdot)$ has full support on $\mathbb{N}$, then Theorem 6 in Barrientos et al. (2012) ensures that

$$P\left\{\omega \in \Omega : N_{ik_1}(\omega) \leq \frac{\epsilon}{2Mk_1}, i = 1, \ldots, T, k(\omega) = k_1\right\} > 0,$$

which completes the proof of the theorem. $\square$

Proof of Theorem 3.7

The following Lemma, proved below, is used in the proof of the theorem.

Lemma B.1. *If $\mathscr{X} \subset \mathbb{R}^p$ is a compact set and $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \tilde{\mathscr{D}}\left([0,1]\right)^{\mathscr{X}}$, then, for every $\epsilon > 0$ there exists $k_0 \in \mathbb{N}$ such that for $k > k_0$,*

$$\sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{y\in[0,1]} |\text{bp}(y \mid k, G_{\boldsymbol{x}}) - g_{\boldsymbol{x}}(y)| < \epsilon.$$

Now, let $\{Q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \tilde{\mathscr{D}}([0,1])^{\mathscr{X}}$, with density functions $\{q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$. An application of the triangle inequality implies that, for every $\boldsymbol{x} \in \mathscr{X}$, $y \in [0,1]$ and almost every $\omega \in \Omega$,

$$|g(\boldsymbol{x},\omega)(y) - q_{\boldsymbol{x}}(y)| \leq |\mathrm{bp}(y \mid k, Q_{\boldsymbol{x}}) - q_{\boldsymbol{x}}(y)| + |g(\boldsymbol{x},\omega)(y) - \mathrm{bp}(y \mid k, Q_{\boldsymbol{x}})|\,.$$

By Lemma B.1, it follows that there exists $k_0 \in \mathbb{N}$ such that

$$\sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{y\in[0,1]} |\mathrm{bp}(y \mid k_0, Q_{\boldsymbol{x}}) - q_{\boldsymbol{x}}(y)| \leq \frac{\epsilon}{2}.$$

On the other hand, note that for every $\boldsymbol{x} \in \mathscr{X}$, $k_1 \in \mathbb{N}$ and almost every $\omega \in \Omega$,

$$\begin{aligned}
\sup_{y\in[0,1]} |g(\boldsymbol{x},\omega)(y) - \mathrm{bp}(y \mid k_1, Q_{\boldsymbol{x}})| &= \sup_{y\in[0,1]} \Bigg| \sum_{l=1}^{k_1} W_l(\boldsymbol{x},\omega,k_1)\beta\left(y|l, k_1-l+1\right) \\
&\quad - \sum_{l=1}^{k_1} Q_{\boldsymbol{x}}\left(\frac{l}{k_1}, \frac{l-1}{k_1}\right]\beta\left(y|l,k_1-l+1\right)dy \Bigg|, \\
&\leq M k_1 \max_{l\in\{1,\ldots,k_1\}} \left| W_l(\boldsymbol{x},\omega,k_1) - Q_{\boldsymbol{x}}\left(\frac{l}{k_1}, \frac{l-1}{k_1}\right] \right|,
\end{aligned}$$

where, $W_l(\boldsymbol{x},\omega,k_1) = \sum_{i=1}^{\infty} w_i(\boldsymbol{x},\omega) I\{\theta_i(\boldsymbol{x},\omega)\}_{\{\lceil k\theta_i(\boldsymbol{x},\omega)\rceil = l\}}$ and

$$M = \sup_{l\in\{1,\ldots,k_1\}} \sup_{y\in[0,1]} \beta\left(y|l,k_1-l+1\right).$$

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

- For $l = 1, \ldots, k_0$,
$$\sup_{\boldsymbol{x}\in\mathscr{X}} \left| h_{\boldsymbol{x}}\left(z_l(\boldsymbol{x},\omega)\right) - \frac{2l-1}{2k_0} \right| \leq \frac{1}{4k_0}.$$
- For $l = 1$,
$$\sup_{\boldsymbol{x}\in\mathscr{X}} \left| v_{\boldsymbol{x}}\left(\eta_1(\boldsymbol{x},\omega)\right) - Q_{\boldsymbol{x}}\left(\left(0, \frac{l}{k_0}\right]\right) \right| \leq \frac{\epsilon}{2k_0 M(2^{k_0}-1)}.$$

- For $l = 2, \ldots, k_0$,

$$\sup_{\boldsymbol{x}\in\mathscr{X}} \left| v_{\boldsymbol{x}}\left(\eta_l(\boldsymbol{x},\omega)\right) - \frac{Q_{\boldsymbol{x}}\left(\left(\frac{l}{k_0}, \frac{l-1}{k_0}\right]\right)}{1-\sum_{j=1}^{l-1} Q_{\boldsymbol{x}}\left(\left(\frac{j}{k_0}, \frac{j-1}{k_0}\right]\right)} \right| \leq \frac{\epsilon}{2k_0 M(2^{k_0}-1)}.$$

- $k(\omega) = k_0$.

Then for almost every $\omega \in \Omega_0$, it follows that

$$\sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{y\in[0,1]} |g(\boldsymbol{x},\omega)(y) - \mathrm{bp}(y \mid k, Q_{\boldsymbol{x}})| < \frac{\epsilon}{2},$$

and, therefore,

$$\begin{aligned}
P&\left\{\omega\in\Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{y\in[0,1]} |g(\boldsymbol{x},\omega)(y) - q_{\boldsymbol{x}}(y)| \leq \epsilon\right\} \geq \\
&\prod_{l=1}^{k_0} P\left\{\omega\in\Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} \left| h_{\boldsymbol{x}}(z_l(\boldsymbol{x},\omega)) - \frac{2l-1}{2k_0}\right| \leq \frac{1}{4k_0}\right\} \times \\
&P\left\{\omega\in\Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} \left| v_{\boldsymbol{x}}(\eta_1(\boldsymbol{x},\omega)) - Q_{\boldsymbol{x}}\left(\left(0, \frac{1}{k_0}\right]\right)\right| \leq \frac{\epsilon}{2k_0 M(2^{k_0}-1)}\right\} \times \\
&\prod_{l=2}^{k_0} P\left\{\omega\in\Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} \left| v_{\boldsymbol{x}}(\eta_l(\boldsymbol{x},\omega)) - \frac{Q_{\boldsymbol{x}}\left(\left(\frac{l}{k_0}, \frac{l-1}{k_0}\right]\right)}{1-\sum_{j=1}^{l-1} Q_{\boldsymbol{x}}\left(\left(\frac{j}{k_0}, \frac{j-1}{k_0}\right]\right)}\right| \leq \frac{\epsilon}{2k_0 M(2^{k_0}-1)}\right\} \times \\
&P\left\{\omega\in\Omega : k(\omega) = k_0\right\}.
\end{aligned}$$

Now, since

$$\boldsymbol{x} \mapsto \frac{2l-1}{2k_0},$$

$$\boldsymbol{x} \mapsto Q_{\boldsymbol{x}}\left(\left(0, \frac{1}{k_0}\right]\right),$$

and

$$\boldsymbol{x} \mapsto \frac{Q_{\boldsymbol{x}}\left(\left(\frac{l}{k_0}, \frac{l-1}{k_0}\right]\right)}{1-\sum_{j=1}^{l-1} Q_{\boldsymbol{x}}\left(\left(\frac{j}{k_0}, \frac{j-1}{k_0}\right]\right)},$$

$l = 2, \dots, k_0$, are continuous functions, it follows that

$$
\begin{aligned}
&\prod_{l=1}^{k_0} P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} \left| h_{\boldsymbol{x}}(z_l(\boldsymbol{x},\omega)) - \frac{2l-1}{2k_0}\right| \leq \frac{1}{4k_0}\right\} \times \\
&P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} \left| v_{\boldsymbol{x}}(\eta_1(\boldsymbol{x},\omega)) - Q_{\boldsymbol{x}}\left(\left(0, \frac{1}{k_0}\right]\right)\right| \leq \frac{\epsilon}{2k_0 M(2^{k_0}-1)}\right\} \times \\
&\prod_{l=2}^{k_0} P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} \left| v_{\boldsymbol{x}}(\eta_l(\boldsymbol{x},\omega)) - \frac{Q_{\boldsymbol{x}}\left(\left(\frac{l}{k_0}, \frac{l-1}{k_0}\right]\right)}{1 - \sum_{j=1}^{l-1} Q_{\boldsymbol{x}}\left(\left(\frac{j}{k_0}, \frac{j-1}{k_0}\right]\right)}\right| \leq \frac{\epsilon}{2k_0 M(2^{k_0}-1)}\right\} \\
&> 0.
\end{aligned}
$$

Finally, since $k(\cdot)$ has full support on $\mathbb{N}$, then

$$
P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{y\in[0,1]} |g(\boldsymbol{x},\omega)(y) - q_{\boldsymbol{x}}(y)| \leq \epsilon\right\} > 0,
$$

which completes the proof of the theorem. □

Proof of Lemma B.1

Let $\{Q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \tilde{\mathscr{D}}([0,1])^{\mathscr{X}}$, with density functions $\{q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$, and assume, without loss of generality, that $\mathscr{X} = [0,1]$. Notice that $q_{\boldsymbol{x}}(y)$ can be seen as a joint density function defined on $\mathscr{X} \times [0,1]$, whose marginal distribution for $\boldsymbol{x}$ is uniform. Now, let

$$
\sum_{i=1}^{k_1} \int_{(i-1)/k_1}^{i/k_1} \mathrm{bp}(y \mid Q_t, k) dt \beta(\boldsymbol{x} \mid i, k_1 - i + 1)
$$

be the density of the multivariate Bernstein polynomial of the joint distribution $q_{\boldsymbol{x}}(y)$ of degrees k and k_1, $(k, k_1) \in \mathbb{N}^2$. Therefore, an extension of the Weierstrass approximation theorem ensures that for every $\epsilon > 0$, there exists $k_0 \in \mathbb{N}$ such that

$$
\sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{y\in[0,1]} \left| q_{\boldsymbol{x}}(y) - \sum_{i=1}^{k_1} \int_{(i-1)/k_1}^{i/k_1} \mathrm{bp}(y \mid Q_t, k) dt \beta(\boldsymbol{x} \mid i, k_1 - i + 1)\right| \leq \frac{\epsilon}{2},
$$

for every $k > k_0$ and $k_1 > k_0$. On the other hand, note that

$$\left|\mathrm{bp}(y \mid Q_{\boldsymbol{x}}, k) - \sum_{i=1}^{k_1} \int_{(i-1)/k_1}^{i/k_1} \mathrm{bp}(y \mid Q_t, k) dt \beta(\boldsymbol{x} \mid i, k_1 - i + 1)\right|$$
$$\leq \sum_{j=1}^{k} \left| \int_{(j-1)/k}^{j/k} q_{\boldsymbol{x}}(z) dz - \sum_{i=1}^{k_1} \int_{(i-1)/k_1}^{i/k_1} \int_{(j-1)/k}^{j/k} q_t(z) dz dt \beta(\boldsymbol{x} \mid i, k_1 - i + 1) \right| \beta(\boldsymbol{x} \mid j, k - j + 1).$$

Thus, since $\sum_{i=1}^{k_1} \int_{(i-1)/k_1}^{i/k_1} \int_{(j-1)/k}^{j/k} q_t(z) dz dt \beta(\boldsymbol{x} \mid i, k_1 - i + 1)$ is the Bernstein polynomial of

$$\int_{(j-1)/k}^{j/k} q_{\boldsymbol{x}}(z) dz,$$

it follows that there exists $\tilde{k}_0 \in \mathbb{N}$, such that for a fixed k and every $k_1 > \tilde{k}_0$,

$$\sup_{j \in \{1,\dots,k\}} \sup_{\boldsymbol{x} \in \mathscr{X}} \left| \int_{(j-1)/k}^{j/k} q_{\boldsymbol{x}}(z) dz - \sum_{i=1}^{k_1} \int_{(i-1)/k_1}^{i/k_1} \int_{(j-1)/k}^{j/k} q_t(z) dz dt \beta(\boldsymbol{x} \mid i, k_1 - i + 1) \right| \leq \frac{\epsilon}{2k^2}.$$

Finally, if $k = k_0$ and $k_1 = \max\{k_0, \tilde{k}_0\}$, then a simple application of the triangle inequality implies that,

$$\begin{aligned} |\mathrm{bp}(y \mid Q_{\boldsymbol{x}}, k) - g_{\boldsymbol{x}}(y)| \;\leq\; & \left|\mathrm{bp}(y \mid Q_{\boldsymbol{x}}, k) - \sum_{i=1}^{k_1} \int_{(i-1)/k_1}^{i/k_1} \mathrm{bp}(y \mid Q_t, k) dt \beta(\boldsymbol{x} \mid i, k_1 - i + 1)\right| \\ & + \left| q_{\boldsymbol{x}}(y) - \sum_{i=1}^{k_1} \int_{(i-1)/k_1}^{i/k_1} \mathrm{bp}(y \mid Q_t, k) dt \beta(\boldsymbol{x} \mid i, k_1 - i + 1)\right|, \\ \leq \;& \epsilon, \end{aligned}$$

which completes the proof of the lemma. The extension of the proof for the case where $\mathscr{X}$ is identified with $[0,1]^p$, $p > 1$, is straightforward; it is only needed to replace

$$\sum_{i=1}^{k_1} \int_{(i-1)/k_1}^{i/k_1} \int_{(j-1)/k}^{j/k} q_t(z) dz dt \beta(\boldsymbol{x} \mid i, k_1 - i + 1)$$

by the corresponding multivariate Bernstein polynomial. $\square$

Proof of Corollary 3.1

Let $\{Q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \tilde{\mathscr{D}}([0,1])$, with density functions $\{q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$. From Theorem 7, it follows that for every $\epsilon > 0$, there exists a subset $\Omega_0 \subset \Omega$ with positive P–measure, such that for every $y \in [0,1]$, $\boldsymbol{x} \in \mathscr{X}$ and almost every $\omega \in \Omega_0$,

$$|g(\boldsymbol{x}, \omega)(y) - q_{\boldsymbol{x}}(y)| < \epsilon,$$

which implies that

$$\frac{1}{1 + \frac{\epsilon}{q_{\boldsymbol{x}}(y)}} \leq \frac{q_{\boldsymbol{x}}(y)}{g(\boldsymbol{x}, \omega)(y)} \leq \frac{1}{1 - \frac{\epsilon}{q_{\boldsymbol{x}}(y)}}.$$

Now, since $(\boldsymbol{x}, y) \mapsto q_{\boldsymbol{x}}(y)$ is continuous on the compact set $\mathscr{X} \times [0,1]$, it follows that

$$\sup_{\boldsymbol{x} \in \mathscr{X}} \sup_{y \in [0,1]} q_{\boldsymbol{x}}(y) < \infty.$$

Furthermore, if we assume that

$$\inf_{\boldsymbol{x} \in \mathscr{X}} \inf_{y \in [0,1]} q_{\boldsymbol{x}}(y) > 0,$$

it follows that there exist $M_1(\epsilon) > 0$ and $M_2(\epsilon) < \infty$, such that $M_1(\epsilon) \leq \frac{q_{\boldsymbol{x}}(y)}{g(\boldsymbol{x},\omega)(y)} \leq M_2(\epsilon)$. Since the logarithm function defined on $[M_1(\epsilon), M_2(\epsilon)]$ is uniformly continuous and bounded, and $M_1(\epsilon)$ and $M_2(\epsilon)$ are decreasing and increasing as functions of ϵ, respectively, it follows that, for every $\epsilon' > 0$, there exists $\epsilon > 0$ such that

$$P\left\{\omega \in \Omega : \sup_{\boldsymbol{x} \in \mathscr{X}} \mathrm{KL}(q_{\boldsymbol{x}}, g(\boldsymbol{x}, \omega)) < \epsilon'\right\} \geq P\left\{\omega \in \Omega : \sup_{\boldsymbol{x} \in \mathscr{X}} \sup_{y \in [0,1]} |g(\boldsymbol{x}, \omega)(y) - q_{\boldsymbol{x}}(y)| < \epsilon\right\} > 0,$$

where $\mathrm{KL}(q, g) = \int_0^1 q(y) \log\left(\frac{q(y)}{g(y)}\right) dy$. Now consider the case when $\inf_{\boldsymbol{x} \in \mathscr{X}} \inf_{y \in [0,1]} q_{\boldsymbol{x}}(y) = 0$. By using a similar reasoning as in the proof of Theorem 2 of Petrone & Wasserman (2002), it is possible to ensure that for every $\epsilon^* > 0$, there exist $\delta > 0$, $\epsilon' > 0$ and $\Omega_0 \subset \Omega$, such that for every $\boldsymbol{x} \in \mathscr{X}$ and almost every $\omega \in \Omega_0$,

$$\mathrm{KL}(q'_{\boldsymbol{x}}, g(\boldsymbol{x}, \omega)) < \epsilon',$$

implying that

$$\mathrm{KL}(q_{\boldsymbol{x}}, g(\boldsymbol{x}, \omega)) \leq (C+1)\log C + C\left[\mathrm{KL}(q'_{\boldsymbol{x}}, g(\boldsymbol{x}, \omega)) + \sqrt{\mathrm{KL}(q'_{\boldsymbol{x}}, g(\boldsymbol{x}, \omega))}\right] < \epsilon^*,$$

where $q'_{\boldsymbol{x}}(y) = C^{-1} q_{\boldsymbol{x}}(y) \vee \delta$ and $C = \int_{[0,1]} q_{\boldsymbol{x}}(y) \vee \delta \, dy$. Note that by definition

$$\inf_{\boldsymbol{x} \in \mathscr{X}} \inf_{y \in [0,1]} q'_{\boldsymbol{x}}(y) > 0,$$

and, therefore,

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

which completes the proof. □

B.2 Properties of the $w\mathrm{DBPP}$

In this section we adapt the results derived for the general model to the special case of the $w\mathrm{DBPP}$. The proofs closely follow those given in Section B.1.

Theorem B.1. *Let* $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \sim w\mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_2, \mathscr{H}, \alpha)$. *If for every* $j \in \mathbb{N}$, *the stochastic processes* z_j *are* P*–a.s. continuous, then for every* $\{\boldsymbol{x}_j\}_1^\infty \subset \mathscr{X}$, *such that* $\lim_{j \to +\infty} \boldsymbol{x}_j \longrightarrow \boldsymbol{x}_0 \in \mathscr{X}$ *and* $x_{jl} \leq x_{0l}$, $l = 1, \ldots, p$,

$$\lim_{j \to +\infty} \sup_{B \in \mathscr{B}([0,1])} |G_{\boldsymbol{x}_j}(B) - G_{\boldsymbol{x}_0}(B)| = 0, \; P\text{–a.s.},$$

that is, $G_{\boldsymbol{x}_j}$ *converges* P*–a.s. in total variation norm to* $G_{\boldsymbol{x}_0}$, *when* $\boldsymbol{x}_j \longrightarrow \boldsymbol{x}_0^-$. *In addition, for every* $\{\boldsymbol{x}_j\}_1^\infty \subset \mathscr{X}$, *such that* $\lim_{j \to +\infty} \boldsymbol{x}_j \longrightarrow \boldsymbol{x}_0 \in \mathscr{X}$ *and* $x_{jl} \geq x_{0l}$, *for some* $l = 1, \ldots, p$, *there exists a random probability measure on* $([0,1],\mathscr{B}([0,1]))$, $\tilde{G}_{\boldsymbol{x}_0}$, *such that*

$$\lim_{j\to+\infty}\sup_{B\in\mathscr{B}([0,1])}|G_{\boldsymbol{x}_j}(B)-\tilde{G}_{\boldsymbol{x}_0}(B)|=0,\ P\text{–a.s.},$$

that is, $G_{\boldsymbol{x}_j}$ *converges* P*–a.s. in total variation norm to* $\tilde{G}_{\boldsymbol{x}_0}$, *when* $\boldsymbol{x}_j\longrightarrow\boldsymbol{x}_0^+$.

PROOF. Since the elements of $\mathscr{H}$ are continuous functions of $\boldsymbol{x}$ and, for every $j\in\mathbb{N}$, z_j is a P–as continuous stochastic process, it follows that $\boldsymbol{x}\mapsto h_{\boldsymbol{x}}(z_j(\boldsymbol{x},\cdot))$, $j\in\mathbb{N}$, is a P–a.s. continuous function. Also, since the ceiling function is continuous from the left and it has a limit from the right, it follows that, for almost every $\omega\in\Omega$ and every $\left\{\boldsymbol{x}_j^{(l)}\right\}_{j=1}^{\infty}$, with $\boldsymbol{x}_j^{(l)}\in\mathscr{X}$, such that $\lim_{j\to+\infty}\boldsymbol{x}_j^{(l)}=\boldsymbol{x}_0\in\mathscr{X}$ and $x_{jm}^l\le x_{0m}$, $m=1,\ldots,p$,

$$\lim_{j\to+\infty}\left\lceil k(\omega)\theta_i\left(\boldsymbol{x}_j^{(l)},\omega\right)\right\rceil=\lceil k(\omega)\theta_i\left(\boldsymbol{x}_0,\omega\right)\rceil.$$

Furthermore, for almost every $\omega\in\Omega$ and every $\left\{\boldsymbol{x}_j^{(r)}\right\}_{j=1}^{\infty}$, with $\boldsymbol{x}_j^{(r)}\in\mathscr{X}$, such that $\lim_{j\to+\infty}$ $\boldsymbol{x}_j^{(r)}=\boldsymbol{x}_0\in\mathscr{X}$ and $x_{jm}^r\ge x_{0m}$, for some $m=1,\ldots,p$, it follows that

$$\lim_{j\to+\infty}\lceil k(\omega)\theta_i(\boldsymbol{x}_j^{(r)},\omega)\rceil=\lceil k(\omega)\theta_i(\boldsymbol{x}_0,\omega)\rceil^{(r)}:=\begin{cases} j & \text{if}\quad k(\omega)\theta_i(\boldsymbol{x}_0,\omega)\in(j-1,j)\\ j+1 & \text{if}\quad k(\omega)\theta_i(\boldsymbol{x}_0,\omega)=j\end{cases}.$$

Therefore, by the Lebesgue's dominated convergence theorem, it follows that the density w.r.t. Lebesgue measure of $G_{\boldsymbol{x}}$, is P–a.s. continuous from the left and it has a limit from the right. In other words, that for every $y\in[0,1]$,

$$P\left\{\omega\in\Omega:\lim_{j\to+\infty}g(\boldsymbol{x}_j^{(l)},\omega)(y)=g(\boldsymbol{x}_0,\omega)(y),\ \ \lim_{j\to+\infty}g(\boldsymbol{x}_j^{(r)},\omega)(y)=g^{(r)}(\boldsymbol{x}_0,\omega)(y)\right\}=1,$$

where

$$g^{(r)}(\boldsymbol{x}_0,\omega)(y)=\sum_{i=1}^{\infty}w_j(\boldsymbol{x}_0,\omega)\beta(y|\lceil k(\omega)\theta_i(\boldsymbol{x}_0,\omega)\rceil^{(r)},k(\omega)-\lceil k(\omega)\theta_i(\boldsymbol{x}_0,\omega)\rceil^{(r)}+1).$$

Finally, let $G^{(r)}(\boldsymbol{x}_0,\omega)$ be a probability measure with density function $g^{(r)}(\boldsymbol{x}_0,\omega)$. A direct application of Scheffe's theorem implies that

$$P\left\{\omega \in \Omega : \lim_{j\to+\infty} \sup_{B\in\mathscr{B}([0,1])} |G(\boldsymbol{x}_j^{(l)},\omega)(B) - G(\boldsymbol{x}_0,\omega)(B)| = 0,\right.$$
$$\left.\lim_{j\to+\infty} \sup_{B\in\mathscr{B}([0,1])} |G(\boldsymbol{x}_j^{(r)},\omega)(B) - G^{(r)}(\boldsymbol{x}_0,\omega)(B)| = 0\right\} = 1,$$

which completes the proof of the theorem. □

Theorem B.2. *Let* $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \sim w\text{DBPP}(\lambda, \boldsymbol{\Psi}_2, \mathscr{H}, \alpha)$*. If for every* $\{\boldsymbol{x}_j\}_1^\infty \subset \mathscr{X}$*, such that* $\lim_{j\to+\infty} \boldsymbol{x}_j \longrightarrow \boldsymbol{x}_0 \in \mathscr{X}$*, we have* $z_i(\boldsymbol{x}_j,\cdot) \xrightarrow{\mathscr{L}} z_i(\boldsymbol{x}_0,\cdot)$*, as* $j \longrightarrow +\infty$*, then, for all* $y \in (0,1)$*,*

$$\lim_{j\to+\infty} \rho\left[G(\boldsymbol{x}_j,\cdot)(B_y), G(\boldsymbol{x}_0,\cdot)(B_y)\right] = 1,$$

where $\rho(A,B)$ *denotes the Pearson correlation between* A *and* B*, and* $B_y = [0,y]$*.*

PROOF. Notice that for every $y \in [0,1]$ and every $\boldsymbol{x} \in \mathscr{X}$,

$$\begin{aligned} E\left\{G(\boldsymbol{x},\cdot)(B_y) \mid k\right\} &= E\left\{\sum_{l=1}^{k} \left[F^*(\boldsymbol{x},\cdot)(l)\right] \text{BIN}(l \mid k, l) \,\middle|\, k\right\}, \\ &= \sum_{l=1}^{k} E\left\{F^*(\boldsymbol{x},\cdot)(l) \mid k\right\} \text{BIN}(l \mid k, l), \end{aligned}$$

where, $\text{BIN}(\cdot \mid k, y)$ stands for binomial distribution with parameters (k,y), and

$$F^*(\boldsymbol{x},\cdot)(l) = \sum_{i=1}^{\infty} w_i(\cdot) I\{\theta_i(\boldsymbol{x},\cdot)\}_{\{\lceil k\theta_i(\boldsymbol{x},\cdot)\rceil \le l\}}.$$

Now, notice that the independence of the stochastic processes and the i.i.d. property of the corresponding elements, imply that

$$\begin{aligned}
E\left\{F^*(\boldsymbol{x},\cdot)\,(l) \mid k\right\} &= E\left\{\sum_{i=1}^{\infty} w_i(\cdot) I\{\theta_i(\boldsymbol{x},\cdot)\}_{\{\lceil k\theta_i(\boldsymbol{x},\cdot)\rceil \le l\}} \,\middle|\, k\right\}, \\
&= \sum_{i=1}^{\infty} E\left\{w_i(\cdot)\right\} E\left\{I\{\theta_i(\boldsymbol{x},\cdot)\}_{\{\lceil k\theta_i(\boldsymbol{x},\cdot)\rceil \le l\}} \,\middle|\, k\right\}, \\
&= E\left\{I\{\theta_1(\boldsymbol{x},\cdot)\}_{\{\lceil k\theta_1(\boldsymbol{x},\cdot)\rceil \le l\}} \,\middle|\, k\right\}, \\
&= G^*_{0\boldsymbol{x}}\left(A_{l,k}\right),
\end{aligned}$$

where, $A_{l,k} = [0, j/k]$ and $G^*_{0\boldsymbol{x}}$ stands for the marginal probability measure of $\theta_i(\boldsymbol{x},\cdot)$, for every $i \in \mathbb{N}$. It follows that

$$E\left\{G_{\boldsymbol{x}}(B_y) \mid k\right\} = \sum_{l=1}^{k} G^*_{0\boldsymbol{x}}\left(A_{l,k}\right) \mathrm{BIN}(l \mid k, l).$$

