Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Bayesian Nonparametric Models for Single and Related Probability Measures Supported on Compact Spaces

Claudia Andrea Wehrhahn Cortes

Supervisor:
PROF. ALEJANDRO JARA
Co-supervisor
ANDRÉS FELIPE BARRIENTOS

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

Members of the Examination Commitee:
PROF. JORGE GONZÁLEZ
PROF. WESLEY O. JOHNSON
PROF. FERNANDO QUINTANA

December 15, 2015

Acknowledgements

I want to thank several persons who have been part of this process and without which everything would have been much more difficult. First and foremost, I would like to thank my thesis advisor, Alejandro Jara, for his support, guide, and for introducing me in the Bayesian nonparametric world. Thanks for your patience and for encouraging me to do good science.

I am indebted to my thesis co-supervisor, Felipe Barrientos, for being such an important help and guide, from the very first day of this process. Thanks for all the statistical support and discussions, but above all, thanks for the nice conversations and friendship.

I would like to gratefully acknowledge the financial support provided by Comisión Nacional de Investigación Científica y Tecnológica (CONICYT), through the program Becas para Estudios de Doctorado en Chile. This research was also supported by Fondecyt grant 1141193.

I want to thank the members of the Department of Statistics for always receiving me with a nice smile. To all professors, who in different ways have had an important influence on my development as a statistician. To my fellow students for their willingness to discuss statistical issues, but even more for making each day more enjoyable. Special gratitude goes to Daniela for being such a good friend and always be there to share frustrations and satisfactions proper of this journey, and to Ignacia for her friendship and for encouraging me to take this path and be part of it. I also want to thank the administrative support provided by Soledad and Fabiola during these years.

I am very grateful to my family who have been part of this process from the very first day. To my parents, Luis y Patricia, for always supporting and teaching me the importance of studying. Thanks for letting me make my own way and pursue my dreams. To my brothers Luis Alberto, Cristián, Germán, and my sister Karin, for always being there for me, and to my mother in law, Cecila, for being part of this process and always having words of encouragement.

I want to especially thank Pablo, who has been the most important person in this process. Thanks for your unconditional support and always being there when I needed it most. Thanks for your infinite patience, understanding, love, and for being part of my life.

Claudia Wehrhahn

December, 2015

Contents

Abstract

Bayesian nonparametric (BNP) statistics is a relative new area of statistics. The intersection of Bayesian and non-parametric statistics was almost empty until the sixties and seventies where the first advances were made, primarily on the mathematical formulations. It was only in the early nineties with the advent of sampling based methods, in particular Markov chain Monte Carlo methods, that substantial progress has been made in the area. Posterior distributions ranging over functional spaces are highly complex and hence sampling methods play a key role. A number of themes are in continuous development including theory, methodology and applications.

The main objective of this dissertation is to propose and study, theoretical and applied issues, concerning BNP models for single probability measures and collections of predictor-dependent probability measures defined on compact sets. The emphasis is on the following three aspects: (I) the proposal of novel BNP models for single density estimation for compositional data and the study of the large sample behavior of its posterior distribution under independent and identically distributed sampling, (II) the empirical study of the small sample behavior of two BNP models for single probability measures for data supported on compact intervals, (III) the proposal and study of the basic properties of predictor-dependent BNP models for compositional data.

In (I), we study the large sample behavior of a BNP procedure for the density estimation for compositional data, that is, data supported in a m–dimensional simplex Δ_m. The procedure is based on a Dirichlet process mixture model of a specific class of mixture of Dirichlet densities. We derive a posterior convergence rate by assuming that the underlying data generating density belongs to a Hölder class with regularity α, $0 < \alpha \leq 1$. Specifically, we show that, up to a logarithmic factor, the convergence rate of the posterior distribution is $n^{-\alpha/(2\alpha+m)}$ and, thus the procedure is rate–optimal. These results are summarized in Chapter 2.

In (II), we compare a rate-optimal and rate-suboptimal BNP model for density estimation for data supported on a compact interval by means of the analyses of simulated and real data. The convergence rate of the posterior distribution is usually considered as a standard criteria for model comparison and selection; the faster the rate, the better the model. However, the rate-optimal behavior of a model might well trap the unwary into a false sense of security on the performance of the model, by suggesting that it has the best performance in a wide range of settings and sample sizes. The results illustrate that rate-optimal models are not uniformly better, across sample sizes, with respect to the way in which the posterior mass concentrates around a true model and that suboptimal models can outperform the optimal ones, even for relatively large sample sizes. These results are summarized in Chapter 3.

In (III), we proposed BNP procedures for fully nonparametric regression for compositional data. The procedures are based on a modified class of multivariate Bernstein polynomials, which retains the well known approximation properties of the classical versions defined on $[0, 1]^m$ and Δ_m, $m \geq 1$, and on the use of dependent stick-breaking processes. A general model class and two simplified versions of the general class are discussed in detail. Appealing theoretical properties such as continuity, association structure, support and consistency of the posterior distribution are established for all models. These results are summarized in Chapter 4.

List of Figures

Chapter 1

Introduction

1.1 Literature overview

1.1.1 The general context

Statistical models are used to understand and to learn about the data generating mechanism and they correspond to simplifications of the actual phenomena. The remark of the well known statistician George Box (Box, 1979) “all models are wrong but some models are useful” recognizes the fact that the statistical inference depends not only on the data but also on the background knowledge of the situation which is formalized in the assumptions with which the analysis is entered. The data are envisioned as realizations of a collection of random variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, where $\boldsymbol{x}_i$ itself could be a vector of random variables, corresponding to data that are collected on the i-th experimental unit in a sample of n units. It is also assumed that the collection of random variables follow a joint probability distribution F and the statistical problem begins when there exists uncertainty about F, characterized by a p.d.f. f. A statistical model arises when f is known to be a member f_θ from a family $\{f_\theta : \theta \in \Theta\}$, labeled by a set of parameters θ from an index set Θ.

Statistical models that are described through a vector θ of a finite number of, typically, real values are referred to as finite-dimensional or *parametric models*. Parametric models can be described as $\{f_\theta : \theta \in \Theta \subset \Re^p\}$. Constraining the inference to a specific parametric form, however, may limit the scope and type of inferences that can be drawn. In many practical situations a parametric model cannot describe in an appropriate manner the chance mechanism generating an observed dataset, and unrealistic features of some common models (for example, the thin tails of the normal distribution when compared to the distribution of observed data) could lead to unsatisfactory inferences. In these situations, we would like to relax parametric assumptions to allow greater modeling flexibility and robustness against mis-specification of a parametric statistical model. Specifically, we may want to consider models where the class of densities is so large that it can no longer be indexed by a set of finite-dimensional parameter, and we therefore require parameters θ in an infinite-dimensional space, such as the space of all probability measures defined on the sample space. Statistical models that are described through infinite-dimensional parameters are referred to as *non-parametric models*. Infinite-dimensional parameters of interest are usually functions. Functions of common interest, among many others, include the probability distribution (through some specific characteristic) and the conditional mean (regression function).

The definition and study of theoretical properties of probability models defined over infinite-dimensional spaces have received increasing attention in the statistical literature because these models are the basis for the Bayesian nonparametric (BNP) generalization of parametric statistical models (see, e.g., Ghosh & Ramamoorthi, 2003; Müller & Quintana, 2004; Hjort et al., 2010; Müller et al., 2015). BNP models are specified by defining a stochastic process whose trajectories lie in a functional space, $\mathcal{F}$. The law governing such a process is then used as a prior distribution for a functional parameter in a 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 model 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. Müller et al., 2015). Furthermore, a rich theoretical literature showing large support, posterior consistency and concentration rates of the posterior distribution (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 $\mathcal{F}$ 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 $\mathcal{F}$-valued stochastic process, F, defined on an appropriated probability space $(\Omega, \mathcal{A}, P)$, such that for almost every $\omega \in \Omega$, the density function of F is given by

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

where $\psi(\cdot, \theta)$ is a continuous density function on $(S, \mathcal{S})$, for every $\theta \in \Theta \subseteq \mathbb{R}^q$, and G is a DP, whose sample paths are probability measures defined on $(\Theta, \mathcal{B}(\Theta))$, with $\mathcal{B}(\Theta)$ being the corresponding Borel σ-field. If G is DP with parameters (M, G_0), where $M \in \mathbb{R}_0^+$ and G_0 is a probability measure on $(\Theta, \mathcal{B}(\Theta))$, written as $G \mid M, G_0 \sim DP(MG_0)$, then the trajectories of the process can be a.s. represented by the following stick-breaking representation (Sethuraman, 1994): $G(B) = \sum_{i=1}^{\infty} \omega_i \delta_{\theta_i}(B)$, $B \in \mathcal{B}(\Theta)$, where $\delta_\theta(\cdot)$ is the Dirac measure at θ, $\omega_i = V_i \prod_{j<i}(1 - V_j)$, with $V_i \mid M \stackrel{iid}{\sim} Beta(1, M)$, and $\theta_i \mid G_0 \stackrel{iid}{\sim} G_0$. 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), Müller et al. (2015) 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 $G(B) = \sum_{i=1}^{\infty} \omega_i \delta_{\theta_i}(B) + (1 - \sum_{i=1}^{\infty} \omega_i \delta_{\theta_i}(B))\, G_0(B)$, $B \in \mathbb{B}$, where, the atoms θ_i are iid random variables with common distribution G_0, $\theta_i \overset{iid}{\sim} G_0$, which are assumed independent of the non-negative random weights ω_i. The weights ω_i are constrained such that $\sum_{i=1}^{\infty} \omega_i \leq 1$ a.s. The name of the class is motivated by the interpretation of the parameters; the ith weight ω_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 $\sum_{i=1}^{\infty} \omega_i = 1$ then the SSM is called proper and the corresponding 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), the two parameter Poisson-Dirichlet processes (Pitman, 1995, 1996; Pitman & Yor, 1997; Ishwaran & James, 2001) and the beta two-parameter processes (Ishwaran & Zarepour, 2000).

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), random Bernstein polynomials (Petrone, 1999a,b; Ghosal, 2001; Petrone & Wasserman, 2002), logistic Gaussian processes (see, e.g. Tokdar & Ghosh, 2007), and quantile pyramids (Hjort & Walker, 2009).

1.1.4 Models for related probability measures

Generalizations of the models discussed in Section 1.1.1 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

$$f(y \mid x, G_x(\omega)) = \int_{\Theta} \psi(y, \theta) G_x(\omega)(d\theta), \;\; y \in S, \tag{1.2}$$

where $f(y \mid x, G_x)$, $G_x \equiv G_x(\omega)$, is a conditional density indexed by the value of predictors $x \in \mathcal{X} \subset \mathbb{R}^p$, and the dependence is introduce through the mixing probability measure G_x. Thus, the inferential problem is related to the modeling of the collection of predictor-dependent probability measures $\{G_x : x \in \mathcal{X}\}$. 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 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 random measures and to introduce dependence by modifying the stick-breaking representation of each element in the set. Specifically, MacEachern (1999, 2000) generalized the stick-breaking representation by assuming $G_x(B) = \sum_{i=1}^{\infty} \omega_i(x) \delta_{\theta_i(x)}(B)$, $B \in \mathbb{B}$, 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 $\omega_i(x) = V_i(x) \prod_{j<i}[1 - V_j(x)]$, with $V_i(x)$, $i = 1, \ldots$, 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, $G_x(B) = \sum_{i=1}^{\infty} \omega_i \delta_{\theta_i(x)}(B)$. 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). Barrientos et al. (2012) studied the support properties of different versions of the DDP and extensions to more general dependent stick-breaking processes. 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; Rodriguez & 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. (2014). 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, which yield conditional probability measures with density w.r.t. Lebesgue measure without the need of convolutions.

1.2 Outline of this dissertation

This thesis is devoted to a series of chapters that propose and study properties of BNP models for single probability measures and collections of predictor-dependent probability measures defined on compact sets. As a result of different work carried out at several stages of the project, the chapters are self-contained regarding its notation and abbreviations. The organization of the thesis is as follows.

1.2.1 Chapter 2

We propose a BNP model for density estimation for data supported on a m–dimensional simplex, Δ_m, and study its large sample behavior. Motivated by the uniform approximation properties, and the models proposed by Petrone (1999a,b) for data supported on compact intervals, Barrientos et al. (2015a) proposed a BNP procedure for density estimation for data supported on Δ_m based on multivariate Bernstein polynomials (MBP). The proposed model corresponds to a DPM of Dirichlet densities and has desirable support properties. However, they found that for "true" densities in a Hölder class with α regularity, $\alpha \in (0, 2]$, the convergence rate of the posterior distribution associated with the proposed model is, up to a logarithmic factor, $n^{-\alpha/(2\alpha+2m)}$, which is suboptimal from a minimax point of view.

In this chapter we suggest a modification of the model proposed in Barrientos et al. (2015a). The modification follows the ideas used by Kruijer & Van der Vaart (2008), who proposed a modification of the model developed by Petrone (1999a,b). Our proposal is based on a class of MBP that retains the same uniform approximation properties than the original class. The proposed model is a DPM of mixtures of Dirichlet densities. We derive the posterior convergence rate by assuming that the underlying data generating density belongs to a Hölder class with α regularity, $\alpha \in (0, 1]$. We show that, up to a logarithmic factor, the convergence rate of the posterior distribution is $n^{-\alpha/(2\alpha+m)}$, which is the optimal minimax rate for this class of densities.

1.2.2 Chapter 3

We compare rate-optimal and rate-suboptimal BNP models for single probability measures defined on a compact interval by means of the analyses of simulated and real data. The convergence rate of the posterior distribution is usually considered as a standard criteria for model comparison and selection; the faster the rate, the better the model.

We argue that the use of purely asymptotic criteria for models selection is not appropriate and can lead to the choice of a model not well suited for finite sample sizes. Asymptotic results usually depend on assumptions which are unverifiable in practice (e.g., the smoothness of the underlying "true" density). Furthermore, rate-optimal models are not necessarily better with respect to the way in which the posterior mass concentrates around a given true model for all sample sizes. Some reasons for this are that they are derived asymptotically and not for finite sample sizes, and that the derived results are usually an upper bound for the concentration rate and not the concentration rate itself.

In this chapter we compare the Bernstein–Dirichlet model proposed by Petrone (1999a,b) and the DPM of mixtures of beta densities proposed by Kruijer & Van der Vaart (2008), which are suboptimal and optimal, respectively. In the simulated data comparison, we consider underlying distributions generating data that belong to a Hölder class with α regularity, $\alpha \in (0, 1]$, which is the class of densities where the asymptotic results have been obtained. Furthermore, we consider sample sizes ranging from 100 to 10,000. The results illustrate that rate-optimal models are not uniformly better, across sample sizes, and that suboptimal models can outperform the optimal ones, even for relatively large sample sizes.

1.2.3 Chapter 4

We propose BNP procedures for fully nonparametric regression for compositional data, that is, data supported in Δ_m. The propose procedures extend the class of MBP priors defined by Barrientos et al. (2015a) for single density estimation, and are based on the use of dependent stick-breaking processes. Three classes of models are defined; a general one in which both weights and support points are indexed by predictors; a single weight class, where only the support points depend on predictors; and a single support point class where only the weights of the stick–breaking process depend on predictors. Appealing theoretical properties such as continuity, association structure, support and consistency of the posterior distribution are established. In particular, we study the covariance function of the random measures. The support of the processes was studied by considering the weak product, L_∞ product, L_∞, and Kullback–Leibler topologies. Finally, we showed that the posterior distribution associated with the random joint distribution for predictor and responses, induced by the proposed model, is weakly consistent at any joint distribution with the same marginal distribution generating the predictors.

Chapter 2

Posterior convergence rate of a class of Dirichlet process mixture model for compositional data

This chapter has been submitted for publication as:

BARRIENTOS, A.F., JARA, A., WEHRHAHN, C. (2015). Posterior convergence rate of a class of Dirichlet process mixture model for compositional data.

2.1 Introduction

Models based on convex combinations of densities from parametric families underly mainstream approaches to density estimation (Silverman, 1986; Lindsay, 1995; Lo, 1984). Under a Bayesian nonparametric approach, a mixture model provides a convenient set up in that a prior distribution on densities is induced by placing a prior distribution on the mixing measure (see, e.g. Ferguson, 1983; Lo, 1984; Escobar & West, 1995). Extending the work by Petrone (1999a,b) on the unit interval by using Bernstein polynomials, Barrientos et al. (2015a) suggested to construct a probability model for densities on the d–dimensional simplex $\Delta_d = \left\{(y_1, \ldots, y_d) \in [0,1]^d : \sum_{i=1}^d y_i \leq 1\right\}$, by using mixtures of Dirichlet–densities of the form

$$b_{k,G}(\boldsymbol{y}) = \sum_{\mathbf{j} \in \mathscr{H}^0_{k,d}} G\left(\left(\frac{j_1 - 1}{k}, \frac{j_1}{k}\right] \times \ldots \times \left(\frac{j_d - 1}{k}, \frac{j_d}{k}\right]\right) \mathrm{d}(\boldsymbol{y} \mid \alpha(k, \mathbf{j})), \tag{2.1}$$

where k is discrete random variable supported on $\mathbb{N}$, G is a discrete random probability measure defined on $\Delta_d^0 = \{\boldsymbol{y} \in \Delta_d : y_j \geq 0, \ \ j = 1, \ldots, d\}$, $\boldsymbol{j} = (j_1, \ldots, j_d)$,

$$\mathscr{H}^0_{k,d} = \left\{(j_1, \ldots, j_d) \in \{1, \ldots, k\}^d : \sum_{l=1}^d j_l \leq k + d - 1\right\},$$

$d(\cdot \mid (\alpha_1, \ldots, \alpha_m))$ denotes the density function of a m–dimensional Dirichlet distribution with parameters $(\alpha_1, \ldots, \alpha_m)$, and $\alpha\,(k, \mathbf{j}) = \left(\mathbf{j}, \ k + d - \sum_{l=1}^d j_l\right)$.

By assuming that G is a stick–breaking process, Barrientos et al. (2015a) showed that the induced prior distribution has full weak, L_∞ and Kullback–Leibler support. By using results provided by Ghosal et al. (1999), Barrientos et al. (2015a) showed that the posterior distribution is L_1–consistent, under mild conditions on the distribution of k. Furthermore, they showed that if the "true" density belongs to the proposed class of mixture models, then the speed of convergence of the posterior distribution is at most $\log n/\sqrt{n}$. They also showed that when the "true" density belongs to a Hölder class with α–regularity, $\alpha \in (0, 2]$, then the speed of convergence of the posterior distribution is at most $(\log n)^{(2\alpha+d)/(2\alpha+2d)}/n^{\alpha/(2\alpha+2d)}$, which is suboptimal.

Following a similar approach to the one suggested by Kruijer & Van der Vaart (2008) for data supported on $[0, 1]$ and $[0, 1]^2$, we proposed a modification of the mixture model given by expression (2.1), for which we showed that the posterior convergence rate is optimal up to a logarithmic factor, by assuming a Dirichlet process mixing. Specifically, we showed that if the "true" density belongs to a Hölder class with α–regularity, of at most $\alpha = 1$, then convergence of the posterior distribution is at most $(\log n)^{(4\alpha+d)/(4\alpha+2d)}/n^{\alpha/(2\alpha+d)}$.

2.2 The Dirichlet process mixture model and properties

Suppose we have independent observations $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n$ from a common density f supported on Δ_d. We model the common density f using the following Dirichlet process mixture model,

$$\boldsymbol{y}_i \mid k, \boldsymbol{\theta}_i \overset{ind.}{\sim} f_{k,\boldsymbol{\theta}_i}(\cdot) \equiv \sum_{j_1=\left(\mathscr{T}_k\left(\lceil k\boldsymbol{\theta}_{(i,1)}\rceil\right)-1\right)\sqrt{k}+1}^{\mathscr{T}_k\left(\lceil k\boldsymbol{\theta}_{(i,1)}\rceil\right)\sqrt{k}} \cdots \sum_{j_d=\left(\mathscr{T}_k\left(\lceil k\boldsymbol{\theta}_{(i,d)}\rceil\right)-1\right)\sqrt{k}+1}^{\mathscr{T}_k\left(\lceil k\boldsymbol{\theta}_{(i,d)}\rceil\right)\sqrt{k}} C(\mathbf{j})\mathrm{d}\left(\cdot \mid \alpha(k, \mathbf{j})\right) \tag{2.2}$$

$$\boldsymbol{\theta}_i \mid G \overset{i.i.d.}{\sim} G, \tag{2.3}$$

$$G \mid M, G_0 \sim DP(MG_0), \tag{2.4}$$

$$k \sim \rho, \tag{2.5}$$

where

$$C(\mathbf{j}) = \mathbb{I}_{\mathscr{H}^0_{k,d}}(\mathbf{j}) \left(k^{-d/2}\left(1 - \mathbb{I}_{\mathscr{Q}^k_d}(\mathbf{j})\right) + \frac{(d!(\sqrt{k}-1)!}{(\sqrt{k}+d-1)!}\mathbb{I}_{\mathscr{Q}^k_d}(\mathbf{j}) \right),$$

$\mathbb{I}_A(\cdot)$ is the indicator function for the set A, $\mathscr{Q}^k_d = \left\{ \mathbf{j} \in \mathscr{H}^0_{k,d} : \sum_{i=1}^d \mathscr{T}_k(j_i) = \sqrt{k} + d - 1 \right\}$, $\mathscr{T}_k(j) = \sum_{i=1}^{\sqrt{k}} i\, \mathbb{I}_{A(k,i)}(j)$, $A(k, i) = \left\{ (i-1)\sqrt{k} + 1, \ldots, i\sqrt{k} \right\}$, $\lceil \cdot \rceil$ denotes the ceiling function, $\mathbf{j} = (j_1, \ldots, j_d)$, $G \sim DP(MG_0)$ refers to G being a Dirichlet process with baseline distribution G_0 and total mass parameter M, M is a nonnegative constant, G_0 is a probability measure defined on Δ_d^0, and ρ is a probability measure defined on $\mathbb{K} = \left\{l \in \mathbb{N} : l^{1/2} \text{ is an integer}\right\}$. Note that the hierarchical representation of the model given by expressions (2.2) – (2.3), induces the following model for the unknown density

$$\tilde{b}_{k,G}(\boldsymbol{y}) = \int_{\Delta_d} f_{k,\boldsymbol{\theta}_i}(\boldsymbol{y}) G\left(d\boldsymbol{\theta}_i\right),$$

where k and G are unknown parameters. It is easy to see that $b_{k,G}(\cdot)$ is a polynomial function of $\boldsymbol{y}$. This polynomial class can approximate any element in the set of absolutely continuous probability measures defined on Δ_d and with α-Hölder continuous density function, $\alpha \in (0, 1]$. The following theorem provides the order of approximation of $\tilde{b}_{k,G}(\boldsymbol{y})$ as k increases.

Theorem 2.1. *Let P be an absolutely continuous probability measure defined on Δ_d, w.r.t. Lebesgue measure, with α–Hölder continuous density function, p. Then*

$$\|\tilde{b}_{k,P} - p\|_\infty = \mathcal{O}\left(k^{-\alpha/2}\right).$$

The previous result and the weak support properties of the DP allow us to show that the posterior distribution of the model induced by expressions (2.2) – (2.5) is rate–optimal.

Theorem 2.2. *Let $\mathcal{P}\left(\Delta_d\right)$ be the set of all absolutely continuous probability measures defined on Δ_d, w.r.t. Lebesgue measure, and with continuous density. Let $F \in \mathcal{P}\left(\Delta_d\right)$ be the true data generating probability measure define on Δ_d, with strictly positive density f. Assume that f is α–Hölder continuous for $\alpha \in (0, 1]$. Let G_0 be a probability measure defined on Δ_d with strictly positive density w.r.t. Lebesgue measure. Assume that for every $\bar{k} \in \mathbb{K}$, $B_1 \exp(-\beta_1 \bar{k}^{d/2}) \leq \rho(\bar{k}) \leq B_2 \exp(-\beta_2 \bar{k}^{d/2})$, for some positive constants B_1, B_2, β_1 and β_2. Then for a sufficiently large constant C,*

$$\Pi\left\{P \in \mathcal{P}\left(\Delta_d\right) : H(p, f) \geq C \frac{(\log n)^{(4\alpha+d)/(4\alpha+2d)}}{n^{\alpha/(2\alpha+d)}} \middle| \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\right\} \underset{n\to\infty}{\longrightarrow} 0,$$

in F^n-probability, where p is the density of P, $H(p,f)$ is the Hellinger distance between P and F and $\Pi\{\cdot \mid \boldsymbol{y}_1,\ldots,\boldsymbol{y}_n\}$ is the posterior distribution induced by $\tilde{b}_{k,G}$ and expressions (2.4) and (2.5).

2.3 Proofs

2.3.1 Proof of Theorem 1

By the triangle inequality it follows that, for every probability measure P defined on Δ_d and with α–Hölder continuous density p, $\|\tilde{b}_{k,P}-p\|_\infty \le \|b_{k,P}-p\|_\infty + \|b_{k,P}-\tilde{b}_{k,P}\|_\infty$, where $b_{k,P}$ is the class of polynomials functions proposed by Barrientos et al. (2015a). They showed that $\|b_{k,P}-p\|_\infty = \mathcal{O}(k^{-\alpha/2})$. Thus, to prove the theorem it is sufficient to show that $\|b_{k,P}-\tilde{b}_{k,P}\|_\infty = \mathcal{O}(k^{-\alpha/2})$.

Let $\mathbf{z}_j = (z_{j,1},\ldots,z_{j,d+1}) \overset{i.i.d.}{\sim} \text{Multinomial}(1,\boldsymbol{\pi})$, where $\boldsymbol{\pi} = \left(\boldsymbol{y}, 1-\sum_{l=1}^{d} y_l\right)$. Set $\mathbf{S}_k = \sum_{j=1}^{k+d-1}\mathbf{z}_j$ and $\mathbf{W}_k = (S_{k,1},\ldots,S_{k,d})$. By the multivariate version of the mean value theorem proposed by Ash (2008), it is possible to ensure the existence of $(0,1)^d$–valued random vectors $\mathbf{c}_{\mathbf{W}_k}$ and $\tilde{\mathbf{c}}_{\mathbf{W}_k}$ such that,

$$P\left(\left(\frac{\mathscr{T}_k(S_{k,1})}{\sqrt{k}},\frac{\mathscr{T}_k(S_{k,1})+1}{\sqrt{k}}\right]\times\cdots\times\left(\frac{\mathscr{T}_k(S_{k,d})}{\sqrt{k}},\frac{\mathscr{T}_k(S_{k,d})+1}{\sqrt{k}}\right]\right) = \frac{1}{k^{d/2}}p\left(\frac{\mathscr{T}_k(\mathbf{W}_k)+\tilde{\mathbf{c}}_{\mathbf{W}_k}}{\sqrt{k}}\right),$$

and

$$P\left(\left(\frac{S_{k,1}}{k},\frac{S_{k,1}+1}{k}\right]\times\ldots\times\left(\frac{S_{k,d}}{k},\frac{S_{k,d}+1}{k}\right]\right) = k^{-d}p\left(\frac{\mathbf{W}_k+\mathbf{c}_{\mathbf{W}_k}}{k}\right).$$

Now, notice that $\tilde{b}_{k,P}(\boldsymbol{y})$ can be expressed as a mixture of Dirichlet densities of the form

$$\tilde{b}_{k,P}(\boldsymbol{y})=\sum_{\mathbf{j}\in\mathscr{H}^0_{k,d}} P\left(\left(\frac{\mathscr{T}_k(j_1)-1}{\sqrt{k}},\frac{\mathscr{T}_k(j_1)}{\sqrt{k}}\right]\times\ldots\times\left(\frac{\mathscr{T}_k(j_d)-1}{\sqrt{k}},\frac{\mathscr{T}_k(j_d)}{\sqrt{k}}\right]\right) C(\mathbf{j})\mathrm{d}(\boldsymbol{y}\mid\alpha(k,\mathbf{j})).$$

Given the relationship between the kernel of a Dirichlet density and the kernel of a multinomial distribution, $\tilde{b}_{k,P}$ can also be expressed as

$$\begin{aligned}\tilde{b}_{k,P}(\boldsymbol{y}) &= \frac{(k+d-1)!}{(k-1)!} E_{\mathbf{S}_k}\left[C(\mathbf{W}_k) P\left(\left(\frac{\mathscr{T}_k(S_{k,1})}{\sqrt{k}}, \frac{\mathscr{T}_k(S_{k,1})+1}{\sqrt{k}}\right] \times \cdots \times \right.\right. \\ &\qquad \left.\left.\left(\frac{\mathscr{T}_k(S_{k,d})}{\sqrt{k}}, \frac{\mathscr{T}_k(S_{k,d})+1}{\sqrt{k}}\right]\right)\right], \qquad (2.6) \\ &= \frac{(k+d-1)!}{(k-1)!k^{d/2}} E_{\mathbf{S}_k}\left[C(\mathbf{W}_k) p\left(\frac{\mathscr{T}_k(\mathbf{W}_k)+\tilde{\mathbf{c}}_{\mathbf{W}_k}}{\sqrt{k}}\right)\right]. \qquad (2.7)\end{aligned}$$

From expression (2.6) and the fact that $b_{k,P}(\boldsymbol{y}) = \frac{(k+d-1)!}{(k-1)!k^d} E_{\mathbf{S}_k}\left[p\left(\frac{\mathbf{W}_k+\mathbf{c}_{\mathbf{W}_k}}{k}\right)\right]$ (see, Barrientos et al., 2015a, section 6.1), it follows that

$$\begin{aligned}&\|\tilde{b}_{k,P} - b_{k,P}\|_\infty \\ &\le \frac{(k+d-1)!}{(k-1)!k^{d/2}} \sup_{\boldsymbol{y}\in\Delta_d} \left| E_{\mathbf{S}_k}\left[C(\mathbf{W}_k) p\left(\frac{\mathscr{T}_k(\mathbf{W}_k)+\tilde{\mathbf{c}}_{\mathbf{W}_k}}{\sqrt{k}}\right) - k^{-d/2} p\left(\frac{\mathbf{W}_k+\mathbf{c}_{\mathbf{W}_k}}{k}\right)\right]\right|, \\ &\le \frac{(k+d-1)!}{(k-1)!k^{d}} \sup_{\boldsymbol{y}\in\Delta_d} \left| E_{\mathbf{S}_k}\left[\left(\frac{\mathscr{T}_k(\mathbf{W}_k)+\tilde{\mathbf{c}}_{\mathbf{W}_k}}{\sqrt{k}}\right) - p\left(\frac{\mathbf{W}_k+\mathbf{c}_{\mathbf{W}_k}}{k}\right)\right]\right| \\ &\quad + \frac{(k+d-1)!}{(k-1)!k^{d/2}} \sup_{\boldsymbol{y}\in\Delta_d} \left| E_{\mathbf{S}_k}\left[\left|C(\mathbf{W}_k) - k^{-d/2}\right| p\left(\frac{\mathscr{T}_k(\mathbf{W}_k)+\tilde{\mathbf{c}}_{\mathbf{W}_k}}{\sqrt{k}}\right)\right]\right|, \\ &\le \frac{(k+d-1)!}{(k-1)!k^{d}} L_\alpha(p) \sup_{\boldsymbol{y}\in\Delta_d} E_{\mathbf{S}_k}\left[\left\|\frac{\mathscr{T}_k(\mathbf{W}_k)+\tilde{\mathbf{c}}_{\mathbf{W}_k}}{\sqrt{k}} - \frac{\mathbf{W}_k+\mathbf{c}_{\mathbf{W}_k}}{k}\right\|^\alpha\right] \\ &\quad + \frac{(k+d-1)!}{(k-1)!k^{d/2}} \|p\|_\infty \left| \frac{d!(\sqrt{k}-1)!}{(\sqrt{k}+d-1)!} - k^{-d/2}\right|, \\ &\le \frac{(k+d-1)!}{(k-1)!k^{d}} L_\alpha(p) \sup_{\boldsymbol{y}\in\Delta_d} E_{\mathbf{S}_k}\left[\sum_{j=1}^{d}\left(\frac{\mathscr{T}_k(S_{k,j})+\tilde{c}_{S_{k,j}}}{\sqrt{k}} - \frac{S_{k,j}+c_{S_{k,j}}}{k}\right)^2\right]^{\alpha/2} \\ &\quad + \frac{(k+d-1)!}{(k-1)!k^{d}} \|p\|_\infty \left| \frac{d!k^{-d/2}}{\prod_{j=1}^{d}(\sqrt{k}+j-1)!} - 1\right|,\end{aligned}$$

where $L_\alpha(p)$ is the Hölder constant for the function p. The proof of the theorem is completed by noticing that

$$\begin{aligned}\left(\frac{\mathscr{T}_k(S_{k,j})+\tilde{c}_{S_{k,j}}}{\sqrt{k}} - \frac{S_{k,j}+c_{S_{k,j}}}{k}\right)^2 &\le \left(\frac{\sqrt{k}-1}{k}\right)^2 + 2\left(\frac{\sqrt{k}-1}{k}\right)\left(\frac{1}{\sqrt{k}}+\frac{1}{k}\right) + \left(\frac{1}{\sqrt{k}}+\frac{1}{k}\right)^2, \\ &\le \frac{5}{k},\end{aligned}$$

$$\left| \frac{d! k^{-d/2}}{\prod_{j=1}^{d} (\sqrt{k} + j - 1)!} - 1 \right| = \mathcal{O}(k^{-1/2}),$$

and that for k large enough, $(k+d-1)!/((k-1)!k^d) \leq 2$.

2.3.2 Proof of Theorem 2

To prove this theorem it is sufficient to show that all the conditions of Theorem 2.1 in Ghosal (2001) are satisfied. In our context, the conditions of Theorem 2.1 are satisfied if there is a sequence $(\mathcal{F}_n)_{n\geq 1}$ of subsets of the parameter space and there are positive sequences $(\tilde{\epsilon}_n)_{n\geq 1}$ and $(\bar{\epsilon}_n)_{n\geq 1}$ satifying $\tilde{\epsilon}_n, \bar{\epsilon}_n \longrightarrow 0$ and $n \min\{\tilde{\epsilon}_n, \bar{\epsilon}_n\} \longrightarrow \infty$, such that for certain positive constants a_1, a_2, a_3, a_4, the following conditions hold:

$$\Pi \left\{ P \in \mathcal{P}\left(\Delta_d\right) : P \in N(\tilde{\epsilon}_n, f) \right\} \quad \geq \quad a_1 \exp\left(-a_2 n \tilde{\epsilon}_n^2\right), \tag{2.8}$$

$$\Pi \left\{ P \in \mathcal{P}\left(\Delta_d\right) : P \in \mathcal{F}_n^C \right\} \quad \leq \quad a_3 \exp\left(-(a_2+4) n \tilde{\epsilon}_n^2\right), \tag{2.9}$$

$$\log\left(D\left(\bar{\epsilon}_n, \mathcal{F}_n, H\right)\right) \quad \leq \quad a_4 n \bar{\epsilon}_n^2, \tag{2.10}$$

where Π is the prior distribution induced by $\tilde{b}_{k,G}$ and expressions (2.4) – (2.5),

$$N(\epsilon, f) \quad = \quad \left\{ Q \in \mathcal{P}\left(\Delta_d\right) : \int_{\Delta_d} f(\boldsymbol{y}) \log\left(\frac{f(\boldsymbol{y})}{q(\boldsymbol{y})}\right) d\boldsymbol{y} \leq \epsilon, \ \int_{\Delta_d} f(\boldsymbol{y}) \left(\log \frac{f(\boldsymbol{y})}{q(\boldsymbol{y})}\right)^2 d\boldsymbol{y} \leq \epsilon \right\},$$

q is the density function of the probability measure Q, $\mathcal{F}_n^C$ is the complement of the set $\mathcal{F}_n$, and $D\left(\epsilon, \mathcal{F}, H\right)$ is the ϵ–packeting number defined to be the maximum number of points in $\mathcal{F}$ such that the H–distance between each pair is at least ϵ.

