Bernstein polynomial angular densities of multivariate extreme value distributions

Timothy E. Hanson
Department of Statistics, University of South Carolina, Columbia, SC 29208, USA
email: *hansont@stat.sc.edu*

Miguel de Carvalho
School of Mathematics, University of Edinburgh, Edinburgh, UK
email: *miguel.decarvalho@ed.ac.uk*

Yuhui Chen
Department of Mathematics, University of Alabama, Tuscaloosa, AL 35401, USA
email: *ychen164@ua.edu*

Abstract

To model the angular measure of a multivariate extreme value distribution, we develop a mean-constrained Bernstein polynomial over the $(p-1)$-dimensional simplex, along with a generalization that places mass on the simplex boundaries.

Keywords: Angular measure, Multivariate Bernstein polynomials, Multivariate extreme values.

1 Introduction

The angular measure of a multivariate extreme value distribution plays a key role in the statistical modeling of extreme value dependence. In this paper, we propose a model for the angular measure which can be used for an arbitrary number of dimensions, and which allows for a generalization that places mass on the simplex boundaries. To lay the groundwork, let $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ be a sequence of independent identically distributed random vectors in $\mathbb{R}^p$ with unit Fréchet marginal distributions, $F_1(y) = \cdots = F_p(y) = \exp(-1/y)$, for $y > 0$. Statistical theory for modeling multivariate extremes is based on a convergence result which provides the limiting distribution of the componentwise standardized maximum, $\mathbf{M}_n = n^{-1} \max\{\mathbf{Y}_1, \ldots, \mathbf{Y}_n\}$. Pickands (1981) established that

$$P(\mathbf{M}_n \leq \mathbf{y}) \to G_H(\mathbf{y}) = \exp\left\{-p \int_{S_p} \max\left\{\frac{x_1}{y_1}, \ldots, \frac{x_p}{y_p}\right\} H(\mathrm{d}\mathbf{x})\right\}, \quad \mathbf{y} \in (0, \infty)^p, \tag{1}$$

as $n \to \infty$, provided the limit exists and is non-degenerate; see also Coles (2001, Theorem 8.1). Here S_p is the unit simplex, that is $S_p = \{\mathbf{x} \in \mathbb{R}^p : x_1 + \cdots + x_p = 1, x_j \geq 0 \text{ for } j = 1, \ldots, p\}$, and G_H is a so-called multivariate extreme value distribution whose parameter H is the so-called angular measure, which is a distribution function on S_p that needs to obey the moment constraint

$$\int_{S_p} \mathbf{x} H(\mathrm{d}\mathbf{x}) = p^{-1} \mathbf{1}_p, \tag{2}$$

where $\mathbf{1}_p$ is a vector of ones. The more mass concentrates on the barycenter of S_p, $p^{-1}\mathbf{1}_p$, the higher the level of dependence between the extreme values of $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$. If H is absolutely continuous we define the angular density as $h(\mathbf{x}) = \frac{\mathrm{d}}{\mathrm{d}\mathbf{x}} H(\mathbf{x})$, for $\mathbf{x} \in S_p$.

To model the angular measure H, we propose a mean-constrained Bernstein polynomial on the $(p-1)$-dimensional simplex S_p. The mean constraints are built directly into the model and the dimension p can be as large as computing resources allow; the basic idea works the same way for all dimensions p. The proposed model is easily generalized to accommodate degenerate densities with mass on lower-dimensional simplexes, e.g. for $p = 3$ a triple such as $(0.29, 0.00, 0.71)$ or even $(0.00, 0.00, 1.00)$ can have positive probability. Besides arbitrary dimension and degenerate data, other benefits include the fact that the sampling algorithm is remarkably easy to implement (`FORTRAN 90` code is provided in the online supplementary content), and the approach gives good results in simulations and real data analyses.