Applying a similar reasoning, it follows that, for every $\boldsymbol{x}, \boldsymbol{x}_0 \in \mathscr{X}$ and every $y \in [0,1]$,

$$\begin{aligned}
E\left\{G_{\boldsymbol{x}}(B_y) G_{\boldsymbol{x}_0}(B_y) \mid k\right\} &= \sum_{l=1,l_1=1}^{k} \left[\sum_{i=1}^{\infty} E\left\{w_i^2(\cdot)\right\} G^*_{0,\boldsymbol{x},\boldsymbol{x}_0}(A_{l,k} \times A_{l_1,k})\right] \\
&\quad \bar{B}(l, l_1 \mid k, y) \\
&+ \sum_{l=1,l_1=1}^{k} \left[\sum_{i=1, i_1 \neq i}^{\infty} E\left\{w_i(\cdot) w_{i_1}(\cdot)\right\} G^*_{0\boldsymbol{x}}(A_{l,k}) G^*_{0\boldsymbol{x}_0}(A_{l_1,k})\right] \\
&\quad \bar{B}(l, l_1 \mid k, y),
\end{aligned}$$

where, $\bar{B}(l, l_1 \mid k, y) = B(l \mid k, y) B(l_1 \mid k, y)$ and $G^*_{0,\boldsymbol{x},\boldsymbol{x}_0}$ corresponds to the marginal distribution of $(\theta_i(\boldsymbol{x},\cdot), \theta_i(\boldsymbol{x}_0,\cdot))$. In particular, for $\boldsymbol{x} = \boldsymbol{x}_0$,

$$\begin{aligned}
E\left\{G_{\boldsymbol{x}}(B_y)^2 \mid k\right\} &= \sum_{l=1}^{k} \left[\sum_{i=1}^{\infty} E\left\{w_i(\cdot)^2\right\} G_{0\boldsymbol{x}}(A_{l,k})\right] \mathrm{BIN}(l|k,y)^2 + \\
&\quad \sum_{l=1,l_1=1}^{k} \left[\sum_{i=1, i_1 \neq i}^{\infty} E\left\{w_i(\cdot) w_{i_1}(\cdot)\right\} G_{0\boldsymbol{x}}(A_{l,k}) G_{0\boldsymbol{x}}(A_{l_1,k})\right] \\
&\quad \bar{B}(l, l_1 \mid k, y).
\end{aligned}$$

Now, since the elements of $\mathscr{H}$ are continuous functions of $\boldsymbol{x}$ and, for every $i \in \mathbb{N}$ and every $\{\boldsymbol{x}_j\}_1^\infty$, with $\boldsymbol{x}_j \in \mathscr{X}$, such that $\lim_{j\to+\infty} \boldsymbol{x}_j \longrightarrow \boldsymbol{x}_0 \in \mathscr{X}$, $z_i(\boldsymbol{x},\cdot)$ converges in distribution to $z_i(\boldsymbol{x}_0,\cdot)$ as $j \longrightarrow +\infty$, it follows that $\theta_i(\boldsymbol{x}_j,\cdot)$ converges in distribution to $\theta_i(\boldsymbol{x}_0,\cdot)$, as $j \longrightarrow +\infty$. Now a few applications of Lebesgue dominated convergence theorem imply that

$$\begin{aligned}
\lim_{j\longrightarrow+\infty} E\left\{G_{\boldsymbol{x}_j}(B_y)^2\right\} &= \sum_{l=1}^{\infty} \lim_{j\longrightarrow+\infty} E\left\{G_{\boldsymbol{x}_j}(B_y)^2 \;\middle|\; l\right\} P\{\omega\in\Omega : k(\omega)=l\},\\
&= \sum_{l=1}^{\infty} E\left\{G_{\boldsymbol{x}_0}(B_y)^2 \;\middle|\; l\right\} P\{\omega\in\Omega : k(\omega)=l\},\\
&= E\left\{G_{\boldsymbol{x}_0}(B_y)^2\right\},
\end{aligned}$$

$$\begin{aligned}
\lim_{j\longrightarrow+\infty} E\left\{G_{\boldsymbol{x}_j}(B_y)G_{\boldsymbol{x}_0}(B_y)\right\} &= \sum_{l=1}^{\infty} \lim_{j\longrightarrow+\infty} E\left\{G_{\boldsymbol{x}_j}(B_y)G_{\boldsymbol{x}_0}(B_y) \;\middle|\; l\right\} P\{\omega\in\Omega : k(\omega)=l\},\\
&= \sum_{l=1}^{\infty} E\left\{G_{\boldsymbol{x}_0}(B_y)^2 \;\middle|\; l\right\} P\{\omega\in\Omega : k(\omega)=l\},\\
&= E\left\{G_{\boldsymbol{x}_0}(B_y)^2\right\},
\end{aligned}$$

and

$$\lim_{j\longrightarrow+\infty} E\left\{G_{\boldsymbol{x}_j}(B_y)\right\} = E\left\{G_{\boldsymbol{x}_0}(B_y)\right\},$$

which completes the proof of the theorem. $\square$

Theorem B.3. *Let* $\{G_{\boldsymbol{x}} : \boldsymbol{x}\in\mathscr{X}\} \sim w\mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_2, \mathscr{H}, \alpha)$. *Asumme that there exists a positive constant* γ *such that if* $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathscr{X}$ *and* $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| > \gamma$, *then* $Cov\left[\mathbb{I}_{\{z_i(\boldsymbol{x}_1,\cdot)\in A_1\}}, \mathbb{I}_{\{z_i(\boldsymbol{x}_2,\cdot)\in A_2\}}\right] = 0$, *for every* $A_1, A_2 \in \mathscr{B}(\mathbb{R})$. *Assume also that for every* $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathscr{X}$ *such that* $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| > \gamma$, *and for every sequence* $\{(\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j})\}_1^\infty \subset \mathscr{X}^2$, *such that* $\lim_{j\to+\infty}(\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j}) = (\boldsymbol{x}_1, \boldsymbol{x}_2)$, *we have that* $(z_i(\boldsymbol{x}_{1j},\cdot), z_i(\boldsymbol{x}_{2j},\cdot)) \xrightarrow{\mathscr{L}} (z_i(\boldsymbol{x}_1,\cdot), z_i(\boldsymbol{x}_2,\cdot))$, *as* $j \longrightarrow +\infty$. *Then, for every*

$y \in [0,1]$,

$$\lim_{j \to +\infty} Cov\left[G(\boldsymbol{x}_{1j},\cdot)(B_y), G(\boldsymbol{x}_{2j},\cdot)(B_y)\right] = Cov\left[\sum_{l=1}^{k(\cdot)} G^*_{0\boldsymbol{x}_1}\left(A_{l,k(\cdot)}\right)\mathrm{BIN}(l \mid k(\cdot), y),\ \sum_{l=1}^{k(\cdot)} G^*_{0\boldsymbol{x}_2}\left(A_{l,k(\cdot)}\right)\mathrm{BIN}(l \mid k(\cdot), y)\right]$$

where $B_y = [0,y]$, $A_{j,k(\cdot)} = [0, l/k(\cdot)]$, $G^*_{0\boldsymbol{x}}$ stands for the marginal probability measure of $\theta_1(\boldsymbol{x},\cdot)$ and $\mathrm{BIN}(\cdot \mid k, y)$ stands for the probability mass function of the binomial distribution with parameters (k,y).

PROOF. Notice that for every $\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j} \in \mathscr{X}$,

$$\begin{aligned}
&E\left[G(\boldsymbol{x}_{1j},\cdot)(B_y), G(\boldsymbol{x}_{2j},\cdot)(B_y) \mid k\right] - E\left[G(\boldsymbol{x}_{1j},\cdot)(B_y) \mid k\right] E\left\{G(\boldsymbol{x}_{2j},\cdot)(B_y) \mid k\right] \\
&= \sum_{l=1,l_1=1}^{k} \sum_{i=1,i_1=1}^{\infty} E\left[w_i(\cdot) w_{i_1}(\cdot)\right] E\left[\mathbb{I}_{\left\{(\theta_i(\boldsymbol{x}_{1j},\cdot),\theta_{i_1}(\boldsymbol{x}_{2j},\cdot)) \in A_{l,k} \times A_{l_1,k}\right\}}\right] \bar{B}(l, l_1 \mid k, y) \\
&- \sum_{l=1,l_1=1}^{k} \sum_{i=1,i_1=1}^{\infty} E\left[w_i(\cdot)\right] E\left[w_{i_1}(\cdot)\right] E\left[\mathbb{I}_{\left\{\theta_i(\boldsymbol{x}_{1j},\cdot) \in A_{l,k}\right\}}\right] E\left[\mathbb{I}_{\left\{\theta_{i_1}(\boldsymbol{x}_{2j},\cdot) \in A_{l_1,k}\right\}}\right] \bar{B}(l, l_1 \mid k, y).
\end{aligned}$$

where, $\bar{B}(l, l_1 \mid k, y) = \mathrm{BIN}(l \mid k, y) \times \mathrm{BIN}(l_1 \mid k, y)$. Now, since the elements of $\mathscr{H}$ are continuous functions of $\boldsymbol{x}$ and, for every $i \in \mathbb{N}$, it follows that

$$\lim_{j \to \infty} E\left[\mathbb{I}_{\left\{(\theta_i(\boldsymbol{x}_{1j},\cdot),\theta_i(\boldsymbol{x}_{2j},\cdot)) \in A_{l,k} \times A_{l_1,k}\right\}}\right] = E\left[\mathbb{I}_{\left\{\theta_i(\boldsymbol{x}_1,\cdot) \in A_{l,k}\right\}}\right] E\left[\mathbb{I}_{\left\{\theta_i(\boldsymbol{x}_2,\cdot) \in A_{l_1,k}\right\}}\right]$$

and

$$\lim_{j \to \infty} E\left[\mathbb{I}_{\left\{\theta_i(\boldsymbol{x}_{1j},\cdot) \in A_{l,k}\right\}}\right] E\left[\mathbb{I}_{\left\{\theta_i(\boldsymbol{x}_{2j},\cdot) \in A_{l_1,k}\right\}}\right] = E\left[\mathbb{I}_{\left\{\theta_i(\boldsymbol{x}_1,\cdot) \in A_{l,k}\right\}}\right] E\left[\mathbb{I}_{\left\{\theta_i(\boldsymbol{x}_2,\cdot) \in A_{l_1,k}\right\}}\right].$$

Thus, by Lebesgue's dominated convergence theorem

$$
\begin{aligned}
&\lim_{j\to\infty} E\left[G(\boldsymbol{x}_{1j},\cdot)(B_y), G(\boldsymbol{x}_{2j},\cdot)(B_y) \mid k\right] - E\left[G(\boldsymbol{x}_{1j},\cdot)(B_y) \mid k\right] E\left\{G(\boldsymbol{x}_{2j},\cdot)(B_y) \mid k\right] \\
&= \sum_{l=1,l_1=1}^{k} \bar{B}(l,l_1 \mid k,y) \sum_{i=1,i_1=1}^{\infty} E\left[\mathbb{I}_{\left\{\theta_i(\boldsymbol{x}_1,\cdot)\in A_{l,k}\right\}}\right] E\left[\mathbb{I}_{\left\{\theta_{i_1}(\boldsymbol{x}_2,\cdot)\in A_{l_1,k}\right\}}\right] E\left[w_i(\cdot)w_{i_1}(\cdot)\right] \\
&- \sum_{l=1,l_1=1}^{k} \bar{B}(l,l_1 \mid k,y) \sum_{i=1,i_1=1}^{\infty} E\left[\mathbb{I}_{\left\{\theta_i(\boldsymbol{x}_1,\cdot)\in A_{l,k}\right\}}\right] E\left[\mathbb{I}_{\left\{\theta_{i_1}(\boldsymbol{x}_2,\cdot)\in A_{l_1,k}\right\}}\right] E\left[w_i(\cdot)\right]E[w_{i_1}(\cdot)] \\
&= \sum_{l=1,l_1=1}^{k} \bar{B}(l,l_1 \mid k,y) \sum_{i=1,i_1=1}^{\infty} E\left[\mathbb{I}_{\left\{\theta_i(\boldsymbol{x}_1,\cdot)\in A_{l,k}\right\}}\right] E\left[\mathbb{I}_{\left\{\theta_{i_1}(\boldsymbol{x}_2,\cdot)\in A_{l_1,k}\right\}}\right] Cov[w_i(\cdot), w_{i_1}(\cdot)] \\
&= \sum_{l=1,l_1=1}^{k} \bar{B}(l,l_1 \mid k,y) E\left[\mathbb{I}_{\left\{\theta_1(\boldsymbol{x}_1,\cdot)\in A_{l,k}\right\}}\right] E\left[\mathbb{I}_{\left\{\theta_1(\boldsymbol{x}_2,\cdot)\in A_{l_1,k}\right\}}\right] Cov\left[\sum_{i=1}^{\infty} w_i(\cdot), \sum_{i_1=1}^{\infty} w_{i_1}(\cdot)\right] \\
&= 0,
\end{aligned}
$$

for every $k \in \mathbb{N}$, and therefore,

$$
\begin{aligned}
\lim_{j\to\infty} & Cov\left[G(\boldsymbol{x}_{1j},\cdot)(B_y), G(\boldsymbol{x}_{2j},\cdot)(B_y)\right] \\
&= \lim_{j\to\infty} E\left[Cov\left[G(\boldsymbol{x}_{1j},\cdot)(B_y), G(\boldsymbol{x}_{2j},\cdot)(B_y) \mid k\right]\right] \\
&\quad + \lim_{j\to\infty} Cov\left[E\left[G(\boldsymbol{x}_{1j},\cdot)(B_y) \mid k\right], E\left[G(\boldsymbol{x}_{2j},\cdot)(B_y) \mid k\right]\right] \\
&= E\left[\lim_{j\to\infty} Cov\left[G(\boldsymbol{x}_{1j},\cdot)(B_y), G(\boldsymbol{x}_{2j},\cdot)(B_y) \mid k\right]\right] \\
&\quad + Cov\left[\lim_{j\to\infty} E\left[G(\boldsymbol{x}_{1j},\cdot)(B_y) \mid k\right], \lim_{j\to\infty} E\left[G(\boldsymbol{x}_{2j},\cdot)(B_y) \mid k\right]\right] \\
&= Cov\left[E\left[G(\boldsymbol{x}_1,\cdot)(B_y) \mid k\right], E\left[G(\boldsymbol{x}_2,\cdot)(B_y) \mid k\right]\right]
\end{aligned}
$$

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

$$
E\left[G(\boldsymbol{x},\cdot)(B_y) \mid k\right] = \sum_{l=1}^{k} G^*_{0\boldsymbol{x}}\left(A_{l,k}\right) \begin{pmatrix} k \\ l \end{pmatrix} y^l (1-y)^{k-l},
$$

which completes the proof of the theorem. □

Theorem B.4. *Let* $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \sim w\mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_2, \mathscr{H}, \alpha)$. *Assume that for every* $\{(\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j})\}_1^\infty$

$\subset \mathscr{X}^2$*, such that* $\lim_{j\to+\infty}(\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j}) = (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathscr{X}^2$*, we have that* $(z_i(\boldsymbol{x}_{1j}, \cdot), z_i(\boldsymbol{x}_{2j}, \cdot)) \xrightarrow{\mathscr{L}} (z_i(\boldsymbol{x}_1, \cdot), z_i(\boldsymbol{x}_2, \cdot))$*, as* $j \longrightarrow +\infty$*. Then, for every* $y = [0, 1]$*,*

$$\lim_{j\to\infty} \rho\left[G(\boldsymbol{x}_{1j}, \cdot)(B_y), G(\boldsymbol{x}_{2j}, \cdot)(B_y)\right] = \rho\left[G(\boldsymbol{x}_1, \cdot)(B_y), G(\boldsymbol{x}_2, \cdot)(B_y)\right],$$

where $B_y = [0, y]$*.*

PROOF. Since the elements of $\mathscr{H}$ are continuous functions of $\boldsymbol{x}$ and, for every $i \in \mathbb{N}$, and every $\{(\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j})\}_1^\infty$, with $\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j} \in \mathscr{X}$, such that $\lim_{j\to+\infty}(\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j}) = (\boldsymbol{x}_1, \boldsymbol{x}_2)$, with $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathscr{X}$, it follows that $(z_i(\boldsymbol{x}_{1j}, \cdot), z_i(\boldsymbol{x}_{2j}, \cdot))$ and $z_i(\boldsymbol{x}_{lj}, \cdot)$ converges in distribution to $(z_i(\boldsymbol{x}_1, \cdot), z_i(\boldsymbol{x}_2, \cdot))$ and $z_i(\boldsymbol{x}_l, \cdot)$, respectively, as $j \longrightarrow +\infty$, for $l = 1, 2$. Finally, since the correlation, ρ, is a continuous function of $G^*_{0\boldsymbol{x}_l}(A_{j,k})$ and $G^*_{0,\boldsymbol{x}_1,\boldsymbol{x}_2}(A_{j,k} \times A_{j_1,k})$, $k \in \mathbb{N}$, $j, j_1 \in \{1, \ldots, k\}$ and $l = 1, 2$, it follows that $(\boldsymbol{x}_1, \boldsymbol{x}_2) \mapsto \rho\left[G(\boldsymbol{x}_1, \cdot)(B_y), G(\boldsymbol{x}_2, \cdot)(B_y)\right]$ is also a continuous function. $\square$

Theorem B.5. *Let* $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \sim w\mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_2, \mathscr{H}, \alpha)$*. If for every* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d \in \mathscr{X}^d$*,* $d \geq 1$*, the joint distribution of* $(z_i(\boldsymbol{x}_1, \cdot), \ldots, z_i(\boldsymbol{x}_d, \cdot))$ *has full support on* $\mathbb{R}^d$*, and* $k(\cdot)$ *has full support on* $\mathbb{N}$*, then* $\mathscr{P}([0,1])^{\mathscr{X}}$ *is the support of* $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$ *under the weak product topology.*

PROOF. The proof is similar to the one of Theorem 3.5. In this case,

$$W_l(\boldsymbol{x}, \omega, k) = \sum_{i=1}^\infty w_i(\omega) I\{\theta_i(\boldsymbol{x}, \omega)\}_{\{\lceil k\theta_i(\boldsymbol{x},\omega)\rceil = l\}}.$$

Now, the non–singularity of the beta distribution, the assumptions that the joint distribution of

$$(z_i(\boldsymbol{x}_1, \cdot), \ldots, z_i(\boldsymbol{x}_d, \cdot))$$

has full support on $\mathbb{R}^d$, for every $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d \in \mathscr{X}^d$, $d \geq 1$, and that $k(\cdot)$ has full support on $\mathbb{N}$, imply that

$$P\left\{\omega\in\Omega : \left|\int f_{ij}dG(\boldsymbol{x}_i,\omega)-\int f_{ij}dQ_{x_i}\right| < \epsilon_i,\ i=1,\dots,T,\ j=1,\dots,K_i\right\} \geq \sum_{k_1>k_0} P\left\{\omega\in\Omega : N_{ik_1}(\omega)\leq \frac{\epsilon_i}{4M_{ij}k_1}, i=1,\dots,T,\ j=1,\dots,K_i, k(\omega)=k_1\right\} > 0,$$

by Theorem 6 in Barrientos et al. (2012), which completes the proof of the theorem. □

Theorem B.6. *Let* $\{G_{\boldsymbol{x}} : \boldsymbol{x}\in\mathscr{X}\} \sim w\mathrm{DBPP}(\lambda,\boldsymbol{\Psi}_2,\mathscr{H},\alpha)$*. If for every* $\boldsymbol{x}_1,\dots,\boldsymbol{x}_d\in\mathscr{X}^d$*,* $d\geq 1$*, the joint distribution of* $(z_i(\boldsymbol{x}_1,\cdot),\dots,z_i(\boldsymbol{x}_d,\cdot))$ *has full support on* $\mathbb{R}^d$*, and* $k(\cdot)$ *has full support on* $\mathbb{N}$*, then* $\mathscr{D}\left([0,1]\right)^{\mathscr{X}}$ *is the support of* $\{G_{\boldsymbol{x}} : \boldsymbol{x}\in\mathscr{X}\}$ *under the* L_∞ *product topology.*

PROOF. The proof is similar to the one of Theorem 3.6. In this case,

$$W_l(\boldsymbol{x},\omega,k)=\sum_{i=1}^{\infty} w_i(\omega) I\{\theta_i(\boldsymbol{x},\omega)\}_{\{\lceil k\theta_i(\boldsymbol{x},\omega)\rceil = l\}}.$$

Now, the non–singularity of the beta distribution, the assumptions that the joint distribution of

$$(z_i(\boldsymbol{x}_1,\cdot),\dots,z_i(\boldsymbol{x}_d,\cdot))$$

has full support on $\mathbb{R}^d$, for every $\boldsymbol{x}_1,\dots,\boldsymbol{x}_d\in\mathscr{X}^d$, $d\geq 1$, and that $k(\cdot)$ has full support on $\mathbb{N}$, imply that

$$P\left\{\omega\in\Omega : \sup_{y\in[0,1]} |g(\boldsymbol{x}_i,\omega)(y)-q_{\boldsymbol{x}_i}(y)| < \epsilon,\ i=1,\dots,T\right\} \geq \sum_{k_1\geq k_0} P\left\{\omega\in\Omega : N_{ik_1}(\omega)\leq \frac{\epsilon}{4Mk_1}, i=1,\dots,T, k(\omega)=k_1\right\} > 0,$$

by Theorem 6 in Barrientos et al. (2012), which completes the proof of the theorem. □

Theorem B.7. *Let* $\{G_{\boldsymbol{x}} : \boldsymbol{x}\in\mathscr{X}\} \sim w\mathrm{DBPP}(\lambda,\boldsymbol{\Psi}_2,\mathscr{H},\alpha)$*. If* $\mathscr{X}$ *is a compact set,* $k(\cdot)$ *has full support on* $\mathbb{N}$*, and the processes used in the definition of the* wDBPP *are such that, for any*

$[0,1]$–valued continuous function defined on $\mathscr{X}$, f, and $\epsilon > 0$,

$$P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} |h_{\boldsymbol{x}}(z_i(\boldsymbol{x},\omega)) - f(\boldsymbol{x})| < \epsilon\right\} > 0,$$

then $\tilde{\mathscr{D}}\left([0,1]\right)^{\mathscr{X}}$ is the support of $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$ under the L_∞ topology.

PROOF. Let $\{Q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \tilde{\mathscr{D}}([0,1])^{\mathscr{X}}$, with density functions $\{q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$. An application of the triangle inequality implies that, for every $\boldsymbol{x} \in \mathscr{X}$, $y \in [0,1]$ and almost every $\omega \in \Omega$,

$$|g(\boldsymbol{x},\omega)(y) - q_{\boldsymbol{x}}(y)| \leq |\mathrm{bp}(y \mid k, Q_{\boldsymbol{x}}) - q_{\boldsymbol{x}}(y)| + |g(\boldsymbol{x},\omega)(y) - \mathrm{bp}(y \mid k, Q_{\boldsymbol{x}})|\,.$$

By Lemma B.1, it follows that there exists $k_0 \in \mathbb{N}$ such that

$$\sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{y\in[0,1]} |\mathrm{bp}(y \mid k_0, Q_{\boldsymbol{x}}) - q_{\boldsymbol{x}}(y)| \leq \frac{\epsilon}{2}.$$

Now, notice that for every $\boldsymbol{x} \in \mathscr{X}$,

$$\begin{aligned}
\sup_{y\in[0,1]} |g(\boldsymbol{x},\omega)(y) - \mathrm{bp}(y \mid k_0, Q_{\boldsymbol{x}})| &= \sup_{y\in[0,1]} \left| \sum_{l=1}^{k_0} W_l(\boldsymbol{x},\omega,k_0)\beta\left(y|l,k_0-l+1\right) \right. \\
&\quad \left. - \sum_{l=1}^{k_0} Q_{\boldsymbol{x}}\left(\frac{l}{k_0}, \frac{l-1}{k_0}\right] \beta\left(y|l,k_0-l+1\right) dy \right|, \\
&\leq M k_0 \max_{l\in\{1,\ldots,k_0\}} \left| W_l(\boldsymbol{x},\omega,k_0) - Q_{\boldsymbol{x}}\left(\frac{l}{k_0}, \frac{l-1}{k_0}\right] \right|,
\end{aligned}$$

where, $W_l(\boldsymbol{x},\omega,k_0) = \sum_{i=1}^{\infty} w_i(\omega) I\{\theta_i(\boldsymbol{x},\omega)\}_{\{\lceil k\theta_i(\boldsymbol{x},\omega)\rceil = l\}}$ and

$$M = \sup_{l\in\{1,\ldots,k_0\}} \sup_{y\in[0,1]} \beta\left(y|l,k_0-l+1\right).$$

Since $\boldsymbol{x} \mapsto Q_{\boldsymbol{x}}\left(\frac{l}{k_0}, \frac{l-1}{k_0}\right]$ is a continuous function, then for each $l \in \{1,\ldots,k_0\}$ there exist step functions of the form

$$S_l(\boldsymbol{x}) = \sum_{j=1}^{m_l} a_{j,l} \mathbb{I}_{\mathscr{X}_{j,l}}(\boldsymbol{x}),$$

$l = 1, \ldots, k_0$, where, m_l is a natural number, $\{a_{j,l}\}_{j=1}^{m_l}$ are positive numbers, $\{\mathscr{X}_{j,l}\}_{j=1}^{m_l}$ is a partition of $\mathscr{X}$, each $\mathscr{X}_{j,l}$ has positive Lebesgue measure, and $\mathbb{I}_A$ is the indicator function of the set A, such that for every $\boldsymbol{x} \in \mathscr{X}$, $\sum_{l=1}^{k_0} S_l(\boldsymbol{x}) = 1$, and

$$\sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{l\in\{1,\ldots,k_0\}} \left| S_l(\boldsymbol{x}) - Q_{\boldsymbol{x}}\left(\frac{l}{k_0}, \frac{l-1}{k_0}\right]\right| \leq \frac{\epsilon}{4Mk_0}.$$

Notice that the above step functions, in turn, can be expressed as a discrete measure of the form

$$S_l(\boldsymbol{x}) = \sum_{i=1}^{m} \tilde{w}_i I\{\tilde{\theta}_i(\boldsymbol{x})\}_{\{\lceil k_0\tilde{\theta}_i(\boldsymbol{x})\rceil = l\}},$$

where m is a natural number, $\{\tilde{w}_i\}_{i=1}^m$ are positive numbers such that $\tilde{w}_i \in (0, \frac{\epsilon}{8Mk_0})$ and $\sum_{i=1}^m \tilde{w}_i = 1$, $\tilde{\theta}_i(\boldsymbol{x})$ is a continuous function, $i = 1, \ldots, m$, such that the set

$$\left\{\boldsymbol{x} \in \mathscr{X} : k_0\tilde{\theta}_i(\boldsymbol{x}) = l,\ l = 1, \ldots, k_0,\ i = 1, \ldots, m\right\}$$

has zero Lebesgue measure. Therefore, it is possible to ensure the existence of a positive constant, γ, such that if $\sup_{\boldsymbol{x}\in\mathscr{X}} \left|\theta_i(\boldsymbol{x}, \omega) - \tilde{\theta}_i(\boldsymbol{x})\right| < \gamma$, then

$$\left|\lceil k_0\theta_i(\boldsymbol{x}, \omega)\rceil - \lceil k_0\tilde{\theta}_i(\boldsymbol{x})\rceil\right| \leq 1.$$

Furthermore, if for some $i_1, i_2 \in \{1, \ldots, m\}$ and $l \in \{1, \ldots, k_0\}$, there exist $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \in \mathscr{X}$ such that

$$k_0\tilde{\theta}_{i_1}\left(\boldsymbol{x}^{(1)}\right) = k_0\tilde{\theta}_{i_2}\left(\boldsymbol{x}^{(2)}\right) = l,$$

then there exist $\boldsymbol{x}_1^{(1)}, \boldsymbol{x}_2^{(1)}, \boldsymbol{x}_1^{(2)}, \boldsymbol{x}_2^{(2)} \in \mathscr{X}$ such that

$$\tilde{\theta}_{i_1}\left(\boldsymbol{x}_1^{(1)}\right) + \gamma = \tilde{\theta}_{i_1}\left(\boldsymbol{x}_2^{(1)}\right) - \gamma = \tilde{\theta}_{i_2}\left(\boldsymbol{x}_1^{(2)}\right) + \gamma = \tilde{\theta}_{i_2}\left(\boldsymbol{x}_2^{(2)}\right) - \gamma = \frac{l}{k_0},$$

$$\min\{\boldsymbol{x}_1^{(1)}, \boldsymbol{x}_2^{(1)}\} < \boldsymbol{x}_1^{(1)} < \max\{\boldsymbol{x}_1^{(1)}, \boldsymbol{x}_2^{(1)}\},$$

$$\min\{\boldsymbol{x}_1^{(2)}, \boldsymbol{x}_2^{(2)}\} < \boldsymbol{x}_1^{(2)} < \max\{\boldsymbol{x}_1^{(2)}, \boldsymbol{x}_2^{(2)}\}$$

and

$$\left(\min\{\boldsymbol{x}_1^{(1)}, \boldsymbol{x}_2^{(1)}\}, \max\{\boldsymbol{x}_1^{(1)}, \boldsymbol{x}_2^{(1)}\}\right) \bigcap \left(\min\{\boldsymbol{x}_1^{(2)}, \boldsymbol{x}_2^{(2)}\}, \max\{\boldsymbol{x}_1^{(2)}, \boldsymbol{x}_2^{(2)}\}\right) = \emptyset.$$

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

- For $i = 1, \dots, m$,
$$\sup_{\boldsymbol{x} \in \mathscr{X}} \left| h_{\boldsymbol{x}}\left(z_l(\boldsymbol{x}, \omega)\right) - \tilde{\theta}_i(\boldsymbol{x}) \right| \leq \gamma.$$
- For $i = 1$,
$$|v_i(\omega) - \tilde{w}_i| \leq \frac{\epsilon}{8Mk_0(2^m - 1)}.$$
- For $i = 2, \dots, m$,
$$\left| v_i(\omega) - \frac{\tilde{w}_i}{1 - \sum_{j=1}^{i-1} \tilde{w}_j} \right| \leq \frac{\epsilon}{8Mk_0(2^m - 1)}.$$
- $k(\omega) = k_0$.