Let $\tilde{\epsilon}_n = (\log n/n)^{\alpha/(2\alpha+d)}$, $k_n = \left\lceil \max\{M, M^{-1}\}/\tilde{\epsilon}_n^{1/\alpha} \right\rceil - 1$, $\tilde{\boldsymbol{\pi}}^P_{k_n^2} = \left(\tilde{\pi}^P_{k_n^2,\mathbf{j}}\right)_{\mathbf{j}\in\mathscr{H}^0_{k_n^2,d}}$, and

$$\tilde{b}_{k_n^2,P}(\cdot) = \sum_{\mathbf{j}\in\mathscr{H}^0_{k_n^2,d}} C(\mathbf{j})\tilde{\pi}^P_{k_n^2,\mathbf{j}} \mathrm{d}(\cdot \mid \alpha(k_n^2,\mathbf{j})),$$

where

$$\pi^P_{k_n^2,\mathbf{j}} = P\left(\left(\frac{\mathscr{T}_{k_n^2}(j_1)-1}{k_n}, \frac{\mathscr{T}_{k_n^2}(j_1)}{k_n}\right] \times \ldots \times \left(\frac{\mathscr{T}_{k_n^2}(j_d)-1}{k_n}, \frac{\mathscr{T}_{k_n^2}(j_d)}{k_n}\right]\right).$$

By Theorem 1, there exists a positive constant, C_1, such that $\|f - \tilde{b}_{k_n^2,F}\|_\infty \le C_1/k_n^\alpha$. Therefore, if P is such that $\left\|\boldsymbol{\pi}^P_{k_n^2} - \boldsymbol{\pi}^F_{k_n^2}\right\|_1 \le \tilde{\epsilon}_n^{\,1+2d/\alpha}$, then, for some $C_2 > 0$,

$$\|f - \tilde{b}_{k_n^2,P}\|_\infty \;\le\; \|f - \tilde{b}_{k_n^2,F}\|_\infty + \|\tilde{b}_{k_n^2,F} - \tilde{b}_{k_n^2,P}\|_\infty \;\le\; \frac{C_1}{k_n^\alpha} + \; k_n^{2d}\,\tilde{\epsilon}_n^{\,1+2d/\alpha} \;\le\; C_2\tilde{\epsilon}_n.$$

This implies that $H^2(f, \tilde{b}_{k_n^2,P}) \le 4\,\|f^{-1}\|_\infty\, C_2^2\tilde{\epsilon}_n^2$. Thus, this results along with expression (8.6) in Ghosal et al. (2000a) and the assumption that P also satisfies that $\pi^P_{k_n^2,\mathbf{j}} > \tilde{\epsilon}_n^2/2$, for every $\mathbf{j} \in \mathscr{H}^0_{k_n^2,d}$, imply that for some $C_3 > 0$, $\tilde{b}_{k_n^2,P} \in N(C_3\tilde{\epsilon}_n, f)$. Therefore, by Lemma A.1 (Ghosal, 2001),

$$\begin{aligned}
\Pi\left\{P \in \mathcal{P}(\Delta_d) : P \in N(C_3\tilde{\epsilon}_n, f)\right\} &\ge B_1 \exp(-\beta_1 k_n^d) C_4 \exp\left(-c_1 k_n^d \log\left(2/\tilde{\epsilon}_n^{\,1+2d/\alpha}\right)\right), \\
&\ge B_1 C_4 \exp\left(-c_2 k_n^d \log n\right), \\
&\ge B_1 C_4 \exp\left(-c_3 n\tilde{\epsilon}_n^2\right),
\end{aligned}$$

where C_4, c_1, c_2 and c_3, are suitable positive constants. Thus, the condition (2.8) is satisfied. Now, let $\{s_n\}_{n\ge1}$ be a sequence such that

$$L_1\, n^{1/(d+2\alpha)}(\log n)^{2\alpha/(d(d+2\alpha))} \le\; s_n \;\le L_2\, n^{1/(d+2\alpha)}(\log n)^{2\alpha/(d(d+2\alpha))},$$

where L_1 and L_2 are positive constants, and $\mathcal{F}_n := \bigcup_{r=1}^{s_n} \mathcal{B}_r$, where, $\mathcal{B}_r = \left\{\tilde{b}_{\bar{k},P} : \bar{k} = r^2\right\}$. The condition (2.9) is satisfied since, for some $c_4 > 0$,

$$\begin{aligned}
\Pi\left\{P \in \mathcal{P}\left(\Delta_d\right) : P \in \mathcal{F}_n^C\right\} &\leq \sum_{r=s_n+1}^{\infty} \rho(r^2), \\
&\leq B_2 \exp\left(-\beta_2 L_1\, n^{d/(2\alpha+d)} (\log n)^{2\alpha/(2\alpha+d)}\right), \\
&\leq B_3 \exp\left(-(c_4+4) n \tilde{\epsilon}_n^2\right).
\end{aligned}$$

Finally, set $\bar{\epsilon}_n = (\log n)^{(4\alpha+d)/(4\alpha+2d)}/n^{\alpha/(2\alpha+d)}$. Since $\left\|\tilde{\boldsymbol{\pi}}^P_{r^2} - \tilde{\boldsymbol{\pi}}^{P_0}_{r^2}\right\|_1 \leq \epsilon^2/(r^2+d-1)^d$ implies that $H(\tilde{b}_{r^2,P}, \tilde{b}_{r^2,P_0}) \leq \epsilon$, by Lemma A.4 in Ghosal & van der Vaart (2001), it follows that

$$\begin{aligned}
D\left(\bar{\epsilon}_n, \mathcal{F}_n, d_H\right) &\leq \sum_{r=1}^{s_n} D\left(\bar{\epsilon}_n^2, \mathcal{B}_r, \|\cdot\|_1\right) \leq \sum_{r=1}^{s_n} D\left(\frac{\bar{\epsilon}_n^2}{(r^2+d-1)^d}, \Delta_{\frac{(r+d-1)!}{d!(r-1)!}}, \|\cdot\|_1\right) \\
&\leq \sum_{r=1}^{s_n} \left(\frac{5(r^2+d-1)^d}{\bar{\epsilon}_n^2}\right)^{\frac{(r+d-1)!}{(d!(r-1)!}-1} \leq s_n \left(\frac{5(s_n^2+d-1)^d}{\bar{\epsilon}_n^2}\right)^{(s_n+d-1)^d}.
\end{aligned}$$

This implies that

$$\begin{aligned}
\log\left(D\left(\bar{\epsilon}_n, \mathcal{F}_n, d_H\right)\right) &\leq \log s_n + (s_n+d-1)^d \log\left(\frac{5(s_n^2+d-1)^d}{\bar{\epsilon}_n^2}\right), \\
&\leq s_n \log n + C_5\, s_n^d \log\left(\frac{5 c_5 s_n^{2d}}{\bar{\epsilon}_n^2}\right), \\
&\leq C_5 s_n^d \log n, \\
&\leq C_5 n \bar{\epsilon}_n^2,
\end{aligned}$$

where C_5 and c_5 are suitable positive constants. Therefore, condition (2.10) is also satisfied, which implies that the proof of the theorem is completed.

Chapter 3

On the small sample behavior of Dirichlet process mixture models for data supported on compact intervals

This chapter has been submitted for publication as:

WEHRHAHN, C., JARA, A., BARRIENTOS, A.F. (2015). On the small sample behavior of Dirichlet process mixture models for data supported on compact intervals.

3.1 Introduction

The definition and study of theoretical properties of probability models defined over infinite-dimensional spaces have received increasing attention in the statistical literature because they are the basis for the Bayesian nonparametric (BNP) generalization of parametric statistical models and offer a highly flexible model-based treatment for standard nonparametric procedures (see, e.g., Müller et al., 2015). For instance, in a single density estimation problem the parameter space corresponds to the set of all probability measures defined on a given measurable space $(S, \mathcal{S})$ and that admit a density function w.r.t. Lebesgue measure, $\mathcal{P}(S)$. In this setting, a BNP model corresponds to a prior distribution defined on $\mathcal{P}(S)$, which is expected to satisfy minimal desirable properties: the topological support should be large and the induced posterior distribution given a sample of observations should be tractable and consistent.

We say that a posterior distribution is d-consistent at $P_0 \in \mathcal{P}(S)$, usually belonging to a given class, if for every $\epsilon > 0$,

$$\lim_{n \longrightarrow \infty} E_{Y_1,\ldots,Y_n|P_0} \left\{ \Pi \left[P \in \mathcal{P}(S) : d(P, P_0) < \epsilon \mid Y_1, \ldots, Y_n \right] \right\} = 1$$

where d is a metric on $\mathcal{P}(A)$ and $\Pi\{\cdot \mid Y_1, \ldots, Y_n\}$ is the posterior distribution given the random sample $Y_1, \ldots, Y_n$, drawn from P_0. For a d-consistent BNP model, the study of the way in which the posterior distribution concentrates around the true model when the sample size increases allows for a better understanding of the behavior of the BNP model. In this regard, the literature has mainly focused on the so called rates of convergence of the posterior distribution, ϵ_n. We say that a posterior distribution converges to P_0 with rate ϵ_n if, for some sufficiently large constant M,

$$\lim_{n \longrightarrow \infty} E_{Y_1,\ldots,Y_n|P_0} \left\{ \Pi \left[P \in \mathcal{P}(A) : d(P, P_0) < M\epsilon_n \mid Y_1, \ldots, Y_n \right] \right\} = 1,$$

where ϵ_n is a sequence that converges to zero when n goes to infinity. Posterior convergence rates are typically used to compare BNP models and to contrast them with frequentist alternatives, and the usual optimality criterion that is considered is the minimaxity over a given class.

Posterior concentration rates have been derived under extremely weak conditions for many BNP models and in a wide variety of contexts (see, e.g., Ghosal et al., 2000b; Ghosal, 2001; Shen & Wasserman, 2001; Ghosal & Van der Vaart, 2007; Rivoirard & Rousseau, 2012).

We argue that the use of purely asymptotic criteria for models selection is not appropriate and can lead to the choice of a model not well suited for finite sample sizes. More specifically, we argue that the suboptimal behavior of the posterior concentration rate of a BNP model should not be the "cause of death" of the model. The overriding problem is that rate-optimal models are not necessarily better with respect to the way in which the posterior mass concentrates around a given true model for all sample sizes. Some reasons for this are: i) usually the derived concentration rate is rather an upper bound for it because it is not proved that ϵ_n is sharp in the case of a sub-optimal model (which can be explained by the mathematical tools employed in the proofs or the researcher's ability to find it), ii) concentration rates are usually derived asymptotically (for a sufficiently large n) and unknown for finite sample sizes, iii) the constant M above can be quite different for different models and can play an important role in the comparison in which different models assign posterior mass around the truth, and iv) optimal rates are usually derived up to a logarithmic factor, which can also play an important role in the comparison of models for small samples.

In this work we illustrate the previous point by comparing the small sample properties of a rate-optimal and a rate-suboptimal BNP model for single density estimation for data supported on a compact interval. Both models correspond to Dirichlet process mixtures (DPM) of probability density functions defined on the unitary interval. Specifically, we consider the Bernstein-Dirichlet prior proposed by Petrone (1999a,b) and the DPM of mixtures of beta densities proposed by Kruijer & Van der Vaart (2008). Under the Hellinger metric, the posterior distribution of the Bernstein-Dirichlet prior and the DPM of mixtures of beta models is minimax suboptimal and optimal, respectively. We compare the models by means of simulated and real data. In the simulation study, different true models were chosen from the class were the posterior concentration rates of the models hold. For the analyses of the real data, the models were compared from a goodness-of-fit point of view.

The structure of the chapter is the following. In Section 3.2, the two models are introduced and their basic properties are discussed. The main asymptotic properties of the models are described in Section 3.3. The comparison of the models using simulated data is described in Section 3.4. The comparison of the performance of the models using real data is given in Section 3.5.

3.2 The models

3.2.1 The Dirichlet-Bernstein model

The Dirichlet-Bernstein model proposed by Petrone (1999a,b) corresponds to a particular example of random Bernstein polynomials (BP). BP were introduced by Bernstein (1912) to give a proof of Weierstrass' approximation theorem. If $G : [0,1] \longrightarrow \mathbb{R}$, the associated BP of degree k is given by

$$\mathrm{BP}_1(y \mid k, G) = \sum_{j=0}^{k} G(j/k) \left(\begin{array}{c} k \\ j \end{array} \right) y^j (1-y)^{k-j},$$

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

$$f_1(y \mid k, G) = \sum_{j=1}^{k} w_{1,j,k} \beta(y \mid j, k-j+1), \tag{3.1}$$

where $w_{1,j,k} = G(j/k) - G((j-1)/k)$, and $\beta(\cdot|a,b)$ stands for a beta density with parameters a and b. The random BP prior arises by considering a random density given by expression (3.1), where k has probability mass function ρ, and given k, $\boldsymbol{w}_{1,k} = (w_{1,1,k}, \ldots, w_{1,k,k})$ 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 (3.1) as the Bernstein polynomial density with parameters k and $\boldsymbol{w}_{1,k}$, and showed that if ρ assigns positive mass to all naturals $\mathbb{N}$, and the density of H_k is positive for any point in Δ_k, then the weak support of the BP is the space of all probability measures on $([0,1], \mathscr{B}([0,1]))$. Letting $\zeta_{j,k} = M\left(G_0(j/k) - G_0((j-1)/k)\right)$, $j = 1, \ldots, k$, G_0 being a probability distribution on $(0,1]$ and M being a positive constant, Petrone (1999a,b) used the fact that assuming

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

is equivalent to assuming 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 proposed a Markov chain Monte Carlo (MCMC) algorithm to scan its posterior distribution.

The BDP model can be equivalently written as the following DPM of beta densities

$$f_1(y \mid k, G) = \int_{[0,1]} \beta(y \mid \lceil zk \rceil, k - \lceil zk \rceil + 1) G(dz), \tag{3.2}$$

$$k \mid \rho_1 \sim \rho_1, \tag{3.3}$$

and

$$G \mid M, G_0 \sim DP(MG_0), \tag{3.4}$$

where $\lceil \cdot \rceil$ denotes the ceiling function and ρ_1 is a probability measure on $\mathbb{N}$.

3.2.2 The DPM of mixtures of beta densities

The DPM of mixtures of beta models proposed by Kruijer & Van der Vaart (2008), referred to as DPMMB, can be justified in a similar way to BDP construction. In this case, the density function can be seen as the derivative of a particular modified class of BP, referred to as BP$_2$. Such modification results in a mixture of mixtures of beta densities, were the dimension the random weights vector is smaller. If $G : [0,1] \longrightarrow \mathbb{R}$, then the associated modified BP of degree $k \in \mathbb{K} = \{l \in \mathbb{N} : \sqrt{l}$ is an integer$\}$ is given by

$$\begin{aligned} \mathrm{BP}_2(y \mid k, G) &= \sum_{i=0}^{\sqrt{k}-1} \sum_{j=i\sqrt{k}+1}^{(i+1)\sqrt{k}} \left[a_{i,j,k}\, G\left(T_1(j,k)\right) + (1-a_{i,j,k})\, G\left(T_2(j,k)\right)\right] \binom{k}{j} y^j (1-y)^{k-j}, \\ &\quad + G(0)(1-y)^k, \end{aligned} \tag{3.5}$$

$a_{i,j,k} = \left(\frac{j - i\sqrt{k}}{\sqrt{k}}\right)$, $T_1(j,k) = \frac{\lceil j/\sqrt{k}\rceil}{\sqrt{k}}$ and $T_2(j,k) = \frac{\lceil j/\sqrt{k}\rceil - 1}{\sqrt{k}}$. If G is the restriction of the CDF of a probability measure defined on the unit interval, then BP_2 is also the restriction of a CDF on $[0,1]$ and represents the following mixture model

$$\sum_{j=1}^{\sqrt{k}} w_{2,j,k} \left(\sum_{j_1=(j-1)\sqrt{k}+1}^{j\sqrt{k}} \frac{1}{\sqrt{k}} \mathrm{Be}(y \mid j_1, k - j_1 + 1) \right) + G(0),$$

where $\mathrm{Be}(\cdot \mid a, b)$ stands for the CDF of a beta distribution with parameters a and b, $w_{2,j,k} = G(j/\sqrt{k}) - G((j-1)/\sqrt{k})$. In this case, if $G(0) = 0$, its density function is given by the following mixture of mixtures of beta densities

$$f_2(y \mid k, G) = \sum_{j=1}^{\sqrt{k}} w_{2,j,k} \left(\sum_{j_1=(j-1)\sqrt{k}+1}^{j\sqrt{k}} \frac{1}{\sqrt{k}}\, \beta(y \,|\, j_1, k - j_1 + 1) \right). \tag{3.6}$$

Figure 3.1 illustrates the differences between expression (3.1) and (3.6). Specifically, Figure 3.1 displays the bases functions under both models for $k = 9$. Under BDP model the bases functions are beta densities, while under DPMMB model the bases functions are mixtures of beta densities with equal weights. Figure 3.1 also illustrates the partition of the unit interval under the different models, implying a different role of G in each case. For a given degree k, the density under DPMMB model is less flexible than the one arising under BDP model. As a matter of fact, expression (3.6) can be seen as a mixture of the same beta densities of expression (3.1), where some components share the same weights and which are functions of the

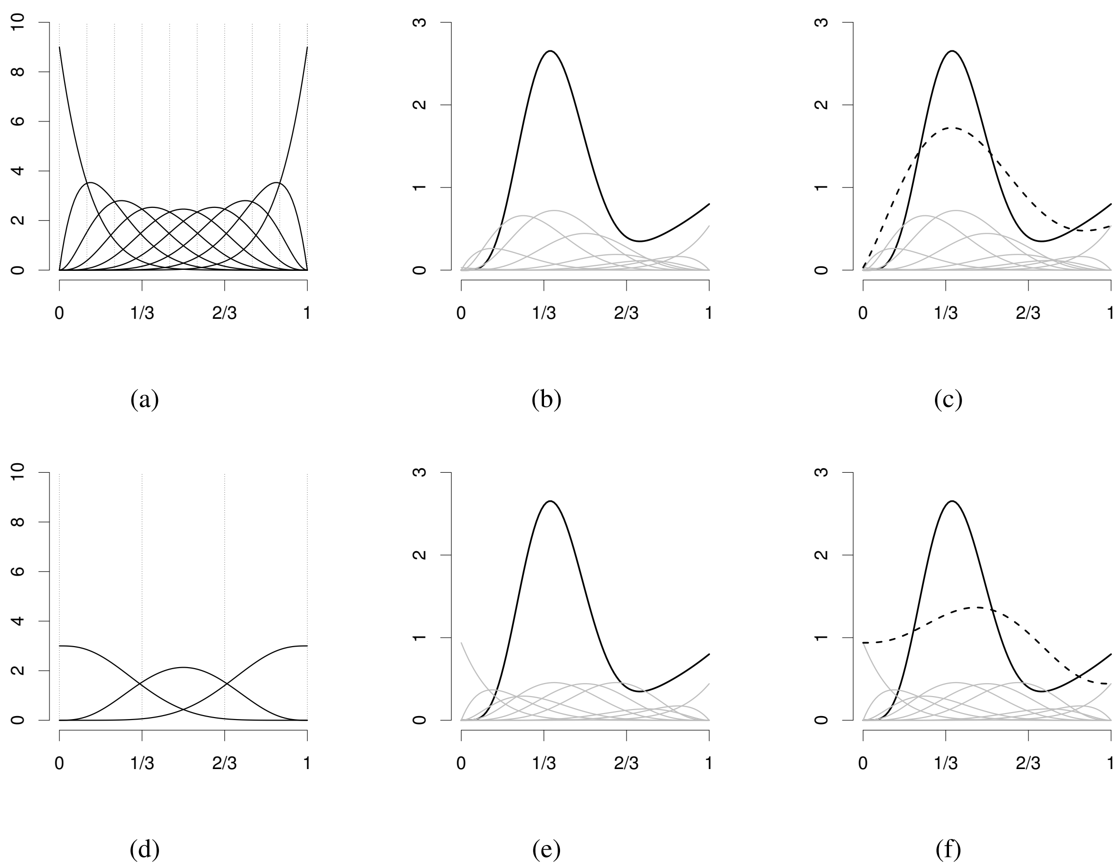


Figure 3.1: The models: Panel (a) and (d) display the bases functions of the mixture model with $k = 9$ arising under BDP and DPMMB model, respectively. In the former case, the bases functions are beta densities. In the latter, the bases functions are mixtures of beta densities. For a given function represented as a solid line, panel (b) and (e) display the weighted beta densities (gray lines) used by the BDP and DPMMB model, respectively. Finally, for a given function represented as a solid line, panel (c) and (f) display the approximation (dashed line) arising under BDP and DPMMB model, respectively, along with the weighted beta densities.

weights considered by BDP model,

$$\begin{aligned} f_2(y \mid k, G) &= \sum_{j=1}^{k} \left(\frac{w_{2,\lceil j/\sqrt{k} \rceil,k}}{\sqrt{k}} \right) \beta(y \,|\, j, k-j+1), \\ &= \sum_{j=1}^{k} \left(\frac{\sum_{j_1=(\lceil j/\sqrt{k} \rceil -1)\sqrt{k}+1}^{\lceil j/\sqrt{k} \rceil \sqrt{k}} w_{1,j_1,k}}{\sqrt{k}} \right) \beta(y \,|\, j, k-j+1). \end{aligned}$$

DPMMB model can be equivalently written as the following DPM of mixtures of beta densities

$$f_2(y \mid k, G) = \int_{[0,1]} \sum_{j_1=(\lceil z\sqrt{k} \rceil -1)\sqrt{k}+1}^{\lceil z\sqrt{k} \rceil \sqrt{k}} \frac{1}{\sqrt{k}} \beta(y \mid j_1, k-j_1+1) G(dz), \tag{3.7}$$

$$k \mid \rho_2 \sim \rho_2, \tag{3.8}$$

and

$$G \mid M, G_0 \sim DP(MG_0), \tag{3.9}$$

where ρ_2 is a probability mass function on $\mathbb{K}$.

3.3 The asymptotic properties of the models

Petrone & Wasserman (2002) showed that under the same assumptions for BDP model to have full weak support, the posterior distribution of the model under iid sampling is weakly consistent if the underlying distribution function generating the data has continuous and bounded Lebesgue density. They also showed that if the model is specified such that the prior distribution on the degree of the polynomial, ρ_1, satisfies an additional tail condition, the posterior distribution of the model is Hellinger consistent. Specifically, if there exists a sequence $k_n \longrightarrow \infty$ such that $k_n = o(n)$ and such that $\sum_{k \geq k_n} \rho_1(k) \leq e^{-cn}$, then the posterior mean of the density under the

BDP model converges in the Hellinger sense to the true density, where c is a positive constant.

Ghosal (2001) studied the concentration rate for BDP model and showed that if the true model admits a density w.r.t. Lebesgue measure that is bounded away from zero and has continuous and bounded second derivatives, then the posterior distribution under iid sampling converges in a Hellinger sense to the true density at rate $\epsilon_n = \log(n)^{5/6}/n^{1/3}$, provided that ρ_1 is specified such that there exist positive constants B_1, B_2, c_1, and c_2, such that $B_1 e^{-c_1 k} \leq \rho_1(k) \leq B_2 e^{-c_2 k}$, for all $k \geq 1$. Ghosal (2001) also showed that if the true density is itself a mixture of beta distributions, then the rate is close to the parametric case: $n^{-1/2}$ up to a logarithmic factor.

Kruijer & Van der Vaart (2008) extended the results obtained by Ghosal (2001) for BDP model by considering an α-smooth class of true models, $\alpha \in (0, 2]$, denoted by $\mathcal{C}_\alpha$. For $\alpha \in (0, 1]$, a real value function defined on $[0, 1]$, f_0, is α-smooth if there exists a constant C such that $|f_0(y_1) - f_0(y_2)| \leq C|y_1 - y_2|^\alpha, \forall\, y_1, y_2 \in [0, 1]$. For $\alpha \in (1, 2]$, f_0 is α-smooth if and only if it is differentiable and its derivative f_0' is $(\alpha - 1)$-smooth. Clearly α regulates the smoothness of the elements of the class $\mathcal{C}_\alpha$: the bigger the α the smoother the function. Under the same assumption on ρ_1 considered by Ghosal (2001), Kruijer & Van der Vaart (2008) showed that the posterior distribution of the BDP model under iid sampling converges in a Hellinger sense to the true density at rate $\epsilon_n = \log(n)^{(1+2\alpha)/(2+2\alpha)}/n^{\alpha/(2+2\alpha)}$, provided that the true density is strictly positive and belongs to $\mathcal{C}_\alpha$, which is suboptimal in a minimax sense for $0 < \alpha \leq 2$.

Kruijer & Van der Vaart (2008) also showed that if ρ_2 is specified such that there exist positive constants B_1, B_2, c_1, and c_2, such that $B_1 e^{-c_1 k} \leq \rho_1(k) \leq B_2 e^{-c_2 k}$, for all $k \geq 1$, then the posterior distribution of DPMMB model under iid sampling converges in a Hellinger sense to the true density at rate $\epsilon_n = \log(n)^{(1+4\alpha)/(2+4\alpha)}/n^{\alpha/(1+2\alpha)}$, provided that the true density belongs to $\mathcal{C}_\alpha$, $\alpha \in (0, 1]$, which is known to be, up to logarithmic factor, the optimal rate for α-smooth densities.

The different behavior of the convergence rate ϵ_n associated with the models is illustrated in Figure 3.2. The figure displays the convergence rate when the true model belongs to the $\mathcal{C}_\alpha$ class, for different values of α. Please notice that the posterior convergence rates described above were obtained using asymptotic arguments and they may not be valid for finite sample sizes.

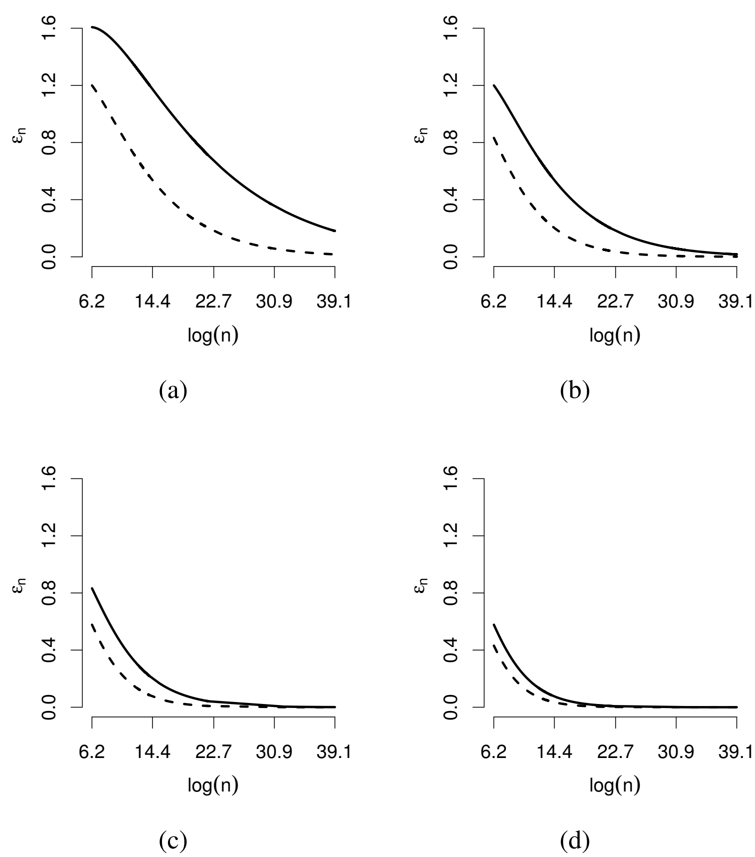


Figure 3.2: Convergence rates: Panel (a), (b), (c), and (d) display the convergence rates, ϵ_n, the models as a function of the sample size n, when the true model belongs to the $\mathcal{C}_\alpha$ class for $\alpha = 0.25$, $\alpha = 0.5$, $\alpha = 1.0$ and $\alpha = 2.0$, respectively. In each plot, the result for BDP and DPMMB model is displayed as continuous and dashed line, respectively.

3.4 An empirical comparison

3.4.1 Simulation settings

Four simulation scenarios were considered. The true density f_0 for Scenario I, II, III and IV is given by

$$f_0(y) = \frac{\sum_{i=1}^{8} 2^{-i\times 0.2}[\cos(2^i y^{0.2}) + 1]}{\sum_{i=1}^{8} 2^{-i\times 0.2}[\int_0^1 \cos(2^i y^{0.2})dy + 1]}, \tag{3.10}$$

$$f_0(y) = \frac{\sum_{i=1}^{5} 2^{-i\times 0.4}[\cos(2^i y^{0.4}) + 1]}{\sum_{i=1}^{5} 2^{-i\times 0.4}[\int_0^1 \cos(2^i y^{0.4})dy + 1]}, \tag{3.11}$$

$$f_0(y) = \begin{cases} 1/2 + 2y, & \text{if } 0 \le y \le 1/2 \\ 5/2 - 2y, & \text{if } 1/2 < y \le 1 \end{cases}, \tag{3.12}$$

and

$$f_0(y) \;= \sum_{j=1}^{\sqrt{25}} w_j \sum_{j_1=(j-1)\sqrt{25}+1}^{j\sqrt{25}} \frac{1}{\sqrt{25}}\, \beta(y \mid j_1, 25 - j_1 + 1), \tag{3.13}$$

respectively, where $w_1 = 0.1$, $w_2 = 0.35$, $w_3 = w_4 = 0$ and $w_5 = 0.55$. Figure 3.3 show the true models under each scenario.

The four probability models satisfy the conditions for the concentration rates described in Section 3.3 to hold. In all cases $f_0(x) > 0$ for all $x \in [0, 1]$. Furthermore, all models belong to the $\mathcal{C}_\alpha$ class. Specifically, the true model of Scenario I is α-smooth with $\alpha = 0.2$, while the true model of Scenario II is α-smooth with $\alpha = 0.4$. As a matter of fact, by using a few trigonometric identities and inequalities, it follows that for $x \in [0, 1]$, $y \in [0, 1]$, and every

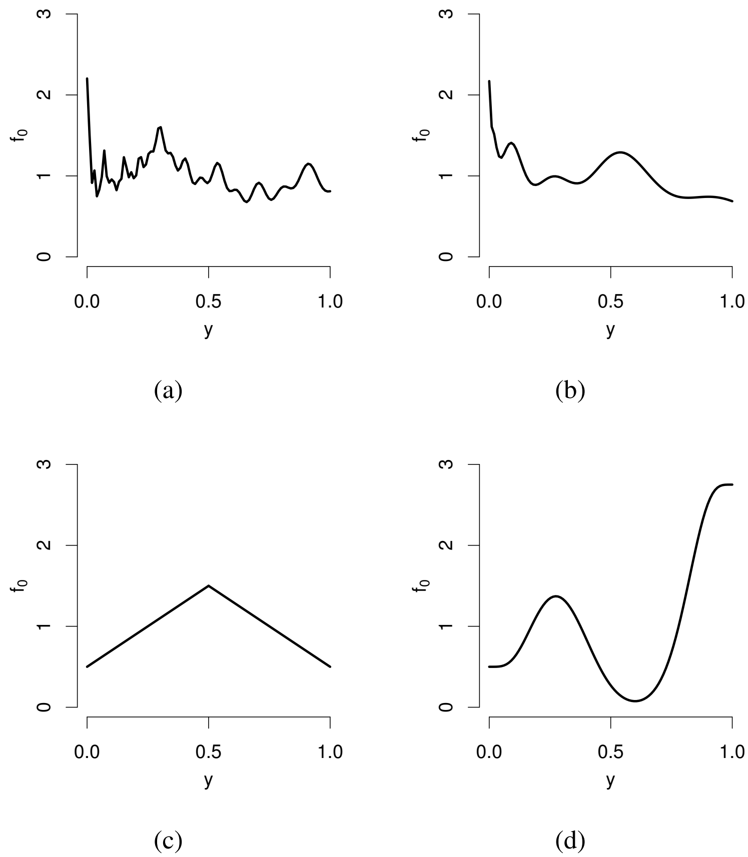


Figure 3.3: Simulation Scenarios: Panel (a), (b), (c), and (d) display the true densities under simulation Scenario I, II, III, and IV, respectively.

$m \in \mathbb{N}$ and $\alpha \in (0,1]$,

$$\begin{aligned}
|f_0(x) - f_0(y)| &= K_{1,\alpha} \left| \sum_{i=1}^{m} 2^{-i\alpha} \left[\cos\left(2^i x^\alpha\right) - \cos\left(2^i y^\alpha\right) \right] \right|, \\
&\leq 2K_{1,\alpha} \sum_{i=1}^{m} 2^{-i\alpha} \left| \sin\left(\frac{2^i x^\alpha + 2^i y^\alpha}{2} \right) \sin\left(\frac{2^i x^\alpha - 2^i y^\alpha}{2} \right) \right|, \\
&\leq 2K_{1,\alpha} \sum_{i=1}^{m} 2^{-i\alpha} \left| \sin\left(\frac{2^i x^\alpha - 2^i y^\alpha}{2} \right) \right|, \\
&\leq K_{1,\alpha} \sum_{i=1}^{m} 2^{-i(\alpha-1)} |x^\alpha - y^\alpha|, \\
&\leq K_\alpha |x-y|^\alpha,
\end{aligned}$$

where $K_{1,\alpha} = \left(\sum_{i=1}^{m} 2^{-i\alpha} \left[\int_0^1 \cos\left(2^i y^\alpha\right) dy + 1 \right] \right)^{-1}$ and $K_\alpha = K_{1,\alpha} \times \left(\frac{1-2^{-m(\alpha-1)}}{2^{\alpha-1}-1} \right)$. Scenario III is also an α-smooth density with $\alpha = 1$, because $|f_0(x) - f_0(y)| \leq 2|x-y|$. Finally, Scenario IV is a particular case of the DPMMB, given by expression (3.6), with $k = 25$ and $w_{2,1,25} = 0.1$, $w_{2,2,25} = 0.35$, $w_{2,3,25} = w_{2,4,25} = 0$ and $w_{2,5,25} = 0.55$. Expression (3.6) is also an α-smooth density with $\alpha = 1$. In fact, by using binomial theorem, it follows that

$$\begin{aligned}
|f_0(x) - f_0(y)| &= \sum_{j=1}^{\sqrt{k}} \frac{w_{2,j,k}}{\sqrt{k}} \sum_{j_1=(j-1)\sqrt{k}+1}^{j\sqrt{k}} |\beta(x \mid j_1, k-j_1+1) - \beta(y \mid j_1, k-j_1+1)|, \\
&\leq k \sum_{j=1}^{\sqrt{k}} \frac{w_{2,j,k}}{\sqrt{k}} \sum_{j_1=(j-1)\sqrt{k}+1}^{j\sqrt{k}} \binom{k}{j_1} \sum_{i=0}^{k-j_1} \binom{k-j_1}{i} |x^{j_1+i-1} - y^{j_1+i-1}|, \\
&\leq 2k^2 \sum_{j=1}^{\sqrt{k}} \frac{w_{2,j,k}}{\sqrt{k}} \sum_{j_1=(j-1)\sqrt{k}+1}^{j\sqrt{k}} \binom{k}{j_1} \sum_{i=0}^{k-j_1} \binom{k-j_1}{i} |x-y|, \\
&\leq C|x-y|,
\end{aligned}$$

where $C = 2^{k+1} k^2 \sum_{j=1}^{\sqrt{k}} \frac{w_{2,j,k}}{\sqrt{k}} \left(\sum_{j_1=(j-1)\sqrt{k}+1}^{j\sqrt{k}} \left(\frac{1}{2}\right)^{j_1} \binom{k}{j_1} \right)$.

For each true model, 15 different sample sizes, n, were considered, ranging from $n = 50$ to $n = 10,000$. For each simulation scenario and sample size, a perfect sample was generated where the data points correspond to the quantiles of equally-spaced probability from the true model.

3.4.2 Model specifications and comparison criteria

Three different versions of BDP model given by expressions (3.2) - (3.4) and DPMMB given by expressions (3.7) - (3.9) were fit to each dataset, by considering different prior distributions on the degree of the polynomial, ρ_1 and ρ_2. In all cases we assume an uniform centering distribution for the underlying Dirichlet process $G_0 = U(0,1)$ and set $M = 1$. Under Prior I, the following restricted negative binomial (NB) prior distributions were assumed,

$$\rho_1(k) \propto \text{NB}(12, 0.3) I(k)_{\{k \in \mathbb{N}: k \geq 1\}},$$

and

$$\rho_2(k) \propto \text{NB}(12, 0.3) I(k)_{\{k \in \mathbb{K}: k \geq 1\}},$$

where $I(x)_A$ is the indicator function for the set A. The prior hyper-parameters, $r = 12$ and $p = 0.3$, were chosen such that the prior mean and variance for the degree of the polynomial was around 5 and 7, respectively, in both models.