Boldi and Davison (2007) develop finite mixtures of k Dirichlet distributions for H that satisfies (2). Both the E-M algorithm and reversible jump MCMC are considered for estimation. The former uses BIC to pick k whereas the latter allows k to be random. Sabourin and Naveau (2014) reparameterize this model and also consider reversible jump. Both of these approaches disallow mass on the simplex boundary. A nonparametric Bayes approach by Giullotte et al. (2011) does allow for mass on the simplex boundary but the proposed prior has been developed with the bivariate extreme value setting in mind and its extension to the p-dimensional is a challenging one.

Bayesian treatments of univariate Bernstein polynomials originate with Petrone (1999a,b). Petrone (1999a) considers the usual Bernstein polynomial over $[0, 1]$ where the weights follow a Dirichlet distribution, whereas Petrone (1999b) considers *random* Bernstein polynomials over $[0, 1]$ where the weights are more flexibly derived from a Dirichlet process. Zheng et al. (2010) extend random Bernstein polynomials to the hypercube $[0, 1]^p$, and Barrientos et al. (2015) extend random Bernstein polynomials to the simplex S_p for compositional data. All of these approaches do not consider mass on lower-dimensional boundaries of S_p and are rather cumbersome to implement. Recently Marcon et al. (2016) proposed an elegant approach based on Bernstein polynomials over $[0, 1]$ for bivariate extremes ($p = 2$) that places mass on the boundaries $\{0, 1\}$ and uses reversible jump to allow for random k; our approach extends theirs by allowing for any $p \geq 2$ but simply fixes k to be as large as is computationally feasible.

Section 2 develops the mean-constrained Bernstein polynomial on S_p and Section 3 discusses Bayesian inference via Markov chain Monte Carlo (MCMC). A more general model placing positive mass on the boundaries of S_p is developed in Section 4. Section 5 presents a short data illustration involving the Leeds air quality data analyzed in other papers and Section 6 concludes the paper.

2 Mean-constrained Bernstein polynomial on S_p

A multivariate Bernstein polynomial expansion for a density on the simplex is a finite mixture of Dirichlet densities with means regularly spread out over the simplex. Let $\mathbb{N} = \{1, 2, 3, \dots\}$ denote the positive integers, and for any $\boldsymbol{\alpha} \in \mathbb{N}^p$, define $|\boldsymbol{\alpha}| = \sum_{j=1}^{p} \alpha_j$. A Dirichlet density on the S_p simplex with parameter $\boldsymbol{\alpha}$ is:

$$d(\mathbf{x} \mid \boldsymbol{\alpha}) = \frac{\Gamma(|\boldsymbol{\alpha}|)}{\prod_{i=1}^{p} \Gamma(\alpha_i)} \prod_{i=1}^{p} x_i^{\alpha_i - 1} I_{S_p}(\mathbf{x}).$$

Recall that for $\mathbf{x} \sim d(\mathbf{x} \mid \boldsymbol{\alpha})$, $E(\mathbf{x} \mid \boldsymbol{\alpha}) = \boldsymbol{\alpha}/|\boldsymbol{\alpha}|$. Fix $J \in \mathbb{N}$ in what follows. A Bernstein polynomial on S_p of order J is written

$$h_{\mathbf{w}}(\mathbf{x}) = \sum_{|\boldsymbol{\alpha}|=J} w_{\boldsymbol{\alpha}} d(\mathbf{x} \mid \boldsymbol{\alpha}), \tag{3}$$

where

$$\sum_{|\boldsymbol{\alpha}|=J} w_{\boldsymbol{\alpha}} = 1, \tag{4}$$

and $\mathbf{w} = \{w_{\boldsymbol{\alpha}} : |\boldsymbol{\alpha}| = J \text{ and } \boldsymbol{\alpha} \in \mathbb{N}\}$; see Lorentz (1986, Section 2.9, Eq. (13)). It is understood that the sum in (3) is only over $\boldsymbol{\alpha} \in \mathbb{N}^p$. The order of the resulting polynomial is $J - p$, therefore $J = p$ gives a uniform distribution; necessarily we only consider $J > p$. For example, setting $J = 5$ and $p = 3$ gives indices $\{113, 131, 311, 122, 212, 221\}$ and the Bernstein polynomial is

$$h_{\mathbf{w}}(\mathbf{x}) = w_{113} 12 x_3^2 + w_{131} 12 x_2^2 + w_{311} 12 x_1^2 + w_{122} 24 x_2 x_3 + w_{212} 24 x_1 x_3 + w_{221} 24 x_1 x_2.$$