Then for almost every $\omega \in \Omega_0$, it follows that

$$\max_{l \in \{1, \dots, k_0\}} \left| W_l(\boldsymbol{x}, \omega, k_0) - \sum_{i=1}^{m} \tilde{w}_i I\{\tilde{\theta}_i(\boldsymbol{x})\}_{\{\lceil k_0 \tilde{\theta}_i(\boldsymbol{x}) \rceil = l\}} \right| \leq \frac{\epsilon}{4Mk_0},$$

wich implies that

$$\sup_{\boldsymbol{x} \in \mathscr{X}} \sup_{y \in [0,1]} |g(\boldsymbol{x}, \omega)(y) - \mathrm{bp}(y \mid k, Q_{\boldsymbol{x}})| < \frac{\epsilon}{2},$$

and, therefore,

$$
\begin{aligned}
&P\left\{\omega\in\Omega:\sup_{\boldsymbol{x}\in\mathscr{X}}\sup_{y\in[0,1]}|g(\boldsymbol{x},\omega)(y)-q_{\boldsymbol{x}}(y)|\le\epsilon\right\}\ge\\
&\qquad\prod_{i=1}^{m}P\left\{\omega\in\Omega:\sup_{\boldsymbol{x}\in\mathscr{X}}\left|h_{\boldsymbol{x}}\left(z_l(\boldsymbol{x},\omega)\right)-\tilde{\theta}_i(\boldsymbol{x})\right|\le\gamma\right\}\times\\
&\qquad P\left\{\omega\in\Omega:|v_i(\omega)-\tilde{w}_i|\le\frac{\epsilon}{8k_0M(2^m-1)}\right\}\times\\
&\qquad\prod_{l=2}^{m}P\left\{\omega\in\Omega:\left|v_i(\omega)-\frac{\tilde{w}_i}{1-\sum_{j=1}^{i-1}\tilde{w}_j}\right|\le\frac{\epsilon}{8k_0M(2^m-1)}\right\}\times\\
&\qquad P\left\{\omega\in\Omega:k(\omega)=k_0\right\}.
\end{aligned}
$$

Now, note that the continuity of the functions $\{\tilde{\theta}_i\}_{i=1}^m$ and the non–singularity of the beta distribution imply that

$$
\begin{aligned}
&\prod_{i=1}^{m}P\left\{\omega\in\Omega:\sup_{\boldsymbol{x}\in\mathscr{X}}\left|h_{\boldsymbol{x}}\left(z_l(\boldsymbol{x},\omega)\right)-\tilde{\theta}_i(\boldsymbol{x})\right|\le\gamma\right\}\times\\
&P\left\{\omega\in\Omega:|v_i(\omega)-\tilde{w}_i|\le\frac{\epsilon}{8k_0M(2^m-1)}\right\}\times\\
&\prod_{l=2}^{m}P\left\{\omega\in\Omega:\left|v_i(\omega)-\frac{\tilde{w}_i}{\prod_{j<l}1-\sum_{j=1}^{i-1}\tilde{w}_j}\right|\le\frac{\epsilon}{8k_0M(2^m-1)}\right\}\\
&>0.
\end{aligned}
$$

Finally, since $k(\cdot)$ has full support on $\mathbb{N}$, it follows that

$$
P\left\{\omega\in\Omega:\sup_{\boldsymbol{x}\in\mathscr{X}}\sup_{y\in[0,1]}|g(\boldsymbol{x},\omega)(y)-q_{\boldsymbol{x}}(y)|\le\epsilon\right\}>0,
$$

which completes the proof of the theorem. □

Theorem B.8. *Let* $\{G_{\boldsymbol{x}}:\boldsymbol{x}\in\mathscr{X}\}\sim w\mathrm{DBPP}(\lambda,\boldsymbol{\Psi}_2,\mathscr{H},\alpha)$*. Assume that* $\mathscr{X}$ *is a compact set,* $k(\cdot)$ *has full support on* $\mathbb{N}$*, and the processes used in the definition of the* wDBPP *are such* *that, for any* $\epsilon > 0$ *and* $[0,1]$*–valued continuous function* f *defined on* $\mathscr{X}$*, we have*

$$P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} |h_{\boldsymbol{x}}(z_i(\boldsymbol{x},\omega)) - f(\boldsymbol{x})| < \epsilon\right\} > 0.$$

Then,

$$P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} \int_0^1 q_{\boldsymbol{x}}(y) \log\left(\frac{q_{\boldsymbol{x}}(y)}{g(\boldsymbol{x},\omega)(y)}\right) dy > \epsilon\right\} > 0,$$

for every $\epsilon > 0$ *and every* $\{Q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \tilde{\mathscr{D}}([0,1])$*, with density functions* $\{q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$*.*

PROOF. This proof is similar to the one of Corollary 3.1.

Theorem B.9. *Let* $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \sim w\mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_2, \mathscr{H}, \alpha)$*. Assume that* $\mathscr{X}$ *is a compact set,* $k(\cdot)$ *has full support on* $\mathbb{N}$*, and the processes used in the definition of the* wDBPP *are such that, for any* $\epsilon > 0$ *and* $[0,1]$*–valued continuous function* f *defined on* $\mathscr{X}$*, we have*

$$P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} |h_{\boldsymbol{x}}(z_i(\boldsymbol{x},\omega)) - f(\boldsymbol{x})| < \epsilon\right\} > 0.$$

Then the posterior distribution associated with the random joint distribution induced by the wDBPP *model,* $m^{(\cdot)}(y,\boldsymbol{x}) = q(\boldsymbol{x})g(\boldsymbol{x},\cdot)(y)$*, where* q *is the density generating the predictors, is weakly consistent at any joint distribution of the form* $m_0(y,\boldsymbol{x}) = q(\boldsymbol{x})q_0(y \mid \boldsymbol{x})$*, where* $\{q_0(\cdot \mid \boldsymbol{x}) : \boldsymbol{x} \in \mathscr{X}\} \in \widetilde{\mathscr{D}}\left([0,1]\right)^{\mathscr{X}}$*.*

PROOF. This proof is similar to the one of Theorem 3.8 (given in the main document).

B.3 Properties of the θDBPP

In this appendix we adapt the results derived for the general model to the special case of the θDBPP. The proofs closely follow those given in Section B.1.

Theorem B.10. *Let* $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \sim \theta\mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_1, \mathscr{V}, G_0)$*. If for every* $\{\boldsymbol{x}_j\}_1^\infty \subset \mathscr{X}$*, such that* $\lim_{j\to+\infty} \boldsymbol{x}_j \longrightarrow \boldsymbol{x}_0 \in \mathscr{X}$*, we have* $\eta_i(\boldsymbol{x}_j,\cdot) \xrightarrow{\mathscr{L}} \eta_i(\boldsymbol{x}_0,\cdot)$*, as* $j \longrightarrow +\infty$*, then, for all* $y \in (0,1)$*,*

$$\lim_{j\to+\infty} \rho\left[G(\boldsymbol{x}_j,\cdot)(B_y), G(\boldsymbol{x}_0,\cdot)(B_y)\right] = 1,$$

where $\rho(A, B)$ *denotes the Pearson correlation between* A *and* B*, and* $B_y = [0, y]$.

PROOF. Notice that for every $y \in [0, 1]$ and every $\boldsymbol{x} \in \mathscr{X}$,

$$\begin{aligned} E\left\{G(\boldsymbol{x}, \cdot)(B_y) \mid k\right\} &= E\left\{\sum_{l=1}^{k} \left[F^*(\boldsymbol{x}, \cdot)(l)\right] \mathrm{BIN}(l \mid k, y) \,\middle|\, k\right\}, \\ &= \sum_{j=1}^{k} E\left\{F^*(\boldsymbol{x}, \cdot)\,(l) \mid k\right\} \mathrm{BIN}(l \mid k, y), \end{aligned}$$

where, $\mathrm{BIN}(\cdot \mid k, y)$ stands for the probability mass function of the binomial distribution with parameters (k, y), and

$$F^*(\boldsymbol{x}, \cdot)\,(l) = \sum_{i=1}^{\infty} w_i(\boldsymbol{x}, \cdot) I\{\theta_i(\cdot)\}_{\{\lceil k\theta_i(\cdot)\rceil \leq l\}}.$$

Now, notice that the independence of the stochastic processes and the i.i.d. property of the corresponding elements, imply that

$$\begin{aligned} E\left\{F^*(\boldsymbol{x}, \cdot)\,(l) \mid k\right\} &= E\left\{\sum_{i=1}^{\infty} w_i(\boldsymbol{x}, \cdot) I\{\theta_i(\cdot)\}_{\{\lceil k\theta_i(\cdot)\rceil \leq l\}} \,\middle|\, k\right\}, \\ &= \sum_{i=1}^{\infty} E\left\{w_i(\boldsymbol{x}, \cdot)\right\} E\left\{I\{\theta_i(\cdot)\}_{\{\lceil k\theta_i(\cdot)\rceil \leq l\}} \,\middle|\, k\right\}, \\ &= E\left\{I\{\theta_1(\cdot)\}_{\{\lceil k\theta_1(\cdot)\rceil \leq l\}} \,\middle|\, k\right\}, \\ &= G_0\left(A_{l,k}\right), \end{aligned}$$

where $A_{l,k} = [0, l/k]$. It follows that

$$E\left\{G_{\boldsymbol{x}}(B_y) \mid k\right\} = \sum_{l=1}^{k} G_0\left(A_{l,k}\right) \mathrm{BIN}(l \mid k, l).$$

Applying a similar reasoning, it follows that for every $\boldsymbol{x}, \boldsymbol{x}_0 \in \mathscr{X}$ and every $y \in [0,1]$,

$$\begin{aligned} E\left\{G_{\boldsymbol{x}}(B_y)G_{\boldsymbol{x}_0}(B_y) \mid k\right\} &= \sum_{l=1}^{k}\left[\sum_{i=1}^{\infty} E\left\{w_i(\boldsymbol{x},\cdot)w_i(\boldsymbol{x}_0,\cdot)\right\} G_0(A_{l,k})\right] \mathrm{BIN}(l \mid k, y)^2 \\ &+ \sum_{l=1,l_1=1}^{k}\left[\sum_{i=1,i_1\neq i}^{\infty} E\left\{w_i(\boldsymbol{x},\cdot)w_{i_1}(\boldsymbol{x}_0,\cdot)\right\} G_0(A_{l,k})G_0(A_{l_1,k})\right] \\ &\quad \bar{B}(l, l_1 \mid k, y), \end{aligned}$$

where, $\bar{B}(l, l_1 \mid k, y) = \mathrm{BIN}(l \mid k, y) \times \mathrm{BIN}(l_1 \mid k, y)$. In particular, for $\boldsymbol{x} = \boldsymbol{x}_0$,

$$\begin{aligned} E\left\{G_{\boldsymbol{x}}(B_y)^2 \,\middle|\, k\right\} &= \sum_{l=1}^{k}\left[\sum_{i=1}^{\infty} E\left\{w_i(\boldsymbol{x},\cdot)^2\right\} G_0(A_{l,k})\right] \mathrm{BIN}(l \mid k, y)^2 + \\ &\quad \sum_{l=1,l_1=1}^{k}\left[\sum_{i=1,i_1\neq i}^{\infty} E\left\{w_i(\boldsymbol{x},\cdot)w_{i_1}(\boldsymbol{x},\cdot)\right\} G_0(A_{l,k})G_0(A_{l_1,k})\right] \\ &\quad \bar{B}(l, l_1 \mid k, y). \end{aligned}$$

Now, since the elements of $\mathscr{V}$ are continuous functions of $\boldsymbol{x}$ and, for every $i \in \mathbb{N}$ and every $\{\boldsymbol{x}_j\}_1^{\infty}$, with $\boldsymbol{x}_j \in \mathscr{X}$, such that $\lim_{j \to +\infty} \boldsymbol{x}_j \longrightarrow \boldsymbol{x}_0 \in \mathscr{X}$, $\eta_i(\boldsymbol{x}, \cdot)$ converges in distribution to $\eta_i(\boldsymbol{x}_0, \cdot)$ as $j \longrightarrow +\infty$, it follows that $w_i(\boldsymbol{x}_j, \cdot)$ converges in distribution to $w_i(\boldsymbol{x}_0, \cdot)$ as $j \longrightarrow +\infty$ and that $\boldsymbol{x} \mapsto E\left\{w_i(\boldsymbol{x},\cdot)\right\}$, $\boldsymbol{x} \mapsto E\left\{w_i(\boldsymbol{x},\cdot)^2\right\}$ and $\boldsymbol{x} \mapsto E\left\{w_i(\boldsymbol{x},\cdot)w_{i_1}(\boldsymbol{x}_0,\cdot)\right\}$ are continuous functions. Now, a few applications of Lebesgue dominated convergence theorem imply that

$$\begin{aligned} \lim_{j \longrightarrow +\infty} E\left\{G_{\boldsymbol{x}_j}(B_y)^2\right\} &= \sum_{l=1}^{\infty} \lim_{j \longrightarrow +\infty} E\left\{G_{\boldsymbol{x}_j}(B_y)^2 \,\middle|\, l\right\} P\{\omega \in \Omega : k(\omega) = l\}, \\ &= \sum_{l=1}^{\infty} E\left\{G_{\boldsymbol{x}_0}(B_y)^2 \,\middle|\, l\right\} P\{\omega \in \Omega : k(\omega) = l\}, \\ &= E\left\{G_{\boldsymbol{x}_0}(B_y)^2\right\}, \end{aligned}$$

$$\begin{aligned}
\lim_{j \longrightarrow +\infty} E\left\{G_{\boldsymbol{x}_j}(B_y) G_{\boldsymbol{x}_0}(B_y)\right\} &= \sum_{l=1}^{\infty} \lim_{j \longrightarrow +\infty} E\left\{G_{\boldsymbol{x}_j}(B_y) G_{\boldsymbol{x}_0}(B_y) \;\middle|\; l\right\} P\{\omega \in \Omega : k(\omega) = l\}, \\
&= \sum_{l=1}^{\infty} E\left\{G_{\boldsymbol{x}_0}(B_y)^2 \;\middle|\; l\right\} P\{\omega \in \Omega : k(\omega) = l\}, \\
&= E\left\{G_{\boldsymbol{x}_0}(B_y)^2\right\},
\end{aligned}$$

and

$$\lim_{j \longrightarrow +\infty} E\left\{G_{\boldsymbol{x}_j}(B_y)\right\} = E\left\{G_{\boldsymbol{x}_0}(B_y)\right\},$$

which completes the proof of the theorem. □

Theorem B.11. *Let* $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \sim \theta\mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_1, \mathscr{V}, G_0)$*. Asumme 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* $Cov\left[\mathbb{I}_{\{\eta_i(\boldsymbol{x}_{1j}, \cdot) \in A_1\}}, \mathbb{I}_{\{\eta_i(\boldsymbol{x}_{2j}, \cdot) \in A_2\}}\right]$ $= 0$*, for every* $A_1, A_2 \in \mathscr{B}(\mathbb{R})$*. Assume also that* $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathscr{X}$ *such that for every* $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| > \gamma$*, and for every sequence* $\{(\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j})\}_1^\infty \subset \mathscr{X}^2$*, such that* $\lim_{j \to +\infty} (\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j}) = (\boldsymbol{x}_1, \boldsymbol{x}_2)$*, we have* $(\eta_i(\boldsymbol{x}_{1j}, \cdot), \eta_i(\boldsymbol{x}_{2j}, \cdot)) \xrightarrow{\mathscr{L}} (\eta_i(\boldsymbol{x}_1, \cdot), \eta_i(\boldsymbol{x}_2, \cdot))$*, as* $j \longrightarrow +\infty$*. Then, for every* $y \in [0, 1]$*,*

$$\begin{aligned}
\lim_{j \to +\infty} Cov\left[G(\boldsymbol{x}_{1j}, \cdot)(B_y), G(\boldsymbol{x}_{2j}, \cdot)(B_y)\right] \quad = \\
\sum_{t=1}^{\infty} P\{\omega \in \Omega : k(\omega) = t\} \\
\sum_{l=1, l_1=1}^{t} \bar{B}(l, l_1 \mid t, y) \sum_{i=1}^{\infty} E\left[w_i(\boldsymbol{x}_1, \cdot)\right] E[w_i(\boldsymbol{x}_2, \cdot)]\, Cov\left[\mathbb{I}_{\left\{\theta_i(\cdot) \in A_{l,t}\right\}}, \mathbb{I}_{\left\{\theta_i(\cdot) \in A_{l_1,t}\right\}}\right] \\
+ Cov\left[\sum_{l=1}^{k(\cdot)} G_0\left(A_{l,k(\cdot)}\right) \mathrm{BIN}(l \mid k(\cdot), y),\ \sum_{l=1}^{k(\cdot)} G_0\left(A_{l,k(\cdot)}\right) \mathrm{BIN}(l \mid k(\cdot), y)\right]
\end{aligned}$$

where $B_y = [0, y]$*,* $A_{j,k(\cdot)} = [0, l/k(\cdot)]$*,* $\mathrm{BIN}(\cdot \mid k, y)$ *stands for the probability mass function of the binomial distribution with parameters* (k, y) *and* $\bar{B}(l, l_1 \mid k, y) = \mathrm{BIN}(l \mid k, y) \times \mathrm{BIN}(l_1 \mid k, y)$*.*

PROOF. Notice that, for every $\boldsymbol{x}_{1j}, \boldsymbol{x}_{2j} \in \mathscr{X}$,

$$
\begin{aligned}
&E\left[G(\boldsymbol{x}_{1j},\cdot)(B_y), G(\boldsymbol{x}_{2j},\cdot)(B_y)|k\right] - E\left[G(\boldsymbol{x}_{1j},\cdot)(B_y) \mid k\right] E\left\{G(\boldsymbol{x}_{2j},\cdot)(B_y) \mid k\right] \\
&= \sum_{l=1,l_1=1}^{k} \sum_{i=1,i_1=1}^{\infty} E\left[w_i(\boldsymbol{x}_{1j},\cdot) w_{i_1}(\boldsymbol{x}_{2j},\cdot)\right] E\left[\mathbb{I}_{\left\{\theta_i(\cdot)\in A_{l,k}\right\}} \mathbb{I}_{\left\{\theta_{i_1}(\cdot)\in A_{l_1,k}\right\}}\right] \bar{B}(y \mid l, l_1, k) \\
&- \sum_{l=1,l_1=1}^{k} \sum_{i=1,i_1=1}^{\infty} E\left[w_i(\boldsymbol{x}_{1j},\cdot)\right] E\left[w_{i_1}(\boldsymbol{x}_{2j},\cdot)\right] E\left[\mathbb{I}_{\left\{\theta_i(\cdot)\in A_{l,k}\right\}}\right] E\left[\mathbb{I}_{\left\{\theta_{i_1}(\cdot)\in A_{l_1,k}\right\}}\right] \bar{B}(y \mid l, l_1, k).
\end{aligned}
$$

Now, since the elements of $\mathscr{V}$ are continuous functions of $\boldsymbol{x}$ and, for every $i, i_1 \in \mathbb{N}$, $w_i(\boldsymbol{x}_{1j},\cdot)w_{i_1}(\boldsymbol{x}_{2j},\cdot)$ is a continuous function of $\{(\eta_i(\boldsymbol{x}_{1j},\cdot), \eta_i(\boldsymbol{x}_{2j},\cdot))\}_1^l$, $l = \max\{i, i_1\}$, it follows that

$$
\lim_{j\to\infty} E\left[w_i(\boldsymbol{x}_{1j},\cdot) w_{i_1}(\boldsymbol{x}_{2j},\cdot)\right] = E\left[w_i(\boldsymbol{x}_1,\cdot)\right] E\left[w_{i_1}(\boldsymbol{x}_2,\cdot)\right]
$$

and

$$
\lim_{j\to\infty} E\left[w_i(\boldsymbol{x}_{1j},\cdot)\right] E\left[w_{i_1}(\boldsymbol{x}_{2j},\cdot)\right] = E\left[w_i(\boldsymbol{x}_1,\cdot)\right] E\left[w_{i_1}(\boldsymbol{x}_2,\cdot)\right].
$$

Thus, by Lebesgue's dominated convergence theorem, it follows that

$$
\begin{aligned}
&\lim_{j\to\infty} Cov\left[G(\boldsymbol{x}_{1j},\cdot)(B_y), G(\boldsymbol{x}_{2j},\cdot)(B_y) \mid k\right] \\
&\quad= \lim_{j\to\infty} E\left[G(\boldsymbol{x}_{1j},\cdot)(B_y), G(\boldsymbol{x}_{2j},\cdot)(B_y) \mid k\right] - E\left[G(\boldsymbol{x}_{1j},\cdot)(B_y)|k\right] E\left\{G(\boldsymbol{x}_{2j},\cdot)(B_y) \mid k\right] \\
&\quad= \sum_{l=1,l_1=1}^{k} \bar{B}(y \mid l, l_1, k) \sum_{i=1,i_1=1}^{\infty} E\left[w_i(\boldsymbol{x}_1,\cdot)\right] E[w_{i_1}(\boldsymbol{x}_2,\cdot)] \, E\left[\mathbb{I}_{\left\{\theta_i(\cdot)\in A_{l,k}\right\}} \mathbb{I}_{\left\{\theta_{i_1}(\cdot)\in A_{l_1,k}\right\}}\right] \\
&\quad- \sum_{l=1,l_1=1}^{k} \bar{B}(y \mid l, l_1, k) \sum_{i=1,i_1=1}^{\infty} E\left[w_i(\boldsymbol{x}_1,\cdot)\right] E\left[w_{i_1}(\boldsymbol{x}_2,\cdot)\right] E\left[\mathbb{I}_{\left\{\theta_i(\cdot)\in A_{l,k}\right\}}\right] E\left[\mathbb{I}_{\left\{\theta_{i_1}(\cdot)\in A_{l_1,k}\right\}}\right] \\
&\quad= \sum_{l=1,l_1=1}^{k} \bar{B}(y \mid l, l_1, k) \sum_{i=1,i_1=1}^{\infty} E\left[w_i(\boldsymbol{x}_1,\cdot)\right] E[w_{i_1}(\boldsymbol{x}_2,\cdot)] \, Cov\left[\mathbb{I}_{\left\{\theta_i(\cdot)\in A_{l,k}\right\}}, \mathbb{I}_{\left\{\theta_{i_1}(\cdot)\in A_{l_1,k}\right\}}\right] \\
&\quad= \sum_{l=1,l_1=1}^{k} \bar{B}(y \mid l, l_1, k) \sum_{i=1}^{\infty} E\left[w_i(\boldsymbol{x}_1,\cdot)\right] E[w_i(\boldsymbol{x}_2,\cdot)] \, Cov\left[\mathbb{I}_{\left\{\theta_i(\cdot)\in A_{l,k}\right\}}, \mathbb{I}_{\left\{\theta_i(\cdot)\in A_{l_1,k}\right\}}\right].
\end{aligned}
$$

On the other hand, the covariance between the conditional expectations of $G(\boldsymbol{x}_{1j},\cdot)(B_y)$ and

$G(\boldsymbol{x}_{2j},\cdot)(B_y)$ given k is of the form,

$$\begin{aligned}
&\lim_{j\to\infty} Cov\left[E\left[G(\boldsymbol{x}_{1j},\cdot)(B_y)|\,k\right], E\left[G(\boldsymbol{x}_{2j},\cdot)(B_y)|\,k\right]\right] \\
&\qquad = Cov\left[\lim_{j\to\infty} E\left[G(\boldsymbol{x}_{1j},\cdot)(B_y)|\,k\right], \lim_{j\to\infty} E\left[G(\boldsymbol{x}_{2j},\cdot)(B_y)\mid k\right]\right] \\
&\qquad = Cov\left[E\left[G(\boldsymbol{x}_{1},\cdot)(B_y)|\,k\right], E\left[G(\boldsymbol{x}_{2},\cdot)(B_y)\mid k\right]\right]
\end{aligned}$$

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

$$E\left[G(\boldsymbol{x},\cdot)(B_y)|\,k\right] = \sum_{l=1}^{k} G^{*}_{0\boldsymbol{x}}\left(A_{l,k}\right)\begin{pmatrix} k \\ l \end{pmatrix} y^l(1-y)^{k-l},$$

which completes the proof of the theorem. □

Theorem B.12. *Let* $\{G_{\boldsymbol{x}} : \boldsymbol{x}\in\mathscr{X}\}\sim\theta\mathrm{DBPP}(\lambda,\boldsymbol{\Psi}_1,\mathscr{V},G_0)$*. Assume for every* $\{(\boldsymbol{x}_{1j},\boldsymbol{x}_{2j})\}_1^{\infty}\subset\mathscr{X}^2$*, such that* $\lim_{j\to+\infty}(\boldsymbol{x}_{1j},\boldsymbol{x}_{2j}) = (\boldsymbol{x}_1,\boldsymbol{x}_2)\in\mathscr{X}^2$*, we have that* $(\eta_i(\boldsymbol{x}_{1j},\cdot),\eta_i(\boldsymbol{x}_{2j},\cdot)) \xrightarrow{\mathscr{L}} (\eta_i(\boldsymbol{x}_1,\cdot),\eta_i(\boldsymbol{x}_2,\cdot))$*, as* $j\longrightarrow+\infty$*. Then, for every* $y=[0,1]$*,*

$$\lim_{j\to\infty}\rho\left[G(\boldsymbol{x}_{1j},\cdot)(B_y), G(\boldsymbol{x}_{2j},\cdot)(B_y)\right] = \rho\left[G(\boldsymbol{x}_{1},\cdot)(B_y), G(\boldsymbol{x}_{2},\cdot)(B_y)\right],$$

where $B_y=[0,y]$*.*

PROOF. Since the elements of $\mathscr{V}$ are continuous functions of $\boldsymbol{x}$ and, for every $i\in\mathbb{N}$ and every $\{(\boldsymbol{x}_{1j},\boldsymbol{x}_{2j})\}_1^{\infty}$, with $\boldsymbol{x}_{1j},\boldsymbol{x}_{2j}\in\mathscr{X}$, such that $\lim_{j\to+\infty}(\boldsymbol{x}_{1j},\boldsymbol{x}_{2j}) = (\boldsymbol{x}_1,\boldsymbol{x}_2)$, with $\boldsymbol{x}_1,\boldsymbol{x}_2\in\mathscr{X}$, it follows that

$$(\eta_i(\boldsymbol{x}_{1j},\cdot),\eta_i(\boldsymbol{x}_{2j},\cdot)),$$

and $\eta_i(\boldsymbol{x}_{lj},\cdot)$ converges in distribution to $(\eta_i(\boldsymbol{x}_1,\cdot),\eta_i(\boldsymbol{x}_2,\cdot))$ and $\eta_i(\boldsymbol{x}_l,\cdot)$, respectively, as $j\longrightarrow+\infty$, for $l=1,2$. It also follows that $(\boldsymbol{x}_1,\boldsymbol{x}_2)\mapsto E\left\{w_i(\boldsymbol{x}_l,\cdot)\right\}$, $(\boldsymbol{x}_1,\boldsymbol{x}_2)\mapsto E\left\{w_i(\boldsymbol{x}_l,\cdot)^2\right\}$, $l=1,2$, and $(\boldsymbol{x}_1,\boldsymbol{x}_2)\mapsto E\left\{w_i(\boldsymbol{x}_1,\cdot)w_{i_1}(\boldsymbol{x}_2,\cdot)\right\}$ are continuous functions, for every $i,i_1\in\mathbb{N}$. Finally, since the correlation, ρ, is a continuous function of $E\left\{w_i(\boldsymbol{x}_l,\cdot)\right\}$, $E\left\{w_i(\boldsymbol{x}_l,\cdot)^2\right\}$ and