Under Prior II, the following restricted NB prior distributions were assumed,

$$\rho_1(k) \propto \text{NB}(65, 0.28) I(k)_{\{k \in \mathbb{N}: k \geq 1\}},$$

and

$$\rho_2(k) \propto \text{NB}(65, 0.28) I(k)_{\{k \in \mathbb{K}: k \geq 1\}}.$$

In this case, the prior hyper-parameters, $r = 65$ and $p = 0.28$, were chosen such that the prior mean and variance for the degree of the polynomial was around 25 and 35, respectively, in both models. Finally, under Prior III, the following restricted discrete uniform prior distributions where assumed,

$$\rho_1(k) \propto I(k)_{\{k \in \mathbb{N}: 1 \leq k \leq 100\}},$$

and

$$\rho_2(k) \propto I(k)_{\{k \in \mathbb{K}: 1 \leq k \leq 625\}}.$$

The prior mean for the degree of the polynomial in both cases was around 46. Priors I and II satisfy the condition required to obtain the concentration rates described in Section 3.3. In fact, it suffices to show that for negative binomial distribution with discrete number of failures the condition holds since ρ_1 and ρ_2 are simply restrictions of a particular member of this family. The proof of the behavior of the negative distribution is given in Appendix A.1.

For each model a single Markov chain was generated. In each case, a conservative total number of 420,000 samples of the posterior distribution were generated. Standard tests (not shown), as implemented in the BOA R library (Smith, 2007), suggested convergence of the chains. Because of storage limitations, the chain was subsampled every 40 iterations and considering a burn-in period of 10,000 samples to give a reduce chain of length 10,000. The performance of the models was evaluated by comparing the posterior mean of the L_1, L_2 and L_∞ distances between the random density and the true model. Models were also compared by considering the log pseudo marginal likelihood (LPML), originally developed by Geisser & Eddy (1979) and further considered by Gelfand & Dey (1994). LPML for model M is defined as 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) have also been 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 Markov chain output.

Functions implementing the Markov chain Monte Carlo (MCMC) algorithms for each model are available upon request to the authors. The implementation for the BDP model is an extension of the one available in DPpackage (Jara et al., 2011).

3.4.3 The results

Figure 3.4 shows the posterior mean of the L_1 distance between the true density and the random probability measure, respectively. The results are displayed for BDP and DPMMB models under the four simulation settings, different samples sizes and prior specification on the degree of the polynomial. As expected, the L_1 distance reduces as the sample size increases for both models.

The results, however, strongly suggest that the sub-optimal BDP model has a better performance regarding the L_1 criteria, even when the assumptions needed to obtain the optimal results for DPMMB model hold. Under Priori I for the degree of the polynomial and the biggest differences in the performance of the models, the posterior mean of the L_1 distance for DPMMB model was 53%, 55%, 53% and 50% bigger than the corresponding value observed for BDP model for Scenario I, II, III, and IV, respectively. Under Priori II for the degree of the polynomial and the biggest differences in the performance of the models, the posterior mean of the L_1 distance for DPMMB model was 60%, 169%, 176% and 14% bigger than the corresponding value observed for BDP model for Scenario I, II, III, and IV, respectively. Moreover, even when the true probability model has the same functional form than DPMMB model, BDP behaves in a similar manner and can outperform DPMMB model for large sample sizes. Similar results are observed when considering the L_2 distance and the results are provided in Appendix A.2.

Figure 3.5 shows the posterior mean of the L_∞ distance between the true density and the random probability measure, respectively. In general, the results under L_∞ are similar to the ones observed under the other distances. The main difference is observed under Scenario III, where DPMMB outperforms BDP in the majority of the cases (35 out of 45). This suggest that for true models with higher regularity, DPMMB produces better estimates under the L_∞ distance.

Figure 3.6 show the results for LPML for each model under the different simulation scenarios, prior distribution for the degree of the polynomial and sample size. The results show that LPML is an adequate but conservative model selection criteria; For instance, if the model selection criteria is base on differences in LPML bigger than 5, it never choose the model with

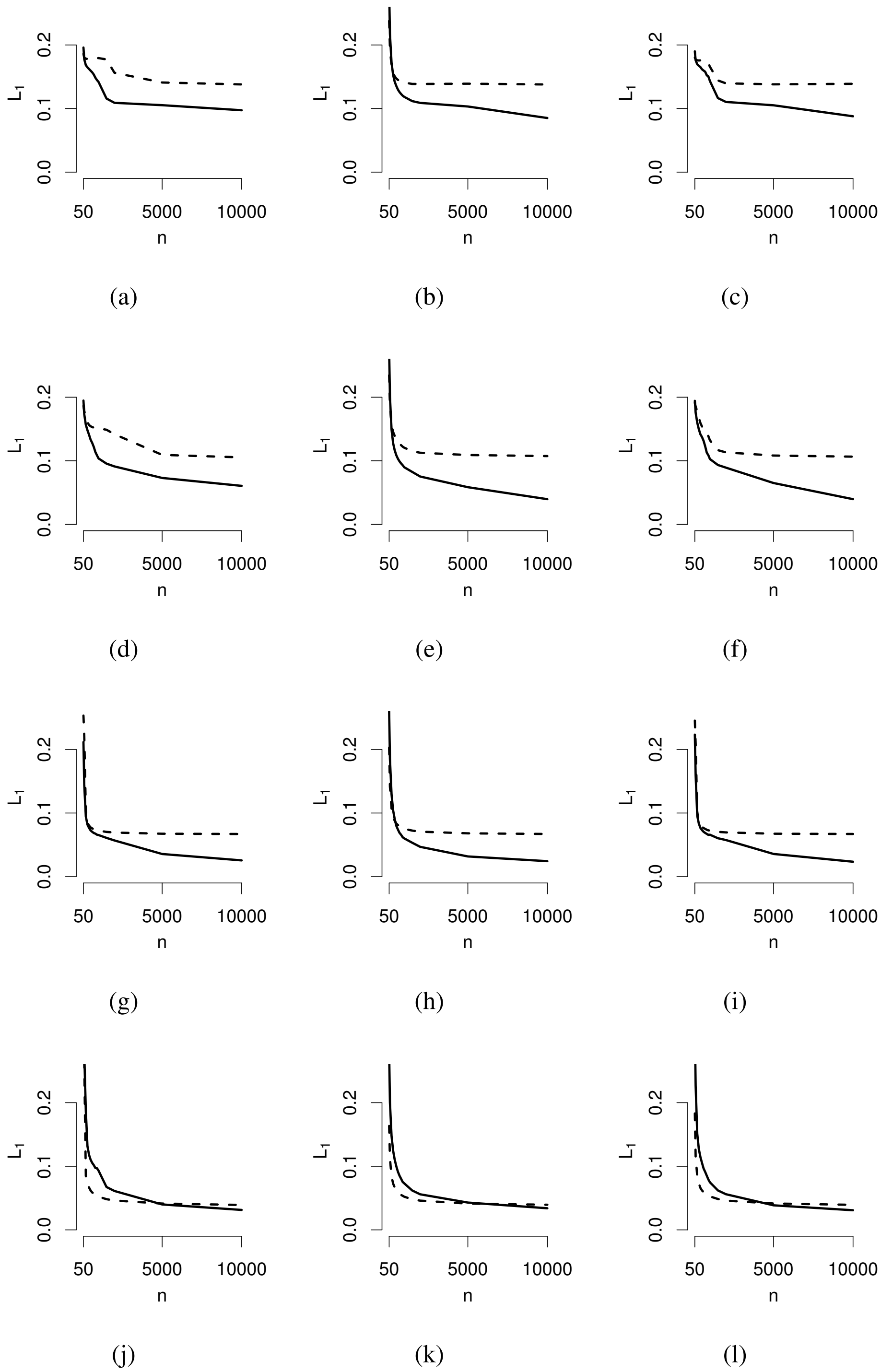


Figure 3.4: Simulated Data: Posterior mean of the L_1 distance between the true density and the random probability measure as a function of the sample size. The results for BDP and DPMMB are shown as a solid and dashed line, respectively. Panel (a) - (c), (d) - (f), (g) - (i), and (j) - (l) display the results under simulation Scenario I, II, III, and IV, respectively. Panel (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l), display the results under Prior I, II and III, respectively.

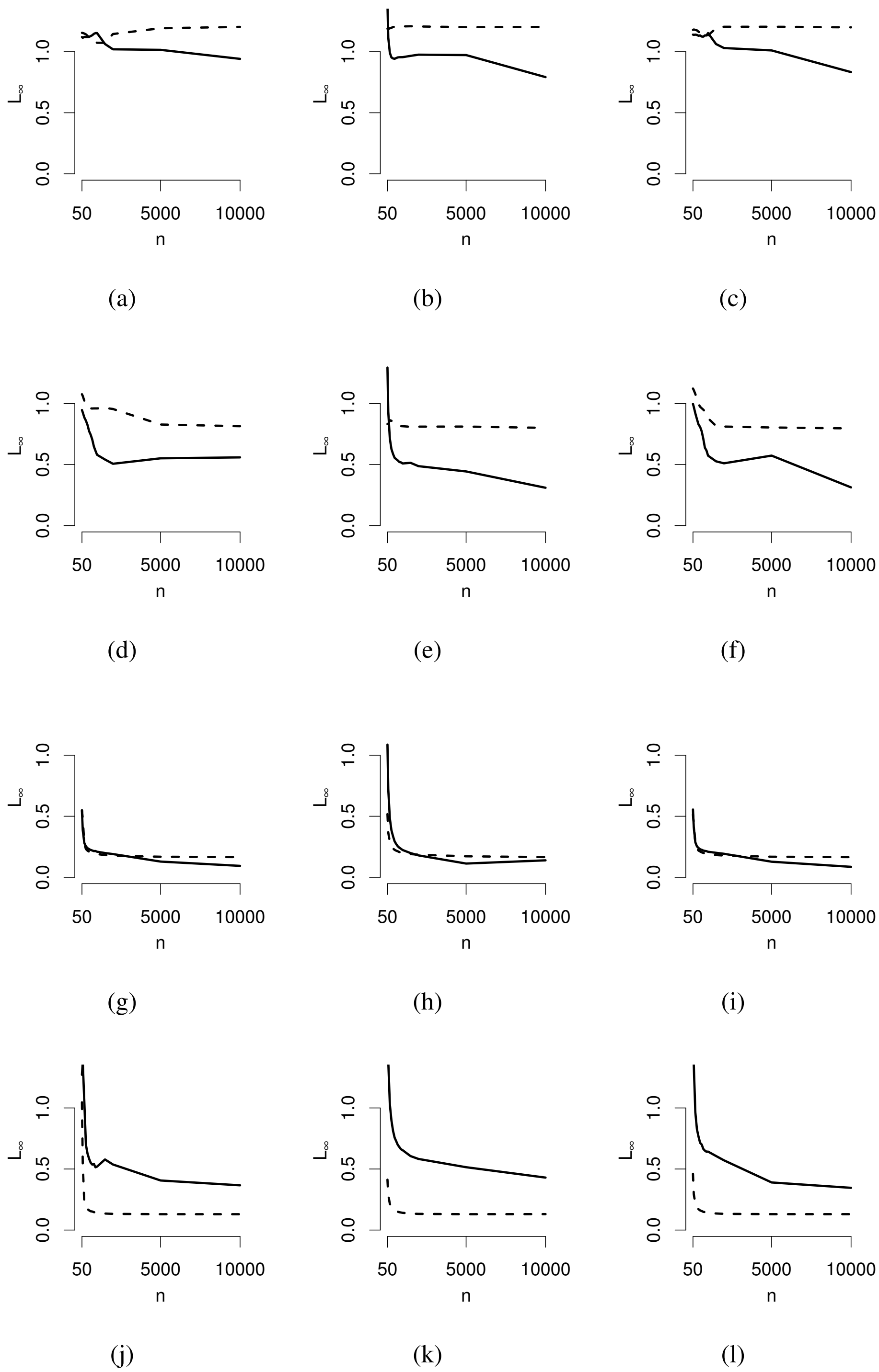


Figure 3.5: Simulated Data: Posterior mean of the L_∞ distance between the true density and the random probability measure as a function of the sample size. The results for the BDP and DPMMB are shown as a solid and dashed line, respectively. Panel (a) - (c), (d) - (f), (g) - (i), and (j) - (l) display the results under simulation Scenario I, II, III, and IV, respectively. Panel (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l), display the results under Prior I, II and III, respectively.

the worst L_1 distance to the truth. However, in 68.3% of the cases where the competing models show a significantly different posterior means of L_1 distance, LPML indicates evidence of no difference in the behavior of the models. Similar agreement results were obtained between LPML and the other distances.

Figures A.6 and A.9 illustrate the different performance of the models. These figures display the posterior mean and point-wise 95% highest posterior density (95% HPD) credible bands for the density under Prior II, and for Scenarios II and III. HPD bands were computed using the method described by Chen & Shao (1999). Additional results for the simulation study are provided in a supplementary material.

The results show that BDP model is highly flexible and is able to capture strong as well as small deviations from standard parametric assumptions. In general under BDP model, the posterior mean estimates with a minimal error and high precision the true density function, even for relatively small sample sizes; the true model was completely covered by 95% point-wise HPD bands and the quality of the estimation improved as the sample size increases. The results also clearly illustrate that poor estimates can be obtained by using DPMMB model.

Similar results were observed under the different priors on the degree of the polynomial in Scenarios I, II and III. Under Scenario IV, both models behaved in a similar manner, yielding adequate posterior inferences on the true model.

3.5 An application to solid waste data

We consider data about residentially generated solid waste in the city of Santiago de Cali, Colombia. The dataset contains information about 261 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 261 block sides was separated in different kinds of materials, including food and hygienic waste. The proportions of these materials were registered for each block side. We refer the reader to Klinger et al. (2009) for more details about these data.

The three versions of BDP and DPMMB models discussed in the previous section were fit to the proportion of food and hygienic waste. In each case, one Markov chain was generated

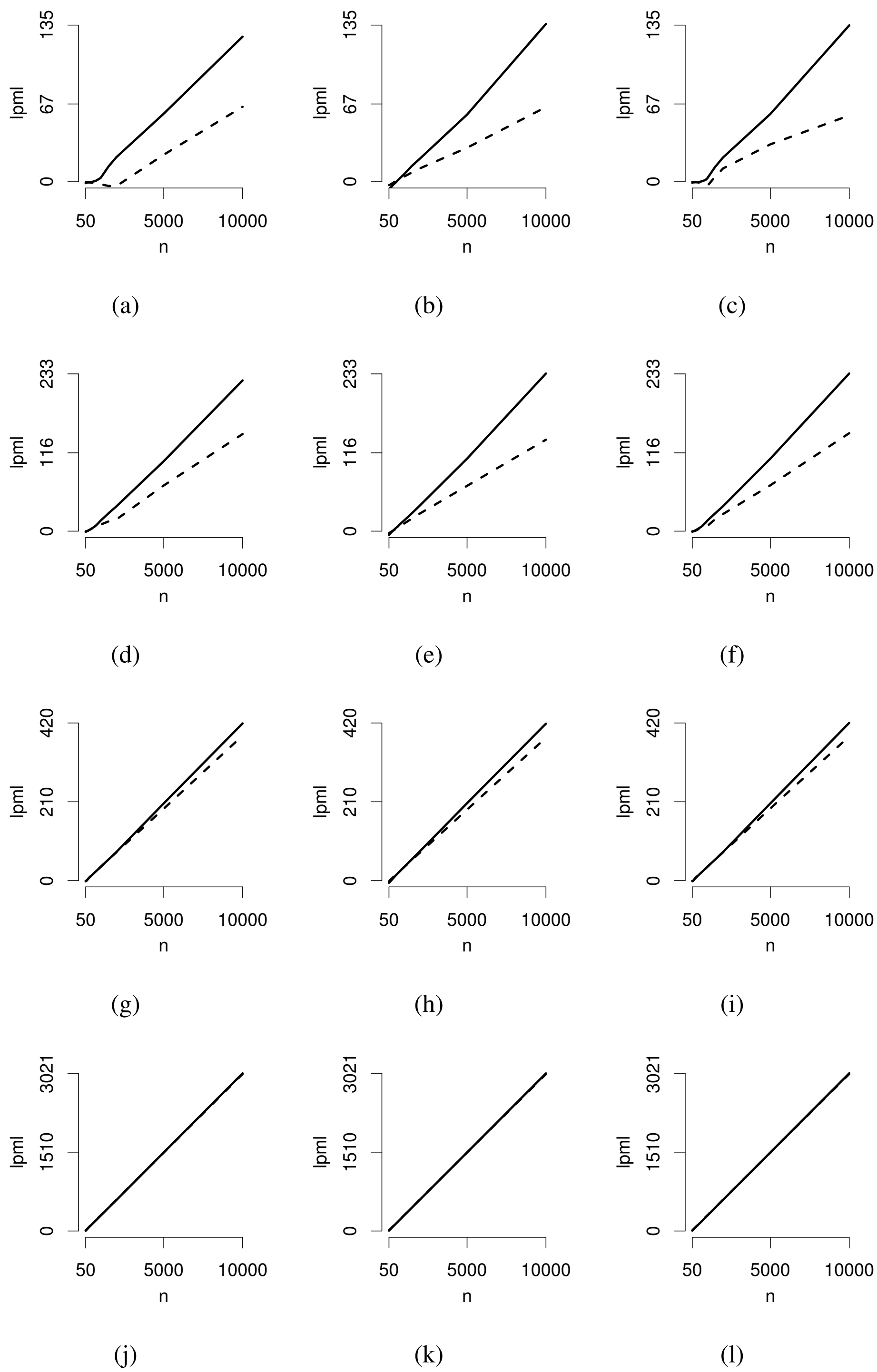
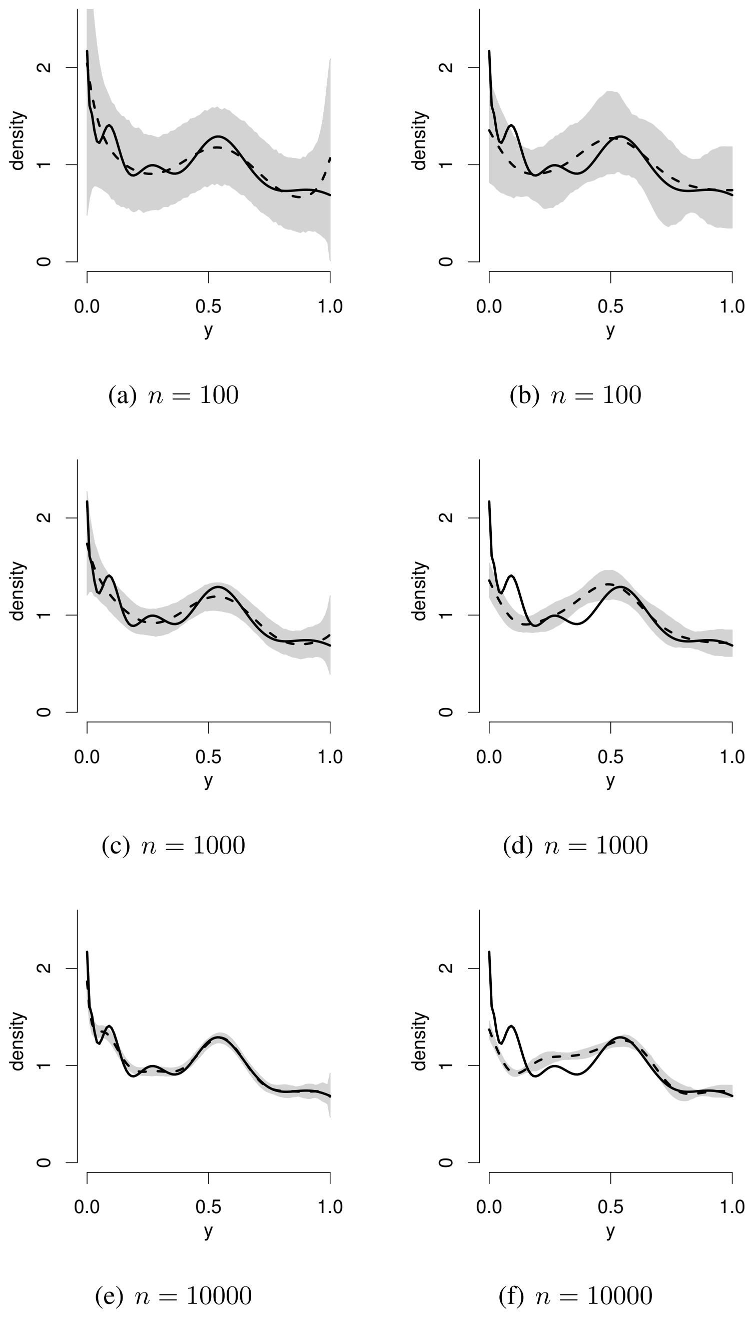


Figure 3.6: Simulated Data: Log pseudo marginal likelihood as a function of the sample size. The results for the BDP and DPMMB are shown as a solid and dashed line, respectively. Panel (a) - (c), (d) - (f), (g) - (i), and (j) - (l) display the results under simulation Scenario I, II, III, and IV, respectively. Panel (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l), display the results under Prior I, II and III, respectively.

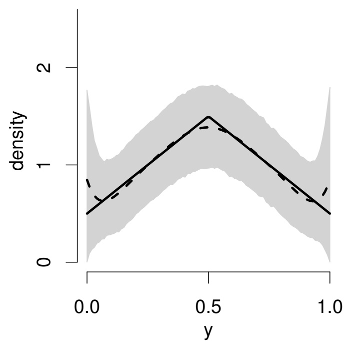


(a) $n = 100$ (b) $n = 100$

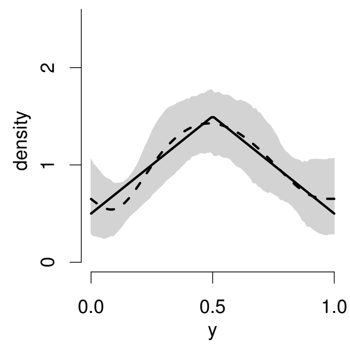
(c) $n = 1000$ (d) $n = 1000$

(e) $n = 10000$ (f) $n = 10000$

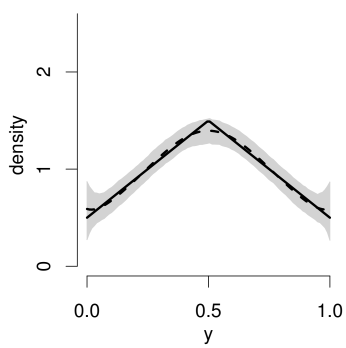
Figure 3.7: Simulated Data: Posterior mean (dashed line), point-wise 95% HPD band (grey area), and true model (continuous line) under Scenario II and Prior II, for selected sample sizes. Panel (a), (c) and (e) display the results for BDP model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively. Panel (b), (d) and (f) display the results for DPMMB model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively.



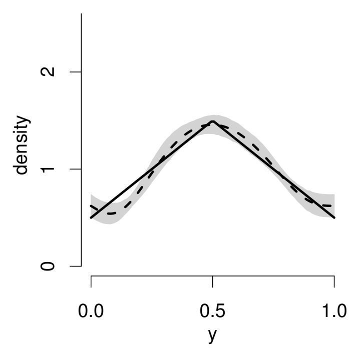
(a) $n = 100$



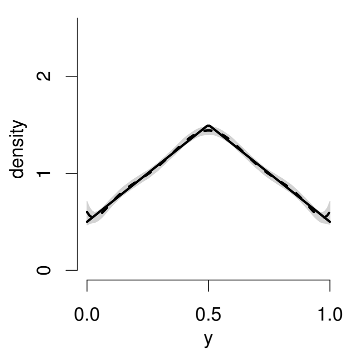
(b) $n = 100$



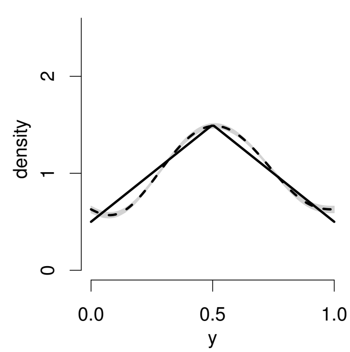
(c) $n = 1000$



(d) $n = 1000$



(e) $n = 10000$



(f) $n = 10000$

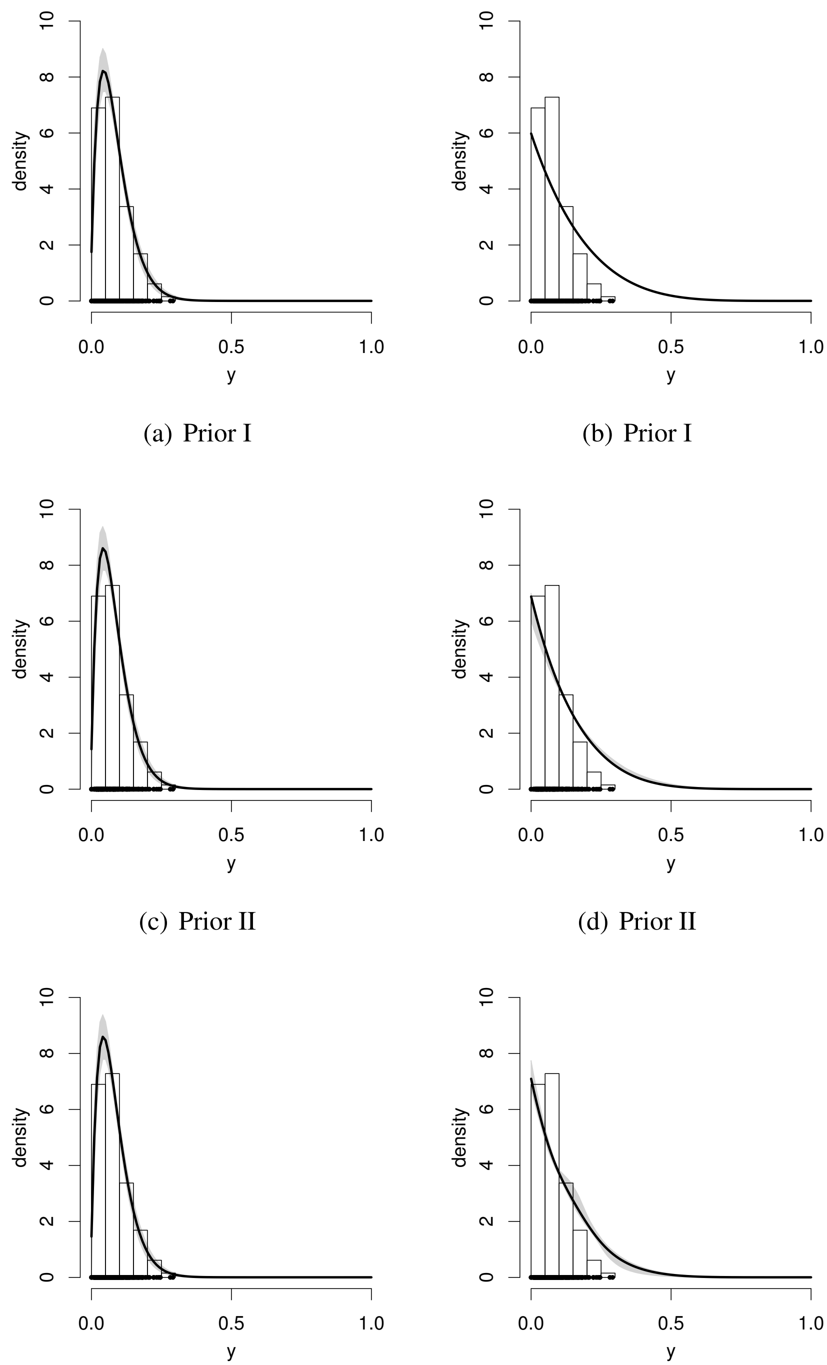
Figure 3.8: Simulated Data: Posterior mean (dashed line), point-wise 95% HPD band (grey area), and true model (continuous line) under Scenario III and Prior II, for selected sample sizes. Panel (a), (c) and (e) display the results for BDP model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively. Panel (b), (d) and (f) display the results for DPMMB model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively.

completing a conservative total number of 620,000 scans of the Markov chain cycle. Because of storage limitations, the full chain was subsampled every 60 iterations, after a burn-in period of 10,000 samples, to give a reduced chain of length 10,000. Model comparison was performed using LPML criteria.

For the proportion of hygienic waste data, BDP model clearly outperformed DPMMB model under LPML criteria. Furthermore, the results for DPMMB model were more sensitive to the prior specification on the polynomial degree. LPML values for BDP model were 422.2, 422.5 and 422.2, under Prior I, II and III, respectively. LPML values for DPMMB model were 353.9, 370.0 and 373.2, under Prior I, II and III, respectively. Figure 3.9 show the posterior inferences for the density under each model and confirm the results obtained under LPML criteria. The results show that poor inferences can be obtain under the rate-optimal DPMMB model for finite samples. Notice also that the results show an important departure from standard parametric assumptions. Specifically, the positive density at zero and the existence of a central mode cannot be obtained from a beta model. As a matter of fact, the positive density observed at zero for the proportion of hygienic waste can be explained by the existence of zero values in the dataset. Because of that, we were not able to fit the beta 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. However, this solution would imply that the density estimate would be equal to zero on the extreme values of the domain, which is clearly not supported by the data.

For the proportion of food the models behaved in a similar way regarding LPML criteria, with the exception of DPMMB model under Prior I, which showed the worst goodness-of-fit performance. LPML values were 213.9, 215.3 and 215.2 for BDP model under Prior I, II and III, respectively. The corresponding LPML values for DPMMB model were 195.5, 218.3 and 218.1, under Prior I, II and III, respectively. Again DPMMB showed to be more sensitive to the prior specification on the degree of the polynomial. Figure 3.10 displays the posterior inferences for the density under each model.

The food waste data also shows an important departure from standard parametric assumptions. Specifically, the proportion of food waste is skew to the left. BDP model outperforms

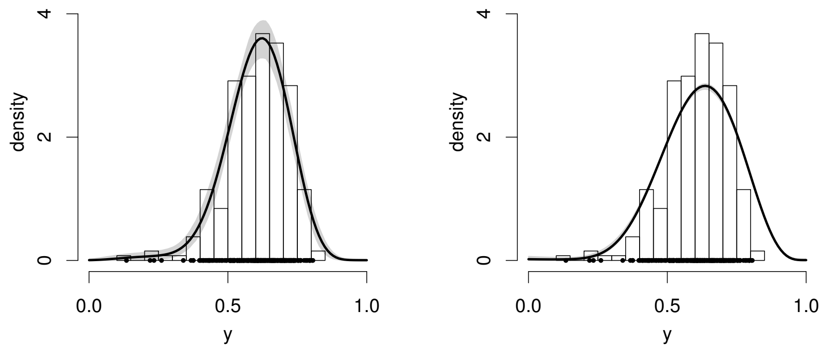


(a) Prior I (b) Prior I

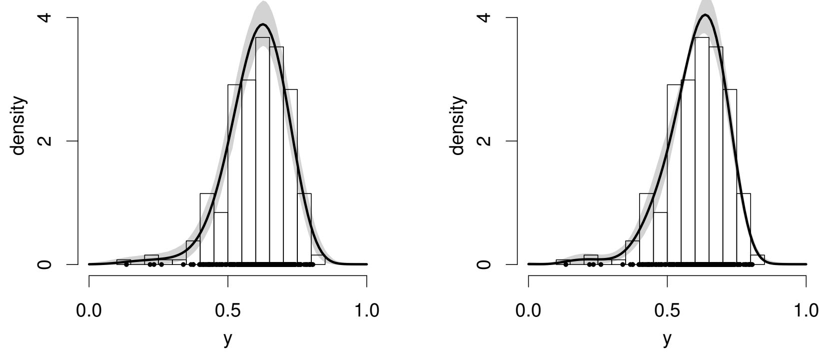
(c) Prior II (d) Prior II

(e) Prior III (f) Prior III

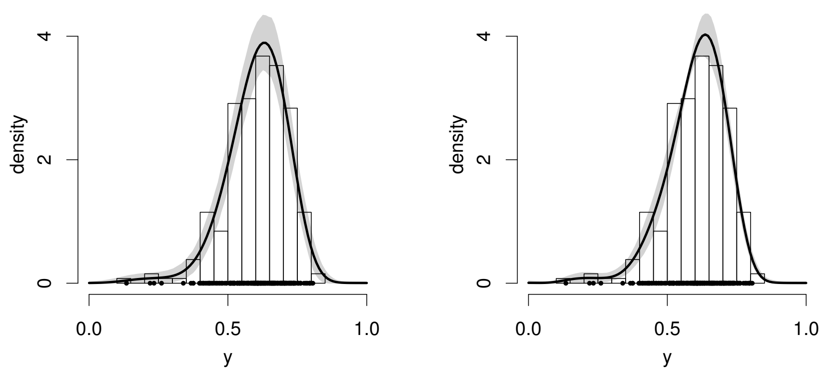
Figure 3.9: Proportion of hygienic waste: Posterior mean (solid line) and histogram of proportion of hygienic waste data. Panels (a), (c) and (e) display the results for BDP model under Prior I, II and III, respectively. Panels (b), (d) and (f) display the results for DPMMB model under Prior I, II and III, respectively.



(a) Prior I (b) Prior I

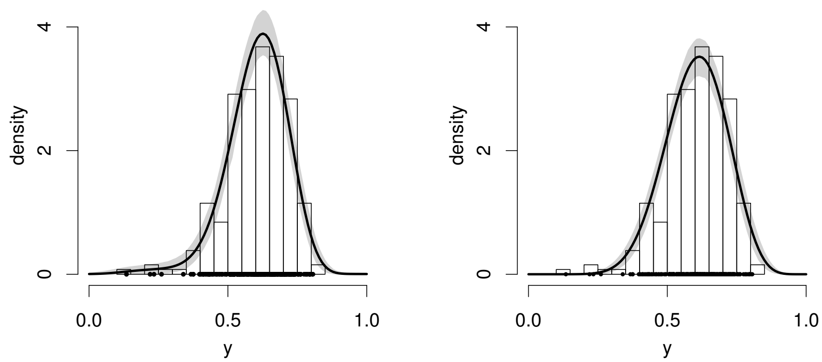


(c) Prior II (d) Prior II



(e) Prior III (f) Prior III

Figure 3.10: Proportion of food: Posterior mean (solid line) and histogram of proportion of food waste data. Panels (a), (c) and (e) display the results for BDP model under Prior I, II and III, respectively. Panels (b), (d) and (f) display the results for DPMMB model under Prior I, II and III, respectively.



(a) BDP (b) Beta

Figure 3.11: Proportion of food: Posterior mean (solid line) and histogram of proportion of food waste data. Panels (a) and (c) and (e) display the results for a beta model and BDP model under prior II for the degree of the polynomial.

a standard beta model for all priors. DPMMB model outperforms the beta model for priors II and III on the degree of the polynomial only. For the beta model using independent $\Gamma(1, 0.01)$ priors for the model parameters LPML was 205.4. Figure 3.11 illustrates the deviation of the data from the parametric model. For comparison purposes, the results for the beta model are displayed along with the best BDP model.

Chapter 4

Dependent Bayesian nonparametric modeling of compositional data using random Bernstein polynomials

This chapter will be submitted for publication as:

WEHRHAHN, C., JARA, A., BARRIENTOS, A.F. (2015). Dependent Bayesian nonparametric modeling of compositional data using random Bernstein polynomials.

4.1 Introduction

Models for probability distributions based on convex combinations of densities from parametric families underly mainstream approaches to single and conditional density estimation, including kernel techniques, nonparametric maximum likelihood and Bayesian nonparametric (BNP) approaches (see, e.g., Müller et al., 2015, and references therein). From a BNP point of view, the mixture model provides a convenient set up for density estimation in that a prior distribution on densities is induced by placing a prior distribution on the mixing measure.

On the real line, mixtures of normal densities induced by a Dirichlet process (DP) (Ferguson, 1973, 1974, 1983) or by a dependent Dirichlet process (MacEachern, 1999, 2000) are often used to model a single smooth densities or a collection of smooth densities indexed by predictors, respectively. Other extensions and alternative constructions for dealing with related probability distributions include the logistic Gaussian process (Tokdar et al., 2010), the ordered-category probit regression model (Karabatsos & Walker, 2012), the dependent beta process (Trippa et al., 2011), the dependent tailfree processes (Jara & Hanson, 2011), the dependent neutral to the right processes and correlated two-parameter Poisson-Dirichlet processes (Epifani & Lijoi, 2010; Leisen & Lijoi, 2011) and the general class of dependent normalized completely random measures (Lijoi et al., 2014).