There are $m = \binom{J-1}{p-1}$ basis functions: the number of ways to place $J - p$ indistiguishable balls into p distinguishable urns where each urn has at least one ball; there are $J - p$ left after placing one ball in each of the p urns. If we add a level to the Bernstein polynomial, we increase the number of basis functions by

$$\binom{J}{p-1} - \binom{J-1}{p-1} = \binom{J-1}{p-2}. \tag{5}$$

There are $p-1$ marginal mean constraints (the pth is implied by the other $p-1$); the jth element x_j of $\mathbf{x}$ is required to satisfy

$$E(Jx_j) = \sum_{i=1}^{J-p+1} i \sum_{\substack{|\boldsymbol{\alpha}|=J \\ \alpha_j=i}} w_{\boldsymbol{\alpha}} = \frac{J}{p}. \tag{6}$$

We define a *Bernstein polynomial angular density*, as a Bernstein polynomial in (3), but obeying the moment constraint (6). For $J=5$ and $p=3$ this boils down to

$$\begin{aligned}
(w_{113}+w_{122}+w_{131})1+(w_{221}+w_{212})2+w_{311}3 &= 5/3, \\
(w_{113}+w_{212}+w_{311})1+(w_{122}+w_{221})2+w_{131}3 &= 5/3, \\
(w_{221}+w_{131}+w_{311})1+(w_{122}+w_{212})2+w_{113}3 &= 5/3.
\end{aligned}$$

Let $\mathbf{e}_j$ be a J-dimensional vector of all zeros except element j is unity. The p vertices of the simplex S_p are at $\mathbf{e}_1,\ldots,\mathbf{e}_p$. Let $\mathbf{a}_j$ be a J-dimensional vector of all ones except element j is $J-p+1$. The Bernstein polynomial basis function that places greatest mass near the vertex $\mathbf{e}_j$ is $d(\mathbf{x} \mid \mathbf{a}_j)$, with corresponding coefficient $w_{\mathbf{a}_j}$; call these the 'vertex coefficients.' Let $\mathcal{V}=\{\mathbf{a}_1,\ldots,\mathbf{a}_p\}$ be the indices for the p vertex coefficients. Ideally, we would like to solve for p of the coefficients in terms of the remaining coefficients; this proves easy for multivariate Bernstein polynomials. Subtracting (4) from (6) gives the vertex coefficients in terms of the remaining, non-vertex coefficients:

$$w_{\mathbf{a}_j} = \frac{1}{p} - \sum_{i=2}^{J-p} \frac{i-1}{J-p} \sum_{\substack{|\boldsymbol{\alpha}|=J \\ \alpha_j=i}} w_{\boldsymbol{\alpha}} = \frac{1}{p} - \sum_{\substack{|\boldsymbol{\alpha}|=J \\ \boldsymbol{\alpha}\neq\mathbf{a}_j}} \frac{\alpha_j-1}{J-p} w_{\boldsymbol{\alpha}}. \tag{7}$$

For $j=1,\ldots,p$, (7) imposes the necessary mean constraint as well as the sum-to-unity constraint; there are $m-p$ free parameters left, the $w_{\boldsymbol{\alpha}}$ indexed by elements of $\mathcal{F}=\{\boldsymbol{\alpha}\in\mathbb{N}^p : |\boldsymbol{\alpha}|=J \text{ and } \boldsymbol{\alpha}\notin\mathcal{V}\}$. For $J=5$ and $p=3$ we have mean-constrained vertex coefficients

$$w_{311}=\tfrac{1}{3}-\tfrac{1}{2}w_{212}-\tfrac{1}{2}w_{221}, \quad w_{131}=\tfrac{1}{3}-\tfrac{1}{2}w_{122}-\tfrac{1}{2}w_{221}, \quad w_{113}=\tfrac{1}{3}-\tfrac{1}{2}w_{122}-\tfrac{1}{2}w_{212}.$$