$E\left\{w_i(\boldsymbol{x}_1,\cdot)w_{i_1}(\boldsymbol{x}_2,\cdot)\right\}$, $i, i_1 \in \mathbb{N}$, and $l = 1, 2$, it follows that

$$(\boldsymbol{x}_1, \boldsymbol{x}_2) \mapsto \rho\left[G(\boldsymbol{x}_1,\cdot)(B_y), G(\boldsymbol{x}_2,\cdot)(B_y)\right]$$

is also a continuous function. □

Theorem B.13. *Let* $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \sim \theta\mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_1, \mathscr{V}, G_0)$*. If for every* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d \in \mathscr{X}^d$*,* $d \geq 1$*, the joint distribution of* $(\eta_i(\boldsymbol{x}_1,\cdot), \ldots, \eta_i(\boldsymbol{x}_d,\cdot))$ *has full support on* $\mathbb{R}^d$*, and,* $k(\cdot)$ *and* G_0 *have full support on* $\mathbb{N}$ *and* $(0,1]$*, respectively, then* $\mathscr{P}\left([0,1]\right)^{\mathscr{X}}$ *is the support of* $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$ *under the weak product topology.*

PROOF. The proof is similar to the one of Theorem 3.5. In this case, $W_l(\boldsymbol{x}, \omega, k) = \sum_{i=1}^{\infty} w_i(\boldsymbol{x}, \omega) I\{\theta_i(\omega)\}_{\{\lceil k\theta_i(\omega)\rceil = l\}}$. Now, since by assumption the joint distribution of $(\eta_i(\boldsymbol{x}_1,\cdot), \ldots, \eta_i(\boldsymbol{x}_d,\cdot))$ has full support on $\mathbb{R}^d$, for every $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d \in \mathscr{X}^d$, $d \geq 1$, and, $k(\cdot)$ and G_0 have full support on $\mathbb{N}$ and $(0,1]$, respectively, then Theorem 6 in Barrientos et al. (2012) ensures that

$$P\left\{\omega \in \Omega : \left|\int f_{ij} dG(\boldsymbol{x}_i, \omega) - \int f_{ij} dQ_{x_i}\right| < \epsilon_i,\ i = 1, \ldots, T,\ j = 1, \ldots, K_i\right\} \geq$$
$$\sum_{k_1 > k_0} P\left\{\omega \in \Omega : N_{ik_1}(\omega) \leq \frac{\epsilon_i}{4M_{ij}k_1}, i = 1, \ldots, T,\ j = 1, \ldots, K_i, k(\omega) = k_1\right\} > 0,$$

which completes the proof of the theorem. □

Theorem B.14. *Let* $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \sim \theta\mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_1, \mathscr{V}, G_0)$*. If for every* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d \in \mathscr{X}^d$*,* $d \geq 1$*, the joint distribution of* $(\eta_i(\boldsymbol{x}_1,\cdot), \ldots, \eta_i(\boldsymbol{x}_d,\cdot))$ *has full support on* $\mathbb{R}^d$*, and,* $k(\cdot)$ *and* G_0 *have full support on* $\mathbb{N}$ *and* $(0,1]$*, respectively, then* $\mathscr{D}\left([0,1]\right)^{\mathscr{X}}$ *is the support of* $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$ *under the* L_∞ *product topology.*

PROOF. The proof is similar to the one of Theorem 3.6. In this case,

$$W_l(\boldsymbol{x}, \omega, k) = \sum_{i=1}^{\infty} w_i(\boldsymbol{x}, \omega) I\{\theta_i(\omega)\}_{\{\lceil k\theta_i(\omega)\rceil = l\}}.$$

Now, since by assumption the joint distribution of $(\eta_i(\boldsymbol{x}_1,\cdot), \ldots, \eta_i(\boldsymbol{x}_d,\cdot))$ has full support on

$\mathbb{R}^d$, for every $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_d \in \mathscr{X}^d$, $d \geq 1$, and, $k(\cdot)$ and G_0 have full support on $\mathbb{N}$ and $(0,1]$, respectively, it follows that

$$P\left\{\omega \in \Omega : \sup_{y\in[0,1]} |g(\boldsymbol{x}_i,\omega)(y) - q_{\boldsymbol{x}_i}(y)| < \epsilon,\ i = 1,\ldots,T\right\} \geq$$
$$\sum_{k_1 \geq k_0} P\left\{\omega \in \Omega : N_{ik_1}(\omega) \leq \frac{\epsilon}{4Mk_1}, i = 1,\ldots,T, k(\omega) = k_1\right\} > 0,$$

by Theorem 6 in Barrientos et al. (2012), which completes the proof of the theorem. □

Theorem B.15. *Let $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \sim \theta\mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_1, \mathscr{V}, G_0)$. If $\mathscr{X}$ is a compact set, G_0 and $k(\cdot)$ have full support on $\mathbb{N}$ and $(0,1]$, respectively, and the processes used in the definition of the θDBPP are such that, for any $\epsilon > 0$ and $[0,1]$–valued continuous function f defined on $\mathscr{X}$, we have that*

$$P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} |v_{\boldsymbol{x}}(\eta_i(\boldsymbol{x},\omega)) - f(\boldsymbol{x})| < \epsilon\right\} > 0,$$

then $\tilde{\mathscr{D}}\left([0,1]\right)^{\mathscr{X}}$ is the support of $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$ under the L_∞ topology.

PROOF. The proof is similar to the one of Theorem 3.7. In this case, it is sufficient to consider a subset $\Omega_0 \subset \Omega$, such that for every $\omega \in \Omega_0$, the following conditions are met:

- For $l = 1,\ldots,k_0$,
$$\lceil \theta_l(\omega) k_0 \rceil = l.$$
- For $l = 1$,
$$\sup_{\boldsymbol{x}\in\mathscr{X}} \left| v_{\boldsymbol{x}}\left(\eta_1(\boldsymbol{x},\omega)\right) - Q_{\boldsymbol{x}}\left(\left(0, \frac{l}{k_0}\right]\right)\right| \leq \frac{\epsilon}{2k_0 M(2^{k_0} - 1)}.$$
- For $l = 2,\ldots,k_0$,
$$\sup_{\boldsymbol{x}\in\mathscr{X}} \left| v_{\boldsymbol{x}}\left(\eta_l(\boldsymbol{x},\omega)\right) - \frac{Q_{\boldsymbol{x}}\left(\left(\frac{l}{k_0}, \frac{l-1}{k_0}\right]\right)}{1 - \sum_{j=1}^{l-1} Q_{\boldsymbol{x}}\left(\left(\frac{j}{k_0}, \frac{j-1}{k_0}\right]\right)}\right| \leq \frac{\epsilon}{2k_0 M(2^{k_0} - 1)}.$$
- $k(\omega) = k_0$.

Then, for almost every $\omega \in \Omega_0$, it follows that

$$\sup_{\boldsymbol{x} \in \mathscr{X}} \sup_{y \in [0,1]} |g(\boldsymbol{x}, \omega)(y) - \mathrm{bp}(y \mid k, Q_{\boldsymbol{x}})| < \frac{\epsilon}{2},$$

and, thus,

$$\begin{aligned}
&P\left\{\omega \in \Omega : \sup_{\boldsymbol{x} \in \mathscr{X}} \sup_{y \in [0,1]} |g(\boldsymbol{x}, \omega)(y) - q_{\boldsymbol{x}}(y)| \leq \epsilon\right\} \geq \\
&\quad \prod_{l=1}^{k_0} P\left\{\omega \in \Omega : \lceil \theta_l(\omega) k_0 \rceil = l\right\} \times \\
&\quad P\left\{\omega \in \Omega : \sup_{\boldsymbol{x} \in \mathscr{X}} \left| v_{\boldsymbol{x}}(\eta_1(\boldsymbol{x}, \omega)) - Q_{\boldsymbol{x}}\left(\left(0, \frac{1}{k_0}\right]\right)\right| \leq \frac{\epsilon}{2k_0 M(2^{k_0} - 1)}\right\} \times \\
&\quad \prod_{l=2}^{k_0} P\left\{\omega \in \Omega : \sup_{\boldsymbol{x} \in \mathscr{X}} \left| v_{\boldsymbol{x}}(\eta_l(\boldsymbol{x}, \omega)) - \frac{Q_{\boldsymbol{x}}\left(\left(\frac{l}{k_0}, \frac{l-1}{k_0}\right]\right)}{1 - \sum_{j=1}^{l-1} Q_{\boldsymbol{x}}\left(\left(\frac{j}{k_0}, \frac{j-1}{k_0}\right]\right)}\right| \leq \frac{\epsilon}{2k_0 M(2^{k_0} - 1)}\right\} \times \\
&\quad P\left\{\omega \in \Omega : k(\omega) = k_0\right\}.
\end{aligned}$$

Now, note that since

$$\boldsymbol{x} \mapsto Q_{\boldsymbol{x}}\left(\left(0, \frac{1}{k_0}\right]\right),$$

and

$$\boldsymbol{x} \mapsto \frac{Q_{\boldsymbol{x}}\left(\left(\frac{l}{k_0}, \frac{l-1}{k_0}\right]\right)}{1 - \sum_{j=1}^{l-1} Q_{\boldsymbol{x}}\left(\left(\frac{j}{k_0}, \frac{j-1}{k_0}\right]\right)},$$

$l = 2, \ldots, k_0$, are continuous functions, it follows that

$$\begin{aligned}
&P\left\{\omega \in \Omega : \sup_{\boldsymbol{x} \in \mathscr{X}} \left| v_{\boldsymbol{x}}(\eta_1(\boldsymbol{x}, \omega)) - Q_{\boldsymbol{x}}\left(\left(0, \frac{1}{k_0}\right]\right)\right| \leq \frac{\epsilon}{2k_0 M(2^{k_0} - 1)}\right\} \times \\
&\prod_{l=2}^{k_0} P\left\{\omega \in \Omega : \sup_{\boldsymbol{x} \in \mathscr{X}} \left| v_{\boldsymbol{x}}(\eta_l(\boldsymbol{x}, \omega)) - \frac{Q_{\boldsymbol{x}}\left(\left(\frac{l}{k_0}, \frac{l-1}{k_0}\right]\right)}{1 - \sum_{j=1}^{l-1} Q_{\boldsymbol{x}}\left(\left(\frac{j}{k_0}, \frac{j-1}{k_0}\right]\right)}\right| \leq \frac{\epsilon}{2k_0 M(2^{k_0} - 1)}\right\} \\
&> 0.
\end{aligned}$$

Finally, since G_0 and $k(\cdot)$ have full support on $(0,1]$ and $\mathbb{N}$, respectively, then

$$P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{y\in[0,1]} |g(\boldsymbol{x},\omega)(y) - q_{\boldsymbol{x}}(y)| \leq \epsilon\right\} > 0,$$

which completes the proof of the theorem. □

Theorem B.16. *Let* $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \sim \theta\mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_1, \mathscr{V}, G_0)$*. Assume that* $\mathscr{X}$ *is a compact set,* G_0 *and* $k(\cdot)$ *have full support on* $\mathbb{N}$ *and* $(0,1]$*, respectively, and the processes used in the definition of the* θDBPP *are such that, for any* $\epsilon > 0$ *and* $[0,1]$*–valued continuous function* f *defined on* $\mathscr{X}$*, we have that*

$$P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} |v_{\boldsymbol{x}}(\eta_i(\boldsymbol{x},\omega)) - f(\boldsymbol{x})| < \epsilon\right\} > 0.$$

Then,

$$P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} \int_0^1 q_{\boldsymbol{x}}(y) \log\left(\frac{q_{\boldsymbol{x}}(y)}{g(\boldsymbol{x},\omega)(y)}\right) dy > \epsilon\right\} > 0,$$

for every $\epsilon > 0$ *and every* $\{Q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \tilde{\mathscr{D}}([0,1])$*, with density functions* $\{q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$*.*

PROOF. This proof is similar to the one of Corollary 3.1.

Theorem B.17. *Let* $\{G_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \sim \theta\mathrm{DBPP}(\lambda, \boldsymbol{\Psi}_1, \mathscr{V}, G_0)$*. Assume that* $\mathscr{X}$ *is a compact set,* G_0 *and* $k(\cdot)$ *have full support on* $\mathbb{N}$ *and* $(0,1]$*, respectively, and the processes used in the definition of the* θDBPP *are such that, for any* $\epsilon > 0$ *and* $[0,1]$*–valued continuous function* f *defined on* $\mathscr{X}$*, we have that*

$$P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} |v_{\boldsymbol{x}}(\eta_i(\boldsymbol{x},\omega)) - f(\boldsymbol{x})| < \epsilon\right\} > 0.$$

Then the posterior distribution associated with the random joint distribution induced by the $theta$DBPP *model,* $m^{(\cdot)}(y,\boldsymbol{x}) = q(\boldsymbol{x})g(\boldsymbol{x},\cdot)(y)$*, where* q *is the density generating the predictors, is weakly consistent at any joint distribution of the form* $m_0(y,\boldsymbol{x}) = q(\boldsymbol{x})q_0(y \mid \boldsymbol{x})$*, where* $\{q_0(\cdot \mid \boldsymbol{x}) : \boldsymbol{x} \in \mathscr{X}\} \in \widetilde{\mathscr{D}}([0,1])^{\mathscr{X}}$*.*

PROOF. This proof is similar to the one of Theorem 3.8 (given in the main document).

B.4 Proof of Theorem 3.9

Since the elements of $\mathscr{V}$ are continuous functions of $\boldsymbol{x}$ and, for every $j \in \mathbb{N}$, η_j is a P–as continuous stochastic process, it follows that $\boldsymbol{x} \mapsto v_{\boldsymbol{x}}(\eta_j(\boldsymbol{x}, \cdot))$ and $\boldsymbol{x} \mapsto w_j(\boldsymbol{x}, \cdot)$, $j \in \mathbb{N}$, are P–a.s. continuous functions. A direct application of Lebesgue's dominated convergence theorem implies that the density, w.r.t. Lebesgue measure, of $G_{\boldsymbol{x}}$, is P–a.s. continuous, i.e., for every $y \in [0, 1]$,

$$P\left\{\omega \in \Omega : \lim_{j \to +\infty} g(\boldsymbol{x}_j, \omega)(y) = g(\boldsymbol{x}_0, \omega)(y), \right\} = 1.$$

Now, by Scheffe's theorem, it follows that

$$P\left\{\omega \in \Omega : \lim_{j \to +\infty} \sup_{B \in \mathscr{B}([0,1])} |G(\boldsymbol{x}_j, \omega)(B) - G(\boldsymbol{x}_0, \omega)(B)| = 0, \right\} = 1,$$

which completes the proof of the theorem. □

B.5 MCMC schemes for DBPP models

In this appendix we provide a description of the MCMC implementation used to draw samples from the posterior distributions of the LDBPP, wLDBPP and θLDBPP models. The computational implementation was based on a finite dimensional approximation to the corresponding dependent stick–breaking processes, where the number of terms in the infinite series representations was truncated to a value N. The MCMC algorithms correspond to a Gibbs sampler, which combines, in the cases where the conditional distributions are not of conjugate, slice (Neal, 2003) and Metropolis–Hastings (Tierney, 1994) algorithms. The specific implementations for each model are given next.

MCMC scheme for the LDBPP model

To update the regression coefficients, $\boldsymbol{\gamma}_1^\eta, \ldots, \boldsymbol{\gamma}_{N-1}^\eta, \boldsymbol{\gamma}_1^z, \ldots, \boldsymbol{\gamma}_N^z$, we used the multivariate slice sampling algorithm proposed by Neal (2003). In this case, the conditional distribution is given by

$$
\begin{aligned}
& f_0\left(\boldsymbol{\gamma}_1^\eta, \ldots, \boldsymbol{\gamma}_{N-1}^\eta, \boldsymbol{\gamma}_1^z, \ldots, \boldsymbol{\gamma}_N^z \mid \ldots\right) \quad \propto \\
& \prod_{i=1}^n \left\{ \sum_{j=1}^{N-1} q\left(x_i, \boldsymbol{\gamma}_j\right) \prod_{l<j} \left[1 - q\left(x_i, \boldsymbol{\gamma}_l\right)\right] \beta\left(y_i \mid \left\lceil k\, h\left(\mathbf{d}_z(x_i)^T \boldsymbol{\gamma}_j^z\right)\right\rceil, k - \left\lceil k\, h\left(\mathbf{d}_z(x_i)^T \boldsymbol{\gamma}_j^z\right)\right\rceil + 1\right) \right. \\
& \qquad \left. + \prod_{l<N} \left[1 - q\left(x_i, \boldsymbol{\gamma}_l\right)\right] \beta\left(y_i \mid \left\lceil k\, h\left(\mathbf{d}_z(x_i)^T \boldsymbol{\gamma}_N^z\right)\right\rceil, k - \left\lceil k\, h\left(\mathbf{d}_z(x_i)^T \boldsymbol{\gamma}_N^z\right)\right\rceil + 1\right) \right\} \\
& \times \exp\left\{ -\frac{1}{2} \sum_{j=1}^{N-1} \left(\boldsymbol{\gamma}_j^\eta - \boldsymbol{\mu}^\eta\right)^T \left(\mathbf{S}^\eta\right)^{-1} \left(\boldsymbol{\gamma}_j^\eta - \boldsymbol{\mu}^\eta\right) - \frac{1}{2} \sum_{j=1}^{N} \left(\boldsymbol{\gamma}_j^z - \boldsymbol{\mu}^z\right)^T \left(\mathbf{S}^z\right)^{-1} \left(\boldsymbol{\gamma}_j^z - \boldsymbol{\mu}^z\right) \right\},
\end{aligned}
$$

where $q\left(x_i, \boldsymbol{\gamma}_j\right) = h\left(\mathbf{d}_\eta(x_i)^T \boldsymbol{\gamma}_j^\eta\right)$. Let $\underline{\boldsymbol{\gamma}}_1^\eta, \ldots, \underline{\boldsymbol{\gamma}}_{N-1}^\eta, \underline{\boldsymbol{\gamma}}_1^z, \ldots, \underline{\boldsymbol{\gamma}}_N^z$ be the current value of the regression coefficients, and w_η, $w_z \in \mathbb{R}^+$. The algorithm begins by drawing a random number $t_0 \sim \mathrm{U}\left(0, f_0\left(\underline{\boldsymbol{\gamma}}_1^\eta, \ldots, \underline{\boldsymbol{\gamma}}_{N-1}^\eta, \underline{\boldsymbol{\gamma}}_1^z, \ldots, \underline{\boldsymbol{\gamma}}_N^z \mid \ldots\right)\right)$, and by defining

$$
\mathbf{L}_j^\eta = \underline{\boldsymbol{\gamma}}_j^\eta - w_j^\eta \mathbf{U}_j^\eta, \quad \mathbf{R}_j^\eta = \underline{\boldsymbol{\gamma}}_j^\eta + w_j^\eta, \quad j = 1, \ldots, N-1,
$$

$$
\mathbf{L}_j^z = \underline{\boldsymbol{\gamma}}_j^z - w_j^z \mathbf{U}_j^z, \quad \mathbf{R}_j^z = \underline{\boldsymbol{\gamma}}_j^z + w_j^z, \quad j = 1, \ldots, N,
$$

where $\mathbf{U}_j^\eta$ and $\mathbf{U}_j^z$ are drawn from the r_1 and r_2-dimensional uniform distributions, $U_{r_1}(0,1)$ and $U_{r_2}(0,1)$, respectively. Then, the following steps are repeated until new coefficient values, $\overline{\boldsymbol{\gamma}}_1^\eta, \ldots, \overline{\boldsymbol{\gamma}}_{N-1}^\eta, \overline{\boldsymbol{\gamma}}_1^z, \ldots, \overline{\boldsymbol{\gamma}}_N^z$, are accepted:

Step (i): Draw $\mathbf{U}_j^\eta$ and $\mathbf{U}_j^z$ from $U_{r_1}(0,1)$ and $U_{r_2}(0,1)$, respectively, and define

$$
\overline{\boldsymbol{\gamma}}_j^\eta = \mathbf{L}_j^\eta + \mathbf{U}_j^\eta \left(\mathbf{R}_j^\eta - \mathbf{L}_j^\eta\right), \quad j = 1, \ldots, N-1,
$$

and

$$
\overline{\boldsymbol{\gamma}}_j^z = \mathbf{L}_j^z + \mathbf{U}_j^z \left(\mathbf{R}_j^z - \mathbf{L}_j^z\right), \quad j = 1, \ldots, N.
$$

If $t_0 < f_0\left(\overline{\gamma}_1^{\eta}, \ldots, \overline{\gamma}_{N-1}^{\eta}, \overline{\gamma}_1^{z}, \ldots, \overline{\gamma}_N^{z} \mid \ldots\right)$, then accept these new coefficient values and stop the algorithm. Otherwise, go to the next step.

Step (ii): Let $\underline{\gamma}_{jl}^{\eta}$, $\overline{\gamma}_{jl}^{\eta}$, $\overline{\gamma}_{jl}^{z}$, $\underline{\gamma}_{jl}^{z}$, R_{jl}^{η}, L_{jl}^{η}, R_{jl}^{z} and L_{jl}^{z} be the l-th component of the vector $\underline{\boldsymbol{\gamma}}_j^{\eta}$,$\overline{\boldsymbol{\gamma}}_j^{\eta}$, $\overline{\boldsymbol{\gamma}}_j^{z}$, $\underline{\boldsymbol{\gamma}}_j^{z}$, $\mathbf{R}_j^{\eta}$, $\mathbf{L}_j^{\eta}$, $\mathbf{R}_j^{z}$ and $\mathbf{L}_j^{z}$, respectively. Then,

- for $j = 1, \ldots, N-1$, and $l = 1, \ldots, r_1$, if $\overline{\gamma}_{jl}^{\eta} < \underline{\gamma}_{jl}^{\eta}$, then set $L_{jl}^{\eta} = \underline{\gamma}_{jl}^{\eta}$; else, set $R_{jl}^{\eta} = \underline{\gamma}_{jl}^{\eta}$.
- for $j = 1, \ldots, N$, and $l = 1, \ldots, r_2$, if $\overline{\gamma}_{jl}^{z} < \underline{\gamma}_{jl}^{z}$, then set $L_{jl}^{z} = \underline{\gamma}_{jl}^{z}$; else, set $R_{jl}^{z} = \underline{\gamma}_{jl}^{z}$.

To update the polynomial degree, k, we used a Metropolis-.Hastings step, where the proposal distribution is given by $Q(k_{new} \mid k_{old}) = p_0\delta_{\{k_{old}-1\}}(k_{new}) + (1-p_0)\delta_{\{k_{old}+1\}}(k_{new})$, $p_0 \in (0,1)$. Finally, the full conditional distributions for the hyper–parameters, $\boldsymbol{\mu}^{\eta}$, $\boldsymbol{\mu}^{z}$, $\mathbf{S}^{\eta}$ and $\mathbf{S}^{z}$ are conjugate. For mean vectors, these are given by

$$\begin{aligned}&\boldsymbol{\mu}^{\eta} \mid \ldots \sim \\ &N_{r_1}\left(\left[(\mathbf{S}_0^{\eta})^{-1} + (N-1)(\mathbf{S}^{\eta})^{-1}\right]^{-1}\left[(\mathbf{S}_0^{\eta})^{-1}\boldsymbol{\mu}_0^{\eta} + (\mathbf{S}^{\eta})^{-1}\sum_{j=1}^{N-1}\boldsymbol{\gamma}_j^{\eta}\right], \left[(\mathbf{S}_0^{\eta})^{-1} + (N-1)(\mathbf{S}^{\eta})^{-1}\right]^{-1}\right),\end{aligned}$$

and

$$\boldsymbol{\mu}^{z} \mid \ldots \sim N_{r_2}\left(\left[(\mathbf{S}_0^{z})^{-1} + N(\mathbf{S}^{z})^{-1}\right]^{-1}\left[(\mathbf{S}_0^{z})^{-1}\boldsymbol{\mu}_0^{z} + (\boldsymbol{S}^{z})^{-1}\sum_{j=1}^{N}\boldsymbol{\gamma}_j^{z}\right], \left[(\mathbf{S}_0^{z})^{-1} + N(\mathbf{S}^{z})^{-1}\right]^{-1}\right),$$

respectively. For covariance matrices, the full conditionals are given by

$$\mathbf{S}^{\eta} \mid \ldots \;\sim\; IW_{r_1}\left((N-1) + \nu^{\eta}, \boldsymbol{\Psi}^{\eta} + \sum_{j=1}^{N-1}\left(\boldsymbol{\mu}^{\eta} - \boldsymbol{\gamma}_j^{\eta}\right)^T\left(\boldsymbol{\mu}^{\eta} - \boldsymbol{\gamma}_j^{\eta}\right)\right)$$

and

$$\mathbf{S}^{z} \mid \ldots \;\sim\; IW_{r_2}\left(N + \nu^{z}, \boldsymbol{\Psi}^{z} + \sum_{j=1}^{N}\left(\boldsymbol{\mu}^{z} - \boldsymbol{\gamma}_j^{z}\right)^T\left(\boldsymbol{\mu}^{z} - \boldsymbol{\gamma}_j^{z}\right)\right),$$

respectively.

MCMC scheme for the wLDBPP model

To update the parameters $v_1, \ldots, v_{N-1}, \boldsymbol{\gamma}_1^z, \ldots, \boldsymbol{\gamma}_N^z$, we used the multivariate slice sampling algorithm proposed by Neal (2003). In this case, the conditional distribution is given by

$$
\begin{aligned}
& f_0\left(v_1, \ldots, v_{N-1}, \boldsymbol{\gamma}_1^z, \ldots, \boldsymbol{\gamma}_N^z \mid \ldots\right) \quad \propto \\
& \prod_{i=1}^{n}\left\{\sum_{j=1}^{N-1} v_j \prod_{l<j}\left(1-v_l\right) \beta\left(y_i \mid \left\lceil k\, h\left(\mathbf{d}_z(x_i)^T \boldsymbol{\gamma}_j^z\right)\right\rceil, k-\left\lceil k\, h\left(\mathbf{d}_z(x_i)^T \boldsymbol{\gamma}_j^z\right)\right\rceil+1\right)\right. \\
& \qquad \left.+\prod_{l<N}\left(1-v_l\right) \beta\left(y_i \mid \left\lceil k\, h\left(\mathbf{d}_z(x_i)^T \boldsymbol{\gamma}_N^z\right)\right\rceil, k-\left\lceil k\, h\left(\mathbf{d}_z(x_i)^T \boldsymbol{\gamma}_N^z\right)\right\rceil+1\right)\right\} \\
& \times \exp\left\{-\frac{1}{2} \sum_{j=1}^{N}\left(\boldsymbol{\gamma}_j^z-\boldsymbol{\mu}^z\right)^T\left(\mathbf{S}^z\right)^{-1}\left(\boldsymbol{\gamma}_j^z-\boldsymbol{\mu}^z\right)\right\} \times \prod_{j=1}^{N-1}\left(1-v_l\right)^{\alpha-1}.
\end{aligned}
$$

Let $\underline{v}_1, \ldots, \underline{v}_{N-1}, \underline{\boldsymbol{\gamma}}_1^z, \ldots, \underline{\boldsymbol{\gamma}}_N^z$ be the current value of the parameters, and take $(w_v,\ w_z) \in (0,1) \times \mathbb{R}^+$. The algorithm begins by drawing a random number $t_0 \sim \mathrm{U}\left(0, f_0\left(v_1, \ldots, v_{N-1}, \boldsymbol{\gamma}_1^z, \ldots, \boldsymbol{\gamma}_N^z \mid \ldots\right)\right)$, and by defining

$$
L_j^v = \underline{v}_j - w_j^v U_j^v, \quad R_j^v = \underline{v}_j + w_j^v, \quad j = 1, \ldots, N-1,
$$

$$
\mathbf{L}_j^z = \underline{\boldsymbol{\gamma}}_j^z - w_j^z \mathbf{U}_j^z, \quad \mathbf{R}_j^z = \underline{\boldsymbol{\gamma}}_j^z + w_j^z, \quad j = 1, \ldots, N,
$$

where U_j^v and $\mathbf{U}_j^z$ are drawn from an $U(0,1)$ and $U_{r_2}(0,1)$ distribution, respectively. Then, the following steps are repeated until new value of the parameters, $\overline{v}_1, \ldots, \overline{v}_{N-1}, \overline{\boldsymbol{\gamma}}_1^z, \ldots, \overline{\boldsymbol{\gamma}}_N^z$, are accepted:

Step (i): Draw U_j^v and $\mathbf{U}_j^z$ from an $U(0,1)$ and $U_{r_2}(0,1)$ distribution, respectively, and define

$$
\overline{v}_j = L_j^v + U_j^v\left(R_j^v - L_j^v\right), \quad j = 1, \ldots, N-1,
$$

and

$$\overline{\gamma}_j^z = \mathbf{L}_j^z + \mathbf{U}_j^z \left(\mathbf{R}_j^z - \mathbf{L}_j^z\right), \quad j = 1, \ldots, N.$$

If $t_0 < f_0\left(v_1, \ldots, v_{N-1}, \boldsymbol{\gamma}_1^z, \ldots, \boldsymbol{\gamma}_N^z \mid \ldots\right)$, then accept these new values and stop the algorithm. Otherwise, go to the next step.