Due to the flexibility and ease in computation, these models are now routinely implemented in a wide variety of applications. While the normal kernel is a sensible choice on the real line, its usefulness is rather limited when considering densities on convex and compact subspaces, such as the closed unit interval or the m–dimensional simplex

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

Although methods based on the normal kernel could be used to deal with data supported on these spaces by using transformations, the resulting model is susceptible to boundary effects.

Motivated by its uniform approximation properties, frequentist and Bayesian methods based on univariate Bernstein polynomials (BP) have been proposed for the estimation of probability distributions supported on bounded intervals and on unit hyper-cubes (see, e.g. Petrone, 1999a,b; Petrone & Wasserman, 2002; Tenbusch, 1994; Babu & Chaubey, 2006; Zheng et al., 2010). Babu & Chaubey (2006) studied a general multivariate version of the bivariate estimator proposed by Tenbusch (1994). Zheng et al. (2010) construct a multivariate Bernstein polynomial (MBP) prior for the spectral density of a random field. Multivariate extensions of Bernstein polynomials defined on Δ_m were considered by Tenbusch (1994) to propose and study a density estimator for the data supported on Δ_2. Tenbusch's estimator arises by taking G to be the restriction of the empirical cumulative distribution function (CDF) to Δ_2, and it is based on the class of MBP given in Definition 4.1.

Definition 4.1. *For a given function $G : \Delta_m \longrightarrow \mathbb{R}$, the associated MBP of degree k on Δ_m is defined by*

$$\begin{aligned}\widetilde{B}_{k,G}(\boldsymbol{y}) &= \sum_{\boldsymbol{j}\in\mathscr{J}_m^k} G\left(\frac{j_1}{k},\ldots,\frac{j_m}{k}\right)\frac{k!}{\left(\prod_{l=1}^m j_l!\right)\left(k-\sum_{l=1}^m j_l\right)!}\left(\prod_{l=1}^m y_l^{j_l}\right)\left(1-\sum_{l=1}^m y_l\right)^{k-\sum_{l=1}^m j_l}, \\ &= \sum_{\boldsymbol{j}\in\mathscr{J}_m^k} G\left(\frac{j_1}{k},\ldots,\frac{j_m}{k}\right)\text{Mult}\left(\mathbf{j}\mid k,\boldsymbol{y}\right),\end{aligned}$$

where $\boldsymbol{j} = (j_1,\ldots,j_m)$, $\mathscr{J}_m^k = \{(j_1,\ldots,j_m)\in\{0,\ldots,k\}^m : \sum_{l=1}^m j_l \le k\}$ and $\text{Mult}(\cdot\mid k,\boldsymbol{y})$ stands for the probability mass function of a multinomial distribution with parameters $(k,\boldsymbol{y})$.

Although Tenbusch's estimator is consistent and optimal at the interior points of the simplex, it is not a valid density function for finite k and finite sample size. Indeed, it is not difficult to show that, under Definition 4.1, if G is the restriction of the CDF of a probability measure on Δ_m, then $\widetilde{B}_{k,G}(\cdot)$ is not the restriction of the CDF of a probability measure defined on Δ_m for a finite k. In this case, $\widetilde{B}_{k,G}(\cdot)$ can be expressed as a linear combination of CDFs of probability measures defined on Δ_m, where the coefficients are nonnegative but do not add up to 1. To avoid this problem, Barrientos et al. (2015a) proposed a modified class of MBP, which retains the well known approximation properties of the original version.

We extend the class of MBP priors of Barrientos et al. (2015a), to deal with sets of predictor-dependents probability distributions for compositional data, based on the use of dependent stick-breaking processes (see, e.g. Barrientos et al., 2012). The proposed methods and the study of its properties complement the results obtained by Barrientos et al. (2015b) for data supported on $[0,1]$. The rest of the chapter is organized as follows. The modified clases of MBP and its main properties are summarized in Section 4.2. The proposed models for collections of probability measures defined on Δ_m are discussed in Section 4.3. The basic properties of the proposed model class are provided in Section 4.4.

4.2 The modified class of MBP on Δ_m

The modified class of MBP proposed by Barrientos et al. (2015a) is obtained by increasing the size of the set $\mathcal{J}_m^k$ and the domain of function G from the original definition.

Definition 4.2. *For a given function $G : \mathbb{R}^m \longrightarrow \mathbb{R}$, the associated MBP of degree $k \in \mathbb{N}$ on Δ_m is defined by*

$$
\begin{aligned}
B_{k,G}(\boldsymbol{y}) &= \sum_{\boldsymbol{j}\in\mathcal{H}_{k,m}} G\left(\frac{j_1}{k},\ldots,\frac{j_m}{k}\right) \frac{k_1!}{\left(\prod_{l=1}^m j_l!\right)\left(k_1-\sum_{l=1}^m j_l\right)!}\left(\prod_{l=1}^m y_l^{j_l}\right)\left(1-\sum_{l=1}^m y_l\right)^{k_1-\sum_{l=1}^m j_l}, \\
&= \sum_{\boldsymbol{j}\in\mathcal{H}_{k,m}} G\left(\frac{j_1}{k},\ldots,\frac{j_m}{k}\right) \mathrm{Mult}\left(\mathbf{j} \mid k_1, \boldsymbol{y}\right),
\end{aligned}
$$

where $k_1 = k+m-1$ and $\mathcal{H}_{k,m} = \{(j_1,\ldots,j_m) \in \{0,\ldots,k\}^m : \sum_{l=1}^m j_l \le k+m-1\}$.

As shown by Barrientos et al. (2015a), the modified class $B_{k,G}$ retains most of the appealing approximation properties of univariate BP and the standard class of MBP, $\tilde{B}_{k,G}$. Specifically, if G is a real-valued function defined on $\mathbb{R}^m$ and $G|_{\Delta_m}$ its restriction on Δ_m, then

$$
\lim_{k\to\infty} B_{k,G}(\boldsymbol{y}) = G|_{\Delta_m}(\boldsymbol{y}),
$$

at each point of continuity $\boldsymbol{y}$ of $G|_{\Delta_m}$. Furthermore, the relation holds uniformly on Δ_m if $G|_{\Delta_m}$ is a continuous function.

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

$$b_{k,G}(\boldsymbol{y}) \;\; = \sum_{\boldsymbol{j} \in \mathcal{H}^0_{k,m}} W_{k,\boldsymbol{j},G} \times \mathrm{d}(\boldsymbol{y} \mid \alpha(k, \boldsymbol{j})), \tag{4.1}$$

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

$$W_{k,\boldsymbol{j},G} = G\left(\left(\frac{j_1 - 1}{k}, \frac{j_1}{k}\right] \times \ldots \times \left(\frac{j_m - 1}{k}, \frac{j_m}{k}\right]\right),$$

$\alpha\,(k, \boldsymbol{j}) = (\boldsymbol{j},\ k + m - \sum_{l=1}^{m} j_l)$, and $\mathrm{d}(\cdot \mid (\alpha_1, \ldots, \alpha_{m+1}))$ denotes the density function of a m–dimensional Dirichlet distribution with parameters $(\alpha_1, \ldots, \alpha_{m+1})$.

4.3 Random MBP for fully nonparametric regression

4.3.1 The inferential problem and motivating ideas

Suppose that we observe regression data $\{(\boldsymbol{x}_i, \boldsymbol{y}_i) : i = 1, \ldots, n\}$, where $\boldsymbol{x}_i \in \mathscr{X} \subseteq \mathbb{R}^p$ is a p-dimensional vector of predictors and $\boldsymbol{y}_i$ is a continuous Δ_m-valued outcome vector. Rather than assuming an unknown functional form for the mean function or another functional, as is usually done in nonparametric regression, under the framework of fully nonparametric regression the problem is cast as inference for a family of conditional distributions

$$\{F_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^p\},$$

where $\boldsymbol{y}_i \mid \boldsymbol{x}_i \overset{ind.}{\sim} F_{\boldsymbol{x}_i}$. Therefore, from a Bayesian point of view, the specification of a fully nonparametric regression model requires of the definition of a probability model for the set of predictor-dependent absolutely continuous probability measures $\{F_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$, allowing the complete shape of the elements of $\{F_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}\}$ to change flexibly with the values of $\boldsymbol{x}$.

To introduce dependence in the continuous random probability measures defined on Δ_m, we replace the mixing distribution G in expression (4.1) by a dependent stick-breaking process, which is defined by using stochastic processes indexed by predictors $\boldsymbol{x} \in \mathscr{X}$. Specifically, by using the fact that expression (4.1) can be equivalently written as the following mixture of Dirichlet densities

$$b_{k,G}(\boldsymbol{y}) = \int_{\Delta_m} \mathrm{d}\left(\boldsymbol{y} \mid |\lceil k\boldsymbol{\theta}\rceil, k+m-\|\lceil k\boldsymbol{\theta}\rceil\|_1\right) G(d\boldsymbol{\theta}),$$

where $\lceil\cdot\rceil$ denotes the ceiling function and $\|\cdot\|$ denotes L^1-norm, we define random dependent densities by considering dependent mixing distributions $G_{\boldsymbol{x}}$,

$$g_{\boldsymbol{x}}(\boldsymbol{y} \mid k, G_{\boldsymbol{x}}) = \int_{\Delta_m} \mathrm{d}\left(\boldsymbol{y} \mid |\lceil k\boldsymbol{\theta}\rceil, k+m-\|\lceil k\boldsymbol{\theta}\rceil\|_1\right) G_{\boldsymbol{x}}(d\boldsymbol{\theta}),$$

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

4.3.2 The formal definition of the models

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

Definition 4.3. *Let $\mathscr{V}$ and $\mathscr{H}$ be two sets of functions as before. Let $F = \{F(\boldsymbol{x},\cdot) : \boldsymbol{x} \in \mathscr{X}\}$ be a $\mathscr{P}(\Delta_m)$-valued stochastic process defined on an appropriate probability space $(\Omega, \mathscr{A}, P)$, such that:*

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

(ii) $\boldsymbol{z}_j : \mathscr{X} \times \Omega \longrightarrow \mathbb{R}^m$, $j \geq 1$, *are independent and identically distributed real-valued stochastic processes with law indexed by a finite-dimensional parameter* $\boldsymbol{\Psi}_z$.

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

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

$$f(\boldsymbol{x}, \omega)(\cdot) = \sum_{j=1}^{\infty} w_j(\boldsymbol{x}, \omega) \mathrm{d}\left(\,\cdot\, \left| \lceil k(\omega)\boldsymbol{\theta}_j(\boldsymbol{x}, \omega) \rceil, k(\omega) + m - \| \lceil k(\omega)\boldsymbol{\theta}_j(\boldsymbol{x}, \omega) \rceil \|_1 \right. \right) \quad (4.2)$$

where $\boldsymbol{\theta}_j(\boldsymbol{x}, \omega) = h_{\boldsymbol{x}}(\boldsymbol{z}_j(\boldsymbol{x}, \omega))$,

$$\lceil k(\omega)\boldsymbol{\theta}_j(\boldsymbol{x}, \omega) \rceil = \left(\lceil k(\omega)\theta_{j1}(\boldsymbol{x}, \omega) \rceil, \ldots, \lceil k(\omega)\theta_{jm}(\boldsymbol{x}, \omega) \rceil \right),$$

and

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

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

In the search of parsimonious models, it is of interest to study two special cases of the general construction given by Definition 4.3. The special involving dependent stick-breaking processes with common support points and predictor-dependent weights is referred to as 'single–atoms' DMBPP. In this simplified version, the real-valued stochastic processes of condition (ii) in Definition 4.3, $\boldsymbol{z}_j = \{\boldsymbol{z}_j(\boldsymbol{x}) : \boldsymbol{x} \in \mathscr{X}\}$, are replaced by independent and identically distributed $\tilde{\Delta}_m$-valued random vectors, $\boldsymbol{\theta}_j$.

Definition 4.4. *Let* $\mathscr{V}$ *and* $\mathscr{H}$ *be two sets of functions as before. Let* $F = \{F(\boldsymbol{x}, \cdot\,) : \boldsymbol{x} \in \mathscr{X}\}$ *be a* $\mathscr{P}(\Delta_m)$*-valued stochastic process defined on an appropriate probability space* $(\Omega, \mathscr{A}, P)$*, such that:*

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

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

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

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

$$f(\boldsymbol{x}, \omega)(\cdot) = \sum_{j=1}^{\infty} w_j(\boldsymbol{x}, \omega) \mathrm{d}\left(\,\cdot\, \middle| \lceil k(\omega)\boldsymbol{\theta}_j(\omega) \rceil, k(\omega) + d - \| \lceil k(\omega)\boldsymbol{\theta}_j(\omega) \rceil \|_1 \right), \quad (4.3)$$

where $w_j(\boldsymbol{x}, \omega)$ are defined as in Definition 4.3 and

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

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

The case involving a dependent stick-breaking process with common weights and predictor-dependent support points is referred to as 'single weights' DMBPP. In this simplified version, the real-valued stochastic processes of condition (i) in Definition 4.3, $\eta_j = \{\eta_j(\boldsymbol{x}) : \boldsymbol{x} \in \mathscr{X}\}$, are replaced by $[0, 1]$-valued independent and identically distributed random variables, v_j.

Definition 4.5. *Let $\mathscr{V}$ and $\mathscr{H}$ be two sets of functions as before. Let $F = \{F(\boldsymbol{x}, \cdot\,) : \boldsymbol{x} \in \mathscr{X}\}$ be a $\mathscr{P}(\Delta_m)$-valued stochastic process defined on an appropriate probability space $(\Omega, \mathscr{A}, P)$, such that:*

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

(ii) $\boldsymbol{z}_j : \mathscr{X} \times \Omega \longrightarrow \mathbb{R}^m$, $j \geq 1$, *are independent and identically distributed real-valued stochastic processes with law indexed by a finite-dimensional parameter $\boldsymbol{\Psi}_z$.*

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

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

$$f(\boldsymbol{x},\omega)(\cdot) = \sum_{j=1}^{\infty} w_j(\omega) \mathrm{d}\left(\,\cdot\, \left| \lceil k(\omega)\boldsymbol{\theta}_j(\boldsymbol{x},\omega)\rceil, k(\omega) + d - \|\lceil k(\omega)\boldsymbol{\theta}_j(\boldsymbol{x},\omega)\rceil\|_1 \right.\right), \tag{4.4}$$

where $\boldsymbol{\theta}_j(\boldsymbol{x},\omega)$ *and* $\lceil k(\omega)\boldsymbol{\theta}_j(\boldsymbol{x},\omega)\rceil$ *are defined as in Definition 4.3, and*

$$w_j(\omega) = v_j(\omega) \prod_{l<j} [1 - v_l(\omega)] .$$

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

Notice that DMBPP are well defined if the mapping induced by (iv) in Definition 4.3, 4.4, and 4.5 is measurable. This will be discussed in detail in Section 4.3.3. Notice also that expressions (4.2), (4.3), and (4.4) are indeed a density w.r.t. Lebesgue measure since, for every $\boldsymbol{x} \in \mathscr{X}$,

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

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

$$f_{\boldsymbol{x}}(\cdot) \;\; = \sum_{\boldsymbol{j} \in \mathcal{H}^0_{k,m}} W_{k,\boldsymbol{j},\boldsymbol{x}} \times \mathrm{d}(\,\cdot\, \mid \alpha(k,\boldsymbol{j})), \tag{4.5}$$

where

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

which has some advantages when the main interest is on single functionals, such as the mean function (Wade et al., 2014).

4.3.3 The measurability of the processes

In this section, we show that the corresponding mappings defining the trajectories of DMBPP, wDMBPP, and θDMBPP are measurable under the Borel σ-field generated by the weak product topology, L_∞ product topology and L_∞ topology, which correspond to generalizations of standard topologies for spaces of single probability measures. A sub-base for the weak product topology for the space $\mathscr{P}(\Delta_m)^{\mathscr{X}} = \prod_{\boldsymbol{x}\in\mathscr{X}} \mathscr{P}(\Delta_m)$, is given by sets of the form $B^W_{f,\epsilon,\boldsymbol{x}_0}(\{\boldsymbol{Q}_{\boldsymbol{x}} : \boldsymbol{x}\in\mathscr{X}\}) = \prod_{\boldsymbol{x}\in\mathscr{X}} \Delta^W_{f,\epsilon,\boldsymbol{x}_0}(\boldsymbol{Q}_{\boldsymbol{x}})$, where

$$\Delta^W_{f,\epsilon,\boldsymbol{x}_0}(\boldsymbol{Q}_{\boldsymbol{x}}) = \begin{cases} \mathscr{P}(\Delta_m), & \text{if } \boldsymbol{x}\in\mathscr{X}, \boldsymbol{x}\neq\boldsymbol{x}_0,\\ \left\{\boldsymbol{M}_{\boldsymbol{x}}\in\mathscr{P}(\Delta_m) : \left|\int_{\Delta_m} f\, d\boldsymbol{M}_{\boldsymbol{x}} - \int_{\Delta_m} f\, d\boldsymbol{Q}_{\boldsymbol{x}}\right| < \epsilon\right\}, & \text{if } \boldsymbol{x}\in\mathscr{X}, \boldsymbol{x}=\boldsymbol{x}_0, \end{cases}$$

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

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

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

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

$M_{\boldsymbol{x}}$ and $Q_{\boldsymbol{x}}$, respectively.

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

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

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

The following theorem, proved in Appendix B.1, summarizes the measurability results for the different versions of the proposed model.

Theorem 4.1. *Let $\mathscr{B}_1$, $\mathscr{B}_2$ and $\mathscr{B}_3$ be the Borel σ-field generated by the weak product topology, L_∞ product topology and L_∞ topology, respectively. If F is a* DMBPP, wDMBPP *or* θDMBPP, *defined on the appropriate measurable space $(\Omega, \mathscr{A})$, then the following mappings are measurable:*

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

4.4 The main properties

In this section, we establish basic properties of the proposed models. They include the characterization of the topological support, continuity, association structure and the asymptotic behavior of the posterior distribution under i.i.d. sampling from responses and predictors.

4.4.1 The support of the processes

Full support is a minimum requirement and almost a "necessary" property for a Bayesian model to be considered "nonparametric". In a fully nonparametric regression model setting, full support implies that the prior probability model assigns positive mass to any neighborhood of every collection of probability measures $\{Q_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$. Therefore, the definition of support strongly depends on the choice of a "distance" defining the basic neighborhoods. The results presented here are based on the weak product topology, L_∞ product topology, and L_∞ topology, and extend the ones provided by Barrientos et al. (2015a) for dependent Bernstein polynomials processes for data supported on compact intervals.

The following theorem, proved in Appendix B.2, provides sufficient conditions for $\mathscr{P}\left(\Delta_m\right)^{\mathscr{X}}$ and $\mathscr{D}\left(\Delta_m\right)^{\mathscr{X}}$ to be the support of DMBPPs under the weak product topology and the L_∞ product topology, respectively.

Theorem 4.2. *Let F be a $\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_z, \mathscr{V}, \mathscr{H})$, a $\theta\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$ or a $w\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_z, \mathscr{H})$. If F is defined such that:*

(i) for every $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L) \in \mathscr{X}^L$, $L \geq 1$, the joint distribution of $(\eta_j(\boldsymbol{x}_1, \cdot), \ldots, \eta_j(\boldsymbol{x}_L, \cdot))$, $j \geq 1$, has full support on $\mathbb{R}^L$,

(ii) for every $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L) \in \mathscr{X}^L$, $L \geq 1$, the joint distribution of $(\boldsymbol{z}_j(\boldsymbol{x}_1, \cdot), \ldots, \boldsymbol{z}_j(\boldsymbol{x}_L, \cdot))$, $j \geq 1$, has full support on $\mathbb{R}^{m \times L}$,

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

(iv) v_j, $j \geq 1$, has full support on $[0, 1]$,

(v) $\boldsymbol{\theta}_j$, $j \geq 1$, has full support on $\tilde{\Delta}_m$,

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

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

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

where $q_{\boldsymbol{x}}$ denotes the density function of $Q_{\boldsymbol{x}}$ w.r.t. Lebesgue measure. The following theorem, proved in Appendix B.3, provides sufficient conditions for $\tilde{\mathscr{D}}\left(\Delta_m\right)^{\mathscr{X}}$ to be in the support of DMBPPs under the L_∞ topology.

Theorem 4.3. *Let F be a* $\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_z, \mathscr{V}, \mathscr{H})$*, a* $\theta\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$ *or a* $w\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_z, \mathscr{H})$*. Assume that $\boldsymbol{x} \in \mathscr{X}$ contains only continuous components and that $\mathscr{X}$ is compact. If F is defined such that:*

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

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

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

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

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

(iv) v_j, $j \geq 1$, has full support on $[0, 1]$,

(v) $\boldsymbol{\theta}_j$, $j \geq 1$, has full support on $\tilde{\Delta}_m$,

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

An important consequence of the previous theorem is that the proposed processes can assign positive mass to arbitrarily small neighborhoods of any collection of probability measures $\{\boldsymbol{Q}_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \tilde{\mathscr{D}}\left(\Delta_m\right)^{\mathscr{X}}$, based on the supremum over the predictor space of Kullback-Leibler (KL) divergences between the predictor-dependent probability measures. The proof of the following theorem is given in Appendix B.4.

Theorem 4.4. *Let F be a* $\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_z, \mathscr{V}, \mathscr{H})$*, a* $\theta\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$ *or a* $w\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_z, \mathscr{H})$*. Under the same assumptions of Theorem 4.3, it follows that*

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

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

4.4.2 The continuity and association structure of the processes

The characteristics of the stochastic processes used in the definitions of a DMBPP determine important properties of the resulting model. Regardless of the specific choice of the stochastic processes used in its definition, the use of almost surely (a.s.) continuous stochastic processes ensures that DMBPP and $w\mathrm{DMBPP}$ have a.s. a limit. The following theorem is proved in Appendix B.5.

Theorem 4.5. *Let F be $\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_z, \mathscr{V}, \mathscr{H})$ or $w\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_z, \mathscr{H})$, defined such that $\mathscr{V}$ and $\mathscr{H}$ are sets of equicontinuous functions of $\boldsymbol{x}$, and for every $i \geq 1$, the stochastic processes η_i and $\boldsymbol{z}_i$ have a.s. continuous trayectories. Then for every $\{\boldsymbol{x}_l\}_{l=0}^{\infty}$, with $\boldsymbol{x}_l \in \mathscr{X}$, such that $\lim_{l\to\infty} \boldsymbol{x}_l = \boldsymbol{x}_0$, $F(\boldsymbol{x}, \cdot)$ has a.s. a limit with the total variation norm.*

An interesting property of the $\theta\mathrm{DMBPP}$ compared to the other version, and the general model, is that the use of a.s. continuous stochastic processes in the weights guarantees a.s. continuity of the 'single atoms' DMBPP. The following theorem is proved in Appendix B.6.

Theorem 4.6. *Let F be a $\theta\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$, defined such that $\mathscr{V}$ is a set of equicontinuous functions, and such that for every $j \geq 1$, the stochastic process η_j is a.s. continuous. Then, for every $\{\boldsymbol{x}_l\}_{l=0}^{\infty}$, with $\boldsymbol{x}_l \in \mathscr{X}$, such that $\lim_{l\to\infty} \boldsymbol{x}_l = \boldsymbol{x}_0$,*

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

That is, $F(\boldsymbol{x}_l, \cdot)$ converges a.s. in total variation norm to $F(\boldsymbol{x}_0, \cdot)$, when $\boldsymbol{x}_l \longrightarrow \boldsymbol{x}_0$.

The dependence structure of DMBPPs is completely determined by the association structure of the stochastic processes used in their definition. For instance, under mild conditions on the stochastic processes defining the DMBPPs, the correlation between the corresponding random measures approaches to one as the predictor values get closer. The proof of the following theorem is given in Appendix B.7.

Theorem 4.7. *Let F be a $\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_z, \mathscr{V}, \mathscr{H})$, a $\theta\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$ or a $w\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_z, \mathscr{H})$, defined such that $\mathscr{V}$ and $\mathscr{H}$ are sets of equicontinuous functions,* *and such that for every* $\{\boldsymbol{x}_l\}_{l=0}^{\infty}$*, with* $\boldsymbol{x}_l \in \mathscr{X}$*, such that* $\lim_{l\to\infty} \boldsymbol{x}_l = \boldsymbol{x}_0$*, we have* $\eta_j(\boldsymbol{x}_l, \cdot) \xrightarrow{\mathscr{L}} \eta_j(\boldsymbol{x}_0, \cdot)$ *and* $\boldsymbol{z}_j(\boldsymbol{x}_l, \cdot) \xrightarrow{\mathscr{L}} \boldsymbol{z}_j(\boldsymbol{x}_0, \cdot)$*, as* $l \to \infty$*,* $j \geq 1$*. Then, for every* $\boldsymbol{y} \in \tilde{\Delta}_m$*,*

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

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

If the stochastic processes defining the DMBPP and $w\mathrm{DMBPP}$ are such that the pairwise finite-dimensional distributions converge to the product of the corresponding marginal distributions as the Euclidean distance between the predictors grows larger, then under mild conditions the correlation between the corresponding random measures can approach zero. The following theorem, shows that under the assumptions previously discussed, the marginal covariance between the random measures is equal to the covariance between the conditional expectations of the random measures, given the degree of the MBP. The proof of the following theorem is given in Appendix B.8.

Theorem 4.8. *Let F be a* $\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_z, \mathscr{V}, \mathscr{H})$ *or a* $w\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_z, \mathscr{H})$*, defined such that* $\mathscr{V}$ *and* $\mathscr{H}$ *are sets of equicontinuous functions and such that there exists a constant* $\gamma > 0$ *such that if* $(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathscr{X}^2$ *and* $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| > \gamma$*, then*

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

for every $A_1, A_2 \in \mathscr{B}(\mathbb{R})$*, and*

$$Cov\left[\mathbb{I}_{\{\boldsymbol{z}_j(\boldsymbol{x}_1,\cdot)\in A_3\}}, \mathbb{I}_{\{\boldsymbol{z}_j(\boldsymbol{x}_2,\cdot)\in A_4\}}\right] = 0,$$

for every $A_3, A_4 \in \mathscr{B}(\mathbb{R}^m)$*,* $j \geq 1$*. Assume also that for every* $(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathscr{X}^2$ *and for every sequence* $\{(\boldsymbol{x}_{1l}, \boldsymbol{x}_{2l})\}_{l=1}^{\infty}$*, with* $(\boldsymbol{x}_{1l}, \boldsymbol{x}_{2l}) \in \mathscr{X}^2$ *and such that* $\lim_{l\to\infty}(\boldsymbol{x}_{1l}, \boldsymbol{x}_{2l}) = (\boldsymbol{x}_1, \boldsymbol{x}_2)$*, we have that* $(\eta_j(\boldsymbol{x}_{1l}, \cdot), \eta_j(\boldsymbol{x}_{2l}, \cdot)) \xrightarrow{\mathscr{L}} (\eta_j(\boldsymbol{x}_1, \cdot), \eta_j(\boldsymbol{x}_2, \cdot))$*, and*

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

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

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

with

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

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

From Theorem 4.8 it is easy to see that if DMBPP or $w\mathrm{DMBPP}$ are specified such that the marginal distribution of k is degenerated, then the correlation between the corresponding random measures goes to zero, since $\lim_{l\to+\infty} Cov\left[F(\boldsymbol{x}_{1l},\cdot)(B_{\boldsymbol{y}}), F(\boldsymbol{x}_{2l},\cdot)(B_{\boldsymbol{y}})\right] = 0$. For $\theta\mathrm{DMBPP}$ the correlation between the associated random measures when the predictor values are far apart reaches a different limit. In such case it is difficult to establish conditions on the prior specification ensuring that the limit is zero. The proof of the following theorem is given in Appendix B.9.

Theorem 4.9. *Let F be a $\theta\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$. Asume that $\mathscr{V}$ is a set of equicontinuous functions and that there exists a constant $\gamma > 0$, such that if $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathscr{X}$ and $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| > \gamma$, then $Cov\left[\mathbb{I}_{\{\eta_j(\boldsymbol{x}_1,\cdot)\in A_1\}}, \mathbb{I}_{\{\eta_j(\boldsymbol{x}_2,\cdot)\in A_2\}}\right] = 0$, for every $A_1, A_2 \in \mathscr{B}(\mathbb{R})$, $j \geq 1$. Assume also that for every $(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathscr{X}^2$ and for every sequence $\{(\boldsymbol{x}_{1l}, \boldsymbol{x}_{2l})\}_{l=1}^{\infty}$, with $(\boldsymbol{x}_{1l}, \boldsymbol{x}_{2l}) \in \mathscr{X}^2$, such that $\lim_{l\to\infty}(\boldsymbol{x}_{1l}, \boldsymbol{x}_{2l}) = (\boldsymbol{x}_1, \boldsymbol{x}_2)$, we have $(\eta_j(\boldsymbol{x}_{1l},\cdot), \eta_j(\boldsymbol{x}_{2l},\cdot)) \xrightarrow{\mathscr{L}} (\eta_j(\boldsymbol{x}_1,\cdot), \eta_j(\boldsymbol{x}_2,\cdot))$, $j \geq 1$, as $l \to \infty$. Then, for every $\boldsymbol{y} \in \Delta_m$,*

$$\begin{aligned}
\lim_{l\to\infty} & Cov\left[F(\boldsymbol{x}_{1l},\cdot)(B_{\boldsymbol{y}}), F(\boldsymbol{x}_{2l},\cdot)(B_{\boldsymbol{y}})\right] = \\
&\sum_{k_1=1}^{\infty} P\{\omega\in\Omega : k(\omega) = k_1\} \sum_{\mathbf{j}_1,\, \mathbf{j}_2\in\mathcal{H}_{k_1,m}} \bar{M}(\mathbf{j}_1, \mathbf{j}_2 \mid k_1+m-1, \boldsymbol{y}) \\
&\times \sum_{j=1}^{\infty} E\left[w_j(\boldsymbol{x}_1,\cdot)\right] E[w_j(\boldsymbol{x}_2,\cdot)]\, Cov\left[\mathbb{I}_{\left\{\boldsymbol{\theta}_j(\cdot)\in A_{\mathbf{j}_1,k_1}\right\}}, \mathbb{I}_{\left\{\boldsymbol{\theta}_j(\cdot)\in A_{\mathbf{j}_2,k_1}\right\}}\right] \\
&+ Cov\left[E\left\{F(\boldsymbol{x}_1,\cdot)(B_{\boldsymbol{y}}) \mid k(\cdot)\right\}, E\left\{F(\boldsymbol{x}_2,\cdot)(B\boldsymbol{y}) \mid k(\cdot)\right\}\right],
\end{aligned}$$

with

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

where $B_{\boldsymbol{y}} = [0,y_1]\times\ldots\times[0,y_m]$*,* $A_{\mathbf{j},k} = [0,j_1/k]\times\ldots\times[0,j_m/k]$*,* $G_{0,\boldsymbol{x}}$ *is the marginal probability measure of* $\boldsymbol{\theta}_j(\boldsymbol{x},\cdot)$ *defined on* $\tilde{\Delta}_m$*, and* $\bar{M}(\mathbf{j},\mathbf{j}_1 \mid k+m-1,\boldsymbol{y}) = \mathrm{Mult}(\mathbf{j} \mid k+m-1,\boldsymbol{y})\times\mathrm{Mult}(\mathbf{j}_1 \mid k+m-1,\boldsymbol{y})$.

Finally, although the trajectories of the DMBPP and wDMBPP are a.s. pseudo-continuous only, the autocorrelation function of all versions of the model are continuous under mild conditions on the elements defining the processes. The proof of the following theorem is given in Appendix B.10.

Theorem 4.10. *Let* F *be a* $\mathrm{DMBPP}(\lambda,\boldsymbol{\Psi}_\eta,\boldsymbol{\Psi}_z,\mathscr{V},\mathscr{H})$*, a* $\theta\mathrm{DMBPP}(\lambda,\boldsymbol{\Psi}_\eta,\mathscr{V},\boldsymbol{\Psi}_{\boldsymbol{\theta}})$ *or a* $w\mathrm{DMBPP}(\lambda,\boldsymbol{\Psi}_v,\boldsymbol{\Psi}_z,\mathscr{H})$*, defined such that* $\mathscr{V}$ *and* $\mathscr{H}$ *are sets of equicontinuous functions. Assume that for every* $(\boldsymbol{x}_1,\boldsymbol{x}_2)\in\mathscr{X}^2$ *and for every sequence* $\{(\boldsymbol{x}_{1l},\boldsymbol{x}_{2l})\}_{l=1}^{\infty}$*, with* $(\boldsymbol{x}_{1l},\boldsymbol{x}_{2l})\in\mathscr{X}^2$*, such that* $\lim_{l\to\infty}(\boldsymbol{x}_{1l},\boldsymbol{x}_{2l}) = (\boldsymbol{x}_1,\boldsymbol{x}_2)$*, we have that*

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

and

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

as $l\to\infty$*,* $j\geq 1$*. Then, for every* $\boldsymbol{y}\in\tilde{\Delta}_m$*,*

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

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

4.4.3 The asymptotic behavior of the posterior distribution

In this section we study the asymptotic behavior of the posterior distribution of DMBPPs. Here we assume that we observe a random sample $(\boldsymbol{y}_i, \boldsymbol{x}_i)$, $i = 1, \ldots, n$. As is common in regression settings, we assume that the predictor vector $\boldsymbol{x}_i$ contains only exogenous covariates. Notice that the exogeneity assumption allows us to focusing on the conditional density estimation problem, regardless the data generating mechanism of the predictors, that is, if they are randomly generated or fixed by design (see, e.g. Barndorff-Nielsen, 1973, 1978; Florens et al., 1990). Let Q be the true probability measure generating the predictors, with density w.r.t. a corresponding σ-additive measure denoted by q. By the exogeneity assumption, the true probability model for the response variable and predictors takes the form $h_0(\boldsymbol{y}, \boldsymbol{x}) = q(\boldsymbol{x})q_0(\boldsymbol{y} \mid \boldsymbol{x})$, where both q and $\{q_0(\cdot \mid \boldsymbol{x}) : \boldsymbol{x} \in \mathscr{X}\}$ are in free variation, with $q_0(\boldsymbol{y} \mid \boldsymbol{x})$ denoting a conditional density defined on Δ_m, $\boldsymbol{x} \in \mathscr{X}$. The proof of the following theorem is given in Appendix B.11.

Theorem 4.11. *Let F be a* $\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_z, \mathscr{V}, \mathscr{H})$*, a* $\theta\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$ *or a* $w\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_z, \mathscr{H})$*. If the assumptions of Theorem 4.3 are satisfied, then the posterior distribution associated with the random joint distribution induced by the corresponding* DMBPP *model, $h(\cdot)(\boldsymbol{y}, \boldsymbol{x}) = q(\boldsymbol{x})f(\boldsymbol{x}, \cdot)(\boldsymbol{y})$, where q is the density generating the predictors, is weakly consistent at any joint distribution of the form $h_0(\boldsymbol{y}, \boldsymbol{x}) = q(\boldsymbol{x})q_0(\boldsymbol{y} \mid \boldsymbol{x})$, where $\{q_0(\cdot \mid \boldsymbol{x}) : \boldsymbol{x} \in \mathscr{X}\} \in \tilde{\mathscr{D}}(\Delta_m)^{\mathscr{X}}$.*

Although Theorem 4.11 assumes that $\boldsymbol{x}$ contains only continuous predictors, a similar result can be obtained when $\boldsymbol{x}$ contains only predictors with finite support (e.g., categorical, ordinal and discrete predictors) or mixed continuous and predictors with finite support.

Chapter 5

Conclusions and future work

In this dissertation, we have addressed three different topics in the context of Bayesian nonparametric (BNP) models for single and predictor–dependent probability measures. This Chapter summarizes the main conclusions of this dissertation and gives some directions of future work.

5.1 Conclusions

In Chapter 2, we have proposed a class of MBP priors for density estimation for data supported on a m–dimensional simplex, which has optimal posterior convergence rate. We also showed that this class, being a modification of the class of MBP priors proposed by Barrientos et al. (2015a), preserves appealing properties of Bernstein polynomials such as approximation of functions that belong to a Hölder class with α regularity, $\alpha \in (0, 1]$.

In Chapter 3, we have compared two BNP models for density estimation for data supported on a compact interval using simulated and real data: a posterior convergence rate-optimal and a rate-suboptimal model. The results show that the rate-suboptimal model can outperform the rate-optimal model when finite samples are considered. Furthermore, the results show that poor inferences can be obtained from the rate-optimal model, even for relatively large sample sizes and when the conditions to obtain the asymptotic results hold. The results reported in Chapter 3 strongly suggest that the suboptimal behavior w.r.t. the posterior concentration rate of a BNP model should not imply the "death" of the model and their merits should be evaluated on a case-by-case basis.