Bernstein polynomials enjoy many appealing properties. Given a Bernstein polynomial of order J and dimension p, the marginal distribution of any subset of $\mathbf{x}$, say $(x_{q_1}, \ldots, x_{q_k})$ where $\{q_1, \ldots, q_k\} \subseteq \{1, \ldots, p\}$, $k \leq p$, is clearly a Bernstein polynomial of order J as well. Bernstein polynomials 'reproduce' in the sense that any Bernstein polynomial of order J can be expressed exactly as a Bernstein polynomial of order $J + 1$; indeed, for $|\boldsymbol{\alpha}| = J + 1$,

$$w^*_{\boldsymbol{\alpha}} = \sum_{j=1}^{p} \frac{\alpha_j - 1}{J + 1} w_{\boldsymbol{\alpha} - \mathbf{e}_j}.$$

See Sauer (1999, Proposition 2.3). The classes of densities generated by lower order Bernstein polynomials are formally nested within higher orders. This facilitates fitting in that only one Bernstein polynomial of a reasonably high order need be fitted.

3 Model specification and posterior inference

Due to the formal nesting property of the Bernstein polynomial, it is only necessary to fit the model once with a J that is as large but practical—keeping in mind (5). In our MCMC scheme the choice of J affects the amount of time necessary to achieve posterior inference, along with the dimension p and the sample size n. As a rule of thumb we have found bounding $m \leq n$ to work well, although m much smaller than n can provide similar estimates depending on how 'localized' the data are. An important problem in fitting multivariate Bernstein polynomials is the delineation of the m elements in the index set $\{\boldsymbol{\alpha} \in \mathbb{N}^p : |\boldsymbol{\alpha}| = J\}$. We use the `nexcom` algorithm in Nijenhuis and Wilf (1978, Chapter 5).

After much experimentation with various approaches to model fitting (including the E-M algorithm, iterative fitting, and various Bayesian approaches), we have found a componentwise adaptive Markov chain Monte Carlo (MCMC) to provide consistently good results. A generalized logit transformation is considered for the free parameters $\{w_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \mathcal{F}\}$;

implicitly define $\{v_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \mathcal{F}\}$ through

$$w_{\boldsymbol{\alpha}} = \frac{e^{v_{\boldsymbol{\alpha}}}}{p + \sum_{\tilde{\boldsymbol{\alpha}} \in \mathcal{F}} e^{v_{\tilde{\boldsymbol{\alpha}}}}}. \tag{8}$$

A common prior on the coefficient vector $\mathbf{w}$ is Dirichlet$(c\mathbf{1}_m)$ (Petrone, 1999a; Chen et al., 2014). The Dirichlet density on the $m-p$ free parameters $\{w_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \mathcal{F}\}$ incorporating the mean constraint is given by

$$p(w_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \mathcal{F}) \propto d(\mathbf{w} \mid c\mathbf{1}_m) \prod_{j=1}^{p} I \left\{ \sum_{i=1}^{J-p+1} i \sum_{\substack{|\boldsymbol{\alpha}|=J \\ \alpha_j = i}} w_{\boldsymbol{\alpha}} = \frac{J}{p} \right\}, \tag{9}$$

where $I\{\cdot\}$ is the indicator function. Thus, the prior on the $\{v_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \mathcal{F}\}$ is simply

$$p(v_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \mathcal{F}) \propto \prod_{\boldsymbol{\alpha} \in \mathcal{F}} \left[\frac{e^{v_{\boldsymbol{\alpha}}}}{p + \sum_{\tilde{\boldsymbol{\alpha}} \in \mathcal{F}} e^{v_{\tilde{\boldsymbol{\alpha}}}}} \right]^c \prod_{j=1}^{p} I \left\{ \sum_{i=1}^{J-p+1} i \sum_{\substack{|\boldsymbol{\alpha}|=J \\ \alpha_j = i}} \frac{e^{v_{\boldsymbol{\alpha}}}}{p + \sum_{\tilde{\boldsymbol{\alpha}} \in \mathcal{F}} e^{v_{\tilde{\boldsymbol{\alpha}}}}} = \frac{J}{p} \right\}.$$