Step (ii): Let $\overline{\gamma}_{jl}^z$, $\underline{\gamma}_{jl}^z$, R_{jl}^η, L_{jl}^η, R_{jl}^z and L_{jl}^z be the l-th component of the vector $\overline{\boldsymbol{\gamma}}_j^z$, $\underline{\boldsymbol{\gamma}}_j^z$, $\mathbf{R}_j^\eta$, $\mathbf{L}_j^\eta$, $\mathbf{R}_j^z$ and $\mathbf{L}_j^z$, respectively. Then,

- for $j = 1, \ldots, N-1$, if $\overline{v}_j < \underline{v}_j$, then set $L_j^v = \underline{v}_j$; else, set $R_j^v = \underline{v}_j$.
- for $j = 1, \ldots, N$, and $l = 1, \ldots, r_2$, if $\overline{\gamma}_{jl}^z < \underline{\gamma}_{jl}^z$, then set $L_{jl}^z = \underline{\gamma}_{jl}^z$; else, set $R_{jl}^z = \underline{\gamma}_{jl}^z$.

Finally, the polynomial degree, k, and the hyper–parameters, $\boldsymbol{\mu}^z$ and $\mathbf{S}^z$, were updated using the same steps described for the LDBPP model.

MCMC scheme for the θLDBPP model

To update the parameters $\boldsymbol{\gamma}_1^\eta, \ldots, \boldsymbol{\gamma}_{N-1}^\eta, \theta_1, \ldots, \theta_N$, we used the multivariate slice sampling algorithm proposed by Neal (2003). In this case, the conditional distribution is given by

$$\begin{aligned}
&f_0\left(\boldsymbol{\gamma}_1^\eta, \ldots, \boldsymbol{\gamma}_{N-1}^\eta, \theta_1, \ldots, \theta_N \mid \ldots\right) \quad \propto \\
&\prod_{i=1}^n \left\{ \sum_{j=1}^{N-1} q\left(x_i, \boldsymbol{\gamma}_j\right) \prod_{l<j} \left[1 - q\left(x_i, \boldsymbol{\gamma}_l\right)\right] \beta\left(y_i \mid \lceil k\, \theta_j \rceil, k - \lceil k\, \theta_j \rceil + 1\right) \right. \\
&\qquad \left. + \prod_{l<N} \left[1 - q\left(x_i, \boldsymbol{\gamma}_l\right)\right] \beta\left(y_i \mid \lceil k\, \theta_N \rceil, k - \lceil k\, \theta_N \rceil + 1\right) \right\} \\
&\times \exp\left\{ -\frac{1}{2} \sum_{j=1}^{N-1} \left(\boldsymbol{\gamma}_j^\eta - \boldsymbol{\mu}^\eta\right)^T \left(\mathbf{S}^\eta\right)^{-1} \left(\boldsymbol{\gamma}_j^\eta - \boldsymbol{\mu}^\eta\right) \right\} \prod_{j=1}^{N-1} \theta_j^{a-1} \left(1 - \theta_l\right)^{b-1}
\end{aligned}$$

where $q\left(x_i, \boldsymbol{\gamma}_j\right) = h\left(\mathbf{d}_\eta(x_i)^T \boldsymbol{\gamma}_j^\eta\right)$. Let $\underline{\boldsymbol{\gamma}}_1^\eta, \ldots, \underline{\boldsymbol{\gamma}}_{N-1}^\eta, \underline{\theta}_1, \ldots, \underline{\theta}_N$ be the current values of the parameters and take $(w_\eta, w_\theta) \in \mathbb{R}^+ \times (0, 1)$. The algorithm begins by drawing a random number

$t_0 \sim \mathrm{U}\left(0, f_0\left(\boldsymbol{\gamma}_1^\eta, \ldots, \boldsymbol{\gamma}_{N-1}^\eta, \theta_1, \ldots, \theta_N \mid \ldots\right)\right)$, and by defining

$$\mathbf{L}_j^\eta = \underline{\boldsymbol{\gamma}}_j^\eta - w_j^\eta \mathbf{U}_j^\eta, \quad \mathbf{R}_j^\eta = \underline{\boldsymbol{\gamma}}_j^\eta + w_j^\eta, \quad j = 1, \ldots, N-1,$$

$$L_j^\theta = \underline{\theta}_j - w_j^z U_j^\theta, \quad R_j^\theta = \underline{\theta}_j + w_j^z, \quad j = 1, \ldots, N,$$

where $\mathbf{U}_j^\eta$ and U_j^θ are drawn from an $U_{r_1}(0,1)$ and $U(0,1)$ distribution, respectively. Then, the following steps are repeated until new value of the parameters, $\overline{\boldsymbol{\gamma}}_1^\eta, \ldots, \overline{\boldsymbol{\gamma}}_{N-1}^\eta, \overline{\theta}_1, \ldots, \overline{\theta}_N$, are accepted:

Step (i): Draw $\mathbf{U}_j^\eta$ and U_j^θ from an $U_{r_1}(0,1)$ and $U(0,1)$ distribution, respectively, and define

$$\overline{\boldsymbol{\gamma}}_j^\eta = \mathbf{L}_j^\eta + \mathbf{U}_j^\eta \left(\mathbf{R}_j^\eta - \mathbf{L}_j^\eta\right), \quad j = 1, \ldots, N-1,$$

and

$$\overline{\theta}_j = L_j^\theta + U_j \left(R_j^\theta - L_j^\theta\right), \quad j = 1, \ldots, N.$$

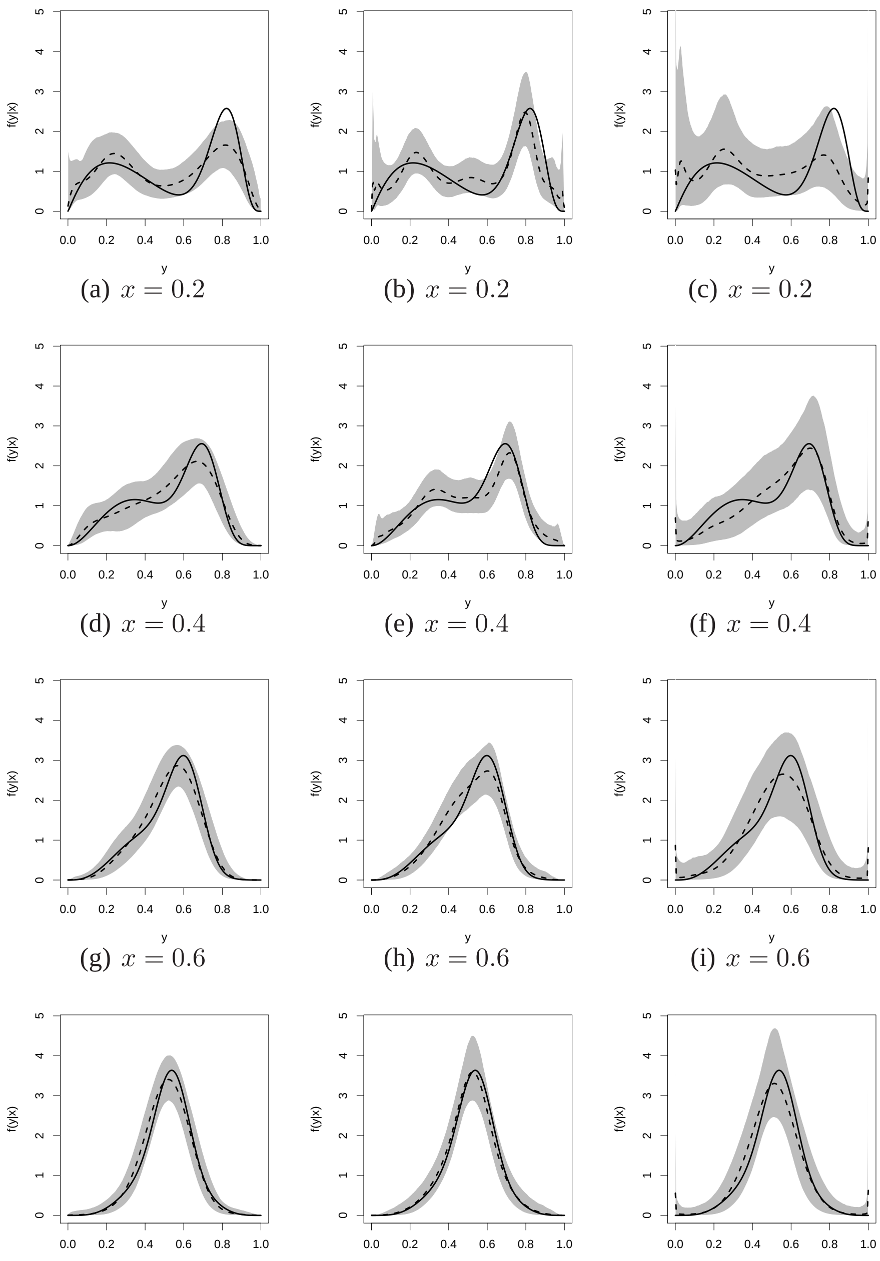
If $t_0 < f_0\left(\boldsymbol{\gamma}_1^\eta, \ldots, \boldsymbol{\gamma}_{N-1}^\eta, \theta_1, \ldots, \theta_N \mid \ldots\right)$, then accept these new values and stop the algorithm. Otherwise, go to the next step.

Step (ii): Let $\underline{\gamma}_{jl}^\eta$, $\overline{\gamma}_{jl}^\eta$, R_{jl}^η and L_{jl}^η be the l-th component of the vector $\underline{\boldsymbol{\gamma}}_j^\eta$,$\overline{\boldsymbol{\gamma}}_j^\eta$, $\mathbf{R}_j^\eta$ and $\mathbf{L}_j^\eta$, respectively. Then,

- for $j = 1, \ldots, N-1$, and $l = 1, \ldots, r_1$, if $\overline{\gamma}_{jl}^\eta < \underline{\gamma}_{jl}^\eta$, then set $L_{jl}^\eta = \underline{\gamma}_{jl}^\eta$; else, set $R_{jl}^\eta = \underline{\gamma}_{jl}^\eta$.
- for $j = 1, \ldots, N$, and $l = 1, \ldots, r_2$, if $\overline{\theta}_j < \underline{\theta}_j$, then set $L_{jl}^z = \underline{\theta}_j$; else, set $R_j^\theta = \underline{\theta}_j$.

Finally, the polynomial degree, k, and the hyper-parameters, $\boldsymbol{\mu}^\eta$ and $\mathbf{S}^\eta$, were updated using the same steps described for the LDBPP model.

B.6 Additional simulation results



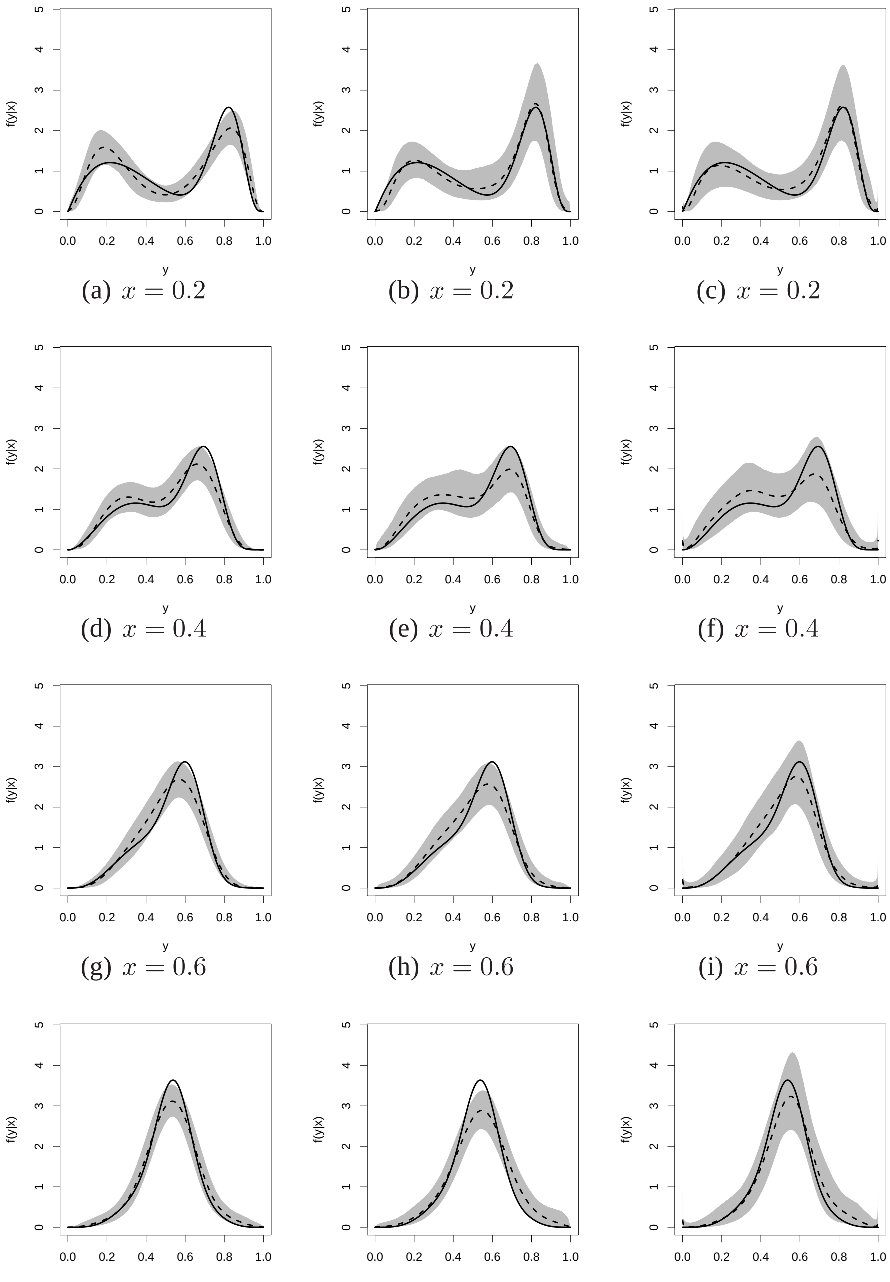
(a) $x = 0.2$ (b) $x = 0.2$ (c) $x = 0.2$

(d) $x = 0.4$ (e) $x = 0.4$ (f) $x = 0.4$

(g) $x = 0.6$ (h) $x = 0.6$ (i) $x = 0.6$

(j) $x = 0.8$ (k) $x = 0.8$ (l) $x = 0.8$

Figure B.1: Simulated data - Scenario I ($n = 250$): True (continuous line) and posterior mean (dotted line) for the conditional density. A band constructed using the 95%point–wise HPD intervals is presented in gray. Panels (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l) display the results for the best DBPP model (LDBPP1), the best LDDP model (LDDP1), both regarding the estimated L_1 distance, and the weight dependent DP for four values of the predictor, respectively.



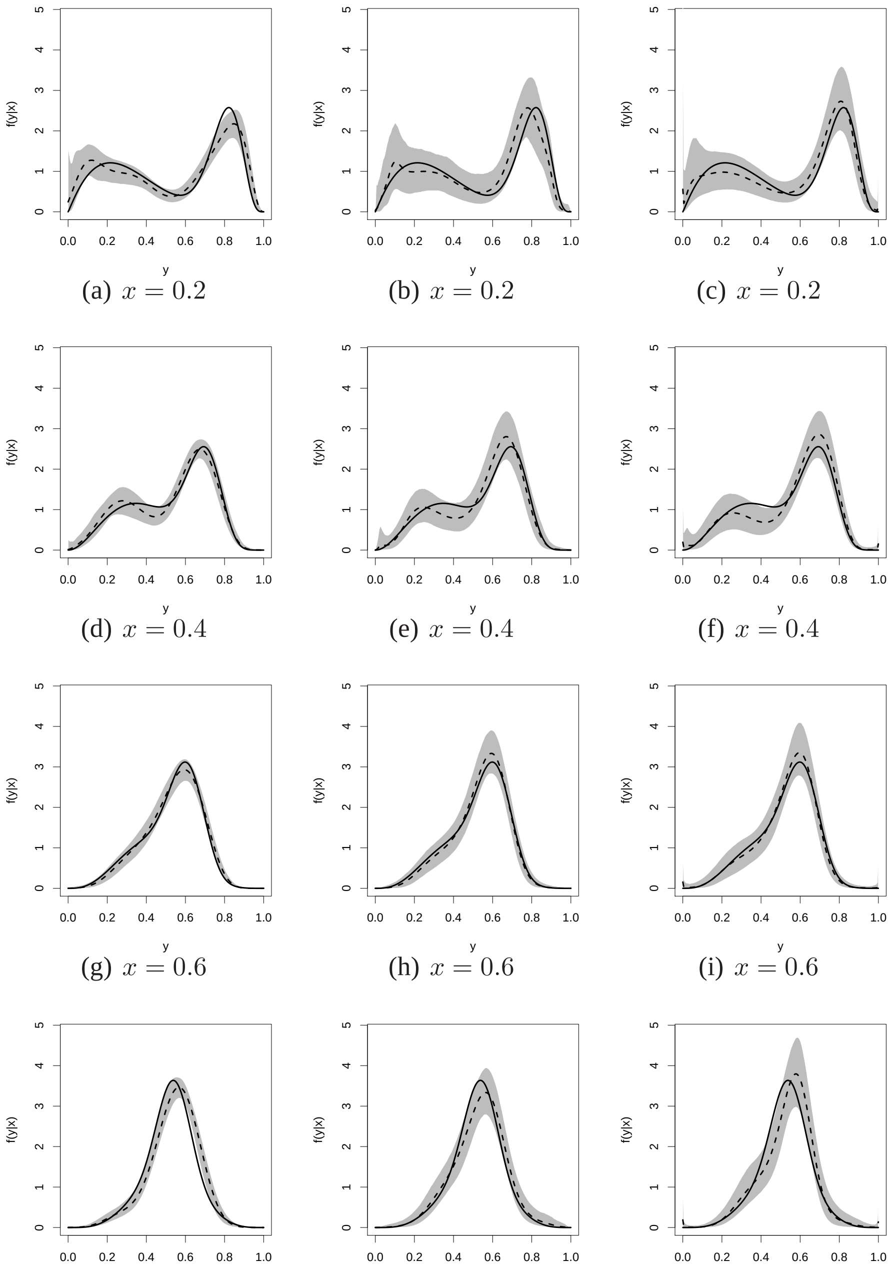
(a) $x = 0.2$ (b) $x = 0.2$ (c) $x = 0.2$

(d) $x = 0.4$ (e) $x = 0.4$ (f) $x = 0.4$

(g) $x = 0.6$ (h) $x = 0.6$ (i) $x = 0.6$

(j) $x = 0.8$ (k) $x = 0.8$ (l) $x = 0.8$

Figure B.2: Simulated data - Scenario I ($n = 500$): True (continuous line) and posterior mean (dotted line) for the conditional density. A band constructed using the 95%point–wise HPD intervals is presented in gray. Panels (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l) display the results for the best DBPP model (LDBPP1), the best LDDP model (LDDP2), both regarding the estimated L_1 distance, and the weight dependent DP for four values of the predictor, respectively.



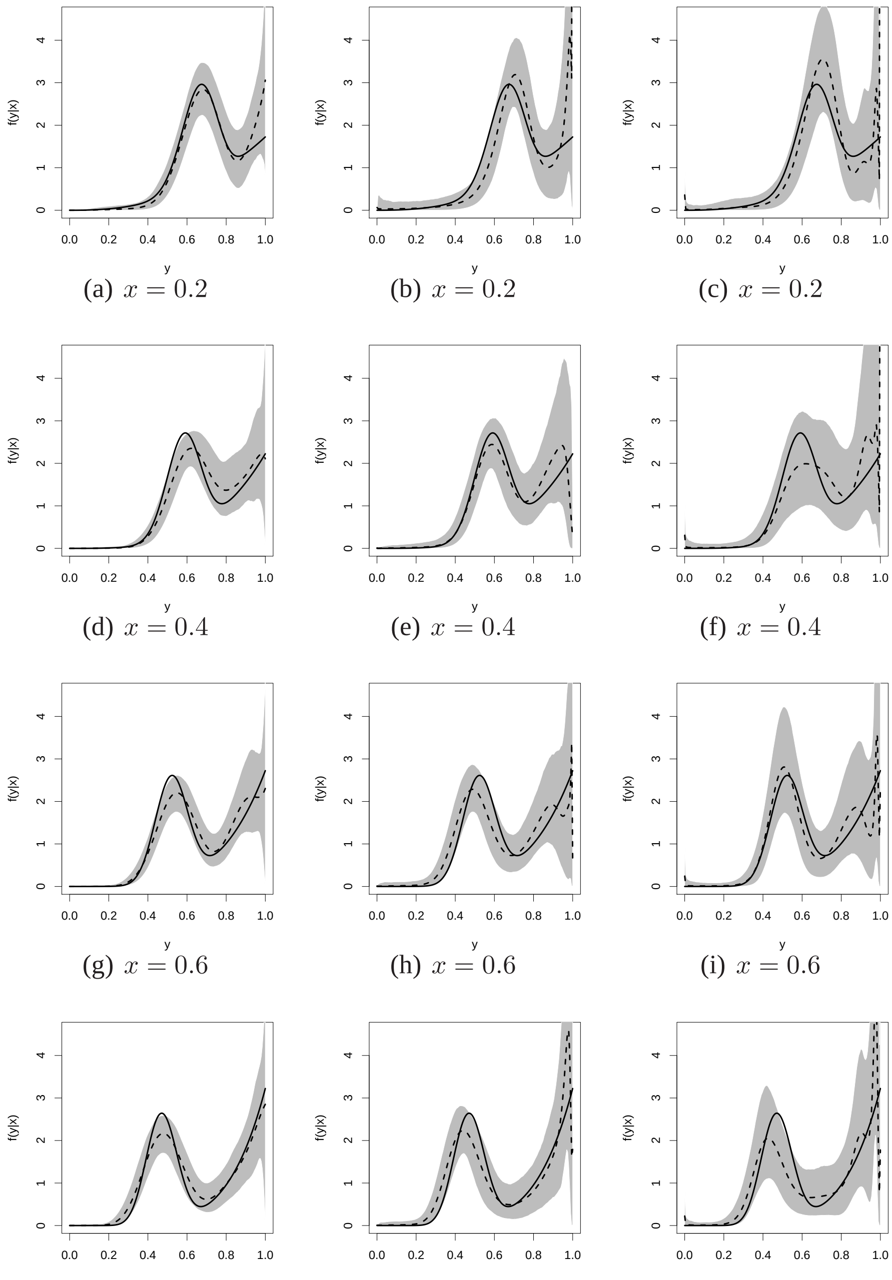
(a) $x = 0.2$ (b) $x = 0.2$ (c) $x = 0.2$

(d) $x = 0.4$ (e) $x = 0.4$ (f) $x = 0.4$

(g) $x = 0.6$ (h) $x = 0.6$ (i) $x = 0.6$

(j) $x = 0.8$ (k) $x = 0.8$ (l) $x = 0.8$

Figure B.3: Simulated data - Scenario I ($n = 1000$): True (continuous line) and posterior mean (dotted line) for the conditional density. A band constructed using the 95%point–wise HPD intervals is presented in gray. Panels (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l) display the results for the best DBPP model (LDBPP1), the best LDDP model (LDDP2), both regarding the estimated L_1 distance, and the weight dependent DP for four values of the predictor, respectively.



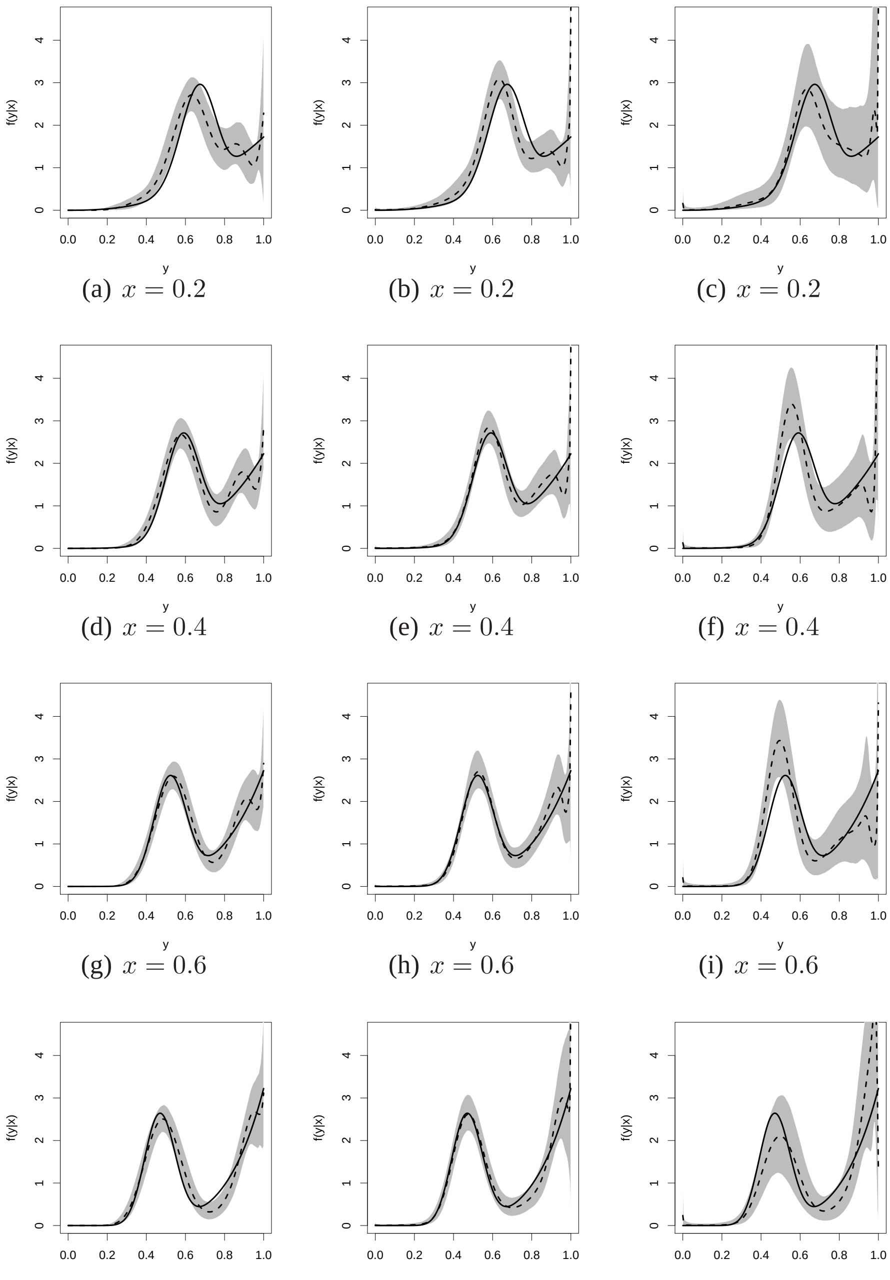
(a) $x = 0.2$ (b) $x = 0.2$ (c) $x = 0.2$

(d) $x = 0.4$ (e) $x = 0.4$ (f) $x = 0.4$

(g) $x = 0.6$ (h) $x = 0.6$ (i) $x = 0.6$

(j) $x = 0.8$ (k) $x = 0.8$ (l) $x = 0.8$

Figure B.4: Simulated data - Scenario II ($n = 250$): True (continuous line) and posterior mean (dotted line) for the conditional density. A band constructed using the 95%point–wise HPD intervals is presented in gray. Panels (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l) display the results for the best DBPP model (wLDBPP1), the best LDDP model (LDDP2), both regarding the estimated L_1 distance, and the weight dependent DP for four values of the predictor, respectively.