In Chapter 4, we have proposed a novel class of probability models for sets of predictor-dependent probability distributions supported on simplex spaces. The proposal corresponds to an extension of the dependent univariate Bernstein polynomial process proposed by Barrientos et al. (2012) and is based on the modified class of MBP proposed by Barrientos et al. (2015a).

We showed that the proposed model class, called dependent multivariate Bernstein polynomial process (DMBPP) has appealing theoretical properties such as full support, continuity, well behaved correlation function and consistent posterior distribution. We also considered two simplified versions of the model, where only weights or support points of the dependent stick–breaking process depend on predictors, and proved they have the same support and posterior consistency properties as the general model, and share continuity and well behaved correlation function properties from the general model. Since all the versions of the model can be represented by a dependent Dirichlet process mixture of dirichlet densities, their use becomes attractive in the context of regression modeling for compositional data.

5.2 Future work

Probability models for sets of predictor-dependent probability distributions supported on simplex spaces could be developed based on the model proposed in Chapter 2. However, the results obtained In Chapter 3 suggest that the models proposed in Chapter 4 would outperform such models from a small sample point of view. This is the subject of future research.

A rather weak consistency result was established for the model proposed in Chapter 4. Specifically, we showed that the posterior distribution associated with the random joint distribution for predictor and responses, induced by a DMBPP model, is weakly consistent at any joint distribution with the same marginal distribution generating the predictors. The study of the asymptotic behavior of the posterior distribution under stronger topologies is part of ongoing research. Specifically, the study of the behavior of the models under a similar construction to the one used in Theorem 12 by Barrientos et al. (2015b) is being currently considered. This is based on the use of specific probit stick-breaking processes (Pati et al., 2013, Section 5).

The study of practicable special cases of the models proposed in Chapter 4 and its computational implementation is subject of ongoing research. User-friendly functions implementing these methods will be written in compiled language and incorporated into the R library DPpackage (Jara, 2007; Jara et al., 2011).

The proof of Theorem 4.3 in Chapter 4 involved a novel class of polynomial functions defined on $\Delta_m \times [0,1]^p$. Specifically, the proof of Lemma B.3 involved a class of polynomials that can approximate uniformly the cumulative distribution function and the density function of absolutely continuous, w.r.t. Lebesgue measure, probability measures defined on $\Delta_m \times [0,1]^p$, and which admit continuous density functions. This novel class of functions can be used to define novel BNP for joint probability measures for data supported on $\Delta_m \times [0,1]^p$. This class can be also used to define BNP regression models for data supported on Δ_m and with predictors supported on a unit-hypercube, or for response vectors supported in a unit-hypercube and with predictors supported on Δ_m. This is the subject of ongoing research.

Appendix A

Supplementary material for Chapter 2

A.1 Proof of the behavior of the NB distribution

It is necessary to prove that there exist positive constants B_1, B_2, c_1 and c_2, such that, for all $k \geq 1$,

$$B_1 e^{-c_1 k} \leq \frac{\Gamma(k+r)}{\Gamma(k+1)\Gamma(r)} p^k (1-p)^r \leq B_2 e^{-c_2 k}, \tag{A.1}$$

where $r > 0$ and $0 < p < 1$. The proof is constructive. In what follows $\rho(k)$ denotes the probability mass function from a negative binomial distribution with parameters r and p.

We will first consider the case where r is an integer. Here, we will first show that for a sufficiently large k, condition (B.11) holds. Notice that $\lim_{k \longrightarrow \infty} \left[\log \Gamma(k+r) - \log \Gamma(k+1)\right]/k = 0$. Therefore, for every $\epsilon > 0$, there exists $k_0 \in \mathbb{N}$, such that, for every $k > k_0$, it follows that

$$-\epsilon < \frac{\log \Gamma(k+r) - \log \Gamma(k+1)}{k} < \epsilon.$$

Take $c_1 > -\log p$ and $c_2 > 0$ such that $c_1 + c_2 < -2\log p$. Now, by noticing that $\log p + c_1 <$

$-\log p - c_2$ and taking $\epsilon = \log p + c_1$, it follows that

$$-k \log p - kc_1 < \log \Gamma(k+r) - \log \Gamma(k+1) < -k \log p - kc_2. \tag{A.2}$$

Thus, by adding $r \log(1-p) - \log \Gamma(r)$ and applying exponential function to expression (A.2), it follows that, for every $k > k_0$,

$$\frac{(1-p)^r}{\Gamma(r)} e^{-c_1 k} < \rho(k) < \frac{(1-p)^r}{\Gamma(r)} e^{-c_2 k},$$

which in turn shows that condition (B.11) holds for every $k \geq k_0$. The proof is completed by noticing that since

$$0 < \rho(k) < 1,$$

there exists

$$B_1 < (1-p)^r / \Gamma(r),$$

and

$$B_2 > (1-p)^r / \Gamma(r),$$

such that condition (B.11) holds for every $k \geq 1$.

A.2 Additional results for the simulation study

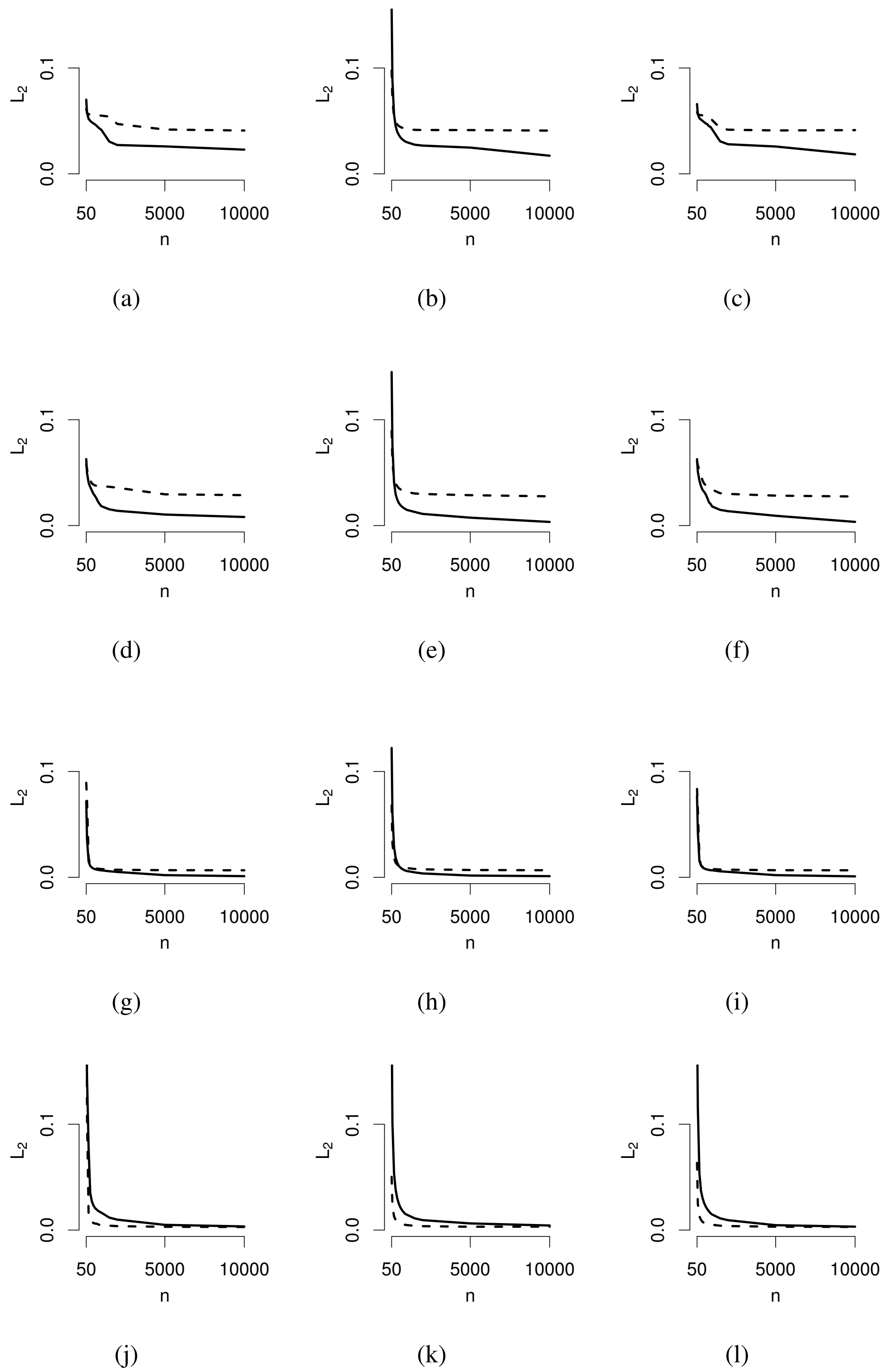
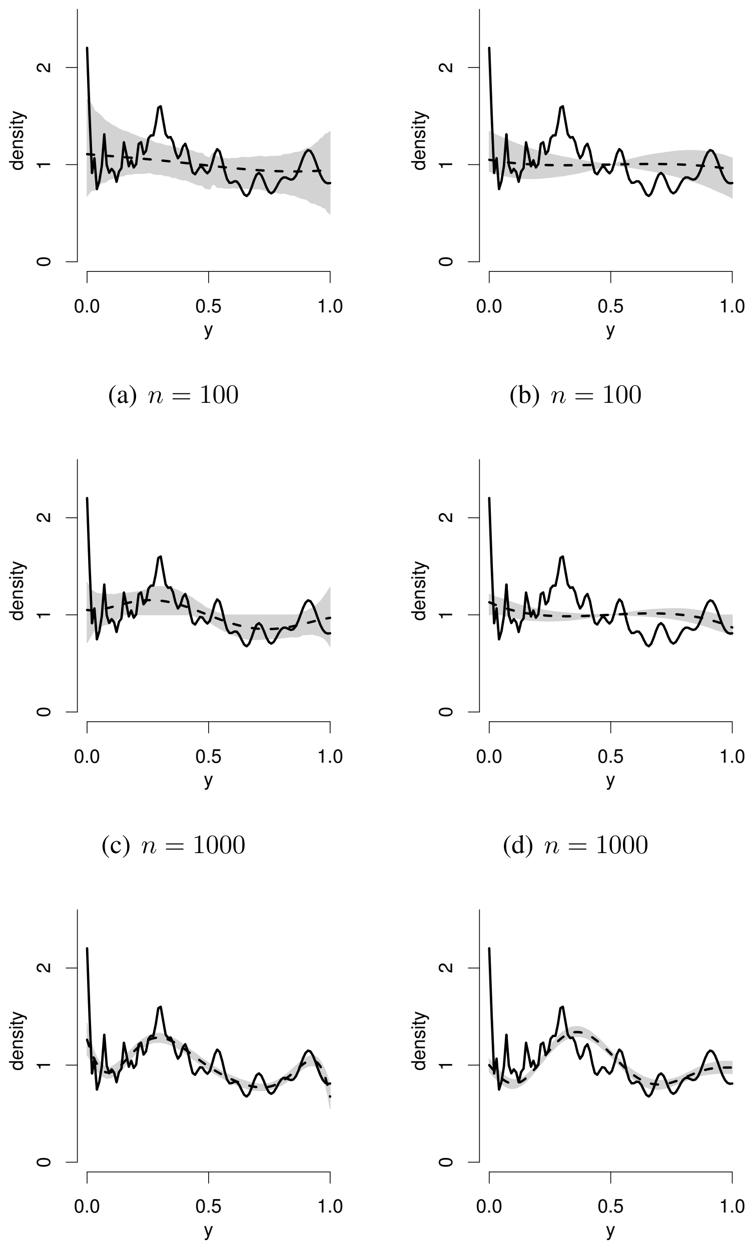


Figure A.1: Simulated Data: Posterior mean of the L_2 distance between the true density and the random probability measure as a function of the sample size. The results for BDP and DPMMB are shown as a solid and dashed line, respectively. Panel (a) - (c), (d) - (f), (g) - (i), and (j) - (l) display the results under simulation Scenario I, II, III, and IV, respectively. Panel (a), (d), (g) and (j), (b), (e), (h) and (k), and (c), (f), (i) and (l), display the results under Prior I, II and III, respectively.

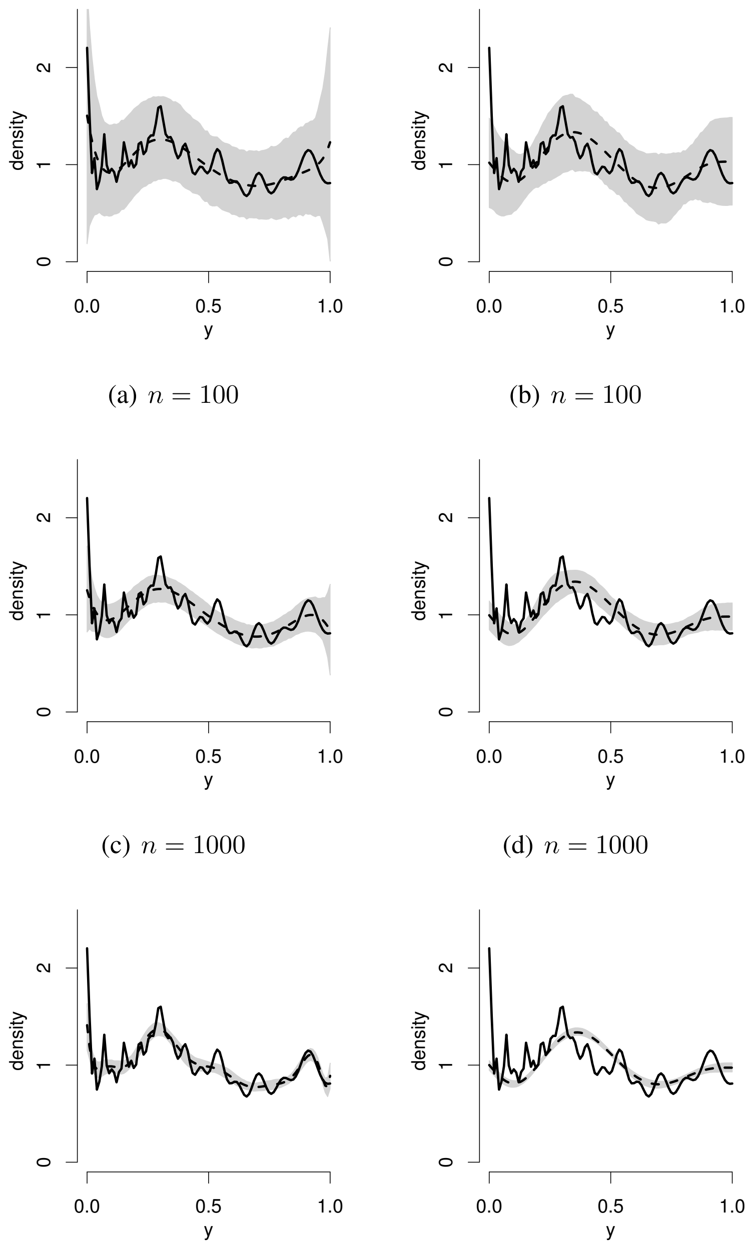


(a) $n = 100$ (b) $n = 100$

(c) $n = 1000$ (d) $n = 1000$

(e) $n = 10000$ (f) $n = 10000$

Figure A.2: Simulated Data: Posterior mean (dashed line), point-wise 95% HPD band (grey area), and true model (continuous line) under Scenario I and Prior I, for selected sample sizes. Panel (a), (c) and (e) display the results for BDP model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively. Panel (b), (d) and (f) display the results for DPMMB model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively.

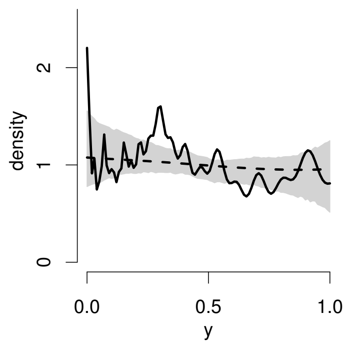


(a) $n = 100$ (b) $n = 100$

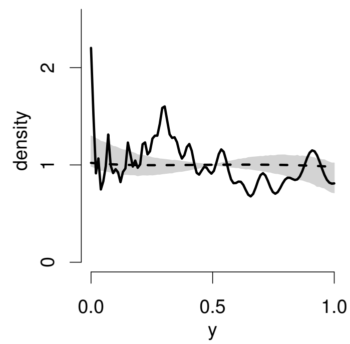
(c) $n = 1000$ (d) $n = 1000$

(e) $n = 10000$ (f) $n = 10000$

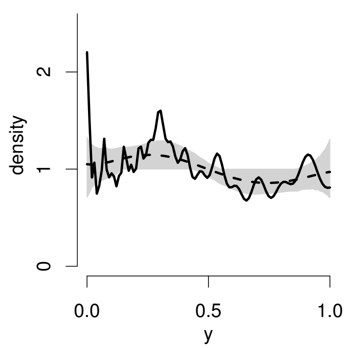
Figure A.3: Simulated Data: Posterior mean (dashed line), point-wise 95% HPD band (grey area), and true model (continuous line) under Scenario I and Prior II, for selected sample sizes. Panel (a), (c) and (e) display the results for BDP model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively. Panel (b), (d) and (f) display the results for DPMMB model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively.



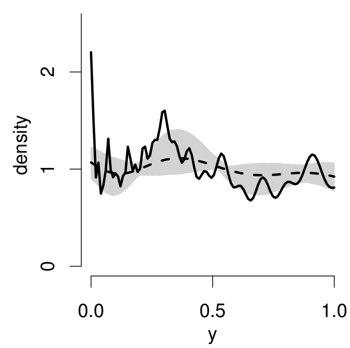
(a) $n = 100$



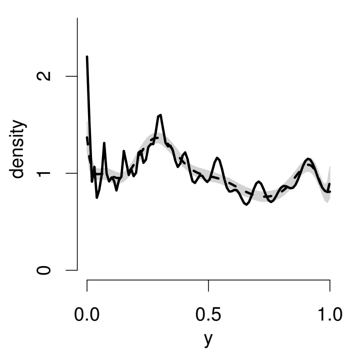
(b) $n = 100$



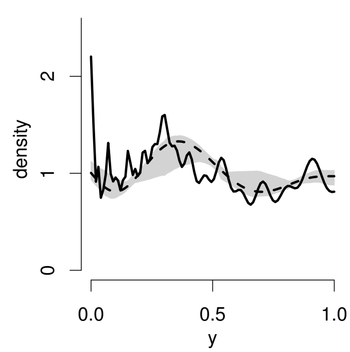
(c) $n = 1000$



(d) $n = 1000$

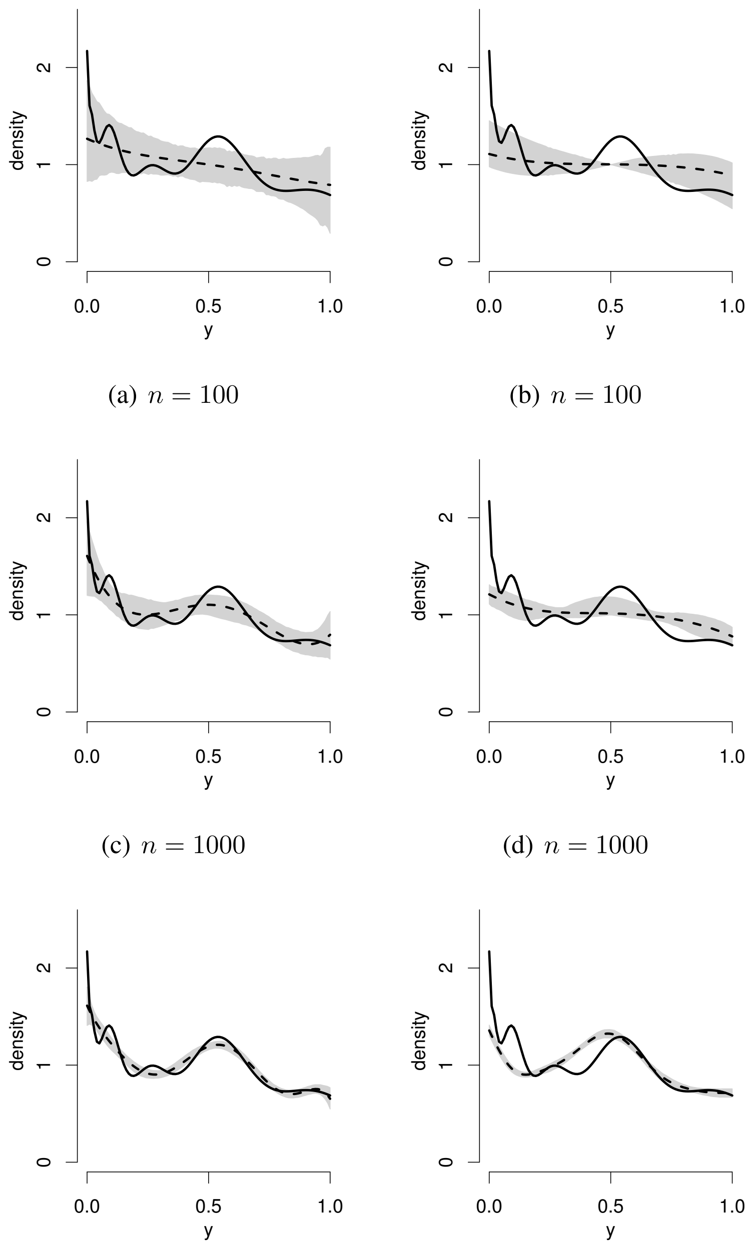


(e) $n = 10000$



(f) $n = 10000$

Figure A.4: Simulated Data: Posterior mean (dashed line), point-wise 95% HPD band (grey area), and true model (continuous line) under Scenario I and Prior I, for selected sample sizes. Panel (a), (c) and (e) display the results for BDP model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively. Panel (b), (d) and (f) display the results for DPMMB model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively.

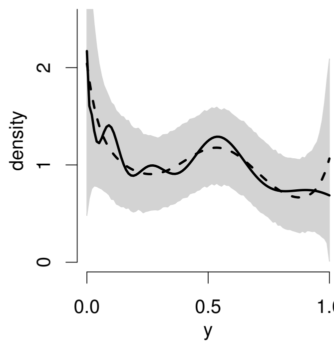


(a) $n = 100$ (b) $n = 100$

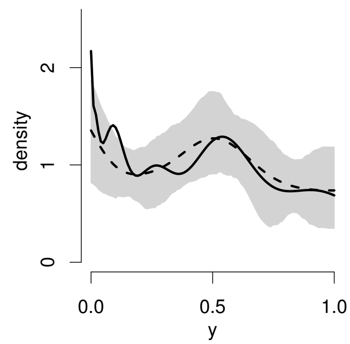
(c) $n = 1000$ (d) $n = 1000$

(e) $n = 10000$ (f) $n = 10000$

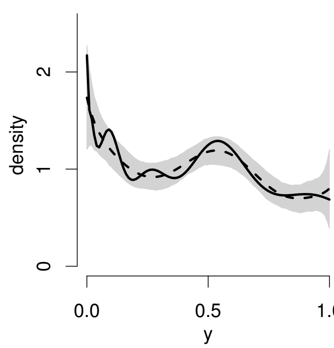
Figure A.5: Simulated Data: Posterior mean (dashed line), point-wise 95% HPD band (grey area), and true model (continuous line) under Scenario II and Prior I, for selected sample sizes. Panel (a), (c) and (e) display the results for BDP model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively. Panel (b), (d) and (f) display the results for DPMMB model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively.



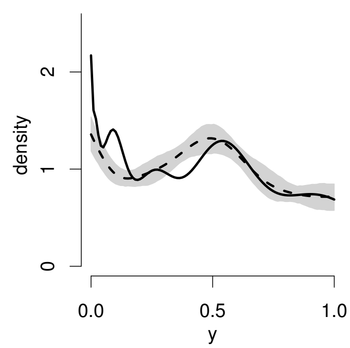
(a) $n = 100$



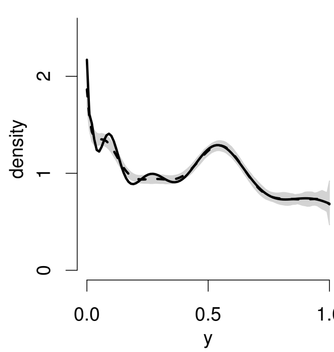
(b) $n = 100$



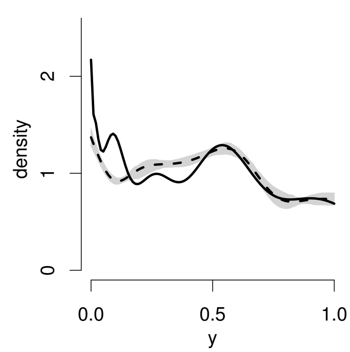
(c) $n = 1000$



(d) $n = 1000$

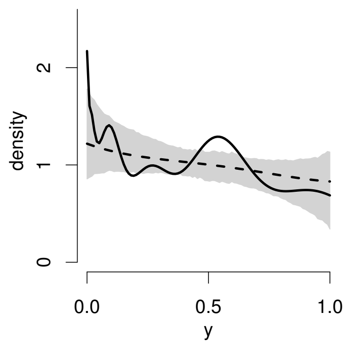


(e) $n = 10000$

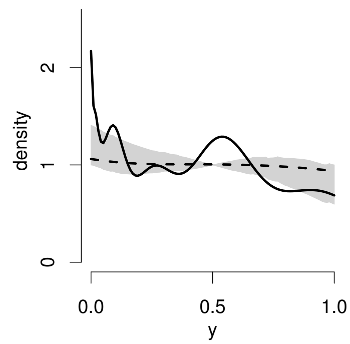


(f) $n = 10000$

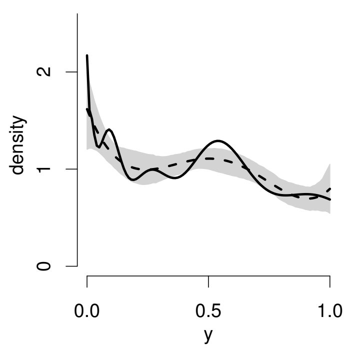
Figure A.6: Simulated Data: Posterior mean (dashed line), point-wise 95% HPD band (grey area), and true model (continuous line) under Scenario II and Prior II, for selected sample sizes. Panel (a), (c) and (e) display the results for BDP model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively. Panel (b), (d) and (f) display the results for DPMMB model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively.



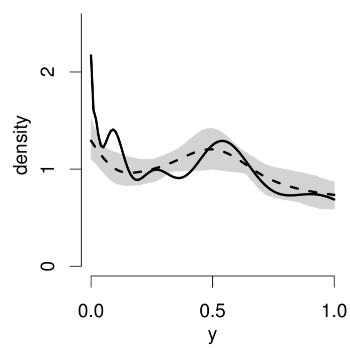
(a) $n = 100$



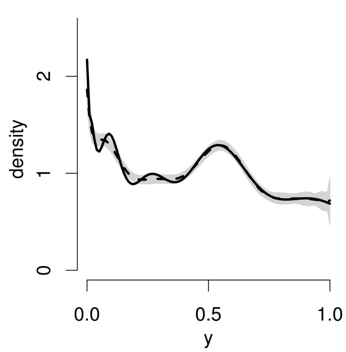
(b) $n = 100$



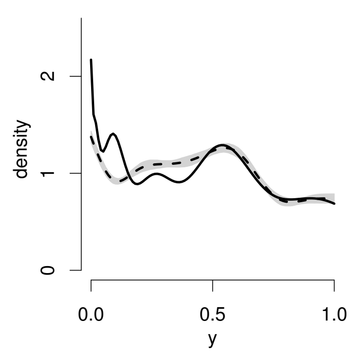
(c) $n = 1000$



(d) $n = 1000$

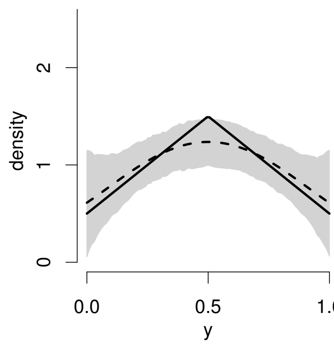


(e) $n = 10000$

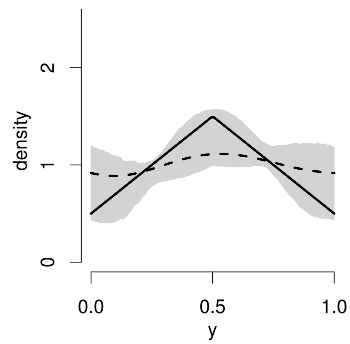


(f) $n = 10000$

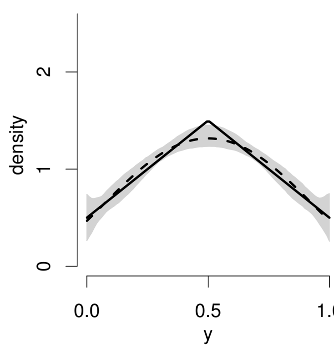
Figure A.7: Simulated Data: Posterior mean (dashed line), point-wise 95% HPD band (grey area), and true model (continuous line) under Scenario II and Prior I, for selected sample sizes. Panel (a), (c) and (e) display the results for BDP model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively. Panel (b), (d) and (f) display the results for DPMMB model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively.



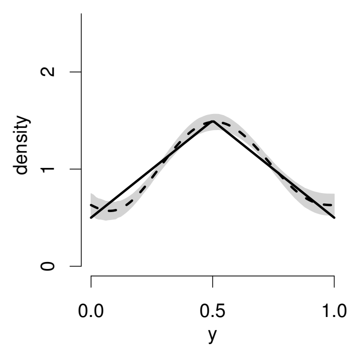
(a) $n = 100$



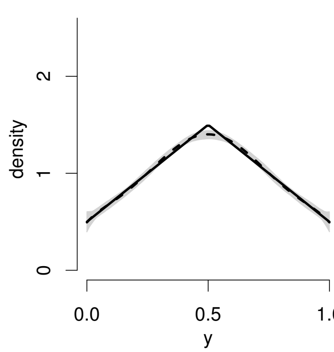
(b) $n = 100$



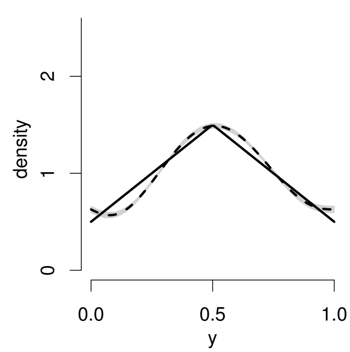
(c) $n = 1000$



(d) $n = 1000$

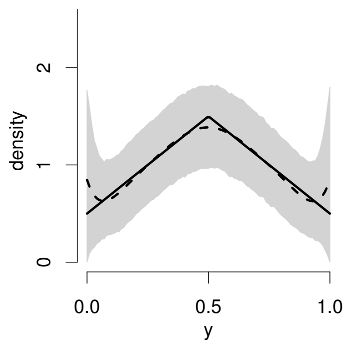


(e) $n = 10000$

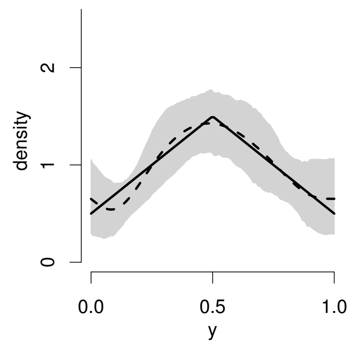


(f) $n = 10000$

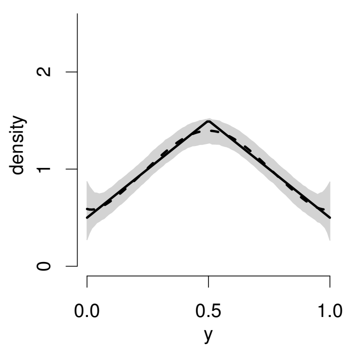
Figure A.8: Simulated Data: Posterior mean (dashed line), point-wise 95% HPD band (grey area), and true model (continuous line) under Scenario III and Prior I, for selected sample sizes. Panel (a), (c) and (e) display the results for BDP model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively. Panel (b), (d) and (f) display the results for DPMMB model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively.



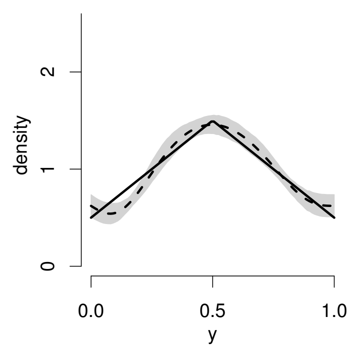
(a) $n = 100$



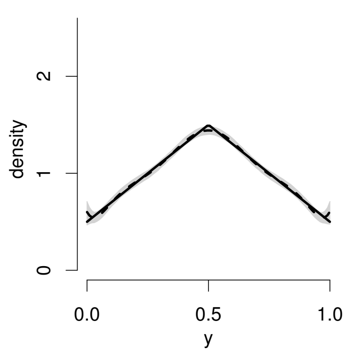
(b) $n = 100$



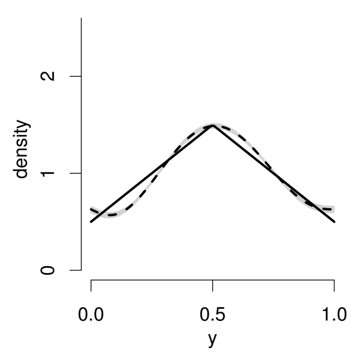
(c) $n = 1000$



(d) $n = 1000$

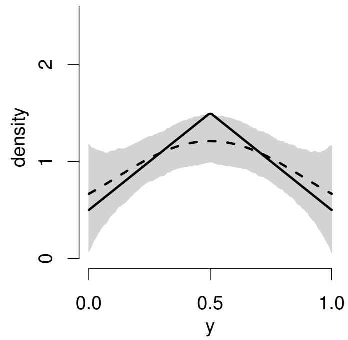


(e) $n = 10000$

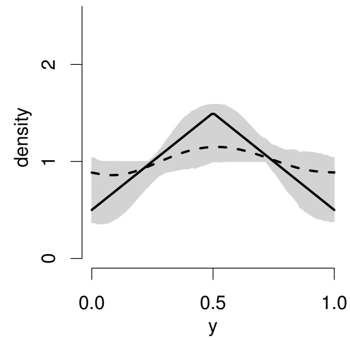


(f) $n = 10000$

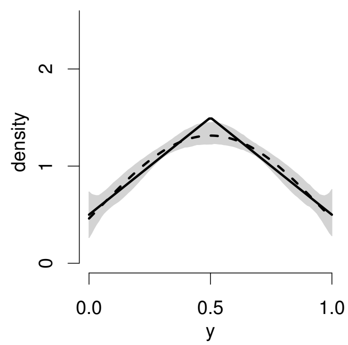
Figure A.9: Simulated Data: Posterior mean (dashed line), point-wise 95% HPD band (grey area), and true model (continuous line) under Scenario III and Prior II, for selected sample sizes. Panel (a), (c) and (e) display the results for BDP model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively. Panel (b), (d) and (f) display the results for DPMMB model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively.