The posterior is proportional to

$$p(v_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \mathcal{F} \mid \mathbf{x}_1, \ldots, \mathbf{x}_n) \propto p(v_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \mathcal{F}) \prod_{i=1}^{n} \sum_{|\boldsymbol{\alpha}|=J} w_{\boldsymbol{\alpha}} d(\mathbf{x}_i \mid \boldsymbol{\alpha}),$$

where the $w_{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha} \in \mathcal{F}$ are given through (8) and the $w_{\mathbf{a}_j}$ for $j = 1, \ldots, p$ are given by (7). The adaptive componentwise random-walk Metropolis–Hastings algorithm of Haario et al. (2005) has worked very well in applications. Initialize $v_{\boldsymbol{\alpha}}^0 = 0$ for $\boldsymbol{\alpha} \in \mathcal{F}$; this corresponds to the uniform distribution on the simplex: $w_{\boldsymbol{\alpha}}^0 = 1/\binom{J-1}{p-1}$ for $|\boldsymbol{\alpha}| = J$. At each iteration of the Gibbs sampler, we cycle through all elements $\boldsymbol{\alpha} \in \mathcal{F}$, updating each $v_{\boldsymbol{\alpha}}^s$ given the current value of the remaining values. Propose $v_{\boldsymbol{\alpha}}^* \sim N(v_{\boldsymbol{\alpha}}^s, a_{\boldsymbol{\alpha}}^s)$. Note that changing only $v_{\boldsymbol{\alpha}}^{s-1}$ to $v_{\boldsymbol{\alpha}}^*$ but leaving the other $\{\tilde{\boldsymbol{\alpha}} \in \mathcal{F} : \tilde{\boldsymbol{\alpha}} \neq \boldsymbol{\alpha}\}$ unchanged changes every $\{w_{\boldsymbol{\alpha}} : |\boldsymbol{\alpha}| = J\}$; call this new collection of weights $\mathbf{w}^*$. The proposal is accepted with probability

$$\rho = \min\left\{1, \frac{p(\mathbf{v}^*)\prod_{i=1}^n \sum_{|\boldsymbol{\alpha}|=J} w^*_{\boldsymbol{\alpha}} d(\mathbf{x}_i \mid \boldsymbol{\alpha})}{p(\mathbf{v}^{j-1})\prod_{i=1}^n \sum_{|\boldsymbol{\alpha}|=J} w^{j-1}_{\boldsymbol{\alpha}} d(\mathbf{x}_i \mid \boldsymbol{\alpha})}\right\} \prod_{j=1}^p I\{0 < w^*_{\mathbf{a}_j} < 1\}.$$

We make explicit here that the $w^*_{\mathbf{a}_j}$ need to be between zero and one, although this is implied by the support of the Dirichlet prior in the simplex. The Metropolis–Hastings algorithm 'automatically' enforces the mean constraint and is valid as long as the support of the proposals is as least as great as the posterior, which we have here. This kind of nonlinear constraint accept/reject approach was used by Jones et al. (2010) to force probabilities in contingency tables subject to nonlinear constraints to be between zero and one.

Let $\mathbf{w}^s$ be the sampled coefficients at iteration s of the MCMC scheme, where $s = 1, \ldots, S$. The density that generated $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is estimated by discarding the first, say, M iterates (termed the burn-in), taking the mean of those remaining $\bar{\mathbf{w}} = (S-M)^{-1}\sum_{s=M+1}^S \mathbf{w}^s$, and using $h_{\bar{\mathbf{w}}}(\mathbf{x})$ through (3).

