Journal of Statistical Planning and Inference 166 (2015) 116–125

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Bayesian density estimation for compositional data using random Bernstein polynomials

Andrés F. Barrientos, Alejandro Jara *, Fernando A. Quintana

Department of Statistics, Pontificia Universidad Católica de Chile, Casilla 306, Correo 22, Santiago, Chile

ARTICLE INFO

Article history:
Received 11 December 2013
Received in revised form 29 January 2015
Accepted 30 January 2015
Available online 14 February 2015

Keywords:
Simplex
Random Bernstein polynomials
Dirichlet process
Bayesian nonparametrics

ABSTRACT

We propose a Bayesian nonparametric procedure for density estimation for data in a d-dimensional simplex. To this aim, we propose a prior distribution on probability measures based on a modified class of multivariate Bernstein polynomials. The model for the probability distribution corresponds to a mixture of Dirichlet distributions, with random weights and a random number of components. Theoretical properties of the proposal are provided, including posterior consistency and concentration rates of the posterior distribution.

1. Introduction

This paper deals with the problem of Bayesian density estimation for data supported on the d-dimensional simplex $\Delta_d = \left\{(w_1, \ldots, w_d) \in [0,1]^d : \sum_{i=1}^d w_i \leq 1\right\}$, where methods based on transformations of the data and the normal kernel are susceptible to boundary effects. For data supported on a convex and compact set, the literature has mainly concentrated on bounded intervals and hyper-cubes. Motivated by its uniform approximation properties, frequentist and Bayesian methods based on univariate Bernstein polynomials and more general discrete mixtures of beta distributions have been proposed for density estimation for data supported on bounded intervals (Vitale, 1975; Petrone, 1999a,b; Kruijer and Van der Vaart, 2008; Rousseau, 2010). If $G : [0,1] \longrightarrow \mathbb{R}$, the associated Bernstein polynomial of degree k is given by

$$\sum_{j=0}^{k} G(j/k) \binom{k}{j} y^j (1-y)^{k-j}, \quad y \in [0,1]. \tag{1}$$

If G is the restriction of the cumulative density function (CDF) of a probability measure defined on the unit interval, then (1) 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

$$\sum_{j=1}^{k} w_{j,k} \beta(y \mid j, k-j+1), \tag{2}$$

where $w_{j,k} = G(j/k) - G((j-1)/k)$, and $\beta(\cdot|a,b)$ stands for a beta density with parameters a and b. Petrone (1999a,b) proposed a hierarchical prior for distribution functions on $[0,1]$, called the Bernstein polynomial prior (BPP). This consists

* Corresponding author.
E-mail addresses: afbarrie@uc.cl (A.F. Barrientos), atjara@uc.cl (A. Jara), quintana@mat.puc.cl (F.A. Quintana).

http://dx.doi.org/10.1016/j.jspi.2015.01.006
0378-3758/

of a random density given by expression (2), where k has probability mass function ρ, and given k, $\boldsymbol{w}_k = (w_{1,k}, \ldots, w_{k,k})$ has distribution H_k on the $(k-1)$-dimensional simplex. Petrone (1999a,b) referred to expression (2) as the Bernstein polynomial density with parameters k and $\boldsymbol{w}_k$, and showed that if ρ assigns positive mass to all naturals, and the density of H_k is positive for any point in Δ_k, then the weak support of the BPP is the space of all probability measures on $([0,1], \mathscr{B}([0,1]))$. Letting $\zeta_{j,k} = M\,(G_0(j/k) - G_0((j-1)/k))$, $j = 1, \ldots, k$, G_0 being a probability distribution on $(0,1]$ and M being a positive constant, Petrone (1999a,b) used the fact that assuming

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

is equivalent to assuming that G follows a Dirichlet process (DP) prior, $G \mid M,\ G_0 \sim DP(MG_0)$. Petrone (1999a,b) refers to the latter model as the Bernstein–Dirichlet prior (BDP), and discussed a Markov chain Monte Carlo algorithm to scan its posterior distribution. Petrone and Wasserman (2002) studied the consistency of the posterior distribution for the BPP and Ghosal (2001) provides rates of convergence for the BDP model. A different class of BPP has been considered by Trippa et al. (2011), where the prior distribution on the weights is defined by means of an auxiliary reinforced urn process.