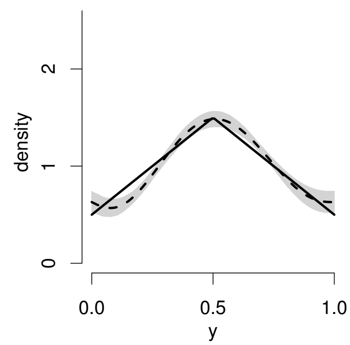
(a) $n = 100$



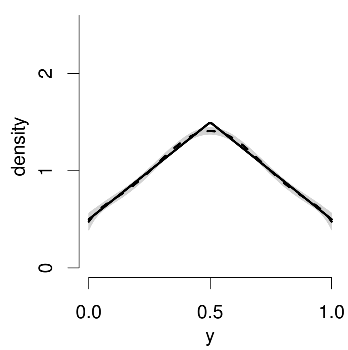
(b) $n = 100$



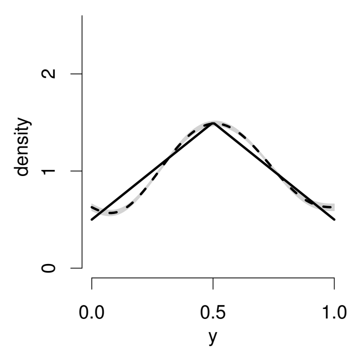
(c) $n = 1000$



(d) $n = 1000$

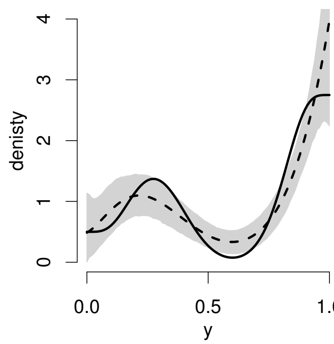


(e) $n = 10000$

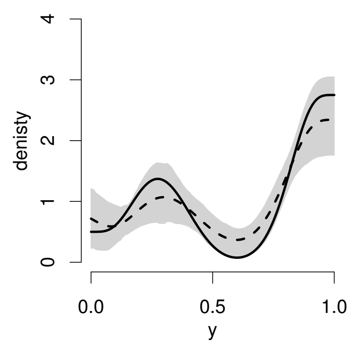


(f) $n = 10000$

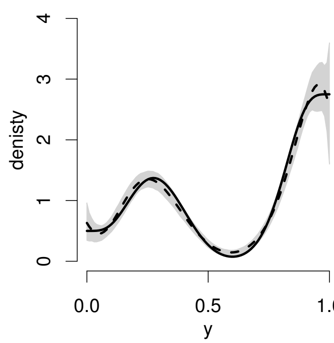
Figure A.10: Simulated Data: Posterior mean (dashed line), point-wise 95% HPD band (grey area), and true model (continuous line) under Scenario III and Prior I, for selected sample sizes. Panel (a), (c) and (e) display the results for BDP model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively. Panel (b), (d) and (f) display the results for DPMMB model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively.



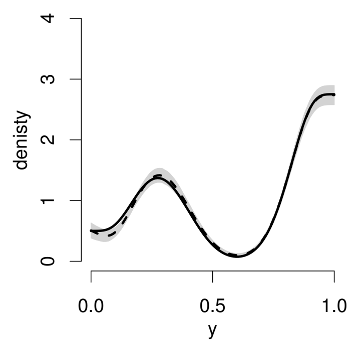
(a) $n = 100$



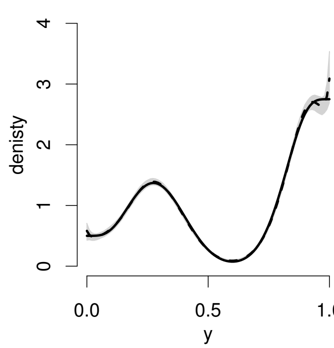
(b) $n = 100$



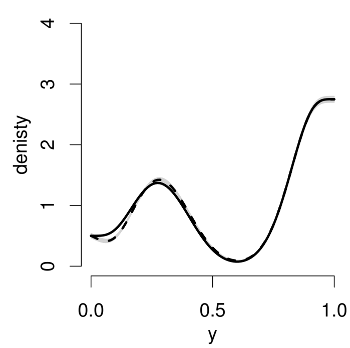
(c) $n = 1000$



(d) $n = 1000$

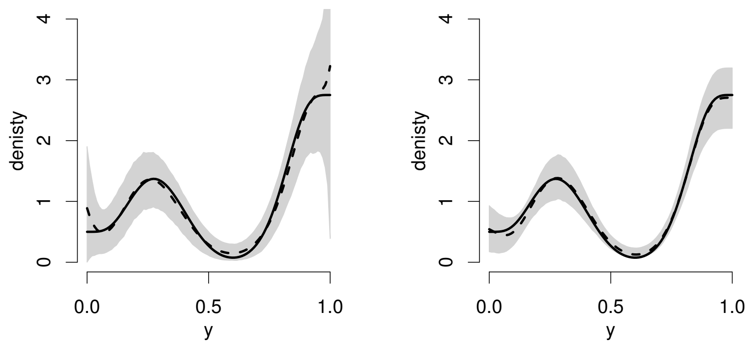


(e) $n = 10000$

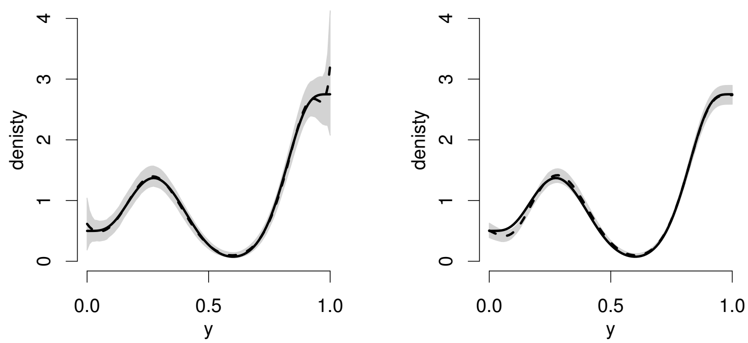


(f) $n = 10000$

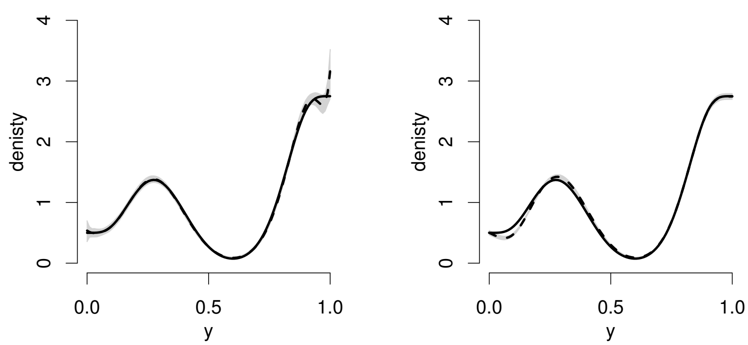
Figure A.11: Simulated Data: Posterior mean (dashed line), point-wise 95% HPD band (grey area), and true model (continuous line) under Scenario IV and Prior I, for selected sample sizes. Panel (a), (c) and (e) display the results for BDP model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively. Panel (b), (d) and (f) display the results for DPMMB model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively.



(a) $n = 100$ (b) $n = 100$

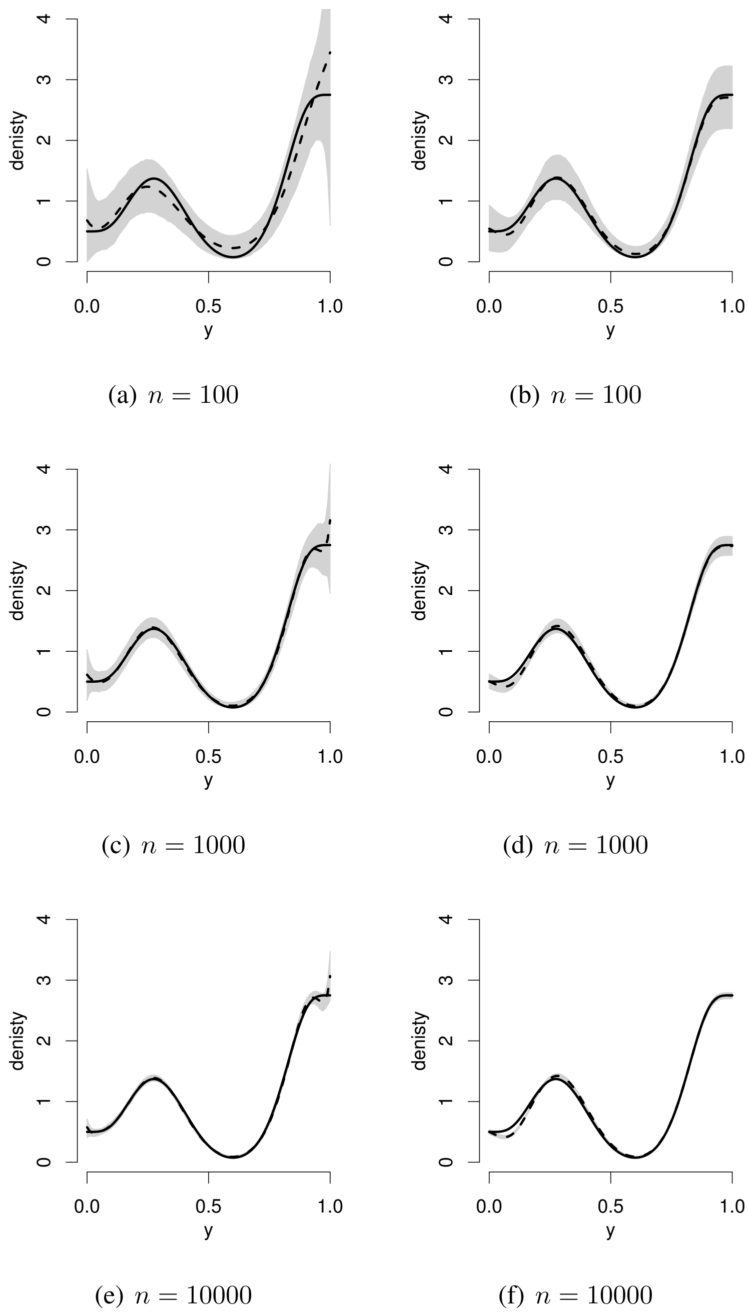


(c) $n = 1000$ (d) $n = 1000$



(e) $n = 10000$ (f) $n = 10000$

Figure A.12: Simulated Data: Posterior mean (dashed line), point-wise 95% HPD band (grey area), and true model (continuous line) under Scenario IV and Prior II, for selected sample sizes. Panel (a), (c) and (e) display the results for BDP model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively. Panel (b), (d) and (f) display the results for DPMMB model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively.



(a) $n = 100$ (b) $n = 100$

(c) $n = 1000$ (d) $n = 1000$

(e) $n = 10000$ (f) $n = 10000$

Figure A.13: Simulated Data: Posterior mean (dashed line), point-wise 95% HPD band (grey area), and true model (continuous line) under Scenario IV and Prior I, for selected sample sizes. Panel (a), (c) and (e) display the results for BDP model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively. Panel (b), (d) and (f) display the results for DPMMB model for sample size $n = 100$, $n = 1000$ and $n = 10000$, respectively.

Appendix B

Supplementary Material for Chapter 3

B.1 Proof of Theorem 4.1

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

- $\mathbf{T} = \{T_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$, where $T_{\boldsymbol{x}}(\omega, \cdot) = \sum_{j=1}^{\infty} w_j(\boldsymbol{x}, \omega)\delta_{\boldsymbol{\theta}_j(\boldsymbol{x},\omega)}(\cdot)$, where $w_j(\boldsymbol{x}, \omega)$ and $\boldsymbol{\theta}_j(\boldsymbol{x}, \omega)$ are define as in Definition 4.3.

- $\mathbf{T}^\theta = \{T^\theta_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$, where $T^\theta_{\boldsymbol{x}}(\omega, \cdot) = \sum_{j=1}^{\infty} w_j(\boldsymbol{x}, \omega)\delta_{\boldsymbol{\theta}_j(\omega)}(\cdot)$, where $w_j(\boldsymbol{x}, \omega)$ and $\boldsymbol{\theta}_j(\omega)$ are define as in Definition 4.4.

- $\mathbf{T}^w = \{T^w_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$, where $T^w_{\boldsymbol{x}}(\omega, \cdot) = \sum_{j=1}^{\infty} w_j(\omega)\delta_{\boldsymbol{\theta}_j(\boldsymbol{x},\omega)}(\cdot)$, where $w_j(\omega)$ and $\boldsymbol{\theta}_j(\boldsymbol{x}, \omega)$ are define as in Definition 4.5.

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

$$S(k_0, \mathcal{Q}) := \{H(k_0, \boldsymbol{Q}_{\boldsymbol{x}}) : \boldsymbol{x} \in \mathscr{X}\}, \tag{B.1}$$

where $k_0 \in \mathbb{N}$, $\mathcal{Q} = \{\boldsymbol{Q}_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}}$ and $H(k_0, \boldsymbol{Q}_{\boldsymbol{x}})$ is the probability measure associated to the Bernstein polynomial of degree k_0 of the measure $\boldsymbol{Q}_{\boldsymbol{x}}$. As stated in (Barrientos et al., 2015b), *F* can thus be expressed as $S(k, \mathbf{T})$, $S(k, \mathbf{T}^\theta)$ or $S(k, \mathbf{T}^w)$, when *F* corresponds to $\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_{\boldsymbol{z}}, \mathscr{V}, \mathscr{H})$, $\theta\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$ and $w\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_{\boldsymbol{z}}, \mathscr{H})$, respectively. Since $\mathbf{T}$, $\mathbf{T}^\theta$ and $\mathbf{T}^w$ are well-defined stochastic processes, to prove the measurability of *F*, it suffices to prove the measurability of S which is prove by showing that mapping S is continuous. For this, it is necessary to consider some topologies in the space where the mapping is valued and defined. This topologies and spaces are described below.

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

$$\tilde{\Delta}^{L_\infty}_{\epsilon,N}(\mathcal{Q}) = \left\{\{M_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}} : \max_{\mathbf{j} \in \mathcal{H}^0_{N,m}} \sup_{\boldsymbol{x} \in \mathscr{X}} |M_{\boldsymbol{x}}(A_{\mathbf{j},N}) - \boldsymbol{Q}_{\boldsymbol{x}}(A_{\mathbf{j},N})| < \epsilon\right\}, \tag{B.2}$$

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

The following Lemma states that mapping S defined as B.1, is continuous under $\mathscr{T}_1$, $\mathscr{T}_2$ and $\mathscr{T}_3$ in the space where S is valued, ensuring thus that *F* is measurable under $\mathscr{B}_1$, $\mathscr{B}_2$ and $\mathscr{B}_3$, respectively.

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

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

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

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

are continuous.

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

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

$$\begin{aligned}
&\left| \int_{\Delta_m} f_{ij} \, dH(k_0, M_{\boldsymbol{x}_i}) - \int_{\Delta_m} f_{ij} \, dH(k_0, \boldsymbol{Q}_{\boldsymbol{x}_i}) \right| \\
&\qquad \leq \left| \int_{\Delta_m} f_{ij}(\boldsymbol{y}) \sum_{\mathbf{j} \in \mathcal{H}^0_{k_0,m}} M_{\boldsymbol{x}_i}\left(A_{\mathbf{j},k_0}\right) \mathrm{d}\left(\, \boldsymbol{y} \mid \mathbf{j}, k_0 + m - \|\mathbf{j}\|_1\right) \right. \\
&\qquad\quad \left. - \int_{\Delta_m} f_{ij}(\boldsymbol{y}) \sum_{\mathbf{j} \in \mathcal{H}^0_{k_0,m}} \boldsymbol{Q}_{\boldsymbol{x}_i}\left(A_{\mathbf{j},k_0}\right) \mathrm{d}\left(\, \boldsymbol{y} \mid \mathbf{j}, k_0 + m - \|\mathbf{j}\|_1\right) \right|, \\
&\qquad \leq \int_{\Delta_m} |f_{ij}(\boldsymbol{y})| \sum_{\mathbf{j} \in \mathcal{H}^0_{k_0,m}} \left| M_{\boldsymbol{x}_i}\left(A_{\mathbf{j},k_0}\right) - \boldsymbol{Q}_{\boldsymbol{x}_i}\left(A_{\mathbf{j},k_0}\right) \right| \mathrm{d}\left(\, \boldsymbol{y} \mid \mathbf{j}, k_0 + m - \|\mathbf{j}\|_1\right), \\
&\qquad \leq \frac{M_0(k_0 + m - 1)!}{m!(k_0 - 1)!} \, N_{k_0}(\mathcal{M}, \mathcal{Q}),
\end{aligned}$$

where

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

$M_0 = \max_{i \in \{1,\ldots,L\}} \max_{j \in \{1,\ldots,K_i\}} \sup_{\boldsymbol{y} \in \Delta_m} |f_{ij}(\boldsymbol{y})|$, $||\quad\cdot\quad||_1$ denotes the l_1–norm, and $A_{\mathbf{j},k_0} = \left(\frac{j_1-1}{k_0}, \frac{j_1}{k_0}\right] \times \ldots \times \left(\frac{j_m-1}{k_0}, \frac{j_m}{k_0}\right]$. From Lemma 1 in (Barrientos et al., 2012), there exists $\mathcal{Q}' = \{\boldsymbol{Q}'_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}}$ such that for every $\boldsymbol{x} \in \mathscr{X}$, $\boldsymbol{Q}'_{\boldsymbol{x}}$ is absolutely continuous w.r.t Lebesgue measure and such that,

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

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

$$\begin{aligned} U'(\mathcal{Q}'; \tilde{\epsilon}) &= \bigcap_{i=1}^{L} \left\{ M_{\boldsymbol{x}_i} \in \mathscr{P}(\tilde{\Delta}_m) : \max_{\mathbf{j} \in \mathcal{H}^0_{k_0,m}} \left| M_{\boldsymbol{x}_i}\left(A_{\mathbf{j},k_0}\right) - \boldsymbol{Q}'_{\boldsymbol{x}_i}\left(A_{\mathbf{j},k_0}\right) \right| \leq \tilde{\epsilon} \right\}, \\ &= \left\{ \mathcal{M} \in \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}} : N_{k_0}(\mathcal{M}, \mathcal{Q}') \leq \tilde{\epsilon} \right\}, \end{aligned} \tag{B.3}$$

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

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

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

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

Notice that for every $\mathcal{M} = \{M_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}}$,

$$
\begin{aligned}
\sup_{\boldsymbol{y}\in\Delta_m} \left|\mathrm{bp}(\boldsymbol{y} \mid k_0, M_{\boldsymbol{x}_i}) - \mathrm{bp}(\boldsymbol{y} \mid k_0, \boldsymbol{Q}_{\boldsymbol{x}_i})\right| & \\
\leq \sup_{\boldsymbol{y}\in\Delta_m} \Bigg| \sum_{\mathbf{j}\in\mathcal{H}^0_{k_0,m}} & M_{\boldsymbol{x}_i}(A_{\mathbf{j},k_0})\,\mathrm{d}\left(\,\boldsymbol{y} \mid \mathbf{j}, k_0 + m - \|\mathbf{j}\|_1\right) \\
& - \sum_{\mathbf{j}\in\mathcal{H}^0_{k_0,m}} \boldsymbol{Q}_{\boldsymbol{x}_i}(A_{\mathbf{j},k_0})\,\mathrm{d}\left(\,\boldsymbol{y} \mid \mathbf{j}, k_0 + m - \|\mathbf{j}\|_1\right)\Bigg|, \\
\leq \sup_{\boldsymbol{y}\in\Delta_m} \sum_{\mathbf{j}\in\mathcal{H}^0_{k_0,m}} & \left|M_{\boldsymbol{x}_i}(A_{\mathbf{j},k_0}) - \boldsymbol{Q}_{\boldsymbol{x}_i}(A_{\mathbf{j},k_0})\right| \mathrm{d}\left(\,\boldsymbol{y} \mid \mathbf{j}, k_0 + m - \|\mathbf{j}\|_1\right), \\
\leq \frac{M_0(k_0+m-1)!}{m!(k_0-1)!} & \, N_{k_0}(\mathcal{M}, \mathcal{Q}),
\end{aligned}
$$

where $M_0 = \max_{\mathbf{j}\in\mathcal{H}^0_{k_0,m}} \sup_{\boldsymbol{y}\in\Delta_m} \mathrm{d}\left(\,\boldsymbol{y} \mid \mathbf{j}, k_0 + m - \|\mathbf{j}\|_1\right)$, $\mathrm{bp}(\boldsymbol{y} \mid k_0, M_{\boldsymbol{x}_i})$ stands for the density function of the multivariate Bernstein polynomial of function $M_{\boldsymbol{x}_i}$ of degree k_0, and $N_{k_0}(\mathcal{M}, \mathcal{Q})$ and $A_{\mathbf{j},k_0}$ are defined as in part (i) of the proof. By the same arguments from part (i), it follows that if $U = \{k_0\} \times U'(\mathcal{Q}'; \tilde{\epsilon})$, where $U'(\mathcal{Q}'; \tilde{\epsilon})$ is defined as in (B.3), with $\tilde{\epsilon} = \frac{m!(k_0-1)!}{2M_0(k_0+m-1)!}\epsilon$, then $U \in \mathscr{L}_1$, $(k_0, \mathcal{Q}) \in U$ and $S(U) \subseteq V(S(k_0, \mathcal{Q}), \epsilon)$, which completes the proof of (ii) in Lemma B.1.

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

$$
\begin{aligned}
\sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{\boldsymbol{y}\in\Delta_m} & \left|\mathrm{bp}(\boldsymbol{y} \mid k_0, M_x) - \mathrm{bp}(\boldsymbol{y} \mid k_0, \boldsymbol{Q}_x)\right| \\
& \leq \sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{\boldsymbol{y}\in\Delta_m} \sum_{\mathbf{j}\in\mathcal{H}^0_{k_0,m}} \left|M_{\boldsymbol{x}}(A_{\mathbf{j},k_0}) - \boldsymbol{Q}_{\boldsymbol{x}}(A_{\mathbf{j},k_0})\right| \mathrm{d}\left(\,\boldsymbol{y} \mid \mathbf{j}, k_0 + m - \|\mathbf{j}\|_1\right), \\
& \leq \frac{M_0(k_0+m-1)!}{m!(k_0-1)!} \sup_{\boldsymbol{x}\in\mathscr{X}} \max_{\mathbf{j}\in\mathcal{H}^0_{k_0,m}} \left|M_{\boldsymbol{x}}(A_{\mathbf{j},k_0}) - \boldsymbol{Q}_{\boldsymbol{x}}(A_{\mathbf{j},k_0})\right|,
\end{aligned}
$$

where $M_0 = \max_{\mathbf{j}\in\mathcal{H}^0_{k_0,m}} \sup_{\boldsymbol{y}\in\Delta_m}, \mathrm{d}\left(\,\boldsymbol{y} \mid \mathbf{j}, k_0 + m - \|\mathbf{j}\|_1\right)$, and $A_{\mathbf{j},k_0}$ are defined as in the proof of (i). Then, if $U = \{k_0\} \times \tilde{\Delta}^{L_\infty}_{\epsilon,k_0}(\mathcal{Q})$, where $\tilde{\Delta}^{L_\infty}_{\tilde{\epsilon},k_0}(\mathcal{Q})$ is defined as in (B.2), with $\tilde{\epsilon} = \frac{m!(k_0-1)!}{M_0(k_0+m-1)!}\epsilon$, then $U \in \mathscr{L}_2$, $(k_0, \mathcal{Q}) \in U$ and $S(U) \subseteq V(S(k_0, \mathcal{Q}), \epsilon)$, which completes the proof of (iii) in Lemma B.1.

B.2 Proof of Theorem 4.2

First we prove that $\mathscr{P}(\Delta_m)^{\mathscr{X}}$ is the support of *F* under the weak product topology. Then we prove that $\mathscr{D}(\Delta_m)^{\mathscr{X}}$ is the support of *F* under the L_∞ product topology. In each case all three versions of *F* are considered.

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

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

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

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

then $V(S(k_0, \mathcal{Q}'); \epsilon/4) \subset V(\mathcal{Q}'; \epsilon/2)$. By Lemma B.1 part (i), there exists $U = \{k_0\} \times U'(\mathcal{Q}'; \tilde{\epsilon}) \in \mathscr{L}_1$, with $\tilde{\epsilon} = \frac{m!(k_0-1)!}{4M_0(k_0+m-1)!}\epsilon$, $k_0 \in \mathbb{N}$ and $U'(\mathcal{Q}'; \tilde{\epsilon}) \in \mathscr{T}_4$, such that $S(U) \subset V(S(k_0, \mathcal{Q}'); \epsilon/4)$. Thus, to prove this theorem, it suffices to prove that $P \circ F^{-1}(V(\mathcal{Q}; \epsilon)) \geq P\{\omega \in \Omega : (k(\omega), \overline{\mathbf{T}}) \in U\} > 0$, where $U = \{k_0\} \times U'(\mathcal{Q}'; \tilde{\epsilon})$, with $U'(\mathcal{Q}'; \tilde{\epsilon})$ defined as in (B.3) and $\overline{\mathbf{T}}$ is either $\mathbf{T}$, $\mathbf{T}^\theta$ or $\mathbf{T}^w$.

First assume that $\overline{\mathbf{T}}$ is $\mathbf{T}$. Notice that there are $N = \frac{(k_0+m-1)!}{m!(k_0-1)!}$ disjoint sets in $\mathcal{H}^0_{k_0,m}$. Denote each of this sets by $A_{[l],N}$, $l = 1, \ldots, N$. If for each $i = 1, \ldots, L$,

$$\boldsymbol{\theta}_l(\boldsymbol{x}_i, \omega) \in A_{[l],N}, \;\; l = 1, \ldots, N,$$
$$\boldsymbol{\theta}_{l_1}(\boldsymbol{x}_i, \omega) \notin A_{[l],N}, \;\; l_1 \neq l, \;\; l, l_1 = 1, \ldots, N,$$
$$\boldsymbol{\theta}_j(\boldsymbol{x}_i, \omega) \in \tilde{\Delta}_m, \;\; j > N,$$
$$\left| V_1(\boldsymbol{x}_i, \omega) - \boldsymbol{Q}'_{\boldsymbol{x}_i}(A_{[1],N}) \right| \leq \frac{\tilde{\epsilon}}{4(N-1)}, \tag{B.4}$$
$$\frac{\boldsymbol{Q}'_{\boldsymbol{x}_i}(A_{[l],N}) - \tilde{\epsilon}/4(N-1)}{\prod_{l_1<l} [1 - V_{l_1}(\boldsymbol{x}_i, \omega)]} \leq V_l(\boldsymbol{x}_i, \omega) \leq \frac{\boldsymbol{Q}'_{\boldsymbol{x}_i}(A_{[l],N}) - \tilde{\epsilon}/4(N-1)}{\prod_{l_1<l} [1 - V_{l_1}(\boldsymbol{x}_i, \omega)]},$$
$$l = 2, \ldots, N-1, \tag{B.5}$$
$$\frac{1 - \sum_{l=1}^{N-1} \boldsymbol{Q}'_{\boldsymbol{x}_i}(A_{[l],N}) - \tilde{\epsilon}/3}{\prod_{l_1<N} [1 - V_{l_1}(\boldsymbol{x}_i, \omega)]} \leq V_N(\boldsymbol{x}_i, \omega) \leq \frac{1 - \sum_{l=1}^{N-1} \boldsymbol{Q}'_{\boldsymbol{x}_i}(A_{[l],N}) - \tilde{\epsilon}/4}{\prod_{l_1<N} [1 - V_{l_1}(\boldsymbol{x}_i, \omega)]}, \tag{B.6}$$

where $V_j(\boldsymbol{x}, \omega) = v_{\boldsymbol{x}}(\eta_j(\boldsymbol{x}, \omega))$, $j \geq 1$, then $\left| M_{\boldsymbol{x}_i}(A_{\mathbf{j},k_0}) - \boldsymbol{Q}'_{\boldsymbol{x}_i}(A_{\mathbf{j},k_0}) \right| \leq \tilde{\epsilon}$, for every $\mathbf{j} \in \mathcal{H}^0_{k_0,m}$ and every $i = 1, \ldots, L$. Finally, since the stochastic processes η_j and $\boldsymbol{z}_j$ are well defined and have full support,

$$\begin{aligned}
P \circ \boldsymbol{F}^{-1}(V(\mathcal{Q}; \epsilon)) \geq\; & P\{\omega \in \Omega : k(\omega) = k_0\} \\
& \times P\left\{\omega \in \Omega : \max_{i \in \{1,\ldots,L\}} \max_{\mathbf{j} \in \mathcal{H}^0_{k_0,m}} \left| M_{\boldsymbol{x}_i}(A_{\mathbf{j},k_0}) - \boldsymbol{Q}'_{\boldsymbol{x}_i}(A_{\mathbf{j},k_0}) \right| \leq \tilde{\epsilon} \right\}, \\
\geq\; & P\{\omega \in \Omega : k(\omega) = k_0\} \\
& \times \prod_{l=1}^{N} P\left\{\omega \in \Omega : (\boldsymbol{\theta}_l(\boldsymbol{x}_1, \omega), \ldots, \boldsymbol{\theta}_l(\boldsymbol{x}_L, \omega)) \in A^L_{[l],N}\right\} \\
& \times P\left\{\omega \in \Omega : (V_l(\boldsymbol{x}_1, \omega), \ldots, V_l(\boldsymbol{x}_L, \omega)) \in B^L_l, l \in \{1, \ldots, N\}\right\} \\
& \times \prod_{l=N+1}^{\infty} P\left\{\omega \in \Omega : (\boldsymbol{\theta}_l(\boldsymbol{x}_1, \omega), \ldots, \boldsymbol{\theta}_l(\boldsymbol{x}_L, \omega)) \in \tilde{\Delta}^L_m\right\} \\
& \times \prod_{l=N+1}^{\infty} P\left\{\omega \in \Omega : (V_l(\boldsymbol{x}_1, \omega), \ldots, V_l(\boldsymbol{x}_L, \omega)) \in [0,1]^L\right\}, \\
>\; & 0,
\end{aligned}$$

where

$$
\begin{aligned}
B_1^L &= \bigotimes_{i=1}^{L} \left\{ \mathcal{Q}'_{\boldsymbol{x}_i}(A_{[1],N}) - \frac{\tilde{\epsilon}}{4(N-1)} \; ; \; \mathcal{Q}'_{\boldsymbol{x}_i}(A_{[1],N}) + \frac{\tilde{\epsilon}}{4(N-1)} \right\}, \\
B_l^L &= \bigotimes_{i=1}^{L} \left\{ \frac{\mathcal{Q}'_{\boldsymbol{x}_i}(A_{[l],N}) - \frac{\tilde{\epsilon}}{4(N-1)}}{\prod_{l_1<l}[1 - V_{l_1}(\boldsymbol{x}_i, \omega)]} \; ; \; \frac{\mathcal{Q}'_{\boldsymbol{x}_i}(A_{[l],N}) + \frac{\tilde{\epsilon}}{4(N-1)}}{\prod_{l_1<l}[1 - V_{l_1}(\boldsymbol{x}_i, \omega)]} \right\}, \;\; l = 2, \ldots, N-1, \\
B_N^L &= \bigotimes_{i=1}^{L} \left\{ \frac{\mathcal{Q}'_{\boldsymbol{x}_i}(A_{[N],N}) - \frac{\tilde{\epsilon}}{3}}{\prod_{l_1<N}[1 - V_{l_1}(\boldsymbol{x}_i, \omega)]} \; ; \; \frac{\mathcal{Q}'_{\boldsymbol{x}_i}(A_{[N],N}) - \frac{\tilde{\epsilon}}{4}}{\prod_{l_1<N}[1 - V_{l_1}(\boldsymbol{x}_i, \omega)]} \right\},
\end{aligned}
$$

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

Now assume that $\overline{\mathbf{T}}$ is $\mathbf{T}^\theta$. The proof follows the same arguments used when $\overline{\mathbf{T}}$ is $\mathbf{T}$. Here, to ensure that $\left| M_{\boldsymbol{x}_i}(A_{\mathbf{j},k_0}) - \mathcal{Q}'_{\boldsymbol{x}_i}(A_{\mathbf{j},k_0}) \right| \leq \tilde{\epsilon}$, for every $\mathbf{j} \in \mathcal{H}_{k_0,m}^0$ and every $i = 1, \ldots, L$, consider , for each $l = 1, \ldots, N$,

$$
\begin{aligned}
&\boldsymbol{\theta}_l(\omega) \in A_{[l],N}, \\
&\boldsymbol{\theta}_{l_1}(\omega) \notin A_{[l_1],N}, \;\; l_1 \neq l, \;\; l_1 = 1, \ldots, N, \\
&\boldsymbol{\theta}_j(\omega) \in \tilde{\Delta}_m, \;\; j > N,
\end{aligned}
$$

and remain conditions (B.4), (B.5), and (B.6) as before. Finally, since the stochastic processes η_j, and the random vectors $\boldsymbol{\theta}_j$ are well defined and have full support, it follows that

$$
\begin{aligned}
P \circ \boldsymbol{F}^{-1}(V(\mathcal{Q}; \epsilon)) \geq{}& P\{\omega \in \Omega : k(\omega) = k_0\} \\
& \times \prod_{l=1}^{N} P\left\{\omega \in \Omega : (\boldsymbol{\theta}_l(\omega), \ldots, \boldsymbol{\theta}_l(\omega)) \in A_{[l],N}^L\right\} \\
& \times P\left\{\omega \in \Omega : (V_l(\boldsymbol{x}_1, \omega), \ldots, V_l(\boldsymbol{x}_L, \omega)) \in B_l^L, l \in \{1, \ldots, N\}\right\}, \\
>{}& 0,
\end{aligned}
$$

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

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

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

Then $U'(\tilde{Q}; \tilde{\epsilon}/2) \subset U'(Q'; \tilde{\epsilon})$, where $U'(\mathcal{Q}; \epsilon)$ is defined as in (B.3). Thus, it suffices to prove that

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

Consider $\left\{\tilde{A}_{[\tilde{l}],M}\right\}_{\tilde{l}=1}^{M}$, a finer partition of $\mathcal{H}^0_{k_0,m}$ than $\left\{A_{[l],N}\right\}_{l=1}^{N}$, such that $A_{[1],N} = \bigcup_{\tilde{l}=1}^{n_1} \tilde{A}_{[\tilde{l}],M}$ and $A_{[l],N} = \bigcup_{\tilde{l}=n_{l-1}+1}^{n_l} \tilde{A}_{\tilde{l},M}$, $l = 1, \ldots, N$, where $\sum_{l=1}^{N} n_l = M$. To ensure that $\left|M_{\boldsymbol{x}_i}(A_{\mathbf{j},k_0}) - \tilde{Q}_{\boldsymbol{x}_i}(A_{\mathbf{j},k_0})\right| \leq \tilde{\epsilon}/2$, for every $\mathbf{j} \in \mathcal{H}^0_{k_0,m}$ and every $i = 1, \ldots, L$, consider, for each $i = 1, \ldots, L$,

$$\begin{aligned}
&\Big(\lceil k_0 \boldsymbol{\theta}_1(\boldsymbol{x}_i, \omega)\rceil - \lceil k_0 \tilde{\boldsymbol{\theta}}_j(\boldsymbol{x}_i)\rceil\Big) = \mathbf{0}, \;\; j = 1, \ldots, n_1,\\
&\Big(\lceil k_0 \boldsymbol{\theta}_l(\boldsymbol{x}_i, \omega)\rceil - \lceil k_0 \tilde{\boldsymbol{\theta}}_j(\boldsymbol{x}_i)\rceil\Big) = \mathbf{0}, \;\; j = n_{l-1}+1, \ldots, n_l, \;\; l = 2, \ldots, N,
\end{aligned}$$

and consider,

$$\left| v_1(\omega) - \sum_{j=1}^{n_1} \tilde{w}_j \right| \leq \frac{\tilde{\epsilon}}{8(N-1)},$$

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

$$\frac{1 - \sum_{j=1}^{n_{N-1}} \tilde{w}_j - \frac{\tilde{\epsilon}}{6}}{\prod_{l_1<N} [1 - V_{l_1}(\omega)]} \leq v_N(\omega) \leq \frac{1 - \sum_{j=1}^{n_{N-1}} \tilde{w}_j - \frac{\tilde{\epsilon}}{8}}{\prod_{l_1<N} [1 - V_{l_1}(\omega)]}.$$

Finally, since the stochastic processes $\boldsymbol{z}_j$ and the random variables v_j are well defined and have full support, it follows that

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

where

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

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

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

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

Now we will prove that $\mathscr{D}\left(\Delta_m\right)^{\mathscr{X}}$ is the support of *F* under the L_∞ product toplogy. For this we use the following Lemma, stating that if *Q* is an absolutely continuous measure, w.r.t. Lebesgue measure, defined on Δ_m with continuous density, q, then the density function of the Bernstein polynomial of function *Q* of degree k converges uniformly to q as $k \to \infty$.