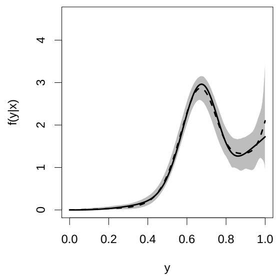
(a) $x = 0.2$ (b) $x = 0.2$ (c) $x = 0.2$

(d) $x = 0.4$ (e) $x = 0.4$ (f) $x = 0.4$

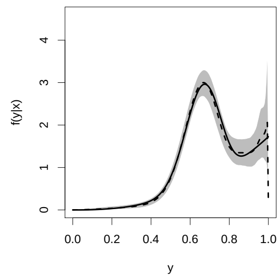
(g) $x = 0.6$ (h) $x = 0.6$ (i) $x = 0.6$

(j) $x = 0.8$ (k) $x = 0.8$ (l) $x = 0.8$

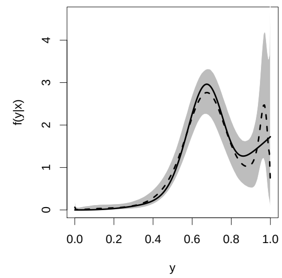
Figure B.5: Simulated data - Scenario II ($n = 500$): True (continuous line) and posterior mean (dotted line) for the conditional density. A band constructed using the 95%point–wise HPD intervals is presented in gray. Panels (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l) display the results for the best DBPP model (wLDBPP1), the best LDDP model (LDDP1), both regarding the estimated L_1 distance, and the weight dependent DP for four values of the predictor, respectively.



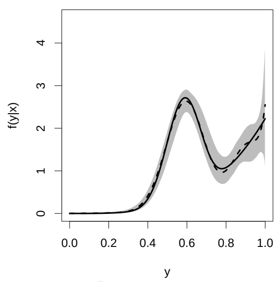
(a) $x = 0.2$



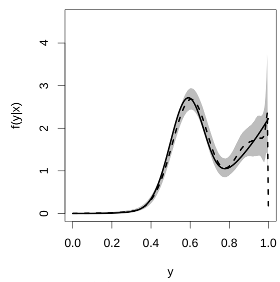
(b) $x = 0.2$



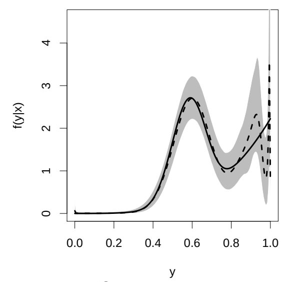
(c) $x = 0.2$



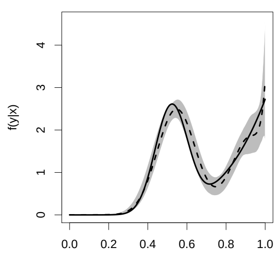
(d) $x = 0.4$



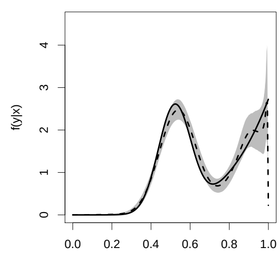
(e) $x = 0.4$



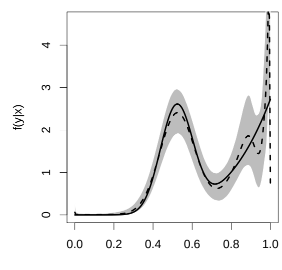
(f) $x = 0.4$



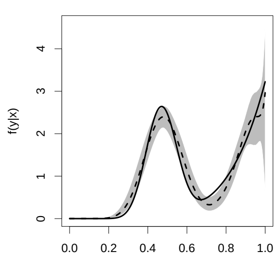
(g) $x = 0.6$



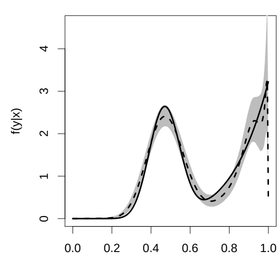
(h) $x = 0.6$



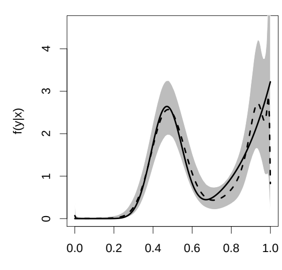
(i) $x = 0.6$



(j) $x = 0.8$

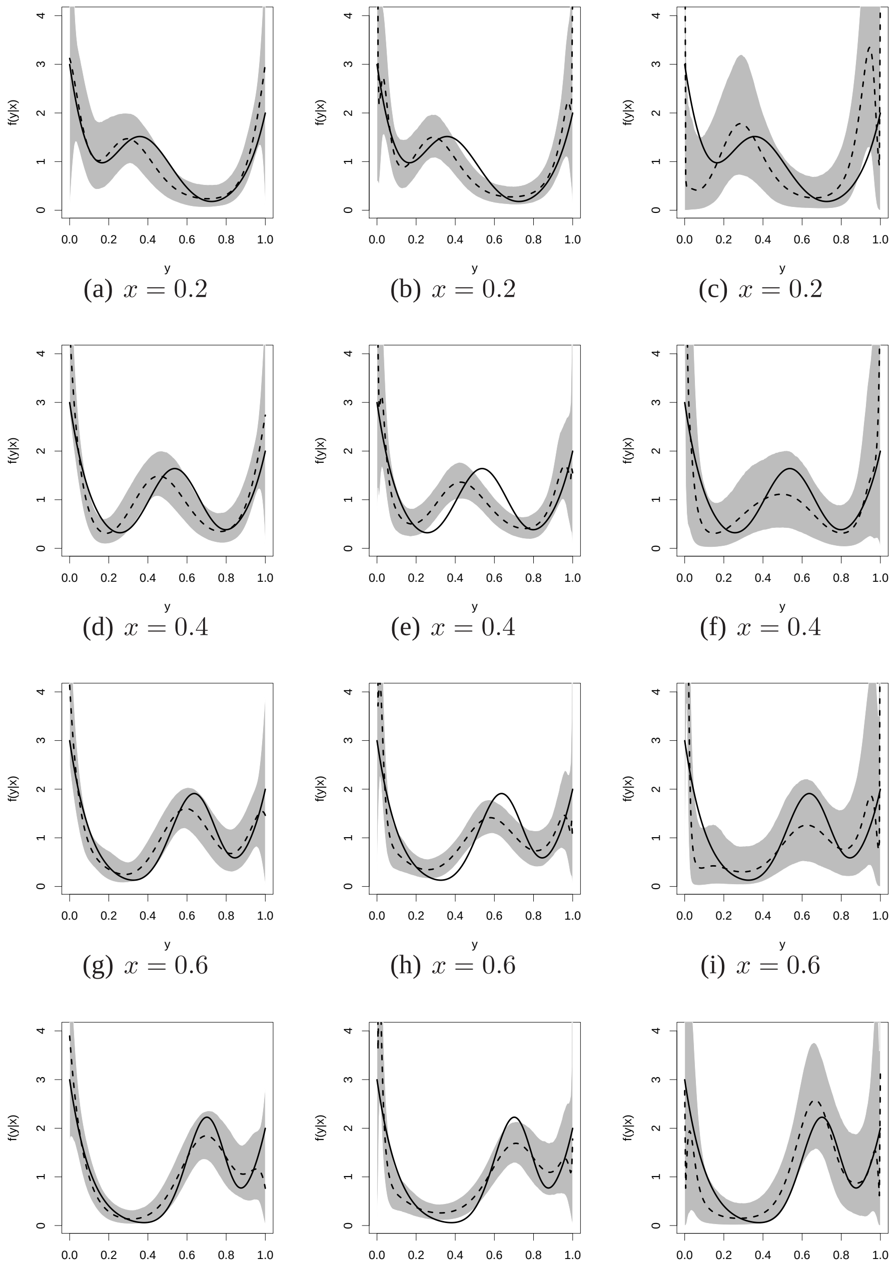


(k) $x = 0.8$



(l) $x = 0.8$

Figure B.6: Simulated data - Scenario II ($n = 1000$): True (continuous line) and posterior mean (dotted line) for the conditional density. A band constructed using the 95%point–wise HPD intervals is presented in gray. Panels (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l) display the results for the best DBPP model (LDBPP1), the best LDDP model (LDDP1), both regarding the estimated L_1 distance, and the weight dependent DP for four values of the predictor, respectively.



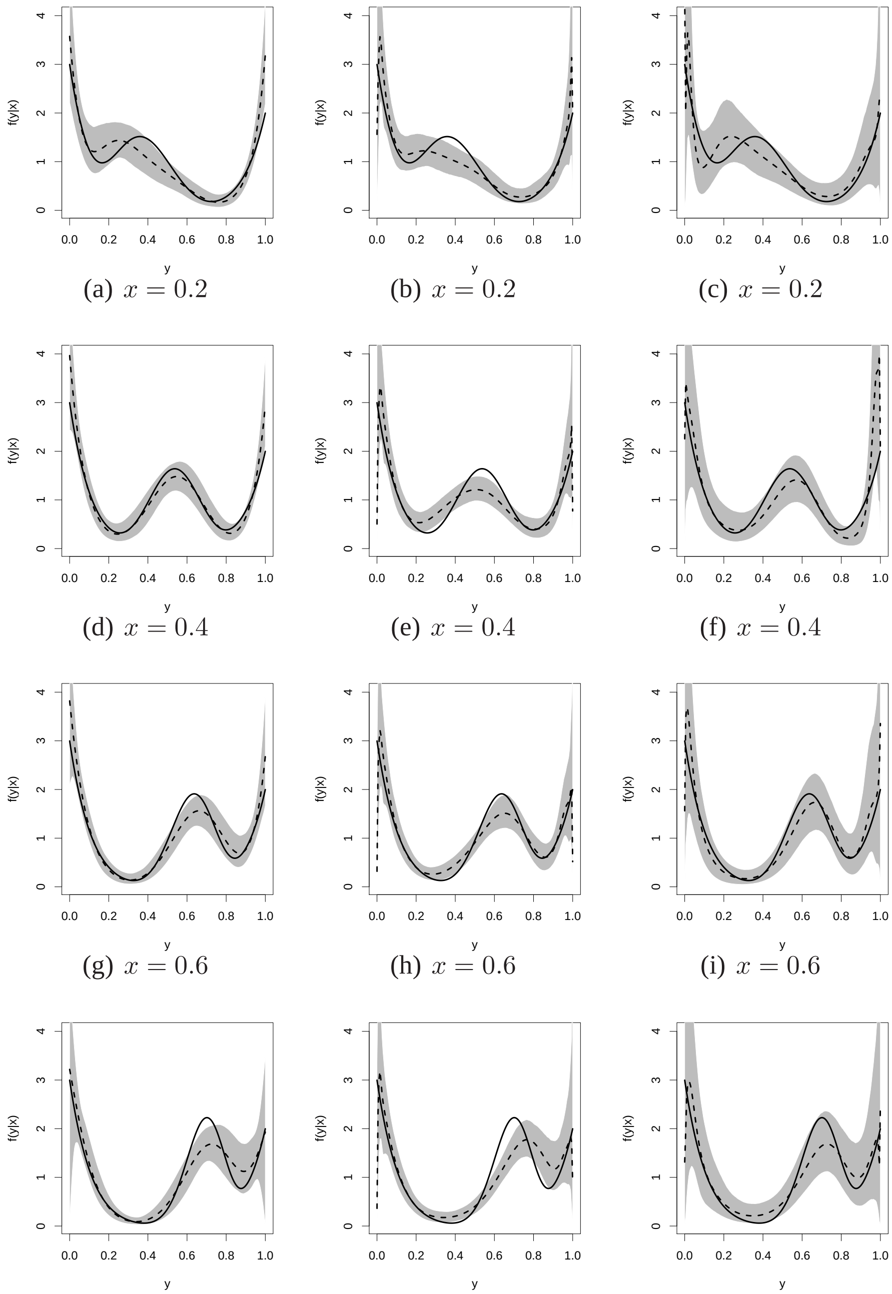
(a) $x = 0.2$ (b) $x = 0.2$ (c) $x = 0.2$

(d) $x = 0.4$ (e) $x = 0.4$ (f) $x = 0.4$

(g) $x = 0.6$ (h) $x = 0.6$ (i) $x = 0.6$

(j) $x = 0.8$ (k) $x = 0.8$ (l) $x = 0.8$

Figure B.7: Simulated data - Scenario III ($n = 250$): True (continuous line) and posterior mean (dotted line) for the conditional density. A band constructed using the 95%point–wise HPD intervals is presented in gray. Panels (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l) display the results for the best DBPP model (LDBPP1), the best LDDP model (LDDP1), both regarding the estimated L_1 distance, and the weight dependent DP for four values of the predictor, respectively.



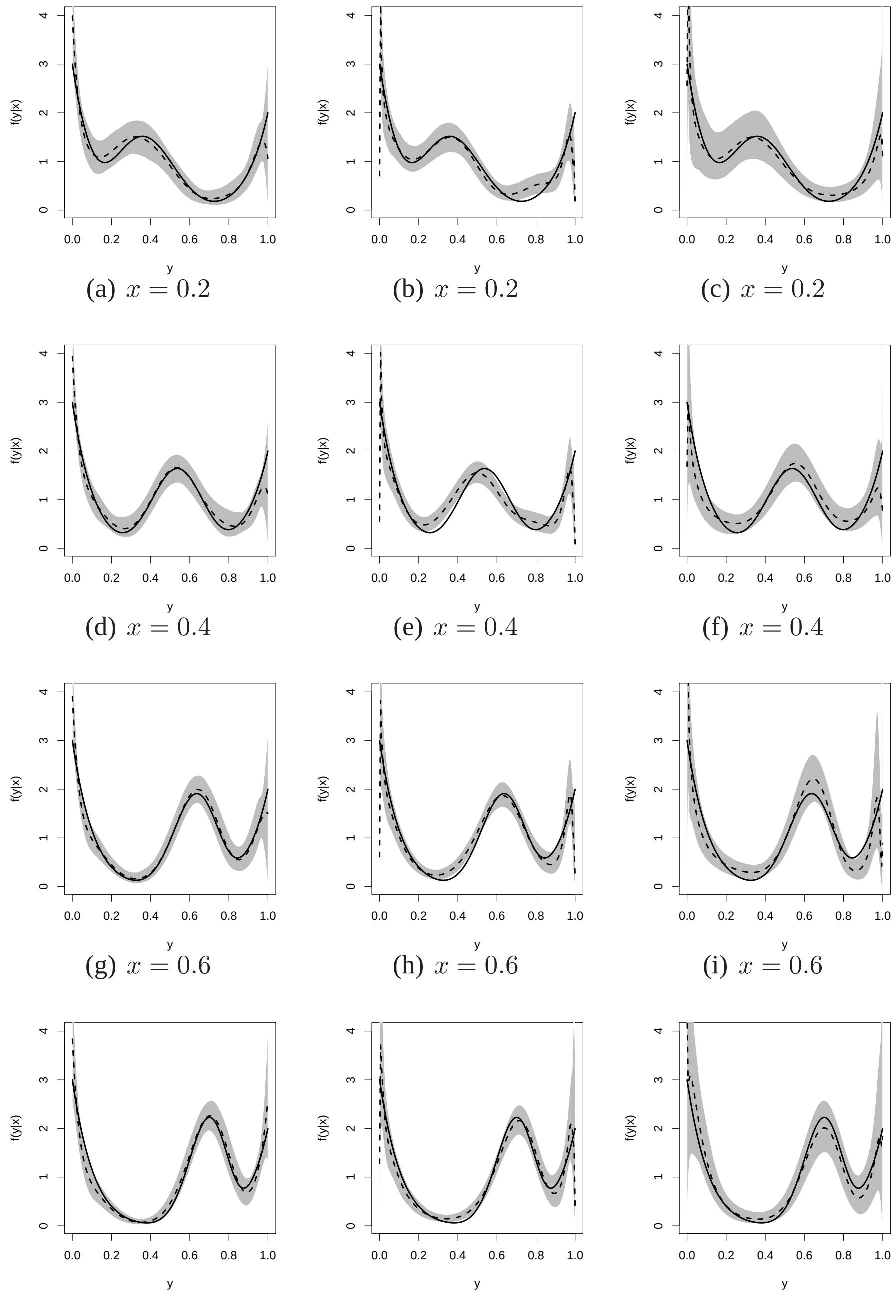
(a) $x = 0.2$ (b) $x = 0.2$ (c) $x = 0.2$

(d) $x = 0.4$ (e) $x = 0.4$ (f) $x = 0.4$

(g) $x = 0.6$ (h) $x = 0.6$ (i) $x = 0.6$

(j) $x = 0.8$ (k) $x = 0.8$ (l) $x = 0.8$

Figure B.8: Simulated data - Scenario III ($n = 500$): True (continuous line) and posterior mean (dotted line) for the conditional density. A band constructed using the 95%point–wise HPD intervals is presented in gray. Panels (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l) display the results for the best DBPP model (LDBPP1), the best LDDP model (LDDP1), both regarding the estimated L_1 distance, and the weight dependent DP for four values of the predictor, respectively.



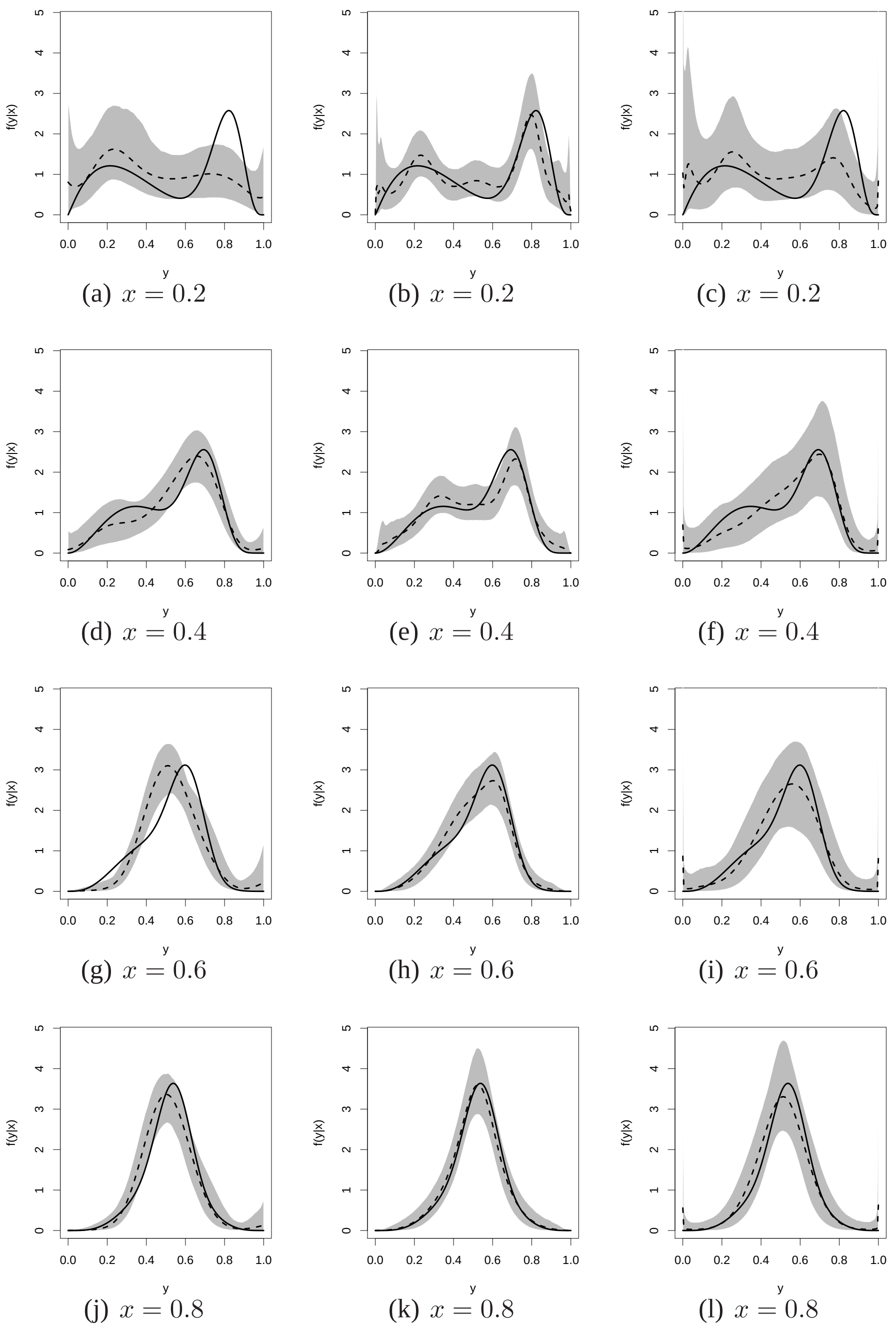
(a) $x = 0.2$ (b) $x = 0.2$ (c) $x = 0.2$

(d) $x = 0.4$ (e) $x = 0.4$ (f) $x = 0.4$

(g) $x = 0.6$ (h) $x = 0.6$ (i) $x = 0.6$

(j) $x = 0.8$ (k) $x = 0.8$ (l) $x = 0.8$

Figure B.9: Simulated data - Scenario III ($n = 1000$): True (continuous line) and posterior mean (dotted line) for the conditional density. A band constructed using the 95%point–wise HPD intervals is presented in gray. Panels (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l) display the results for the best DBPP model (LDBPP1), the best LDDP model (LDDP1), both regarding the estimated L_1 distance, and the weight dependent DP for four values of the predictor, respectively.



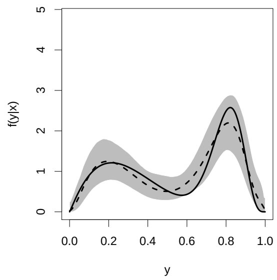
(a) $x = 0.2$ (b) $x = 0.2$ (c) $x = 0.2$

(d) $x = 0.4$ (e) $x = 0.4$ (f) $x = 0.4$

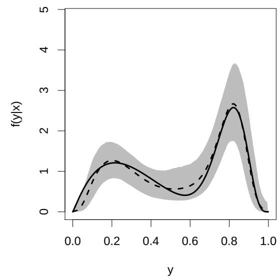
(g) $x = 0.6$ (h) $x = 0.6$ (i) $x = 0.6$

(j) $x = 0.8$ (k) $x = 0.8$ (l) $x = 0.8$

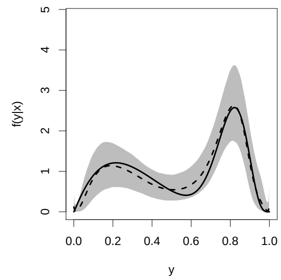
Figure B.10: Simulated data - Scenario I ($n = 250$): True (continuous line) and posterior mean (dotted line) for the conditional density. A band constructed using the 95%point–wise HPD intervals is presented in gray. Panels (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l) display the results for the best DBPP model (θLDBPP2), the best LDDP model (LDDP1), both regarding the estimated L_∞ distance, and the weight dependent DP for four values of the predictor, respectively.



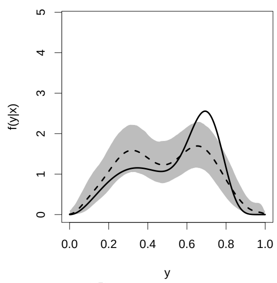
(a) $x = 0.2$



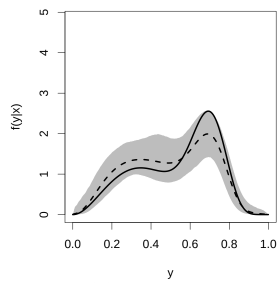
(b) $x = 0.2$



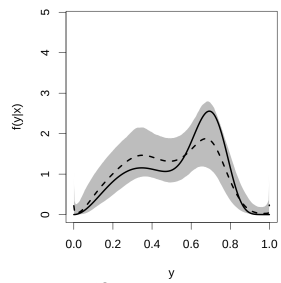
(c) $x = 0.2$



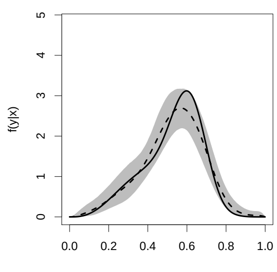
(d) $x = 0.4$



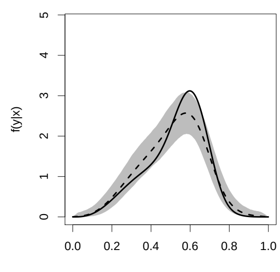
(e) $x = 0.4$



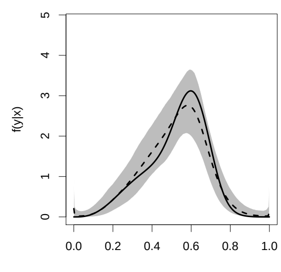
(f) $x = 0.4$



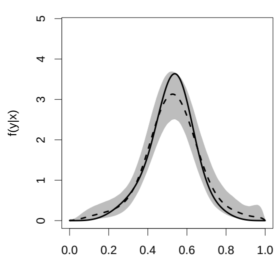
(g) $x = 0.6$



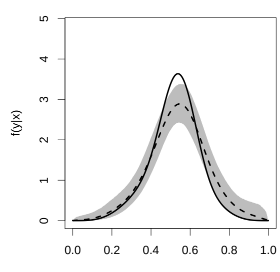
(h) $x = 0.6$



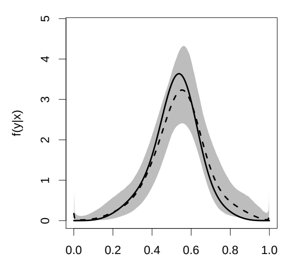
(i) $x = 0.6$



(j) $x = 0.8$

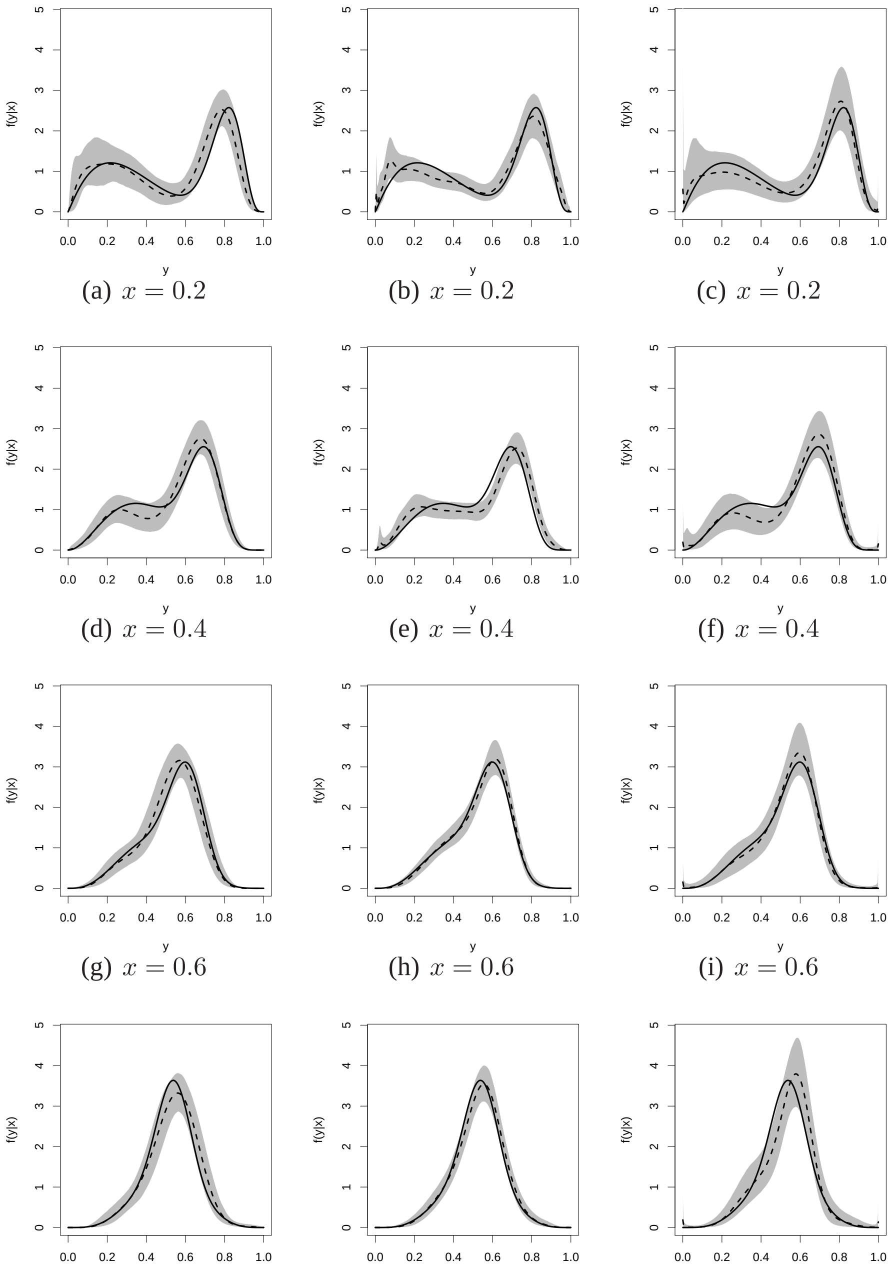


(k) $x = 0.8$



(l) $x = 0.8$

Figure B.11: Simulated data - Scenario I ($n = 500$): True (continuous line) and posterior mean (dotted line) for the conditional density. A band constructed using the 95%point–wise HPD intervals is presented in gray. Panels (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l) display the results for the best DBPP model (θLDBPP2), the best LDDP model (LDDP2), both regarding the estimated L_∞ distance, and the weight dependent DP for four values of the predictor, respectively.