4 Mass on the simplex boundaries

The Bernstein approach naturally extends to allow for densities on the boundary of the simplex—boundaries are simply lower dimensional simplexes. Now instead of requiring that each urn have at least one ball, we allow some urns to be empty. For basis functions with one or more empty urns, we simply define Dirichlet distributions on the lower dimensional simplex. Let's reconsider $J = 5$ and $p = 3$. Our original formulation delineates the basis functions as $\{113, 131, 311, 122, 212, 221\}$. Adding in lower dimensional densities adds $\{005, 014, 023, 032, 041, 050, 140, 230, 320, 410, 500, 401, 302, 203, 104\}$. Any Dirichlet distribution with only one non-zero element (which must equal J) is Dirac measure, e.g.

$$D(\cdot \mid 5, 0, 0) = \delta_{(1,0,0)}(\cdot).$$

In fact, the Dirac distributions occur at the new vertex indices $\boldsymbol{\alpha} \in \{J\mathbf{e}_j : j = 1, \ldots, p\}$. Any Dirichlet distribution with more than one but fewer than J zero elements has a density over the simplex corresponding to the non-zero elements, e.g.

$$d(p_1, p_2 \mid 2, 3, 0) = \frac{\Gamma(5)}{\Gamma(2)\Gamma(3)} p_1^{2-1} p_2^{3-1} I\{p_1 + p_2 = 1, 0 \leq p_1, p_2 \leq 1\}.$$

Note that the means of these two distributions are $(\frac{5}{5}, \frac{0}{5}, \frac{0}{5}) = (1, 0, 0)$ and $(\frac{2}{5}, \frac{3}{5}, \frac{0}{5}) = (0.4, 0.6, 0)$ respectively.

The mean constraint (6) now becomes

$$E(Jx_j) = \sum_{i=1}^{J} i \sum_{\substack{|\boldsymbol{\alpha}|=J \\ \alpha_j = i}} w_{\boldsymbol{\alpha}} = \frac{J}{p}, \tag{10}$$

where the sum is taken over $\boldsymbol{\alpha} \in \mathbb{N}_0^p$ where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Equation (10) immediately gives the new vertex coefficients (with Dirac measure) in terms of the remaining, non-vertex coefficients:

$$w_{J\mathbf{e}_j} = \frac{1}{p} - \sum_{i=1}^{J-1} \frac{i}{J} \sum_{\substack{|\boldsymbol{\alpha}|=J \\ \alpha_j = i}} w_{\boldsymbol{\alpha}} = \frac{1}{J} - \sum_{\substack{|\boldsymbol{\alpha}|=J \\ \boldsymbol{\alpha} \neq J\mathbf{e}_j}} \frac{\alpha_j}{J} w_{\boldsymbol{\alpha}}. \tag{11}$$

The likelihood is now a weighted sum of densities over $2^p - 1$ $(p-1)$-dimensional or lower-dimensional simplexes, each conveniently indexed by a binary number. Let $E = \{0, 1\}^p \backslash \{\mathbf{0}_p\}$, the set of all p-dimensional binary numbers except for the zero vector. For $\boldsymbol{\epsilon} \in E$, let $A_{\boldsymbol{\epsilon}} = \{\boldsymbol{\alpha} \in \mathbb{N}_0^p : |\boldsymbol{\alpha}| = J, I\{\alpha_j > 0\} = \epsilon_j, j = 1, \ldots, p\}$ and $S_{\boldsymbol{\epsilon}} = \{\mathbf{x} \in [0, 1]^p : \sum_{j=1}^{p} x_j I\{\epsilon_j = 1\} = 1\}$. For example, if $p = 3$, then $S_{001} = \{(0, 0, 1)\}$, $S_{010} = \{(0, 1, 0)\}$, $S_{011} = \{(0, x_2, x_3) \in [0, 1]^3 : x_2 + x_3 = 1\}$, and $S_{111} = \{(x_1, x_2, x_3) \in [0, 1]^3 : x_1 + x_2 + x_3 = 1\}$. If $J = 4$ then $A_{(0,1,1)} = \{(0, 1, 3), (0, 2, 2), (0, 3, 1)\}$. The density $h_{\mathbf{w}}(\mathbf{x})$ is now given through the law of total probability as