Lemma B.2. *Let Q be an absolutely continuous measure, w.r.t. Lebesgue measure, defined on Δ_m, with continuous density q. Consider $b_{k,Q}(\boldsymbol{y})$ the density function, w.r.t. Lebesgue measure, of the Bernstein polynomial of function Q of degree $k \in \mathbb{N}$. Then for every $\epsilon > 0$, there exists $k_0 \in \mathbb{N}$ such that for every $k \geq k_0$,*

$$\sup_{\boldsymbol{y} \in \Delta_m} \mid b_{k,Q}(\boldsymbol{y}) - q(\boldsymbol{y}) \mid < \epsilon.$$

Proof of Lemma B.2

Notice that the density function, w.r.t. Lebesgue measure, of the Bernstein polynomial of function Q of degree k, can be written as

$$b_{k,Q}(\boldsymbol{y}) = \prod_{l=1}^{m}(k+m-l)\sum_{\boldsymbol{j}\in\mathcal{H}_{k-1,m}} Q\left(\left(\frac{j_1}{k},\frac{j_1+1}{k}\right]\times\ldots\times\left(\frac{j_m}{k},\frac{j_m+1}{k}\right]\right)\mathrm{Mult}(\boldsymbol{j}\mid k-1,\boldsymbol{y}),$$

where $\mathcal{H}_{k-1,m} = \{(j_1,\ldots,j_m)\in\{0,\ldots,k-1\}^m:\ \sum_{l=1}^{m} j_l \le k-1\}$, and $\mathrm{Mult}(\boldsymbol{j}\mid k-1,\boldsymbol{y})$ stands for a multinomial distribution with parameters $k-1$ and $\boldsymbol{y}$. Since Q is continuous and defined on Δ_m, by the multivariate mean value theorem, there exists $\boldsymbol{c}\in\left(\frac{j_1}{k},\frac{j_1+1}{k}\right]\times\ldots\times\left(\frac{j_m}{k},\frac{j_m+1}{k}\right]$, such that

$$Q\left(\left(\frac{j_1}{k},\frac{j_1+1}{k}\right]\times\ldots\times\left(\frac{j_m}{k},\frac{j_m+1}{k}\right]\right) = \frac{q(\boldsymbol{c})}{k^m}.$$

Notice that if $\boldsymbol{c}\in\left(\frac{j_1}{k},\frac{j_1+1}{k}\right]\times\ldots\times\left(\frac{j_m}{k},\frac{j_m+1}{k}\right]$, then $||\boldsymbol{c}-\mathbf{j}/k||^2 < m/k$, where $||\cdot||$ denotes the l_1 norm. Thus, by the uniform continuity of q and considering large enough k, it follows that $|q(\boldsymbol{c}) - q(\mathbf{j}/k)| < \epsilon/2$. In addition, since q is a continuous density function and $\boldsymbol{J}$ follows a multinomial distribution with parameters $k-1$ and $\boldsymbol{y}$, then for large enough k, $\sup_{\boldsymbol{y}\in\Delta_m} E_J\left|q\left(\frac{\boldsymbol{J}}{k}\right) - q(\boldsymbol{y})\right| < \epsilon/2$. Also, for large enough k, $C(k)/k^m \approx 1$, where $C(k) = \prod_{l=1}^{m}(k+m-l)$. Therefore, for large enough k,

$$\begin{aligned}
\sup_{\boldsymbol{y}\in\Delta_m}|b_{k,Q}(\boldsymbol{y}) - q(\boldsymbol{y})| &\le \sup_{\boldsymbol{y}\in\Delta_m}\sum_{\boldsymbol{j}\in\mathcal{H}_{k-1,m}}|q(\boldsymbol{c}) - q(\boldsymbol{y})|\,\mathrm{Mult}(\boldsymbol{j}\mid k-1,\boldsymbol{y}),\\
&\le \sup_{\boldsymbol{y}\in\Delta_m}\left|\sum_{\mathbf{j}\in\mathcal{H}_{k-1,m}}\left[q(\boldsymbol{c}) - q\left(\frac{\mathbf{j}}{k}\right)\right]\mathrm{Mult}(\mathbf{j}\mid k-1,\boldsymbol{y})\right|\\
&\quad + \sup_{\boldsymbol{y}\in\Delta_m}\sum_{\mathbf{j}\in\mathcal{H}_{k-1,m}}\left|q\left(\frac{\mathbf{j}}{k}\right) - q(\boldsymbol{y})\right|\mathrm{Mult}(\mathbf{j}\mid k-1,\boldsymbol{y}),\\
&< \epsilon.
\end{aligned}$$

Which completes the proof of Lemma B.2. □

Now we will prove the L_∞ product support of *F*. For this, it suffices to prove that any open set of the L_∞ product topology has positive $P \circ F^{-1}$–measure. Let $\mathcal{Q} \in \mathscr{D}(\Delta_m)^{\mathscr{X}}$ and $V(\mathcal{Q};\epsilon) = \bigcap_{i=1}^{L} \Delta_{\epsilon,\boldsymbol{x}_i}^{L_\infty}(\mathcal{Q})$, where L is a positive integer, $\epsilon > 0$, and $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_L) \in \mathscr{X}^L$. Recall that for every $\boldsymbol{x} \in \mathscr{X}$, $\boldsymbol{Q}_{\boldsymbol{x}} \in \mathscr{D}(\Delta_m)$ is an absolutely continuous measures, w.r.t. Lebesgue measure, with continuous density, $q_{\boldsymbol{x}}$. By Lemma B.2, for every $\epsilon > 0$, there exists large enough $k_0 \in \mathbb{N}$, such that for every $\boldsymbol{x} \in \mathscr{X}$,

$$\sup_{\boldsymbol{y}\in\Delta_m} |\mathrm{bp}(\boldsymbol{y} \mid k_0, \boldsymbol{Q}_{\boldsymbol{x}}) - q_{\boldsymbol{x}}(\boldsymbol{y})| < \frac{\epsilon}{2},$$

where $\mathrm{bp}(\boldsymbol{y} \mid k, \boldsymbol{Q}_{\boldsymbol{x}})$ stands for the density function of the multivariate Bernstein polynomial of degree k of function $\boldsymbol{Q}_{\boldsymbol{x}}$. Then $V(S(k_0, \mathcal{Q}); \epsilon/2) \subset V(\mathcal{Q};\epsilon)$. By Lemma B.1 part (ii), there exists $U = \{k_0\} \times U'(\mathcal{Q}'; \tilde{\epsilon}) \in \mathscr{L}_1$, with $\tilde{\epsilon} = \frac{m!(k_0-1)!}{2M_0(k_0+m-1)!}\epsilon$, $k_0 \in \mathbb{N}$ and $U'(\mathcal{Q}'; \tilde{\epsilon}) \in \mathscr{T}_4$, such that $S(U) \subset V(S(k_0,\mathcal{Q}); \epsilon/2)$. In analogy with the weak product support proof, it suffices to prove that $P \circ F^{-1}(V(\mathcal{Q};\epsilon)) \geq P\{\omega \in \Omega : (k(\omega), \overline{\mathbf{T}}) \in \{k_0\} \times U'(\mathcal{Q}; \tilde{\epsilon})\} > 0$, where $\overline{\mathbf{T}}$ is either $\mathbf{T}$, $\mathbf{T}^\theta$ or $\mathbf{T}^w$. By the same arguments used to prove the weak product support of *F*, it follows that $P \circ F^{-1}(V(\mathcal{Q};\epsilon)) > 0$, when *F* is considered as DMBPP, $\theta\mathrm{DMBPP}$, or $w\mathrm{DMBPP}$. This completes the proof that $\mathscr{D}(\Delta_m)^{\mathscr{X}}$ is the support of *F* under the L_∞ product toplogy, and thus completes the proof of Theorem 4.2. □

B.3 Proof of Theorem 4.3

The following Lemma is used in the proof of this theorem.

Lemma B.3. *Let* $\{\boldsymbol{Q}_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\} \in \tilde{\mathscr{D}}(\Delta_m)^{\mathscr{X}}$ *be an absolutely continuous measure, w.r.t. Lebesgue measure, such that the mapping* $(\boldsymbol{x}, \boldsymbol{y}) \mapsto q_{\boldsymbol{x}}(\boldsymbol{y})$ *is continuous and consider* $\mathscr{X}$ *a compact space on* $\mathbb{R}^p$. *Denote* $b_{k,\boldsymbol{Q}_{\boldsymbol{x}}}(\boldsymbol{y})$, *the density function, w.r.t. Lebesgue measure, of the multivariate Bernstein polynomial of degree* k *of function* $\boldsymbol{Q}_{\boldsymbol{x}}$. *Then for every* $\epsilon > 0$, *there exists* $k_0 \in \mathbb{N}$ *such that for every* $k \geq k_0$,

$$\sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{\boldsymbol{y}\in\Delta_m} |\, b_{k,\boldsymbol{Q}_{\boldsymbol{x}}}(\boldsymbol{y}) - q_{\boldsymbol{x}}(\boldsymbol{y}) \,| < \epsilon.$$

Proof of Lemma B.3

Consider $\mathscr{X} = [0,1]^p$, and a uniform marginal distribution for $\boldsymbol{x}$ on $\mathscr{X}$. Then, $q_{\boldsymbol{x}}(\boldsymbol{y})$ is a joint density on $\Delta_m \times \mathscr{X}$. Note that $b_{k,Q_{\boldsymbol{x}}}(\boldsymbol{y})$ can be written as

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

where $\mathcal{H}^0_{k,m} = \{(j_1, \ldots, j_m) \in \{1, \ldots, k\}^m : \sum_{l=1}^m j_l \leq k + m - 1\}$, $A_{\mathbf{j},k} = \left(\frac{j_1 - 1}{k}, \frac{j_1}{k}\right] \times \ldots \times \left(\frac{j_m - 1}{k}, \frac{j_m}{k}\right]$, $\mathbf{j} = (j_1, \ldots, j_m)$, $\alpha(k, \boldsymbol{j}) = (\boldsymbol{j},\ k + m - \sum_{l=1}^m j_l)$, and $\mathrm{d}(\cdot \mid (\alpha_1, \ldots, \alpha_m))$ denotes the density function of a m–dimensional Dirichlet distribution with parameters $(\alpha_1, \ldots, \alpha_m)$. Now, set

$$b_{k,\boldsymbol{l},Q}(\boldsymbol{y}, \boldsymbol{x}) =$$
$$\sum_{\boldsymbol{j} \in \mathcal{H}^0_{k,m}} \sum_{i_1=1}^{l_1} \cdots \sum_{i_p=1}^{l_p} \left[\int_{B_{i_1}} \cdots \int_{B_{i_p}} \int_{A_{\mathbf{j},k}} q_{\boldsymbol{x}}(\boldsymbol{y}) d\boldsymbol{y} dx_p \ldots dx_1 \right] \times \prod_{s=1}^p \beta(x_s \mid a_s, b_s) \mathrm{d}(\boldsymbol{y} \mid \alpha(k, \boldsymbol{j})),$$

where $B_{i_s} = \left(\frac{i_s - 1}{l_s}, \frac{i_s}{l_s}\right]$, $a_s = i_s$, $b_s = l_s - i_s + 1$, $s = 1, \ldots, p$, and $\beta(\cdot \mid a, b)$ stands for a beta density with parameters a and b. Given that $(\boldsymbol{x}, \boldsymbol{y}) \mapsto q_{\boldsymbol{x}}(\boldsymbol{y})$ is a continuous mapping, it is not difficult to show that $b_{k,\boldsymbol{l},Q}(\boldsymbol{y}, \boldsymbol{x})$ can approximate uniformly any continuous density function defined on $\Delta_m \times \mathscr{X}$. Thus, for large enough $k, l_1, \ldots, l_p$, it follows that

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

Now, noting that

$$\sum_{i_1=1}^{l_1} \cdots \sum_{i_p=1}^{l_p} \left[\int_{B_{i_1}} \cdots \int_{B_{i_p}} \int_{A_{\mathbf{j},k}} q_{\boldsymbol{x}}(\boldsymbol{y}) d\boldsymbol{y}\, dx_p \ldots dx_1 \right] \times \prod_{s=1}^p \beta(x_s \mid a_s, b_s), \tag{B.7}$$

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

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

defined on $\mathscr{X}$, it follows that (B.7) converges uniformly to $\int_{A_{\mathbf{j},k}} q_{\boldsymbol{x}}(\boldsymbol{y}) d\,\boldsymbol{y}$, as $(l_1, \ldots, l_p) \to \infty$. Therefore, for large enough $(l_1, \ldots, l_p)$,

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

where $\tilde{\epsilon} = \frac{m!(k-1)!}{M_0(k+m-1)!}\epsilon$, with $M_0 = \max_{\mathbf{j}\in\mathcal{H}^0_{k,m}} \sup_{\boldsymbol{y}\in\Delta_m} \mathrm{d}(\boldsymbol{y} \mid \alpha(k, \boldsymbol{j}))$. Finally, there exists large enough $k, l_1, \ldots, l_p$, such that,

$$\sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{\boldsymbol{y}\in\Delta_m} |b_{k,\mathcal{Q}_{\boldsymbol{x}}}(\boldsymbol{y}) - q_{\boldsymbol{x}}(\boldsymbol{y})| < \epsilon,$$

which completes to proof of the Lemma. □

Now we will prove that $\tilde{\mathscr{D}}(\Delta_m)^{\mathscr{X}}$ is contained in the support of *F* under the L_∞ topology. For this, it suffices to prove that any open set of the L_∞ topology has positive $P \circ F^{-1}$–measure. Let $\mathcal{Q} \in \tilde{\mathscr{D}}(\Delta_m)^{\mathscr{X}}$ and $V(\mathcal{Q}; \epsilon) = B^{L_\infty}_{\epsilon}(\mathcal{Q})$, $\epsilon > 0$. Recall that $\mathscr{X}$ is compact, and $\boldsymbol{Q}_{\boldsymbol{x}} \in \tilde{\mathscr{D}}(\Delta_m)$ is an absolutely continuous measures, w.r.t. Lebesgue measure, with continuous density, $q_{\boldsymbol{x}}$, sucht that $(\boldsymbol{x}, \boldsymbol{y}) \mapsto q_{\boldsymbol{x}}(\boldsymbol{y})$ is continuous. From Lemma B.3, there exists large enough k_0, such that,

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

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

$P\{\omega \in \Omega : (k(\omega), \overline{\mathbf{T}}) \in \{k_0\} \times U^\star(\mathcal{Q};\tilde{\epsilon})\} > 0$, where

$$U^\star(\mathcal{Q};\tilde{\epsilon}) = \left\{ \mathcal{M} \in \mathscr{P}(\tilde{\Delta}_m)^{\mathscr{X}} : \sup_{\boldsymbol{x}\in\mathscr{X}} \max_{\mathbf{j}\in\mathcal{H}^0_{k_0,m}} \left| M_{\boldsymbol{x}}\left(A_{\mathbf{j},k_0}\right) - Q_{\boldsymbol{x}}\left(A_{\mathbf{j},k_0}\right)\right| \leq \tilde{\epsilon} \right\}, \tag{B.8}$$

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

First assume that $\overline{\mathbf{T}}$ is $\mathbf{T}$. Following a similar reasoning as in the proof of Theorem 4.2, and if for each $l = 1, \ldots, N$,

$$\begin{aligned}
&\sup_{\boldsymbol{x}\in\mathscr{X}} |\boldsymbol{\theta}_l(\boldsymbol{x},\omega)| \in A_{[l],N},\\
&\sup_{\boldsymbol{x}\in\mathscr{X}} |\boldsymbol{\theta}_{l_1}(\boldsymbol{x},\omega)| \notin A_{[l],N},\ \ l \neq l_1,\ \ l_1 = 1,\ldots,N,\\
&\sup_{\boldsymbol{x}\in\mathscr{X}} |\boldsymbol{\theta}_j(\boldsymbol{x},\omega)| \in \tilde{\Delta}_m, j > N,\\
&\sup_{\boldsymbol{x}\in\mathscr{X}} \left|V_1(\boldsymbol{x},\omega) - Q_{\boldsymbol{x}}(A_{[1],N})/2\right| < \frac{\tilde{\epsilon}}{2N}, &\text{(B.9)}\\
&\sup_{\boldsymbol{x}\in\mathscr{X}} \left| V_l(\boldsymbol{x},\omega) - \frac{Q_{\boldsymbol{x}}(A_{[l],N})/2}{\prod_{l_1<l}\left[1 - V_{l_1}(\boldsymbol{x},\omega)\right]}\right| < \frac{\tilde{\epsilon}}{2N}, &\text{(B.10)}
\end{aligned}$$

where $V_j(\boldsymbol{x},\omega) = v_{\boldsymbol{x}}(\eta_j(\boldsymbol{x},\omega))$, $j \geq 1$, then $\sup_{\boldsymbol{x}\in\mathscr{X}} \left|M_{\boldsymbol{x}}\left(A_{\mathbf{j},k_0}\right) - Q_{\boldsymbol{x}}\left(A_{\mathbf{j},k_0}\right)\right| \leq \tilde{\epsilon}$, for every $\mathbf{j} \in \mathcal{H}^0_{k_0,m}$. Finally, since the stochastic processes η_j and $\boldsymbol{z}_j$ are well defined and have full support, $A_{[l],N} \in \mathscr{B}(\Delta_m)$ and the mappings

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

are continuous, it follows that,

$$
\begin{aligned}
P \circ F^{-1}&(V(\mathcal{Q};\epsilon)) \\
&\geq P\{\omega \in \Omega : k(\omega) = k_0\} \\
&\quad \times \prod_{l=1}^{N} P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} |\boldsymbol{\theta}_l(\boldsymbol{x},\omega)| \in A_{[l],N}\right\} \\
&\quad \times P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} \left|V_1(\boldsymbol{x},\omega) - \boldsymbol{Q}_{\boldsymbol{x}}(A_{[l],N})/2\right| < \frac{\tilde{\epsilon}}{2N}\right\} \\
&\quad \times P\left\{\omega \in \Omega : \sup_{\boldsymbol{x}\in\mathscr{X}} \left|V_l(\boldsymbol{x},\omega) - \frac{\boldsymbol{Q}_{\boldsymbol{x}}(A_{[l],N})/2}{\prod_{l_1<l}[1 - V_{l_1}(\boldsymbol{x},\omega)]}\right| < \frac{\tilde{\epsilon}}{2N}, l = 2,\ldots,N\right\}, \\
&> 0,
\end{aligned}
$$

which completes the proof that *F* considered as $\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_{\boldsymbol{z}}, \mathscr{V}, \mathscr{H})$ has L_∞ support. Now assume that $\overline{\mathbf{T}}$ is $\mathbf{T}^\theta$. The proof follows the same as arguments used when $\overline{\mathbf{T}}$ is $\mathbf{T}$. Here, to ensure that $\sup_{\boldsymbol{x}\in\mathscr{X}} |M_{\boldsymbol{x}}(A_{\mathbf{j},k_0}) - \boldsymbol{Q}_{\boldsymbol{x}}(A_{\mathbf{j},k_0})| \leq \tilde{\epsilon}$, for every $\mathbf{j} \in \mathcal{H}^0_{k_0,m}$, we consider for each $l = 1,\ldots,N$,

$$
\begin{aligned}
&\boldsymbol{\theta}_l(\omega) \in A_{[l],N}, \\
&\boldsymbol{\theta}_{l_1}(\omega) \notin A_{[l],N}, \;\; l \neq l_1, \;\; l_1 = 1,\ldots,N, \\
&\boldsymbol{\theta}_j(\omega) \in \tilde{\Delta}_m, j > N,
\end{aligned}
$$

and remain conditions (B.9) and (B.10) as before. Then $P \circ F^{-1}(V(\mathcal{Q};\epsilon)) > 0$, which completes the proof that *F* considered as $\theta\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$ has L_∞ support.
Finally, assume that $\overline{\mathbf{T}}$ is $\mathbf{T}^w$. Consider the partition $\left\{A_{[l],N}\right\}_{l=1}^N$ of Δ_m and for each $l = 1,\ldots,N$, consider $\{\mathscr{X}_{l,j}\}_{j=1}^{N_l}$, a partition of space $\mathscr{X}$, $N_l \in \mathbb{N}$, $N_l > N$. Since $\boldsymbol{Q}_{\boldsymbol{x}} \in \tilde{\mathscr{D}}(\Delta_m)$ are such that $(\boldsymbol{y},\boldsymbol{x}) \mapsto q_{\boldsymbol{x}}(\boldsymbol{y})$ are continuous, then $(\boldsymbol{y},\boldsymbol{x}) \mapsto \boldsymbol{Q}_{\boldsymbol{x}}(\boldsymbol{y})$ are continuous and can be approximated by functions of the form,

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

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

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

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

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

Then $U^\star(\tilde{\mathcal{Q}}; \tilde{\epsilon}/2) \subset U^\star(\mathcal{Q}; \tilde{\epsilon})$, where $U^\star(\mathcal{Q}; \epsilon)$ is defined as (B.8). Thus, in analogy with the previous proofs, it suffices to prove that $P\left\{\omega \in \Omega : (k(\omega), \mathbf{T}^w) \in \{k_0\} \times U^\star(\tilde{\mathcal{Q}}; \tilde{\epsilon}/2)\right\} > 0$. Following a similar reasoning as in the proof of Theorem 4.2 when $\overline{\mathbf{T}}$ was considered as $\mathbf{T}^w$, and considering,

$$\begin{aligned}
&\sup_{\boldsymbol{x} \in \mathscr{X}} \left|\lceil k_0 \boldsymbol{\theta}_1(\boldsymbol{x}, \omega)\rceil - \lceil k_0 \tilde{\boldsymbol{\theta}}_{1,j}(\boldsymbol{x})\rceil\right| = \mathbf{0}, \;\; j = 1, \ldots, n_1, \\
&\sup_{\boldsymbol{x} \in \mathscr{X}} \left|\lceil k_0 \boldsymbol{\theta}_l(\boldsymbol{x}, \omega)\rceil - \lceil k_0 \tilde{\boldsymbol{\theta}}_{l,j}(\boldsymbol{x})\rceil\right| = \mathbf{0}, \;\; j = n_{l-1} + 2, \ldots, n_l, \;\; l = 2, \ldots, N, \\
&\left|v_1(\omega) - \sum_{j=1}^{n_1} \tilde{w}_{1,j}/2\right| < \frac{\tilde{\epsilon}}{2N}, \\
&\left|v_l(\omega) - \frac{\sum_{j=1}^{n_l} \tilde{w}_{l,j}/2}{\prod_{l_2 < l} [1 - v_{l_2}(\omega)]}\right| < \frac{\tilde{\epsilon}}{2N}, \;\; l = 2, \ldots, N,
\end{aligned}$$

then, $\sup_{\boldsymbol{x} \in \mathscr{X}} \left|M_{\boldsymbol{x}}\left(A_{\mathbf{j},k_0}\right) - \tilde{\mathcal{Q}}_{\boldsymbol{x}}\left(A_{\mathbf{j},k_0}\right)\right| \le \tilde{\epsilon}$, for every $\mathbf{j} \in \mathcal{H}_{k_0,m}^0$. Finally, since the stochastic processes η_j and $\boldsymbol{z}_j$ are well defined and have full support, $A_{[l],N} \in \mathscr{B}(\Delta_m)$, and the mappings

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

are continuous, it follows that,

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

which completes the proof that F considered as $w\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_{\boldsymbol{z}}, \mathscr{H})$ has L_∞ support. Thus the proof of the Theorem is completed. $\square$

B.4 Proof of Theorem 4.4

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

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

where $KL(q,p) = \int_{\Delta_m} q(\boldsymbol{y}) \log\left(\frac{q(\boldsymbol{y})}{p(\boldsymbol{y})}\right) d\boldsymbol{y}$. Since $\mathscr{X}$ and Δ_m are compact sets and $(\boldsymbol{x},\boldsymbol{y}) \mapsto q_{\boldsymbol{x}}(\boldsymbol{y})$ is a continuous mapping, it follows that $\inf_{\boldsymbol{x}\in\mathscr{X}} \inf_{\boldsymbol{y}\in\Delta_m} q_{\boldsymbol{x}}(\boldsymbol{y})$ exists and is bounded. First, suppose that $\inf_{\boldsymbol{x}\in\mathscr{X}} \inf_{\boldsymbol{y}\in\Delta_m} q_{\boldsymbol{x}}(\boldsymbol{y}) > 0$. If for every $\epsilon > 0$, $\sup_{\boldsymbol{x}\in\mathscr{X}} \sup_{\boldsymbol{y}\in\Delta_m} |f(\boldsymbol{x},\omega)(\boldsymbol{y}) - q_{\boldsymbol{x}}(\boldsymbol{y})| < \epsilon$, then $\inf_{\boldsymbol{x}\in\mathscr{X}} \inf_{\boldsymbol{y}\in\Delta_m} f(\boldsymbol{x},\omega)(\boldsymbol{y}) > 0$ and for every $\epsilon' > 0$, there exists $\epsilon > 0$ such that for every $\boldsymbol{x} \in \mathscr{X}$ and every $\boldsymbol{y} \in \Delta_m$,

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

This in turn implies that $\sup_{\boldsymbol{x}\in\mathscr{X}} KL(q_{\boldsymbol{x}}, f(\boldsymbol{x},\omega)) < \epsilon'$. From Theorem 4.3, it follows that

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

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

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

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

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

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

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

which completes the proof. $\square$

B.5 Proof of Theorem 4.5

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

Lemma B.4. *Let $\mathscr{F} = \{f_{\boldsymbol{x}} : \boldsymbol{x} \in \mathscr{X}\}$ be a set of known biyective continuous functions such that for every $\boldsymbol{x} \in \mathscr{X}$, $f_{\boldsymbol{x}} : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ is such that for every $\boldsymbol{a} \in \mathbb{R}^m$, $f_{\boldsymbol{x}}(\boldsymbol{a})$ is a continuous functions of $\boldsymbol{x}$. In addition asume that $\mathscr{F}$ is an equicontinuous family of functions of $\boldsymbol{a}$ or $\{\boldsymbol{x} \mapsto f_{\boldsymbol{x}}(\boldsymbol{a}) : \boldsymbol{a} \in \mathbb{R}^m, f_x \in \mathscr{F}\}$ is an equicontinuous family of functions of $\boldsymbol{x}$. Let $g_i : \mathscr{X} \times \Omega \longrightarrow \mathbb{R}^m$, $i \geq 1$, be stochastic processes defined on an appropiate probability space $(\Omega, \mathscr{A}, P)$.*

(i) If for every $i \in \mathbb{N}$, the stochastic process g_i is P-a.s. continuous, then $\boldsymbol{x} \mapsto f_{\boldsymbol{x}}\{g_i(\boldsymbol{x}, \cdot)\}$, $i \in \mathbb{N}$ is P-a.s. continuous.

(ii) Consider $\{\boldsymbol{x}_j\}_{j=1}^{\infty} \subset \mathscr{X}$, such that $\lim_{j\to\infty}\boldsymbol{x}_j = \boldsymbol{x}_0 \in \mathscr{X}$. If $g_i(\boldsymbol{x}_j, \cdot) \xrightarrow{\mathscr{L}} g_i(\boldsymbol{x}_0, \cdot)$, as $j \longrightarrow \infty$, then $f_{\boldsymbol{x}_j}\{g_i(\boldsymbol{x}_j, \cdot)\} \xrightarrow{\mathscr{L}} f_{\boldsymbol{x}_0}\{g_i(\boldsymbol{x}_0, \cdot)\}$, as $j \longrightarrow \infty$.

Proof of Lemma B.4

(i) First consider, for every $\boldsymbol{x} \in \mathscr{X}$, $f_{\boldsymbol{x}}$ an equicontinuous of function of $\boldsymbol{a}$. Consider $\boldsymbol{x}_0 \in \mathscr{X}$. Since $f_{\boldsymbol{x}}(g_i(\boldsymbol{x}_0, \omega))$ is a continuous function of $\boldsymbol{x}$, there exists $\delta_1 > 0$ such that for all $\boldsymbol{x} \in B(\boldsymbol{x}_0, \delta_1)$, $|f_{\boldsymbol{x}}(g_i(\boldsymbol{x}_0, \omega)) - f_{\boldsymbol{x}_0}(g_i(\boldsymbol{x}_0, \omega))| < \epsilon/2$. By assumption, g_i, being a P-a.s. continuous stochastic process, implies that, for almost every $\omega \in \Omega$, and for every $\epsilon_2 > 0$, there exists $\delta_2 > 0$ such that for all $\boldsymbol{x} \in B(\boldsymbol{x}_0, \delta_2)$, $|g_i(\boldsymbol{x}, \omega) - g_i(\boldsymbol{x}_0, \omega)| < \epsilon_2$. Hence, by the equicontinuity of $f_{\boldsymbol{x}}$, for almost every $\omega \in \Omega$, and every $g_i(\boldsymbol{x}, \omega) \in B(g_i(\boldsymbol{x}_0, \omega), \epsilon_2)$, $|f_{\boldsymbol{x}}(g_i(\boldsymbol{x}, \omega)) - f_{\boldsymbol{x}}(g_i(\boldsymbol{x}_0, \omega))| < \epsilon/2$. Finally, considering $\delta = \min\{\delta_1, \delta_2\}$ which does not depend on $f_{\boldsymbol{x}}$, by the triangle inequality, it follows that for every $\omega \in \Omega$ and every $\boldsymbol{x} \in B(\boldsymbol{x}_0, \delta)$, $|f_{\boldsymbol{x}}\{g_i(\boldsymbol{x}, \omega)\} - f_{\boldsymbol{x}_0}\{g_i(\boldsymbol{x}_0, \omega)\}| < \epsilon$, which completes this part of the proof.

Now consider $\{\boldsymbol{x} \mapsto f_{\boldsymbol{x}}(\boldsymbol{a}) : \boldsymbol{a} \in \mathbb{R}^m, f_x \in \mathscr{F}\}$ an equicontinuous family of functions of

$\boldsymbol{x}$. The proof is similar to the previous. By the equicontinuity consideration, there exists $\delta_1 > 0$ such that for every $\boldsymbol{x} \in B(\boldsymbol{x}_0, \delta_1)$, $|f_{\boldsymbol{x}}(g_i(\boldsymbol{x}, \omega)) - f_{\boldsymbol{x}_0}(g_i(\boldsymbol{x}, \omega))| < \epsilon/2$. Since g_i is a P-a.s. continuous stochastic process, for almost every $\omega \in \Omega$, and for every $\epsilon_2 > 0$, there exists $\delta_2 > 0$ such that for every $\boldsymbol{x} \in B(\boldsymbol{x}_0, \delta_2)$, $|g_i(\boldsymbol{x}, \omega) - g_i(\boldsymbol{x}_0, \omega)| < \epsilon_2$. Due to continuity of $f_{\boldsymbol{x}}$ as a function for $\boldsymbol{a}$, it follows that for almost every $\omega \in \Omega$, and every $g_i(\boldsymbol{x}, \omega) \in B(g_i(\boldsymbol{x}_0, \omega), \epsilon_2)$, $|f_{\boldsymbol{x}_0}(g_i(\boldsymbol{x}, \omega)) - f_{\boldsymbol{x}_0}(g_i(\boldsymbol{x}_0, \omega))| < \epsilon/2$. Finally, considering $\delta = \min\{\delta_1, \delta_2\}$ which does not depend on $f_{\boldsymbol{x}}$, by the triangle inequality, it follows that for every $\omega \in \Omega$ and every $\boldsymbol{x} \in B(\boldsymbol{x}_0, \delta)$, $|f_{\boldsymbol{x}}\{g_i(\boldsymbol{x}, \omega)\} - f_{\boldsymbol{x}_0}\{g_i(\boldsymbol{x}_0, \omega)\}| < \epsilon$, which completes the proof of the first part of the Lemma. $\square$

(ii) Consider $\mathscr{F}$ an equicontinuous family of functions of $\boldsymbol{a}$ or $\{\boldsymbol{x} \mapsto f_{\boldsymbol{x}}(\boldsymbol{a}) : \boldsymbol{a} \in \mathbb{R}^m, f_x \in \mathscr{F}\}$ an equicontinuous family of functions of $\boldsymbol{x}$. If $g_i(\boldsymbol{x}_j, \cdot) \xrightarrow{\mathscr{L}} g_i(\boldsymbol{x}_0, \cdot)$, as $j \longrightarrow \infty$, then by baby Skorohod's theorem (Resnick, 2013), there exist random variables $\{\tilde{g}_i(\boldsymbol{x}_j, \cdot)\}_{j \geq 0}$ defined on the Lebesgue probability space $([0,1], \mathscr{B}([0,1]), \lambda)$, where λ is the Lebesgue measure, such that for each fixed $j \geq 0$, $g_i(\boldsymbol{x}_j, \cdot) \stackrel{d}{=} \tilde{g}_i(\boldsymbol{x}_j, \cdot)$, and $\tilde{g}_i(\boldsymbol{x}_j, \cdot) \longrightarrow \tilde{g}_i(\boldsymbol{x}_0, \cdot)$ λ-a.s. as $j \longrightarrow \infty$. Since $f_{\boldsymbol{x}}(\boldsymbol{a})$ is a continuous function of $\boldsymbol{a}$, it follows that for $\boldsymbol{x} \in \mathscr{X}$, $f_{\boldsymbol{x}}\{g_i(\boldsymbol{x}_j, \cdot)\} \stackrel{d}{=} f_{\boldsymbol{x}}\{\tilde{g}_i(\boldsymbol{x}_j, \cdot)\}$. In particular, $f_{\boldsymbol{x}_j}\{g_i(\boldsymbol{x}_j, \cdot)\} \stackrel{d}{=} f_{\boldsymbol{x}_j}\{\tilde{g}_i(\boldsymbol{x}_j, \cdot)\}$ and $f_{\boldsymbol{x}_0}\{g_i(\boldsymbol{x}_j, \cdot)\} \stackrel{d}{=} f_{\boldsymbol{x}_0}\{\tilde{g}_i(\boldsymbol{x}_j, \cdot)\}$. Since $\tilde{g}_i(\boldsymbol{x}_j, \cdot) \longrightarrow \tilde{g}_i(\boldsymbol{x}_0, \cdot)$ λ-a.s. as $j \longrightarrow \infty$ then $\tilde{g}_i$ is λ-a.s continuous. Therefore, by Lemma B.4 part *(i)*, $f_{\boldsymbol{x}_j}\{\tilde{g}_i(\boldsymbol{x}_j, \cdot)\} \longrightarrow f_{\boldsymbol{x}_0}\{\tilde{g}_i(\boldsymbol{x}_0, \cdot)\}$ λ-a.s. as $j \longrightarrow \infty$ which implies that $f_{\boldsymbol{x}_j}\{\tilde{g}_i(\boldsymbol{x}_j, \cdot)\} \xrightarrow{\mathscr{L}} f_{\boldsymbol{x}_0}\{\tilde{g}_i(\boldsymbol{x}_0, \cdot)\}$. Thus

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

as $j \longrightarrow \infty$, which completes proof of this part of the Lemma. $\square$

Now we provide the proof of the theorem. Firstly, assume that F is a $\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_z, \mathscr{V}, \mathscr{H})$. Since the elements of $\mathscr{V}$ and $\mathscr{H}$ are equicontinuous functions of $\boldsymbol{x}$, and for every $i \geq 1$, the stochastic processes η_i and $\boldsymbol{z}_i$ are P–a.s. continuous, by Lemma B.4 and continuous mapping theorem, it follows that $\boldsymbol{x} \mapsto v_{\boldsymbol{x}}(\eta_i(\boldsymbol{x}, \cdot))$, $\boldsymbol{x} \mapsto w_i(\boldsymbol{x}, \cdot)$, and $\boldsymbol{x} \mapsto \boldsymbol{\theta}_i(\boldsymbol{x}, \cdot\ )$ are P–a.s. continuous mappings. Now, the ceiling function being continuous from the left and having a limit from the right implies that, for $i \geq 1$, and almost every $\omega \in \Omega$, $\lceil k(\omega)\boldsymbol{\theta}_i(\boldsymbol{x}_l, \omega)\rceil$ has a limit, as $l \longrightarrow \infty$. Note that there exists $M > 0$ such that, for every $\boldsymbol{y} \in \Delta_m$, $i \geq 1$, $\boldsymbol{x} \in \mathscr{X}$ and $\omega \in \Omega$, $\mathrm{d}(\boldsymbol{y} \mid \alpha\left(k(\omega), \lceil k(\omega)\boldsymbol{\theta}_i(\boldsymbol{x}, \omega)\rceil\right)) \leq M$, where $\alpha(k,\ \mathbf{j}) = (\mathbf{j},\ k + m - \sum_{l=1}^{m} j_l)$, and that for every $\boldsymbol{x} \in \mathscr{X}$ and $\omega \in \Omega$, $\sum_{i=1}^{\infty} w_i(\boldsymbol{x}, \omega) = 1$. Then by dominated convergence theorem for series, the density, w.r.t. Lebesgue measure, of $F(\boldsymbol{x}, \cdot)$, $f(\boldsymbol{x}_l, \omega)$, has $a.s.$ a limit, say $\tilde{f}(\boldsymbol{x}_0, \cdot)$. This is, for every $\boldsymbol{y} \in \Delta_m$,

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

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

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

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

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

The proof of the theorem is thus completed. $\square$

B.6 Proof of Theorem 4.6

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

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

By assumption, for every $i \geq 1$, the stochastic processes η_i are $a.s.$ continuous, i.e., for every $i \geq 1$, $\boldsymbol{x} \longmapsto \eta_i(\boldsymbol{x},\cdot)$ is an $a.s.$ continuous function. By Lemma B.4, the equicontinuity assumption of $\mathscr{V}$ as a function of $\boldsymbol{x}$, and continuous mapping theorem, it follows that for every $i \geq 1$, $\boldsymbol{x} \longmapsto w_i(\boldsymbol{x},\cdot)$ is an $a.s.$ continuous function. Therefore for every $i \geq 1$ and every $\{\boldsymbol{x}_l\}_{l=0}^{\infty}$, such that $\lim_{l\to\infty} \boldsymbol{x}_l = \boldsymbol{x}_0$, we have that $\lim_{l\to\infty} w_i(\boldsymbol{x}_l,\cdot) = w_i(\boldsymbol{x}_0,\cdot)$, $a.s.$. Noting that there exists $M > 0$ such that, for every $\boldsymbol{y} \in \Delta_m$, $i \geq 1$, and $\omega \in \Omega$, $\mathrm{d}(\boldsymbol{y} \mid \alpha\left(k(\omega), \lceil k(\omega)\boldsymbol{\theta}_i(\omega)\rceil\right)) \leq M$, and that for every $\boldsymbol{x} \in \mathscr{X}$ and $\omega \in \Omega$, $\sum_{i=1}^{\infty} w_i(\boldsymbol{x},\omega) = 1$, dominated convergence theorem for series implies that the density, w.r.t. Lebesgue measure, of $F(\boldsymbol{x},\cdot)$ is $a.s.$ continuos, i.e., for every $\boldsymbol{y} \in \Delta_m$,

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

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

B.7 Proof of Theorem 4.7

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

$$\lim_{l\to\infty} \frac{E\left\{F(\boldsymbol{x}_l,\cdot)(B_{\boldsymbol{y}})F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}})\right\} - E\left\{F(\boldsymbol{x}_l,\cdot)(B_{\boldsymbol{y}})\right\} E\left\{F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}})\right\}}{\sqrt{Var\left\{F(\boldsymbol{x}_l,\cdot)(B_{\boldsymbol{y}})\right\} Var\left\{F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}})\right\}}} = 1, \tag{B.11}$$

where $B_{\boldsymbol{y}} = [0, y_1] \times \ldots \times [0, y_m]$, and expectations are obtained by the law of total expectation conditioning on the degree of the polynomial. The proof is developed for the three definitions of *F*. In order to reduce the notation, $k(\omega)$ is denoted by k when necessary.