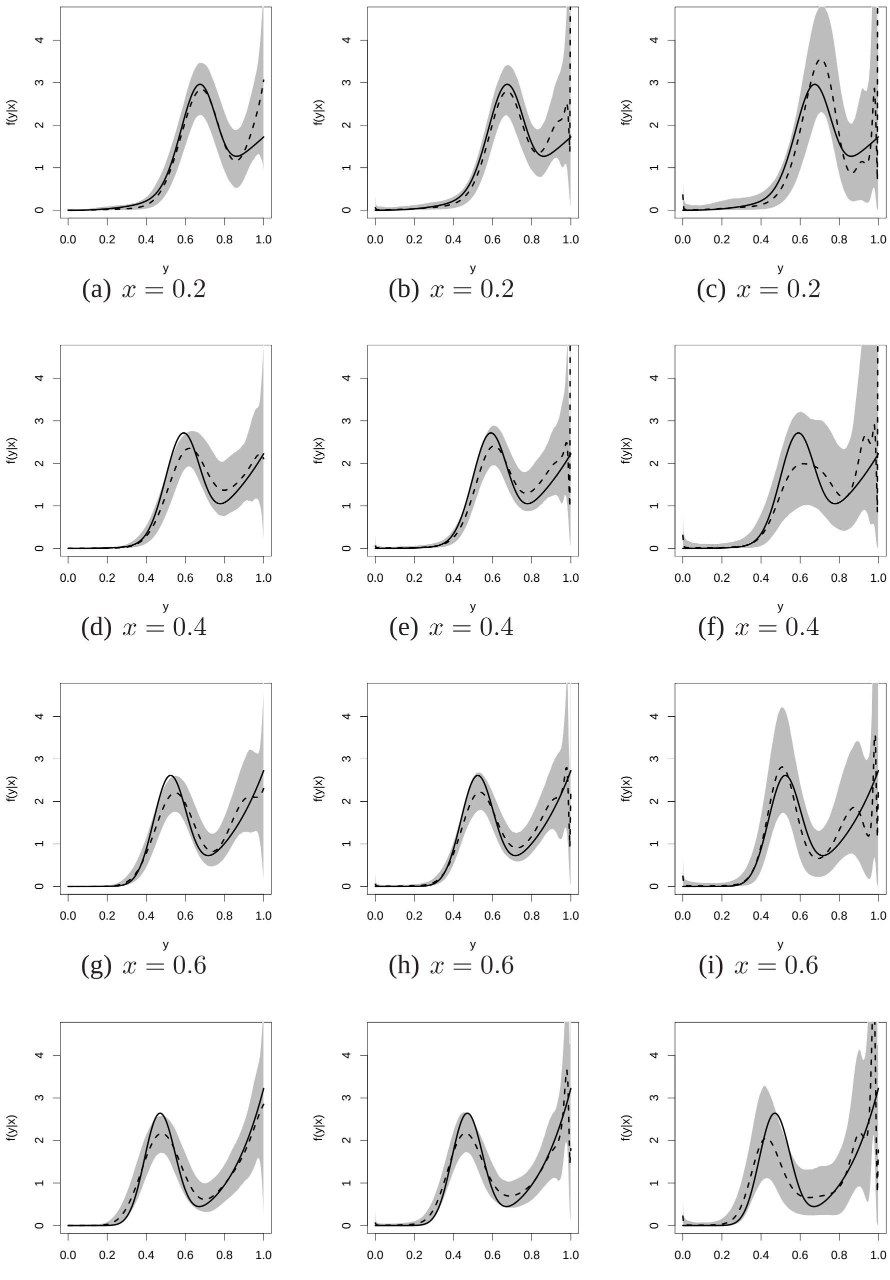
(a) $x = 0.2$ (b) $x = 0.2$ (c) $x = 0.2$

(d) $x = 0.4$ (e) $x = 0.4$ (f) $x = 0.4$

(g) $x = 0.6$ (h) $x = 0.6$ (i) $x = 0.6$

(j) $x = 0.8$ (k) $x = 0.8$ (l) $x = 0.8$

Figure B.12: Simulated data - Scenario I ($n = 1000$): True (continuous line) and posterior mean (dotted line) for the conditional density. A band constructed using the 95%point–wise HPD intervals is presented in gray. Panels (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l) display the results for the best DBPP model (LDBPP2), the best LDDP model (LDDP1), both regarding the estimated L_∞ distance, and the weight dependent DP for four values of the predictor, respectively.



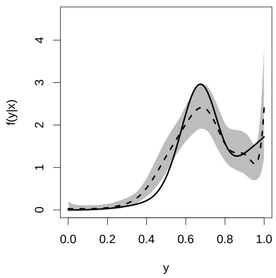
(a) $x = 0.2$ (b) $x = 0.2$ (c) $x = 0.2$

(d) $x = 0.4$ (e) $x = 0.4$ (f) $x = 0.4$

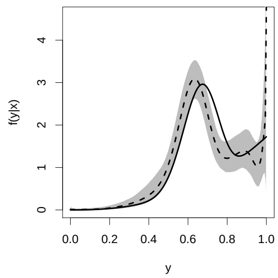
(g) $x = 0.6$ (h) $x = 0.6$ (i) $x = 0.6$

(j) $x = 0.8$ (k) $x = 0.8$ (l) $x = 0.8$

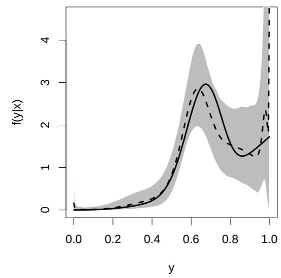
Figure B.13: Simulated data - Scenario II ($n = 250$): True (continuous line) and posterior mean (dotted line) for the conditional density. A band constructed using the 95%point–wise HPD intervals is presented in gray. Panels (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l) display the results for the best DBPP model (wLDBPP1), the best LDDP model (LDDP1), both regarding the estimated L_∞ distance, and the weight dependent DP for four values of the predictor, respectively.



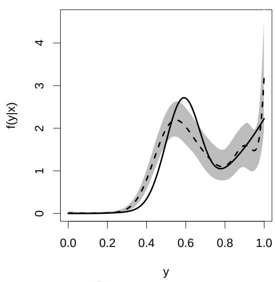
(a) $x = 0.2$



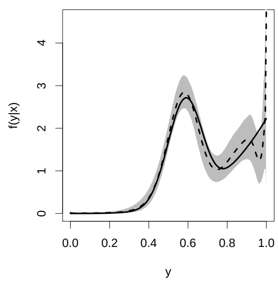
(b) $x = 0.2$



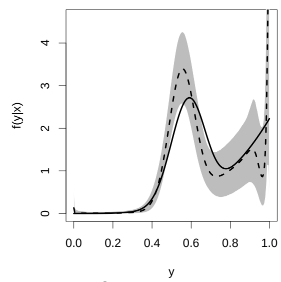
(c) $x = 0.2$



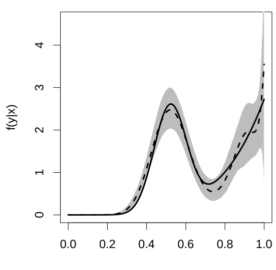
(d) $x = 0.4$



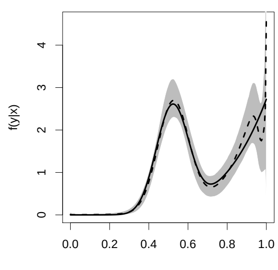
(e) $x = 0.4$



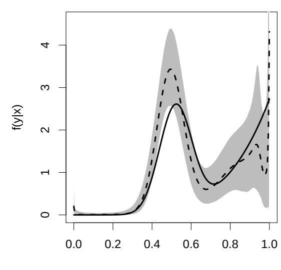
(f) $x = 0.4$



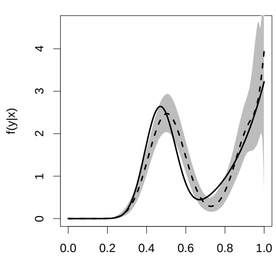
(g) $x = 0.6$



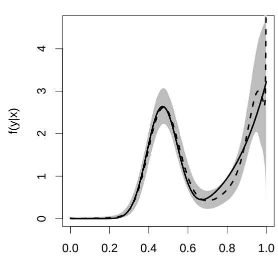
(h) $x = 0.6$



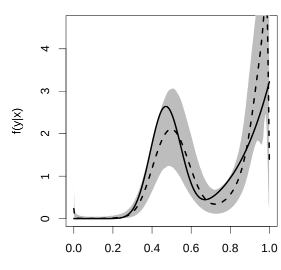
(i) $x = 0.6$



(j) $x = 0.8$

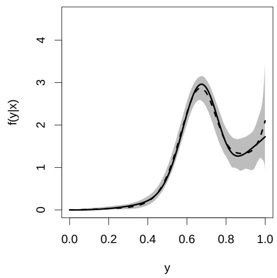


(k) $x = 0.8$

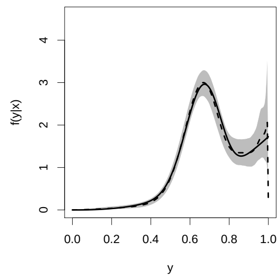


(l) $x = 0.8$

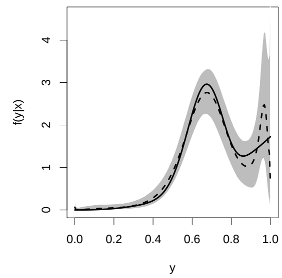
Figure B.14: Simulated data - Scenario II ($n = 500$): True (continuous line) and posterior mean (dotted line) for the conditional density. A band constructed using the 95%point–wise HPD intervals is presented in gray. Panels (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l) display the results for the best DBPP model (θLDBPP1), the best LDDP model (LDDP1), both regarding the estimated L_∞ distance, and the weight dependent DP for four values of the predictor, respectively.



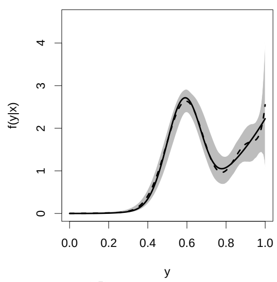
(a) $x = 0.2$



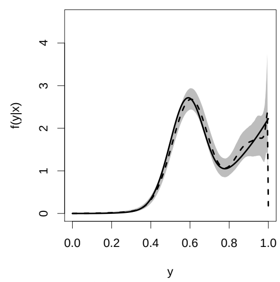
(b) $x = 0.2$



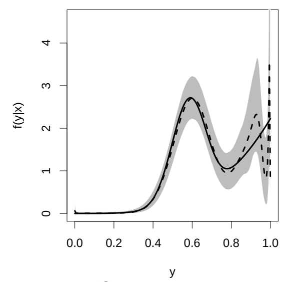
(c) $x = 0.2$



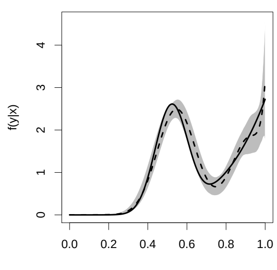
(d) $x = 0.4$



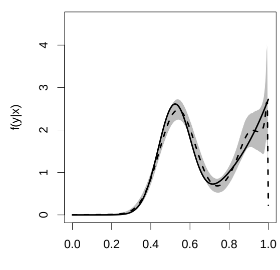
(e) $x = 0.4$



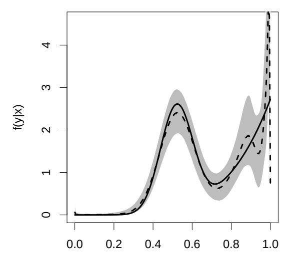
(f) $x = 0.4$



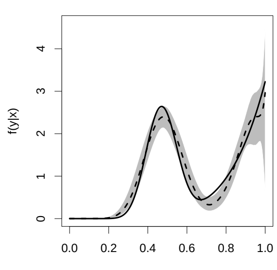
(g) $x = 0.6$



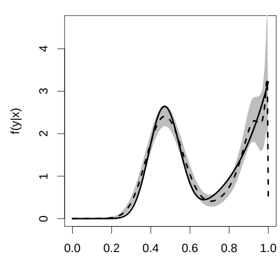
(h) $x = 0.6$



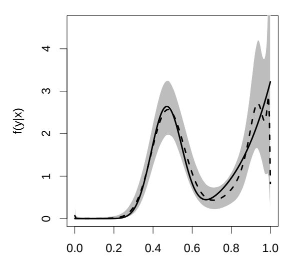
(i) $x = 0.6$



(j) $x = 0.8$

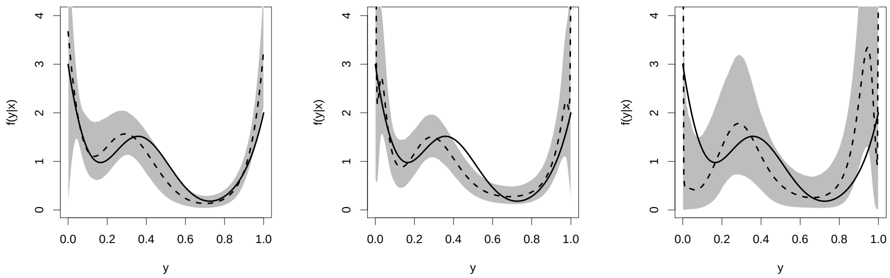


(k) $x = 0.8$

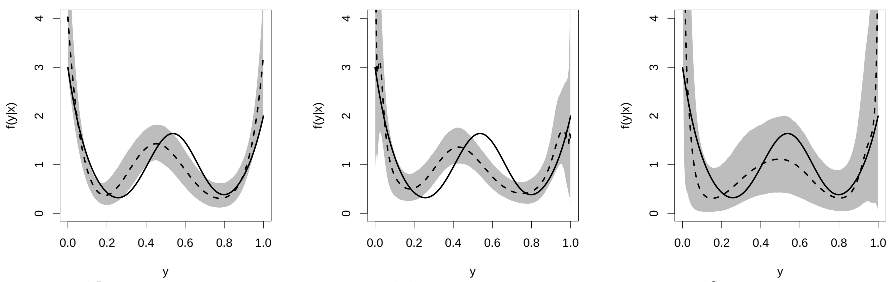


(l) $x = 0.8$

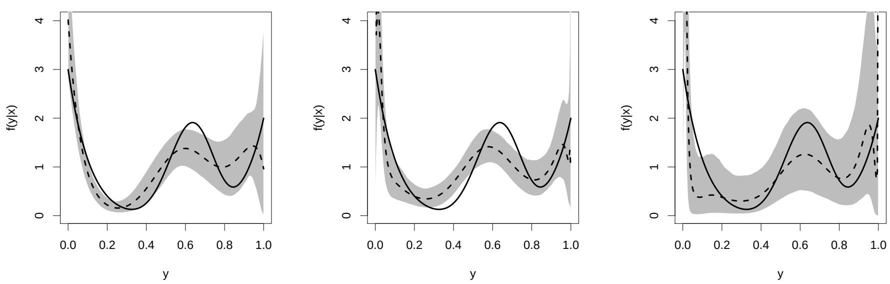
Figure B.15: Simulated data - Scenario III ($n = 1000$): True (continuous line) and posterior mean (dotted line) for the conditional density. A band constructed using the 95%point–wise HPD intervals is presented in gray. Panels (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l) display the results for the best DBPP model (LDBPP1), the best LDDP model (LDDP1), both regarding the estimated L_∞ distance, and the weight dependent DP for four values of the predictor, respectively.



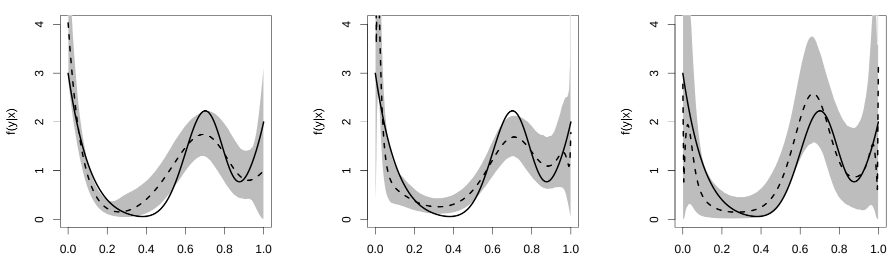
(a) $x = 0.2$ (b) $x = 0.2$ (c) $x = 0.2$



(d) $x = 0.4$ (e) $x = 0.4$ (f) $x = 0.4$

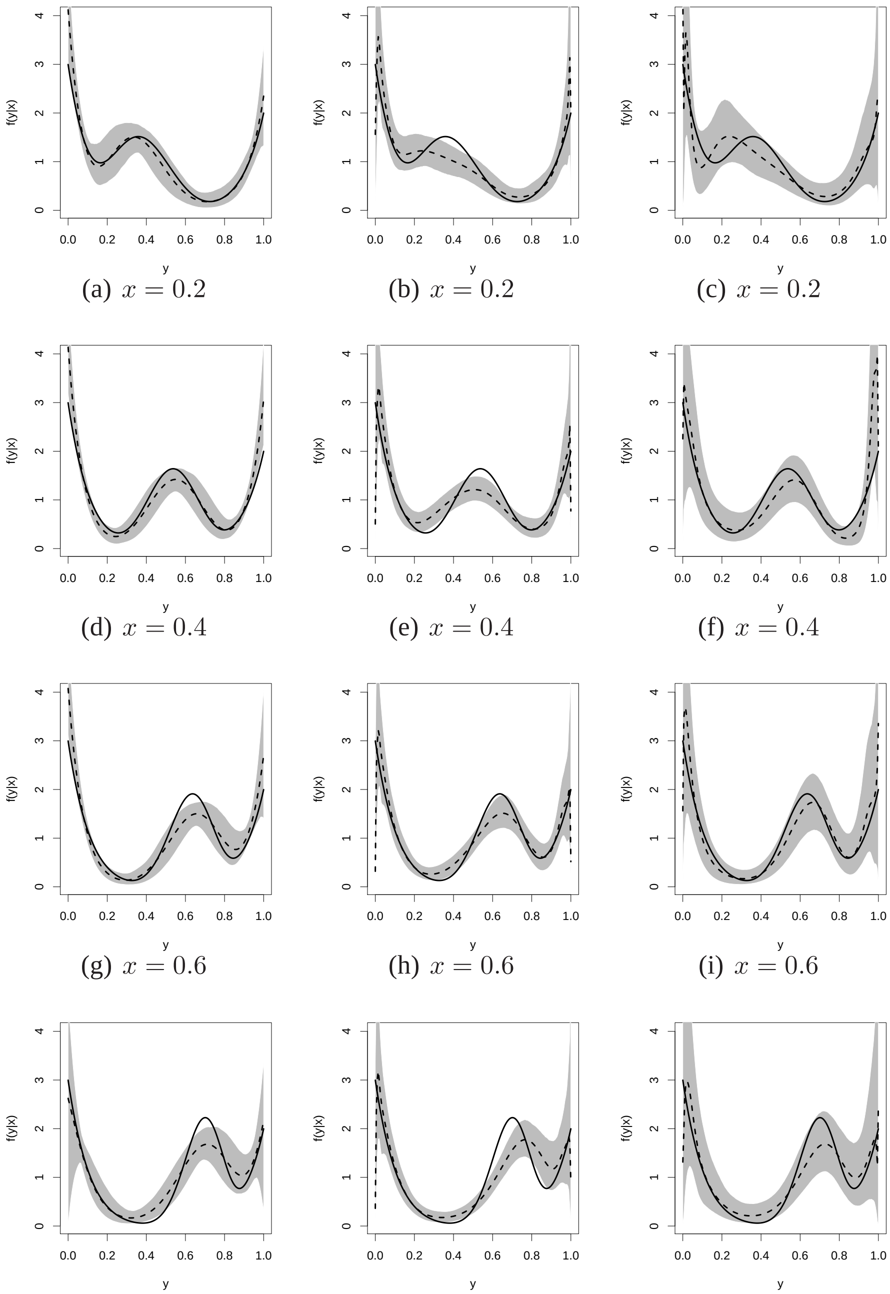


(g) $x = 0.6$ (h) $x = 0.6$ (i) $x = 0.6$



(j) $x = 0.8$ (k) $x = 0.8$ (l) $x = 0.8$

Figure B.16: Simulated data - Scenario III ($n = 250$): True (continuous line) and posterior mean (dotted line) for the conditional density. A band constructed using the 95%point–wise HPD intervals is presented in gray. Panels (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l) display the results for the best DBPP model (wLDBPP1), the best LDDP model (LDDP1), both regarding the estimated L_∞ distance, and the weight dependent DP for four values of the predictor, respectively.



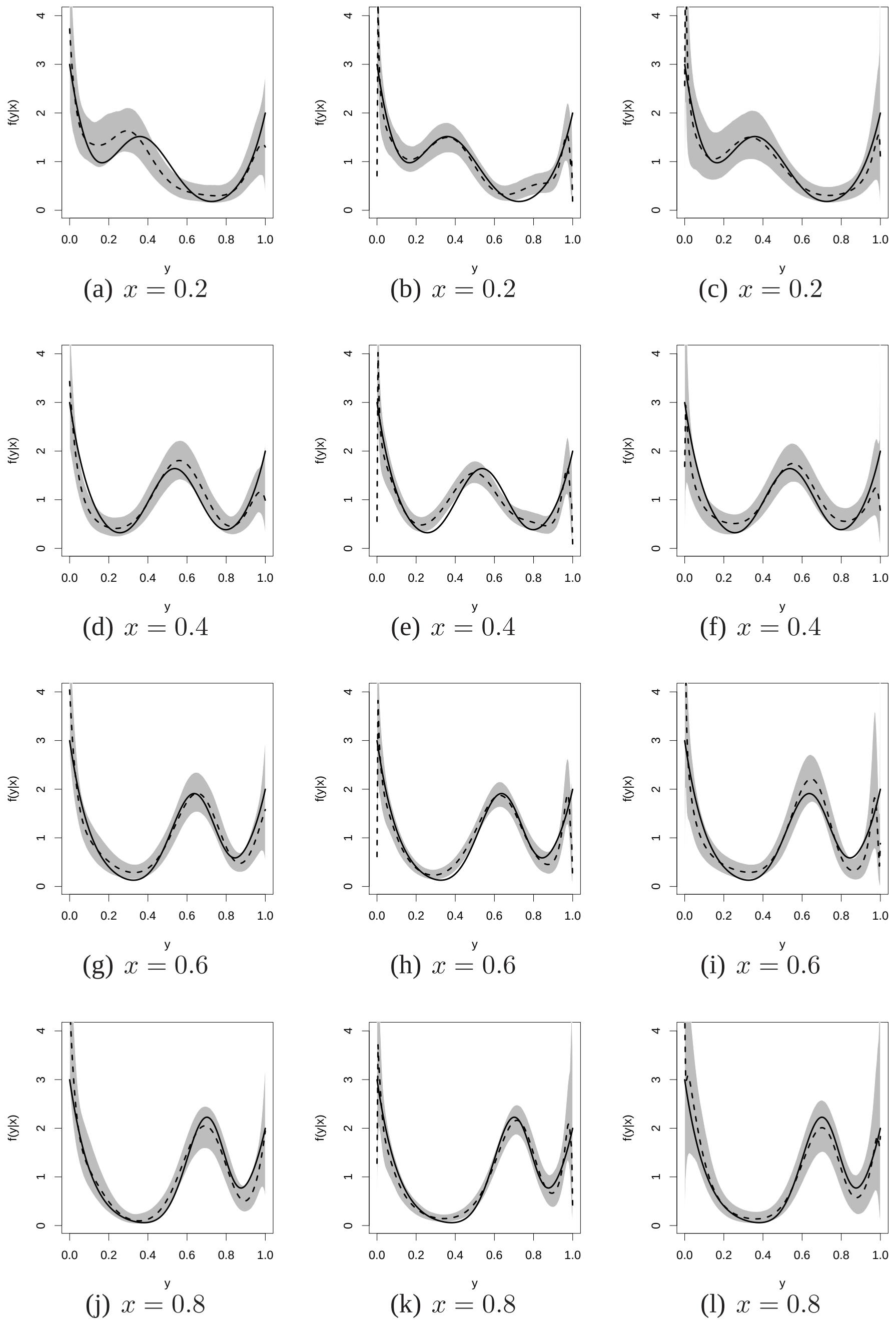
(a) $x = 0.2$ (b) $x = 0.2$ (c) $x = 0.2$

(d) $x = 0.4$ (e) $x = 0.4$ (f) $x = 0.4$

(g) $x = 0.6$ (h) $x = 0.6$ (i) $x = 0.6$

(j) $x = 0.8$ (k) $x = 0.8$ (l) $x = 0.8$

Figure B.17: Simulated data - Scenario III ($n = 500$): True (continuous line) and posterior mean (dotted line) for the conditional density. A band constructed using the 95%point–wise HPD intervals is presented in gray. Panels (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l) display the results for the best DBPP model (wLDBPP2), the best LDDP model (LDDP1), both regarding the estimated L_∞ distance, and the weight dependent DP for four values of the predictor, respectively.



(a) $x = 0.2$ (b) $x = 0.2$ (c) $x = 0.2$

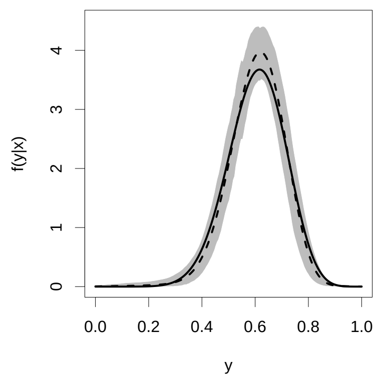
(d) $x = 0.4$ (e) $x = 0.4$ (f) $x = 0.4$

(g) $x = 0.6$ (h) $x = 0.6$ (i) $x = 0.6$

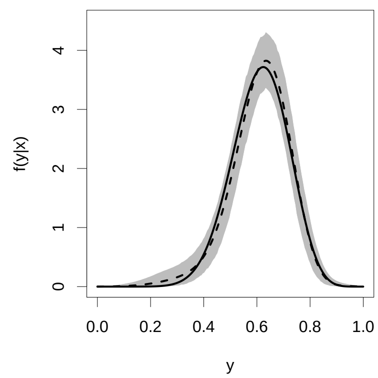
(j) $x = 0.8$ (k) $x = 0.8$ (l) $x = 0.8$

Figure B.18: Simulated data - Scenario III ($n = 1000$): True (continuous line) and posterior mean (dotted line) for the conditional density. A band constructed using the 95%point–wise HPD intervals is presented in gray. Panels (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l) display the results for the best DBPP model (θLDBPP2), the best LDDP model (LDDP1), both regarding the estimated L_∞ distance, and the weight dependent DP for four values of the predictor, respectively.