Extensions based on multivariate Bernstein polynomials (MBP) defined on the unit hyper-cube have been also considered in the literature (see, e.g. Tenbusch, 1994; Babu and Chaubey, 2006; Zheng et al., 2010). Babu and Chaubey (2006) studied a general multivariate version of the bivariate estimator proposed by Tenbusch (1994). Zheng et al. (2010) construct a multivariate Bernstein polynomial prior for the spectral density of a random field. Multivariate extensions of Bernstein polynomials defined on Δ_d 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 CDF to Δ_2, and it is based on the class of MBP given in Definition 1.

Definition 1. For a given function $G : \Delta_d \longrightarrow \mathbb{R}$, the associated MBP of degree k on Δ_d is defined by

$$\widetilde{B}_{k,G}(\mathbf{y}) = \sum_{\boldsymbol{j} \in \mathscr{J}_d^k} G\left(\frac{j_1}{k}, \ldots, \frac{j_d}{k}\right) \text{Mult}\,(\mathbf{j} \mid k, \mathbf{y})\,,$$

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

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 1, if G is the restriction of the CDF of a probability measure on Δ_d, then $\widetilde{B}_{k,G}(\cdot)$ is not the restriction of the CDF of a probability measure defined on Δ_d 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_d, where the coefficients are nonnegative but do not add up to 1.

In this paper, we propose and study the properties of a Bayesian nonparametric model for the density estimation for data support on Δ_d, based on a modified class of MBP, referred to as $B_{k,G}$, that retains the well known approximation properties of the classical univariate and standard multivariate versions. An important property of the modified class of MBP is that if G is the CDF of a probability measure defined on Δ_d, then $B_{k,G}$ is the restriction of the CDF of a probability measure defined on Δ_d. Furthermore, the derivative of $B_{k,G}$ is a particular class of mixtures of Dirichlet distributions, that has appealing approximation properties. The organization of the paper is as follows. In Section 2, the modified class of multivariate Bernstein polynomials is introduced and its main properties are stated. In Section 3, we introduce the proposed model class and establish its main properties. The most relevant technical proofs are deferred to Section 4.

2. The modified class of multivariate Bernstein polynomials

The modified class of MBP is given in Definition 2. The modified class is obtained by increasing the number of elements where the sum runs (from $\mathscr{J}_d^k$ to $\mathscr{H}_d^k$), and increasing the domain of the function G (from Δ_d to $\mathbb{R}^d$), in the original class of MBP given in Definition 1, which implies a change in the parameters in the corresponding multinomial kernel.

Definition 2. For a given function $G : \mathbb{R}^d \longrightarrow \mathbb{R}$, the associated MBP of degree k on Δ_d is defined by

$$B_{k,G}(\mathbf{y}) = \sum_{\boldsymbol{j} \in \mathscr{H}_d^k} G\left(\frac{j_1}{k}, \ldots, \frac{j_d}{k}\right) \text{Mult}\,(\mathbf{j} \mid k+d-1, \mathbf{y})\,, \tag{3}$$

where $\boldsymbol{j} = (j_1, \ldots, j_d)$, and $\mathscr{H}_d^k = \left\{(j_1, \ldots, j_d) \in \{0, \ldots, k\}^d : \sum_{l=1}^d j_l \le k+d-1\right\}$.

The modified class of MBP retains the appealing approximation properties of univariate BP and the class given by Definition 1, that is, point-wise convergence at the continuity points of G and uniform convergence for continuous G. It is also possible to show that if G is the restriction of the CDF of a probability measure defined on Δ_d, then $B_{k,G}(\cdot)$ is also the restriction of the CDF of a probability measure defined on Δ_d. From expression (3), and after some algebra, it can be shown