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

$$E\left\{F(\boldsymbol{x},\cdot)(B_{\boldsymbol{y}})\right| k = k_0\} = \sum_{\mathbf{j}\in\mathcal{H}_{k_0,m}} E\left\{F^*(\boldsymbol{x},\cdot)\left(\frac{\mathbf{j}}{k}\right)\middle| k = k_0\right\} \mathrm{Mult}(\mathbf{j} \mid k_0 + m - 1, \boldsymbol{y}),$$

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

$$F^*(\boldsymbol{x},\cdot)\left(\frac{\mathbf{j}}{k}\right) = \sum_{i=1}^{\infty} w_i(\boldsymbol{x},\cdot)\mathbb{I}\left\{\theta_i(\boldsymbol{x},\cdot)\right\}_{\left\{\theta_{i1}(\boldsymbol{x},\cdot)\leq\frac{j_1}{k},\ldots,\theta_{im}(\boldsymbol{x},\cdot)\leq\frac{j_m}{k}\right\}}.$$

Since the stochastic processes $\{\eta_i\}_{i\geq1}$ and $\{\boldsymbol{z}_i\}_{i\geq1}$ are independent and indentically distributed, it follows that,

$$\begin{aligned}
E\left\{F^*(\boldsymbol{x},\cdot)\left(\frac{\mathbf{j}}{k}\right)\middle| k = k_0\right\} &= \sum_{i=1}^{\infty} E\left\{w_i(\boldsymbol{x},\cdot)\mathbb{I}\left\{\boldsymbol{\theta}_i(\boldsymbol{x},\cdot)\right\}_{\left\{\theta_{i1}(\boldsymbol{x},\cdot)\leq\frac{j_1}{k_0},\ldots,\theta_{im}(\boldsymbol{x},\cdot)\leq\frac{j_m}{k_0}\right\}}\right\},\\
&= \sum_{i=1}^{\infty} E\left\{w_i(\boldsymbol{x},\cdot)\right\} E\left\{\mathbb{I}\left\{\boldsymbol{\theta}_i(\boldsymbol{x},\cdot)\right\}_{\left\{\theta_{i1}(\boldsymbol{x},\cdot)\leq\frac{j_1}{k_0},\ldots,\theta_{im}(\boldsymbol{x},\cdot)\leq\frac{j_m}{k_0}\right\}}\right\},\\
&= \sum_{i=1}^{\infty} E\left\{w_i(\boldsymbol{x},\cdot)\right\} E\left\{\mathbb{I}\left\{\theta_1(\boldsymbol{x},\cdot)\right\}_{\left\{\theta_{i1}(\boldsymbol{x},\cdot)\leq\frac{j_1}{k_0},\ldots,\theta_{im}(\boldsymbol{x},\cdot)\leq\frac{j_m}{k_0}\right\}}\right\},\\
&= G_{0,\boldsymbol{x}}\left(A_{\mathbf{j},k_0}\right),
\end{aligned}$$

where $A_{\mathbf{j},k_0} = [0, j_1/k_0] \times \ldots \times [0, j_m/k_0]$ and $G_{0,\boldsymbol{x}}(A) = G_0(\boldsymbol{x},\cdot)(A)$ denotes the distribution function of $\boldsymbol{\theta}_1(\boldsymbol{x},\cdot)$ defined on $\tilde{\Delta}_m$. Thus

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

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

$$
\begin{aligned}
&E\left\{F(\boldsymbol{x},\cdot)(B_{\boldsymbol{y}})F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}}) \middle| k=k_0\right\} \\
&\qquad = \sum_{\substack{\mathbf{j}_1 \in \mathcal{H}_{k,m},\\ \mathbf{j}_2 \in \mathcal{H}_{k,m}}} E\left\{F^*(\boldsymbol{x},\boldsymbol{x}_0,\cdot)\left(\frac{\mathbf{j}_1}{k},\frac{\mathbf{j}_2}{k}\right)\middle| k=k_0\right\} \times \bar{\mathrm{M}}(\mathbf{j}_1,\mathbf{j}_2 \mid k+m-1,\boldsymbol{y}),
\end{aligned}
$$

where $\bar{M}(\mathbf{j}_1,\mathbf{j}_2 \mid k+m-1,\boldsymbol{y}) = \mathrm{Mult}(\mathbf{j}_1 \mid k+m-1,\boldsymbol{y}) \times \mathrm{Mult}(\mathbf{j}_2 \mid k+m-1,\boldsymbol{y})$, and

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

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

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

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

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

where $A_{\min\{\mathbf{j}_1,\mathbf{j}_2\},k} = [0,\min\{j_{11},j_{21}\}/k]\times\ldots\times[0,\min\{j_{1m},j_{2m}\}/k]$. By assumption, for every $i\geq 1$, and every $\{\boldsymbol{x}_l\}_{l=0}^{\infty}$, with $\boldsymbol{x}_l\in\mathscr{X}$, such that $\lim_{l\to\infty}\boldsymbol{x}_l=\boldsymbol{x}_0$, the processes $\eta_i(\boldsymbol{x}_l,\cdot)$ and $\boldsymbol{z}_i(\boldsymbol{x}_l,\cdot)$ converge in distribution to $\eta_i(\boldsymbol{x}_0,\cdot)$ and $\boldsymbol{z}_i(\boldsymbol{x}_0,\cdot)$, respectively, as $l\to\infty$. Since $\mathscr{V}$ and $\mathscr{H}$ are sets of equicontinuous functions of $\boldsymbol{x}$, by Lemma B.4, and continuous mapping theorem, it follows that $w_i(\boldsymbol{x}_l,\cdot)$ converges in distribution to $w_i(\boldsymbol{x}_0,\cdot)$ and $\boldsymbol{\theta}_i(\boldsymbol{x}_l,\cdot)$ converges in distribution to $\boldsymbol{\theta}_i(\boldsymbol{x}_0,\cdot)$, as $l\to\infty$. Thus, for every $a\in\tilde{\Delta}_m$, $\lim_{l\to\infty} G_{0,\boldsymbol{x}_l}(a) = G_{0,\boldsymbol{x}_0}(a)$. Noting that $w_i(\boldsymbol{x},\cdot)$ are bounded variables, Portmanteau‘s theorem implies that the mappings $\boldsymbol{x}\mapsto E\{w_i(\boldsymbol{x},\cdot)\}$, $\boldsymbol{x}\mapsto E\{w_i(\boldsymbol{x},\cdot)^2\}$ and $\boldsymbol{x}\mapsto E\{w_i(\boldsymbol{x},\cdot)w_i(\boldsymbol{x}_0,\cdot)\}$, are continuous. Now, considering $\boldsymbol{y}\in\tilde{\Delta}_m$, the above expressions and few applications of dominated convergence theorem for series, it follows that,

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

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

and

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

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

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

$$
E\left\{F(\boldsymbol{x},\cdot)(B_{\boldsymbol{y}})\right| k=k_0\} = \sum_{\mathbf{j}\in\mathcal{H}_{k_0,m}} E\left\{F^*(\boldsymbol{x},\cdot)\left(\frac{\mathbf{j}}{k}\right)\middle| k=k_0\right\} \mathrm{Mult}(\mathbf{j}\mid k_0+m-1,\boldsymbol{y}),
$$

where $\mathcal{H}_{k,m}$, $(\mathbf{j}/k)$, and $\mathrm{Mult}(\cdot\mid k_0+m-1,\boldsymbol{y})$ are defined as in the first part of this proof, and

$$
F^*(\boldsymbol{x},\cdot)\left(\frac{\mathbf{j}}{k}\right) = \sum_{i=1}^{\infty} w_i(\boldsymbol{x},\cdot)\mathbb{I}\left\{\theta_i(\cdot)\right\}_{\left\{\theta_{i1}\le\frac{j_1}{k},\ldots,\theta_{im}\le\frac{j_m}{k}\right\}}.
$$

Since $\{\boldsymbol{\theta}_i\}_{i\geq 1}$ are indentically distributed and independent of the stochastic processes $\{\eta_i\}_{i\geq 1}$, it follows that

$$\begin{aligned}
E\left\{F^*(\boldsymbol{x},\cdot)\left(\frac{\mathbf{j}}{k}\right)\middle| k=k_0\right\} &= \sum_{i=1}^{\infty} E\left\{w_i(\boldsymbol{x},\cdot)\mathbb{I}\left\{\boldsymbol{\theta}_i(\cdot)\right\}_{\left\{\boldsymbol{\theta}_{i1}\leq\frac{j_1}{k_0},\ldots,\theta_{im}\leq\frac{j_m}{k_0}\right\}}\right\},\\
&= \sum_{i=1}^{\infty} E\left\{w_i(\boldsymbol{x},\cdot)\right\} E\left\{\mathbb{I}\left\{\theta_1(\cdot)\right\}_{\left\{\theta_{i1}\leq\frac{j_1}{k_0},\ldots,\boldsymbol{\theta}_{im}\leq\frac{j_m}{k_0}\right\}}\right\},\\
&= G_0\left(A_{\mathbf{j},k_0}\right),
\end{aligned}$$

where $A_{\mathbf{j},k_0}$ is defined as in the first part of the proof and G_0 is the distribution function of $\boldsymbol{\theta}_1$. Thus

$$E\left\{F(\boldsymbol{x},\cdot)(B_{\boldsymbol{y}}) \mid k=k_0\right\} = \sum_{\mathbf{j}\in\mathcal{H}_{k_0,m}} G_0\left(A_{\mathbf{j},k_0}\right)\mathrm{Mult}(\mathbf{j}\mid k_0+m-1,\boldsymbol{y}).$$

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

$$\begin{aligned}
&E\left\{F(\boldsymbol{x},\cdot)(B_{\boldsymbol{y}})F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}})\middle| k=k_0\right\}\\
&\qquad= \sum_{\substack{\mathbf{j}_1\in\mathcal{H}_{k,m},\\ \mathbf{j}_2\in\mathcal{H}_{k,m}}} E\left\{F^*(\boldsymbol{x},\boldsymbol{x}_0,\cdot)\left(\frac{\mathbf{j}_1}{k},\frac{\mathbf{j}_2}{k}\right)\middle| k=k_0\right\}\times\bar{\mathrm{M}}(\mathbf{j}_1,\mathbf{j}_2\mid k+m-1,\boldsymbol{y}),
\end{aligned}$$

where $\bar{M}(\mathbf{j}_1,\mathbf{j}_2\mid k+m-1,\boldsymbol{y})$ is defined as in the first part of the proof, and

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

where $A_{\min\{\mathbf{j}_1,\mathbf{j}_2\},k}$ is defined as in the first part of the proof. Thus, it follows that for every $\boldsymbol{x}$, $\boldsymbol{x}_0 \in \mathscr{X}$ and every $\boldsymbol{y} \in \Delta_m$,

$$
\begin{aligned}
&E\left\{F(\boldsymbol{x},\cdot)(B_{\boldsymbol{y}})F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}}) \mid k=k_0\right\} \\
&\qquad = \sum_{\substack{\mathbf{j}_1\in\mathcal{H}_{k_0,m}\\ ,\mathbf{j}_2\in\mathcal{H}_{k_0,m}}} \left\{\sum_{i=1}^{\infty} E\left\{w_i(\boldsymbol{x},\cdot)w_i(\boldsymbol{x}_0,\cdot)\right\} G_0\left(A_{\min\{\mathbf{j}_1,\mathbf{j}_2\},\mathrm{k}_0}\right),\right. \\
&\qquad\qquad \left. + \sum_{\substack{i,i_1=1,\\ i\neq i_1}}^{\infty} E\left\{w_i(\boldsymbol{x},\cdot)w_{i_1}(\boldsymbol{x}_0,\cdot)\right\} G_0\left(A_{\mathbf{j}_2,k_0}\right) G_0\left(A_{\mathbf{j}_2,k_0}\right)\right\}, \\
&\qquad\qquad \times \bar{M}(\mathbf{j}_1,\mathbf{j}_2 \mid k_0+m-1,\boldsymbol{y}).
\end{aligned}
$$

In particular, for $\boldsymbol{x}=\boldsymbol{x}_0$,

$$
\begin{aligned}
&E\left\{F(\boldsymbol{x},\cdot)(B_{\boldsymbol{y}})^2 \;\middle|\; k=k_0\right\} \\
&\qquad = \sum_{\substack{\mathbf{j}_1\in\mathcal{H}_{k_0,m}\\ ,\mathbf{j}_2\in\mathcal{H}_{k_0,m}}} \left\{\sum_{i=1}^{\infty} E\left\{w_i(\boldsymbol{x},\cdot)^2\right\} G_0\left(A_{\min\{\mathbf{j}_1,\mathbf{j}_2\},\mathrm{k}_0}\right),\right. \\
&\qquad\qquad \left. + \sum_{\substack{i,i_1=1,\\ i\neq i_1}}^{\infty} E\left\{w_i(\boldsymbol{x},\cdot)w_{i_1}(\boldsymbol{x},\cdot)\right\} G_0\left(A_{\mathbf{j}_2,k_0}\right) G_0\left(A_{\mathbf{j}_2,k_0}\right)\right\}, \\
&\qquad\qquad \times \bar{M}(\mathbf{j}_1,\mathbf{j}_2 \mid k_0+m-1,\boldsymbol{y}).
\end{aligned}
$$

By the same arguments used when $\boldsymbol{F}$ as $\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_{\boldsymbol{z}}, \mathscr{V}, \mathscr{H})$ was considered, it follows that

$$
\lim_{j\to\infty} E\left\{F(\boldsymbol{x}_j,\cdot)(B_{\boldsymbol{y}})^2\right\} = E\left\{F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}})^2\right\},
$$

$$
\lim_{j\to\infty} E\left\{F(\boldsymbol{x}_j,\cdot)(B_{\boldsymbol{y}})F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}})\right\} = E\left\{F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}})^2\right\},
$$

and

$$\lim_{j\to\infty} E\left\{F(\boldsymbol{x}_j,\cdot)(B_{\boldsymbol{y}})\right\} = E\left\{F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}})\right\},$$

which completes this part of the proof.

Finally, assume that F is a $w\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_z, \mathscr{H})$. Notice that for every $\boldsymbol{y} \in \Delta_m$ and every $\boldsymbol{x} \in \mathscr{X}$,

$$E\left\{F(\boldsymbol{x},\cdot)(B_{\boldsymbol{y}}) \middle| \, k=k_0\right\} = \sum_{\mathbf{j}\in\mathcal{H}_{k_0,m}} E\left\{F^*(\boldsymbol{x},\cdot)\left(\frac{\mathbf{j}}{k}\right)\middle|\, k=k_0\right\}\mathrm{Mult}(\mathbf{j}\mid k_0+m-1,\boldsymbol{y}),$$

where

$$F^*(\boldsymbol{x},\cdot)\left(\frac{\mathbf{j}}{k}\right) = \sum_{i=1}^{\infty} w_i(\cdot)\mathbb{I}\left\{\theta_i(\boldsymbol{x},\cdot)\right\}_{\left\{\boldsymbol{\theta}_{i1}(\boldsymbol{x},\cdot)\le\frac{j_1}{k},\ldots,\theta_{im}(\boldsymbol{x},\cdot)\le\frac{j_m}{k}\right\}}.$$

Since for every $i \geq 1$, the stochastic processes $\boldsymbol{z}_i$ are identically distributed and independent of v_i, it follows that

$$\begin{aligned} E\left\{F^*(\boldsymbol{x},\cdot)\left(\frac{\mathbf{j}}{k}\right)\middle|\, k=k_0\right\} &= \sum_{i=1}^{\infty} E\left\{w_i(\cdot)\right\} E\left\{\mathbb{I}\left\{\boldsymbol{\theta}_1(\boldsymbol{x},\cdot)\right\}_{\left\{\theta_{11}(\boldsymbol{x},\cdot)\le\frac{j_1}{k_0},\ldots,\theta_{1m}(\boldsymbol{x},\cdot)\le\frac{j_m}{k_0}\right\}}\right\}, \\ &= G_{0,\boldsymbol{x}}\left(A_{\mathbf{j},k_0}\right). \end{aligned}$$

Thus

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

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

$$\begin{aligned} &E\left\{F(\boldsymbol{x},\cdot)(B_{\boldsymbol{y}})F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}}) \middle| \, k=k_0\right\} \\ &\qquad = \sum_{\substack{\mathbf{j}_1\in\mathcal{H}_{k,m},\\ \mathbf{j}_2\in\mathcal{H}_{k,m}}} E\left\{F^*(\boldsymbol{x},\boldsymbol{x}_0,\cdot)\left(\frac{\mathbf{j}_1}{k},\frac{\mathbf{j}_2}{k}\right)\middle|\, k=k_0\right\}\times\bar{\mathrm{M}}(\mathbf{j}_1,\mathbf{j}_2\mid k+m-1,\boldsymbol{y}), \end{aligned}$$

and

$$F^*(\boldsymbol{x}, \boldsymbol{x}_0, \cdot)\left(\frac{\mathbf{j}_1}{k}, \frac{\mathbf{j}_2}{k}\right) = \sum_{i=1}^{\infty} w_i(\cdot)^2 \mathbb{I}\left\{\boldsymbol{\theta}_i(\boldsymbol{x}, \cdot)\right\}_{\left\{A_{\mathbf{j}_1,k}\right\}} \mathbb{I}\left\{\boldsymbol{\theta}_i(\boldsymbol{x}_0, \cdot)\right\}_{\left\{A_{\mathbf{j}_2,k}\right\}},$$
$$+ \sum_{\substack{i,i_1=1,\\ i\neq i_1}}^{\infty} w_i(\cdot) w_{i_1}(\cdot) \mathbb{I}\left\{\boldsymbol{\theta}_i(\boldsymbol{x}, \cdot)\right\}_{\left\{A_{\mathbf{j}_1,k}\right\}} \mathbb{I}\left\{\boldsymbol{\theta}_{i_1}(\boldsymbol{x}_0, \cdot)\right\}_{\left\{A_{\mathbf{j}_2,k}\right\}},$$

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

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

In particular, for $\boldsymbol{x} = \boldsymbol{x}_0$,

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

By the same arguments as in the first part of the proof, it follows that

$$\lim_{j\to\infty} E\left\{F(\boldsymbol{x}_j, \cdot)(B_{\boldsymbol{y}})^2\right\} = E\left\{F(\boldsymbol{x}_0, \cdot)(B_{\boldsymbol{y}})^2\right\},$$

$$\lim_{j\to\infty} E\left\{F(\boldsymbol{x}_j,\cdot)(B_{\boldsymbol{y}})F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}})\right\} = E\left\{F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}})^2\right\},$$

and

$$\lim_{j\to\infty} E\left\{F(\boldsymbol{x}_j,\cdot)(B_{\boldsymbol{y}})\right\} = E\left\{F(\boldsymbol{x}_0,\cdot)(B_{\boldsymbol{y}})\right\}.$$

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

B.8 Proof of Theorem 4.8

Here we use the law of total covariance conditioning on the degree of the polynomial. Assume now that F is a $\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_z, \mathscr{V}, \mathscr{H})$. By assumption, for every $i \geq 1$, and every $\{(\boldsymbol{x}_{1l}, \boldsymbol{x}_{2l})\}_{l=1}^{\infty}$ with $(\boldsymbol{x}_{1l}, \boldsymbol{x}_{2l}) \in \mathscr{X}^2$, $(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathscr{X}^2$, such that $\lim_{l\to\infty}(\boldsymbol{x}_{1l}, \boldsymbol{x}_{2l}) = (\boldsymbol{x}_1, \boldsymbol{x}_2)$, the joint processes $(\eta_i(\boldsymbol{x}_{1l},\cdot), \eta_i(\boldsymbol{x}_{2l},\cdot))$ and $(\boldsymbol{z}_i(\boldsymbol{x}_{1l},\cdot), \boldsymbol{z}_i(\boldsymbol{x}_{2l},\cdot))$ converge in distribution to $(\eta_i(\boldsymbol{x}_1,\cdot), \eta_i(\boldsymbol{x}_2,\cdot))$, and $(\boldsymbol{z}_i(\boldsymbol{x}_1,\cdot), \boldsymbol{z}_i(\boldsymbol{x}_2,\cdot))$, as $l \to \infty$, respectively. Since $\mathscr{V}$ and $\mathscr{H}$ are sets of equicontinuous functions of $\boldsymbol{x}$, by Lemma B.4 and continuous mapping theorem, it follows that for every $i \geq 1$, $(w_i(\boldsymbol{x}_{1l},\cdot), w_i(\boldsymbol{x}_{2l},\cdot))$ and $(\boldsymbol{\theta}_i(\boldsymbol{x}_{1l},\cdot), \boldsymbol{\theta}_i(\boldsymbol{x}_{2l},\cdot))$ converge in distribution to $(w_i(\boldsymbol{x}_1,\cdot), w_i(\boldsymbol{x}_2,\cdot))$ and $(\boldsymbol{\theta}_i(\boldsymbol{x}_1,\cdot), \boldsymbol{\theta}_i(\boldsymbol{x}_2,\cdot))$, as $l \to \infty$, respectively. Thus, for every $\boldsymbol{a}_1 \in \tilde{\Delta}_m$ and $\boldsymbol{a}_2 \in \tilde{\Delta}_m$, $\lim_{l\to\infty} G_{0,\boldsymbol{x}_{1l},\boldsymbol{x}_{2l}}(\boldsymbol{a}_1, \boldsymbol{a}_2) = G_{0,\boldsymbol{x}_1,\boldsymbol{x}_2}(\boldsymbol{a}_1, \boldsymbol{a}_2)$, where $G_{0,\boldsymbol{x}_1,\boldsymbol{x}_2}$ denotes the joint distribution function of $(\boldsymbol{\theta}_i(\boldsymbol{x}_1,\cdot), \boldsymbol{\theta}_i(\boldsymbol{x}_2,\cdot))$. Noting that for every $\boldsymbol{x}$, $w_i(\boldsymbol{x},\cdot)$ are bounded variables, Portmanteau's theorem implies that mappings $\boldsymbol{x} \mapsto E\{w_i(\boldsymbol{x},\cdot)\}$ and $(\boldsymbol{x}_1, \boldsymbol{x}_2) \mapsto E\{w_i(\boldsymbol{x}_1,\cdot)w_{i_1}(\boldsymbol{x}_2,\cdot)\}$, $i, i_1 \in \mathbb{N}$, are continuous. In addition, for $(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathscr{X}^2$ such that $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| > \gamma$, the assumption

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

and

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

imply that

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

Therefore, for every $(\boldsymbol{x}_1,\boldsymbol{x}_2)\in\mathscr{X}^2$ such that $\|\boldsymbol{x}_1-\boldsymbol{x}_2\|>\gamma$, it follows that $G_{0,\boldsymbol{x}_1,\boldsymbol{x}_2}(\boldsymbol{a}_1,\boldsymbol{a}_2) = G_{0,\boldsymbol{x}_1}(\boldsymbol{a}_1)G_{0,\boldsymbol{x}_2}(\boldsymbol{a}_2)$, and $E\left\{w_i(\boldsymbol{x}_1,\cdot)w_{i_1}(\boldsymbol{x}_2,\cdot)\right\} = E\left\{w_i(\boldsymbol{x}_1,\cdot)\right\}E\left\{w_{i_1}(\boldsymbol{x}_2,\cdot)\right\}$, $i,i_1\in\mathbb{N}$. Now, considering the expressions from the proof of theorem 4.7, for every $\boldsymbol{y}\in\Delta_m$, $(\boldsymbol{x}_1,\boldsymbol{x}_2)\in\mathscr{X}^2$ such that $\|\boldsymbol{x}_1-\boldsymbol{x}_2\|>\gamma$, and an application of dominated convergence theorem, it follows that,

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

Thus,

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

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

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

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

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

which completes this part of the proof.

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

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

and

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

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

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

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

B.9 Proof of Theorem 4.9

Here we use the law of total covariance conditioning on the degree of the polynomial. Assume that F is a θDMBPP$(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$. By the same arguments as in the proof of the first part of Theorem 4.8, for every $\boldsymbol{y} \in \Delta_m$ and $(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathscr{X}^2$ such that $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| > \gamma$, and few applications of dominated convergence theorem, it follows that

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

and

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

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

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

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

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

which completes the proof of the theorem. □

B.10 Proof of Theorem 4.10

We proof this theorem using the definition of correlation. Expectations are obtained by the law of total expectation, conditioning on the degree of the polynomial. Assume that F is a $\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_\eta, \boldsymbol{\Psi}_z, \mathscr{V}, \mathscr{H})$, a $\theta\mathrm{MDBPP}(\lambda, \boldsymbol{\Psi}_\eta, \mathscr{V}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$ or a $w\mathrm{DMBPP}(\lambda, \boldsymbol{\Psi}_v, \boldsymbol{\Psi}_z, \mathscr{H})$. By assumption, for every $i\geq 1$, and every $\{(\boldsymbol{x}_{1l},\boldsymbol{x}_{2l})\}_{l=1}^{\infty}$ with $(\boldsymbol{x}_{1l},\boldsymbol{x}_{2l})\in\mathscr{X}^2$, $(\boldsymbol{x}_1,\boldsymbol{x}_2)\in\mathscr{X}^2$, such that $\lim_{l\to\infty}(\boldsymbol{x}_{1l},\boldsymbol{x}_{2l}) = (\boldsymbol{x}_1,\boldsymbol{x}_2)$, the joint processes $(\eta_i(\boldsymbol{x}_{1l},\cdot),\eta_i(\boldsymbol{x}_{2l},\cdot))$ and $(\boldsymbol{z}_i(\boldsymbol{x}_{1l},\cdot),\boldsymbol{z}_i(\boldsymbol{x}_{2l},\cdot))$ converge in distribution to $(\eta_i(\boldsymbol{x}_1,\cdot),\eta_i(\boldsymbol{x}_2,\cdot))$, and $(\boldsymbol{z}_i(\boldsymbol{x}_1,\cdot),\boldsymbol{z}_i(\boldsymbol{x}_2,\cdot))$, as $l\to\infty$, respectively. By the same arguments used in the proof of the first part of Theorem 4.8, it follows that for every $\boldsymbol{a}_1\in\tilde{\Delta}_m$ and $\boldsymbol{a}_2\in\tilde{\Delta}_m$, $\lim_{l\to\infty} G_{0,\boldsymbol{x}_{1l},\boldsymbol{x}_{2l}}(\boldsymbol{a}_1,\boldsymbol{a}_2) =$

$G_{0,\boldsymbol{x}_1,\boldsymbol{x}_2}(\boldsymbol{a}_1, \boldsymbol{a}_2)$, where $G_{0,\boldsymbol{x}_1,\boldsymbol{x}_2}$ denotes the joint distribution function of $(\boldsymbol{\theta}_i(\boldsymbol{x}_1, \cdot), \boldsymbol{\theta}_i(\boldsymbol{x}_2, \cdot))$, and mappings $\boldsymbol{x} \mapsto E\{w_i(\boldsymbol{x}, \cdot)\}$ and $(\boldsymbol{x}_1, \boldsymbol{x}_2) \mapsto E\{w_i(\boldsymbol{x}_1, \cdot)w_{i_1}(\boldsymbol{x}_2, \cdot)\}$, $i, i_1 \in \mathbb{N}$, are continuous. By few applications of dominated convergence theorem, it follows that for $m = 1, 2$,

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

and

$$\lim_{l\to\infty} E\{F(\boldsymbol{x}_{1l}, \cdot)(B_{\boldsymbol{y}})F(\boldsymbol{x}_{2l}, \cdot)(B_{\boldsymbol{y}})\} = E\{F(\boldsymbol{x}_1, \cdot)(B_{\boldsymbol{y}})F(\boldsymbol{x}_2, \cdot)(B_{\boldsymbol{y}})\}.$$

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

B.11 Proof of Theorem 4.11

Let $m(\boldsymbol{y}, \boldsymbol{x}) = q(\boldsymbol{x})g_{\boldsymbol{x}}(\boldsymbol{y})$ be the random joint distribution for the response and predictors arising when $\{g_{\boldsymbol{x}}(\boldsymbol{y}) : \boldsymbol{x} \in \mathscr{X}\}$ is a DMBPP, wDMBPP or θDMBPP. Since the KL divergence between m_0 and the implied joint distribution m can be bounded by the supremum over the predictor space of KL divergences between the predictor-dependent probability measures,

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

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

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

under the assumptions of Thorem 4.4. Thus, by Schwartz's theorem (Schwartz, 1965), it follows that the posterior distribution associated with the random joint distribution induced by any of the proposed models is weakly consistent, that is, the posterior measure of any weak neighborhood, of any joint distribution of the form $m_0(\boldsymbol{y}, \boldsymbol{x}) = q(\boldsymbol{x})q_0(\boldsymbol{y} \mid \boldsymbol{x})$, converges to one as the sample size goes to infinity. $\square$

Bibliography

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 nonparametric problems. *The Annals of Statistics* 2 1152–1174.

ASH, J. M. (2008). Mean value theorems for differences. *Elemente der Mathematik* 63 122–125.

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

BARNDORFF-NIELSEN, O. (1973). On M-ancillarity. *Biometrika* 60 447–455.

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

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.

BARRIENTOS, A. F., JARA, A. & QUINTANA, F. A. (2015a). Bayesian density estimation for compositional data using random Bernstein polynomials. *Journal of Statistical Planning and Inference* 166 116–125.

BARRIENTOS, A. F., JARA, A. & QUINTANA, F. A. (2015b). Fully nonparametric regression for bounded data using dependent Bernstein polynomials. Tech. rep., Department of Statistics, Pontificia Universidad Católica de Chile.

BARRIENTOS, A. F., JARA, A. & WEHRHAHN, C. (2015c). Posterior convergence rate of a class of Dirichlet process mixture model for compositional data. Tech. rep., Department of Statistics, Pontificia Universidad Católica de Chile.

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.

BOX, G. E. P. (1979). Robustness in the strategy of scientific model building. In R. L. Launer & G. N. Wilkinson, eds., *Robustness in Statistics*. Amsterdam, The Netherlands: Academic Press, 201–235.

CHEN, M. H. & SHAO, Q. M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics* 8 69–92.

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.

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.

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.

EPIFANI, I. & LIJOI, A. (2010). Nonparametric 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.

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

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.

GHOSAL, S. (2001). Convergence rates for density estimation with Bernstein polynomials. *The Annals of Statistics* 29 1264–1280.

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., GHOSH, J. K. & VAN DER VAART, A. W. (2000a). Convergence rates of posterior distributions. *The Annals of Statistics* 28 500–531.

GHOSAL, S., GHOSH, J. K. & VAN DER VAART, A. W. (2000b). Convergence rates of posterior distributions. *The Annals of Statistics* 28 500531.

GHOSAL, S. & VAN DER VAART, A. W. (2001). Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *The Annals of Statistics* 29 1233–1263.

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.

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. *The Annals of Applied Statistics* 4 2126–2149.

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

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.

KORWAR, R. M. & HOLLANDER, M. (1973). Contributions to the theory of Dirichlet processes. *The Annals of Probability* 1 705–711.

KRUIJER, W. & VAN DER VAART, A. W. (2008). Posterior convergence rates for Dirichlet mixtures of beta densities. *Journal of Statistical Planning and Inference* 138 1981–1992.

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. (2014). Bayesian inference with dependent normalized completely random measures. *Bernoulli* 20 1260–1291.

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.

LINDSAY, B. (1995). *Mixture Models: Theory, Geometry and Applications*. California, USA: IMS Hayward.

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.

MAULDIN, R. D., SUDDERTH, W. D. & WILLIAMS, S. C. (1992). Polya trees and random distributions. *Annals of Statistics* 20 1203–1221.

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* 140 2801–2808.

MÜLLER, P., QUINTANA, F. A., JARA, A. & E, H. T. (2015). *Bayesian Nonparametric Data Analysis*. New York, USA: Springer.

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. & VIDAKOVIC, B. (1998). Bayesian inference with wavelets: Density estimation. *Journal of Computational and Graphical Statistics* 7 456–468.

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.

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.

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

PATI, D., DUNSON, D. B. & TOKDAR, S. T. (2013). Posterior consistency in conditional distribution estimation. *Journal of Multivariate Analysis* 116 456–472.

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. *Chilean Journal of Statistics* 1 35–49.

REN, L., DU, L., CARIN, L. & DUNSON, D. B. (2011). Logistic stick-breaking process. *Journal of Machine Learning Research* 12 203–239.

RESNICK, S. I. (2013). *A Probability Path*. Springer Science & Business Media.

RIVOIRARD, V. & ROUSSEAU, J. (2012). Posterior concentration rates for infinite dimensional exponential families. *Bayesian Analysis* 7 311334.

RODRIGUEZ, A. & DUNSON, D. B. (2011). Nonparametric Bayesian models through probit stick-breaking processes. *Bayesian Analysis* 6 145–178.

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 prior. *Statistica Sinica* 2 639–650.

SHEN, X. & WASSERMAN, L. (2001). Rates of convergence of posterior distributions. *The Annals of Statistics* 29 687714.

SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. New York, USA: Chapman and Hall.

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.

TENBUSCH, A. (1994). Two-dimensional Bernstein polynomial density estimators. *Metrika* 41 233–253.

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.

WADE, S., WALKER, S. G. & PETRONE, S. (2014). A predictive study of Dirichlet process mixture models for curve fitting. *Scandinavian Journal of Statistics* 41 580–605.

WEHRHAHN, C., JARA, A. & BARRIENTOS, A. F. (2015). On the small sample behavior of Dirichlet process mixture models for data supported on compact intervals. Tech. rep., Department of Statistics, Pontificia Universidad Católica de Chile.

ZHENG, Y., ZHU, J. & ROY, A. (2010). Nonparametric bayesian inference for the spectral density function of a random field. *Biometrika* 97 238–245.