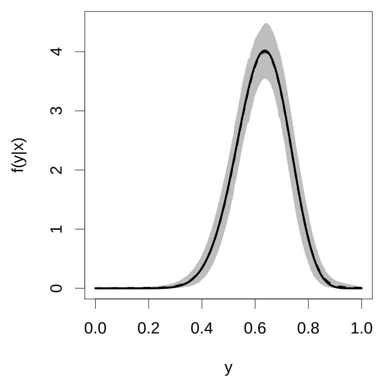
B.7 Additional results for the proportion of food



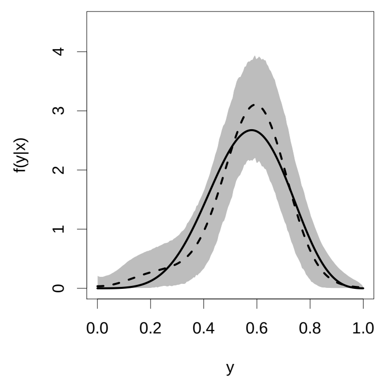
(a) Socioeconomic level: low-low



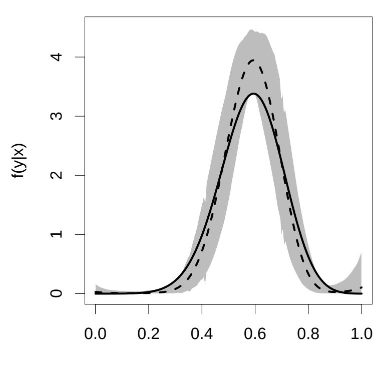
(b) Socioeconomic level: low



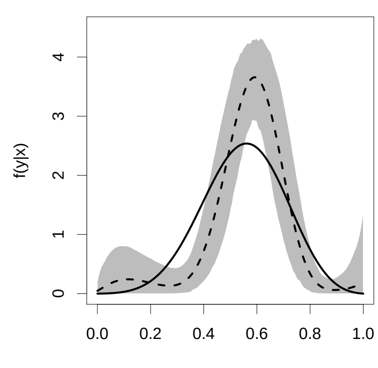
(c) Socioeconomic level medium-low



(d) Socioeconomic level: medium

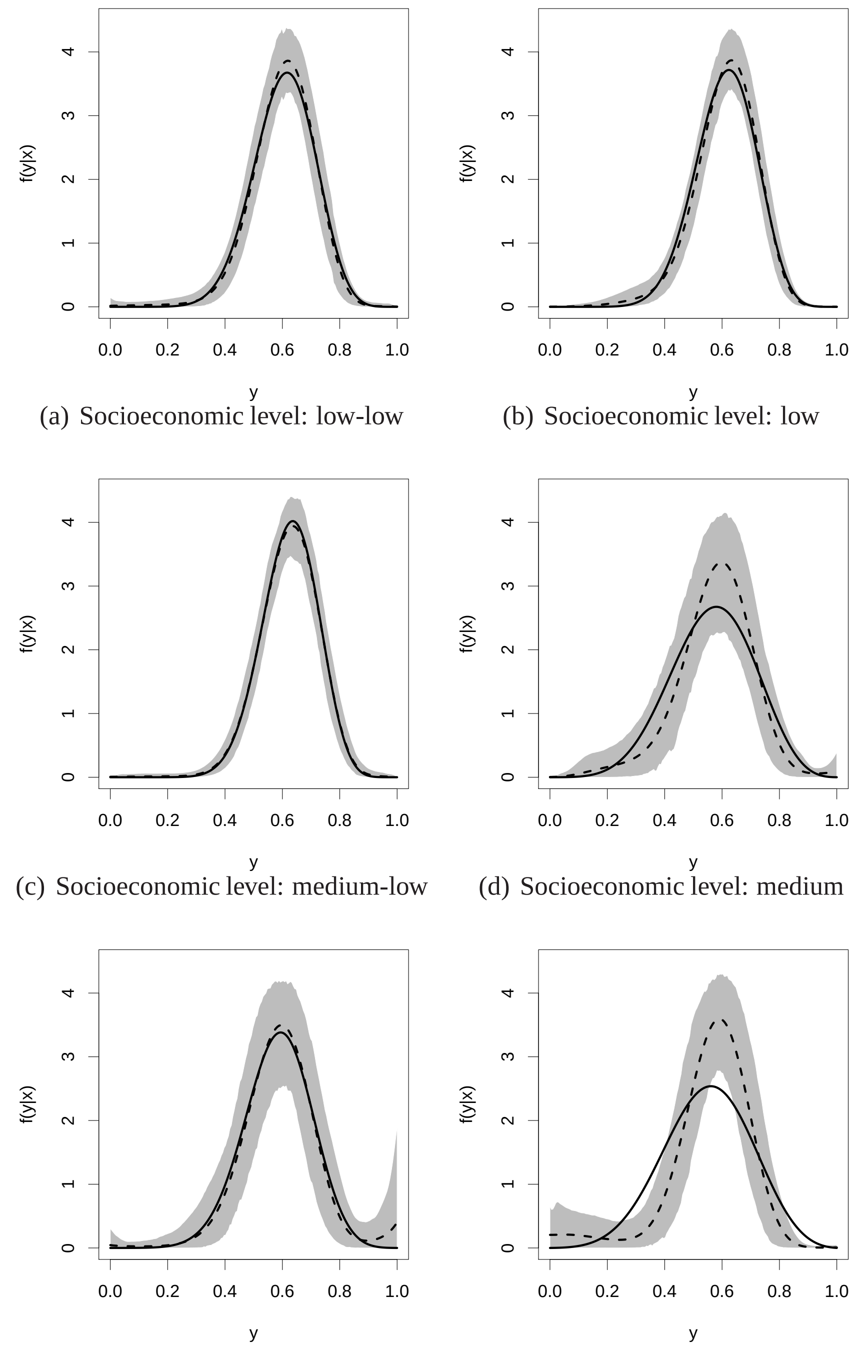


(e) Socioeconomic level: medium-high



(f) Socioeconomic level: high

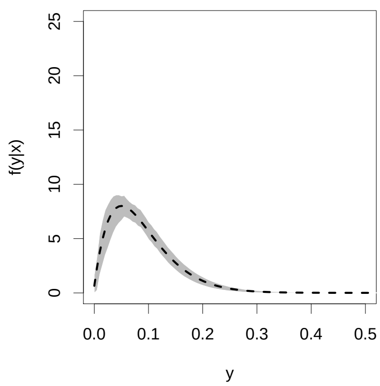
Figure B.19: Proportion of food - LDBPP model. Panels (a), (b), (c), (d), (e) and (f) display the posterior mean (dashed line) and a 95% point-wise HPD band (grey area) for the conditional density at socioeconomic level low-low, low, medium-low, medium, medium-high and high, respectively, under the LDBPP model. The posterior mean under the parametric beta regression model is given as a solid line for comparison purposes.



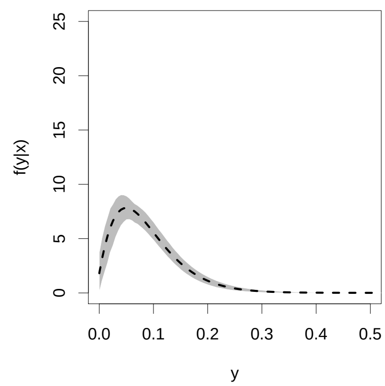
(a) Socioeconomic level: low-low
(b) Socioeconomic level: low
(c) Socioeconomic level: medium-low
(d) Socioeconomic level: medium
(e) Socioeconomic level: medium-high
(f) Socioeconomic level: high

Figure B.20: Proportion of food - wLDBPP model. Panels (a), (b), (c), (d), (e) and (f) display the posterior mean (dashed line) and a 95% point-wise HPD band (grey area) for the conditional density at socioeconomic level low-low, low, medium-low, medium, medium-high and high, respectively, under the wLDBPP model. The posterior mean under the parametric beta regression model is given as a solid line for comparison purposes.

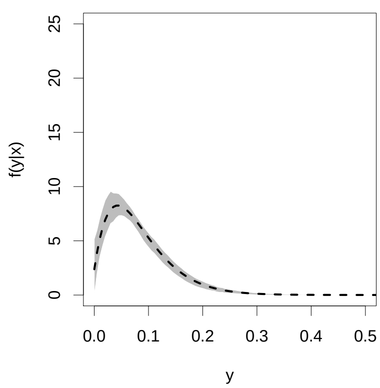
B.8 Additional results for the proportion of hygienic waste



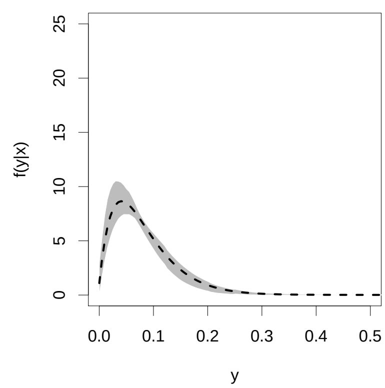
(a) Socioeconomic level: low-low



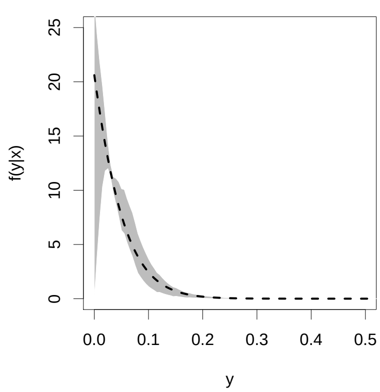
(b) Socioeconomic level: low



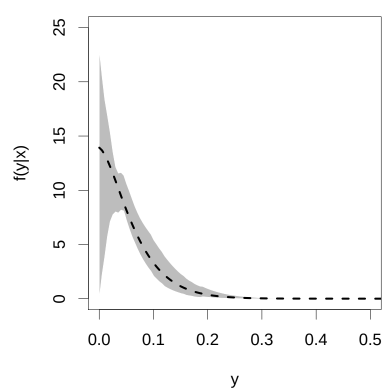
(c) Socioeconomic level: medium-low



(d) Socioeconomic level: medium

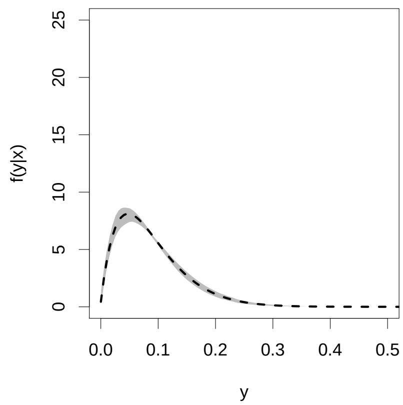


(e) Socioeconomic level: medium-high

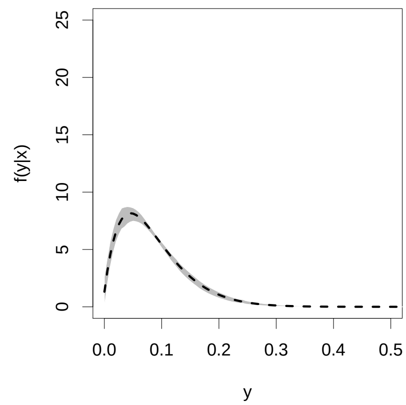


(f) Socioeconomic level: high

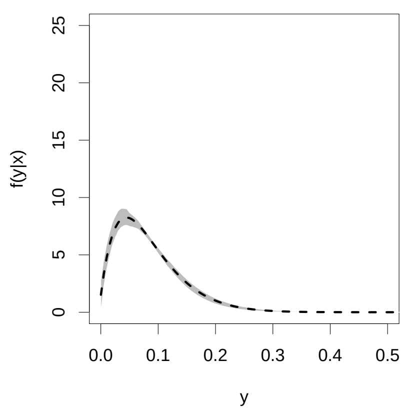
Figure B.21: Proportion of hygienic waste - LDBPP model. Panels (a), (b), (c), (d), (e) and (f) display the posterior mean (dashed line) and a 95% point-wise HPD band (grey area) for the conditional density at socioeconomic level low-low, low, medium-low, medium, medium-high and high, respectively, under the LDBPP model.



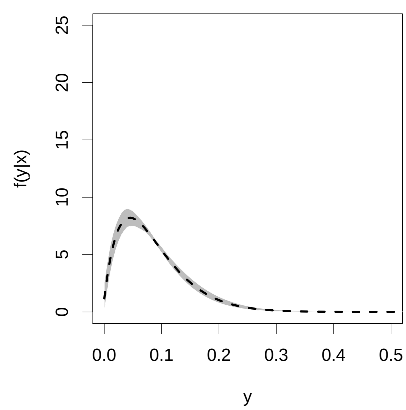
(a) Socioeconomic level low-low



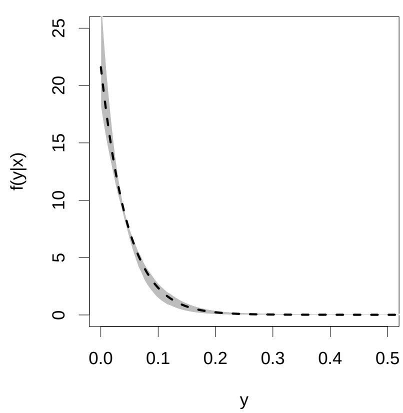
(b) Socioeconomic: level low



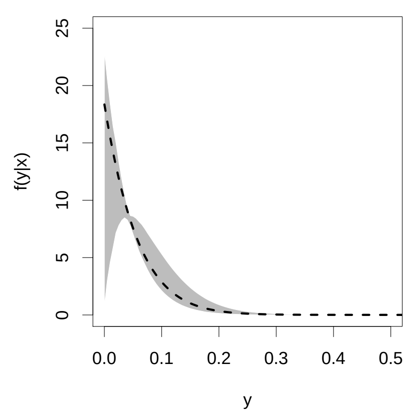
(c) Socioeconomic level: medium-low



(d) Socioeconomic level: medium



(e) Socioeconomic level: medium-high



(f) Socioeconomic: level high

Figure B.22: Proportion of hygienic waste - wLDBPP model. Panels (a), (b), (c), (d), (e) and (f) display the posterior mean (dashed line) and a 95% point-wise HPD band (grey area) for the conditional density at socioeconomic level low-low, low, medium-low, medium, medium-high and high, respectively, under the wLDBPP model.

Bibliography

ABRAMOWITZ, M. & STEGUN, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York, USA: Dover.

ANGERS, J. F. & DELAMPADY, M. (1992). Hierarchical bayesian curve fitting and smoothing. *The Canadian Journal of Statistics* 20 35–49.

ANTONIAK, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian non-parametric problems. *The Annals of Statistics* 2 1152–1174.

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.

BERNSTEIN, S. N. (1912). Démonstration du théoreme de weierstrass fondée sur le calcul des probabilités. *Communications of the Kharkov Mathematical Society* 13 1–2.

BLACKWELL, D. & MACQUEEN, J. (1973). Ferguson distributions via Pólya urn schemes. *The Annals of Statistics* 1 353–355.

CARON, F., DAVY, M., DOUCET, A., DUFLOS, E. & VANHEEGHE, P. (2006). Bayesian inference for dynamic models with Dirichlet process mixtures. In *International Conference on Information Fusion*. Florence, Italia, July 10-13.

CARON, F., DAVY, M., DOUCET, A., DUFLOS, E. & VANHEEGHE, P. (2008). Bayesian inference for linear dynamic models with Dirichlet process mixtures. *IEEE Transactions on Signal Processing* 56 71–84.

CHRISTENSEN, R., HANSON, T. & JARA, A. (2008). Parametric nonparametric statistics: An introduction to mixtures of finite Polya trees. *The American Statistician* 62 296–306.

CHUNG, Y. & DUNSON, D. B. (2009). Nonparametric Bayes conditional distribution modeling with variable selection. *Journal of the American Statistical Association* 104 1646–1660.

CHUNG, Y. & DUNSON, D. B. (2011). The local Dirichlet process. *Annals of the Institute of Statistical Mathematics* 63 59–80.

CIFARELLI, D. & REGAZZINI, E. (1978). Problemi statistici non parametrici in condizioni di scambialbilita parziale e impiego di medie associative. Tech. rep., Quaderni Istituto Matematica Finanziaria, Torino.

CIFARELLI, D. M. & REGAZZINI, E. (1990). Distribution functions of means of a Dirichlet process. *The Annals of Statistics* 18 429–442.

DARSOW, W. F., NGUYEN, B. & OLSEN, E. T. (1992). Copulas and Markov processes. *Illinois Journal of Mathematics* 36 600–642.

DE IORIO, M., JOHNSON, W. O., MÜLLER, P. & ROSNER, G. L. (2009). Bayesian nonparametric non-proportional hazards survival modelling. *Biometrics* 65 762–771.

DE IORIO, M., MÜLLER, P., ROSNER, G. L. & MACEACHERN, S. N. (2004). An ANOVA model for dependent random measures. *Journal of the American Statistical Association* 99 205–215.

DE LA CRUZ, R., QUINTANA, F. A. & MÜLLER, P. (2007). Semiparametric Bayesian classification with longitudinal markers. *Applied Statistics* 56 119–137.

DI LUCCA, M., GUGLIELMI, A., MÜLLER, P. & QUINTANA, F. (2012). A simple class of Bayesian nonparametric autoregression models. Tech. rep., Department of Statistics, Pontificia Universidad Católica de Chile.

DUNSON, D. B. & HERRING, A. H. (2006). Semiparametric Bayesian latent trajectory models. Tech. rep., ISDS Discussion Paper 16, Duke University, Durham, NC, USA.

DUNSON, D. B. & PARK, J. H. (2008). Kernel stick-breaking processes. *Biometrika* 95 307–323.

DUNSON, D. B., PILLAI, N. & PARK, J. H. (2007). Bayesian density regression. *Journal of the Royal Statistical Society, Series B* 69 163–183.

DUNSON, D. B., XUE, Y. & CARIN, L. (2008). The matrix stick-breaking process: Flexible Bayes meta-analysis. *Journal of the American Statistical Association* 103 317–327.

EILERS, P. H. C. & MARX, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* 11 89–121.

EPIFANI, I. & LIJOI, A. (2010). Nonparamtric priors for vectors of survival functions. *Statistica Sinica* 20 1455–1484.

ESCOBAR, M. D. & WEST, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* 90 577–588.

FABIUS, J. (1964). Asymptotic behavior of Bayes' estimates. *The Annals of Mathematical Statistics* 35 846–856.

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics* 1 209–230.

FERGUSON, T. S. (1974). Prior distribution on the spaces of probability measures. *Annals of Statistics* 2 615–629.

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

FERRARI, S. L. P. & CRIBARI-NETO, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics* 31 799–815.

FREEDMAN, D. (1963). On the asymptotic distribution of Bayes' estimates in the discrete case. *Annals of Mathematical Statistics* 34 1386–1403.

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

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

GELFAND, A. E., KOTTAS, A. & MACEACHERN, S. N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association* 100 1021–1035.

GELFAND, A. E. & MALLICK, B. K. (1995). Bayesian analysis of proportional hazards models built from monotone functions. *Biometrics* 51 843–852.

GHOSAL, S., GHOSH, J. K. & RAMAMOORTHI, R. V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *The Annals of Statistics* 27 143–158.

GHOSAL, S. & VAN DER VAART, A. W. (2007). Posterior convergence rates of Dirichlet mixtures at smooth densities. *The Annals of Statistics* 35 697–723.

GHOSH, J. K. & RAMAMOORTHI, R. V. (2003). *Bayesian nonparametrics*. New York, USA: Springer.

GIUDICI, P., MEZZETTI, M. & MULIERE, P. (2003). Mixtures of Dirichlet process priors for variable selection in survival analysis. *Journal of Statistical Planning and Inference* 111 101–115.

GRIFFIN, J. E. & STEEL, M. F. J. (2006). Order-based dependent Dirichlet processes. *Journal of the American Statistical Association* 101 179–194.

HANSON, T. (2006). Inference for mixtures of finite Polya tree models. *Journal of the American Statistical Association* 101 1548–1565.

HANSON, T., BRANSCUM, A. & JOHNSON, W. (2005). Bayesian nonparametric modeling and data analysis: an introduction. In D. K. Dey & C. R. Rao, eds., *Bayesian Thinking: Modeling and Computation (Handbook of Statistics, volume 25)*. Amsterdam, The Netherlands: Elsevier, 245–278.

HANSON, T. & JOHNSON, W. O. (2002). Modeling regression error with a mixture of Polya trees. *Journal of the American Statistical Association* 97 1020–1033.

HJORT, N. L., HOLMES, C., MÜLLER, P. & WALKER, S. (2010). *Bayesian nonparametrics*. Cambridge, UK: Cambridge University Press.

HJORT, N. L. & ONGARO, A. (2005). Exact inference for random Dirichlet means. *Statistical Inference for Stochastic Processes* 8 227–254.

HJORT, N. L. & WALKER, S. (2009). Quantile pyramids for Bayesian nonparametrics. *The Annals of Statistics* 37 105–131.

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

ISHWARAN, H. & ZAREPOUR, M. (2000). Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika* 87 371–390.

JARA, A. (2007). Applied Bayesian non- and semi-parametric inference using DPpackage. *Rnews* 7 17–26.

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

JARA, A., HANSON, T. & LESAFFRE, E. (2009). Robustifying generalized linear mixed models using a new class of mixture of multivariate Polya trees. *Journal of Computational and Graphical Statistics* 18 838–860.

JARA, A., HANSON, T., QUINTANA, F., MÜLLER, P. & ROSNER, G. L. (2011). DPpackage: Bayesian Semi- and Nonparametric Modeling in R. *Journal of Statistical Software* 40 1–30.

JARA, A., LESAFFRE, E., DE IORIO, M. & QUINTANA, F. A. (2010). Bayesian semiparametric inference for multivariate doubly-interval-censored data. *Annals of Applied Statistics* 4 2126–2149.

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

KARABATSOS, G. & WALKER, S. G. (2011). Bayesian unimodal density regression. Tech. rep., Department of Educational Psychology, University of Illinois-Chicago, USA.

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.

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

KORWAR, R. M. & HOLLANDER, M. (1973). Contributions to the theory of Dirichlet processes. *The Annals of Probability* 1 705–711.

KOTTAS, A. (2006). Dirichlet process mixtures of beta distributions, with applications to density and intensity estimation. Tech. rep., Department of Applied Mathematics and Statistics, University of California, Santa Cruz.

LANG, S. & BREZGER, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics* 13 183–212.

LAVINE, M. (1992). Some aspects of Polya tree distributions for statistical modeling. *The Annals of Statistics* 20 1222–1235.

LAVINE, M. (1994). More aspects of Polya tree distributions for statistical modeling. *The Annals of Statistics* 22 1161–1176.

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

LIJOI, A., NIPOTI, B. & PRÜNSTER, I. (2012). Bayesian inference with dependent normalized completely random measures. Tech. rep., Department of Mathematics, University of Pavia.

LIJOI, A., PRÜNSTER, I. & WALKER, S. (2004). Extending Doob's consistency theorem to nonparametric densities. *Bernoulli* 10 651–663.

LIJOI, A., PRÜNSTER, I. & WALKER, S. (2005). On consistency of non-parametric normal mixtures for Bayesian density estimation. *Journal of the American Statistical Association* 100 1292–1296.

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

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.

MALLICK, B. K. & GELFAND, A. E. (1994). Generalized linear models with unknown link functions. *Biometrika* 81 237–245.

MAULDIN, R. D., SUDDERTH, W. D. & WILLIAMS, S. C. (1992). Polya trees and random distributions. *Annals of Statistics* 20 1203–1221.

MENA, R. H., RUGGIERO, M. & WALKER, S. G. (2011). Geometric stick-breaking processes for continuous-time Bayesian nonparametric modeling. *Journal of Statistical Planning and Inference* 141 3217–3230.

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

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

MULIERE, P. & SECCHI, P. (1995). A note on a proper Bayesian Bootstrap. Tech. rep., Università degli Studi di Pavia, Dipartamento di Economia Politica e Metodi Quantitativ.

MULIERE, P. & TARDELLA, L. (1998). Approximating distributions of random functionals of Ferguson-Dirichlet priors. *The Canadian Journal of Statistics* 26 283–297.

MÜLLER, P., ERKANLI, A. & WEST, M. (1996). Bayesian curve fitting using multivariate normal mixtures. *Biometrika* 83 67–79.

MÜLLER, P. & QUINTANA, F. A. (2004). Nonparametric Bayesian data analysis. *Statistical Science* 19 95–110.

MÜLLER, P. & QUINTANA, F. A. (2010). Random partition models with regression on covariates. *Journal of Statistical Planning and Inference* (To appear).

MÜLLER, P., QUINTANA, F. A. & ROSNER, G. (2004). A method for combining inference across related nonparametric Bayesian models. *Journal of the Royal Statistical Society, Series B* 66 735–749.

MÜLLER, P., QUINTANA, F. A. & ROSNER, G. L. (2011). A product partition model with regression on covariates. *Journal of Computational and Graphical Statistics* 20 260–278.

MÜLLER, P., ROSNER, G. L., DE IORIO, M. & MACEACHERN, S. (2005). A nonparametric Bayesian model for inference in related longitudinal studies. *Journal of the Royal Statistical Society, Series C* 54 611–626.

MÜLLER, P. & VIDAKOVIC, B. (1998). Bayesian inference with wavelets: Density estimation. *Journal of Computational and Graphical Statistics* 7 456–468.

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

NIETO-BARAJAS, L. E., PRÜNSTER, I. & WALKER, S. (2004). Normalized random measures driven by increasing additive processes. *The Annals of Statistics* 32 2343–2360.

NIPOTI, B. (2011). *Dependent completely random measures and statistical applications*. Unpublished doctoral thesis, Department of Mathematics, University of Pavia.

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

NORETS, A. & PELENIS, J. (2011). Posterior consistency in conditional density estimation by covariate dependent mixtures. Tech. rep., Princeton University, Economics Department.

O'HAGAN, A. (1992). Some bayesian numerical analysis. In J. M. Bernardo, J. O. Berger & S. A. F. M, eds., *Proceedings of the Fourth Valencia International Meeting*. Clarendon Press, 345–365.

PADDOCK, S. M. (1999). *Randomized Polya trees: Bayesian nonparametrics for multivariate data analysis*. Unpublished doctoral thesis, Inistitute of Statistics and Decision Sciences, Duke University.

PADDOCK, S. M. (2002). Bayesian nonparametric multiple imputation of partially observed data with ignorable nonresponse. *Biometrika* 89 529–538.

PADDOCK, S. M., RUGGERI, F., LAVINE, M. & WEST, M. (2003). Randomized Polya tree models for nonparametric Bayesian inference. *Statistica Sinica* 13 443–460.

PAPASPILIOPOULOS, O. & ROBERTS, G. O. (2008). Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika* 95 169–186.

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

PATI, D., DUNSON, D. & TOKDAR, S. (2011). Posterior consistency in conditional distribution estimation. Tech. rep., University of Duke, Department of Statistical Science.

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

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

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

PITMAN, J. (1995). Exchangeable and partially exchangeable random partitions. *Probability Theory and Related Fields* 102 145–158.

PITMAN, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In T. S. Ferguson, L. S. Shapeley & J. B. MacQueen, eds., *Statistics, Probability and Game Theory. Papers in Honor of David Blackwell*. IMS Lecture Notes - Monograph Series, Hayward, California, 245–268.

PITMAN, J. & YOR, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability* 25 855–900.

QUINTANA, F. A. (2010). Linear regression with a dependent skewed Dirichlet process. Tech. rep., Pontificia Universidad Católica de Chile, Department of Statistics.

REN, L., DU, L., CARIN, L. & DUNSON, D. B. (2011). Logistic stick-breaking process. *Journal of Machine Learning Research* 12 203–239.

ROBERT, C. P. & ROUSSEAU, J. (2003). A mixture appproach to Bayesian goodness of fit. Tech. rep., CEREMADE, Universitè Paris Dauphine.

RODRIGUEZ, A. & DUNSON, D. B. (2009). Nonparametric Bayesian models through probit stick-breaking processes. Tech. rep., University of California, Santa Cruz, Department of Applied Mathematics and Statistics.

RODRÍGUEZ, A. & DUNSON, D. B. (2011). Nonparametric Bayesian models through probit stick-breaking processes. *Bayesian Analysis* 6 145–177.

RODRIGUEZ, A., DUNSON, D. B. & GELFAND, A. (2008). The nested Dirichlet process. *Journal of the American Statistical Association* 103 1131–1154.

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

SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* 2 639–650.

SIMAS, A. B., BARRETO-SOUZA, W. & ROCHA, A. V. (2010). Improved estimators for a general class of beta regression models. *Computational Statistics & Data Analysis* 54 348–366.

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

SMITH, B. J. (2007). Boa: An r package for mcmc output convergence assessment and posterior inference. *Journal of Statistical Software* 21 1–37.

TEH, Y. W., JORDAN, M. I., BEAL, M. J. & BLEI, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association* 101 1566–1581.

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

TOKDAR, S. T. & GHOSH, J. K. (2007). Posterior consistency of logistic Gaussian process priors in density estimation. *Journal of Statistical Planning and Inference* 137 34–42.

TOKDAR, S. T., ZHU, Y. M. & GHOSH, J. K. (2010). Bayesian density regression with logistic Gaussian process and subspace projection. *Bayesian Analysis* 5 1–26.

TRIPPA, L., MÜLLER, P. & JOHNSON, W. (2011). The multivariate beta process and an extension of the Polya tree model. *Biometrika* 98 17–34.

WALKER, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics: Simulation and Computing* 36 45–54.

WU, Y. & GHOSAL, S. (2008). Kullback Leibler property of kernel mixture priors in Bayesian density estimation. *Electronic Journal of Statistics* 2 298–331.