$$h_{\mathbf{w}}(\mathbf{x}) = \sum_{\boldsymbol{\epsilon} \in E} P(\mathbf{x} \in S_{\boldsymbol{\epsilon}}) h(\mathbf{x} \mid \mathbf{x} \in S_{\boldsymbol{\epsilon}}) = \sum_{\boldsymbol{\alpha} \in A_{\boldsymbol{\epsilon}(\mathbf{x})}} w_{\boldsymbol{\alpha}} d(\mathbf{x} \mid \boldsymbol{\alpha}),$$

where

$$d(\mathbf{x} \mid \boldsymbol{\alpha}) = \Gamma(J) \prod_{\alpha_j \neq 0} I\{\alpha_j = 0, x_j = 0\} \left[\frac{p_j^{\alpha_j - 1}}{\Gamma(\alpha_j)} \right]^{I\{\alpha_j > 0, x_j > 0\}},$$

and $\boldsymbol{\epsilon}(\mathbf{x})$ is such that $\mathbf{x} \in S_{\boldsymbol{\epsilon}(\mathbf{x})}$.

Let the free parameters be in $\mathcal{F}_0 = \{\boldsymbol{\alpha} \in \mathbb{N}_0^p : |\boldsymbol{\alpha}| = J \text{ and } \boldsymbol{\alpha} \neq J\mathbf{e}_j \text{ for } j = 1, \dots, p\}$. The posterior is proportional to

$$p(v_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \mathcal{F}_0 \mid \mathbf{x}) \propto p(\mathbf{v}_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \mathcal{F}_0) \prod_{i=1}^{n} \sum_{\alpha \in A_{\boldsymbol{\epsilon}(\mathbf{x}_i)}} w_{\boldsymbol{\alpha}} d(\mathbf{x}_i | \boldsymbol{\alpha}),$$

where the $w_{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha} \in \mathcal{F}_0$ are given through

$$w_{\boldsymbol{\alpha}} = \frac{e^{v_{\boldsymbol{\alpha}}}}{p + \sum_{\tilde{\boldsymbol{\alpha}} \in \mathcal{F}_0} e^{v_{\tilde{\boldsymbol{\alpha}}}}}, \tag{12}$$

and the $w_{J\mathbf{e}_j}$ for $j = 1, \dots, p$ are given by (11).

5 Example

Boldi and Davison (2007) and Sabourin and Naveau (2014) considered $p = 5$ air quality measurements from central Leeds over the years 1994–1998. The measurements are daily ozone levels O_3, nitrogen dioxide NO_2, nitrogen oxide NO, sulfer dioxide SO_2, and particulate matter PM_{10}. We transform the data to unit Fréchet margins using a rank approach and use the same threshold as Boldi and Davison (2007, p. 224) (i.e. $e^{2.5}$). Boldi and Davison (2007) used $n = 247$ extremes whereas Sabourin and Naveau (2014) used $n = 100$; we used $n = 267$.

Since no $\mathbf{x}_i$ had zero elements, the model of Section 2 was used with $c = 0.1$. $S = 6000$ iterates were generated; the burn-in was $M = 2000$, so $\bar{\mathbf{w}}$ was computed from 4000 iterates post-burn-in. Figure 1 shows level curves from fitting these data over S_5 with $J = 11$ and $J = 12$. Taking $J = 11$ leads to $m = 210$ Bernstein polynomial basis functions on the 5-dimensional simplex; $J = 12$ gives $m = 330$ basis functions; there is almost no difference

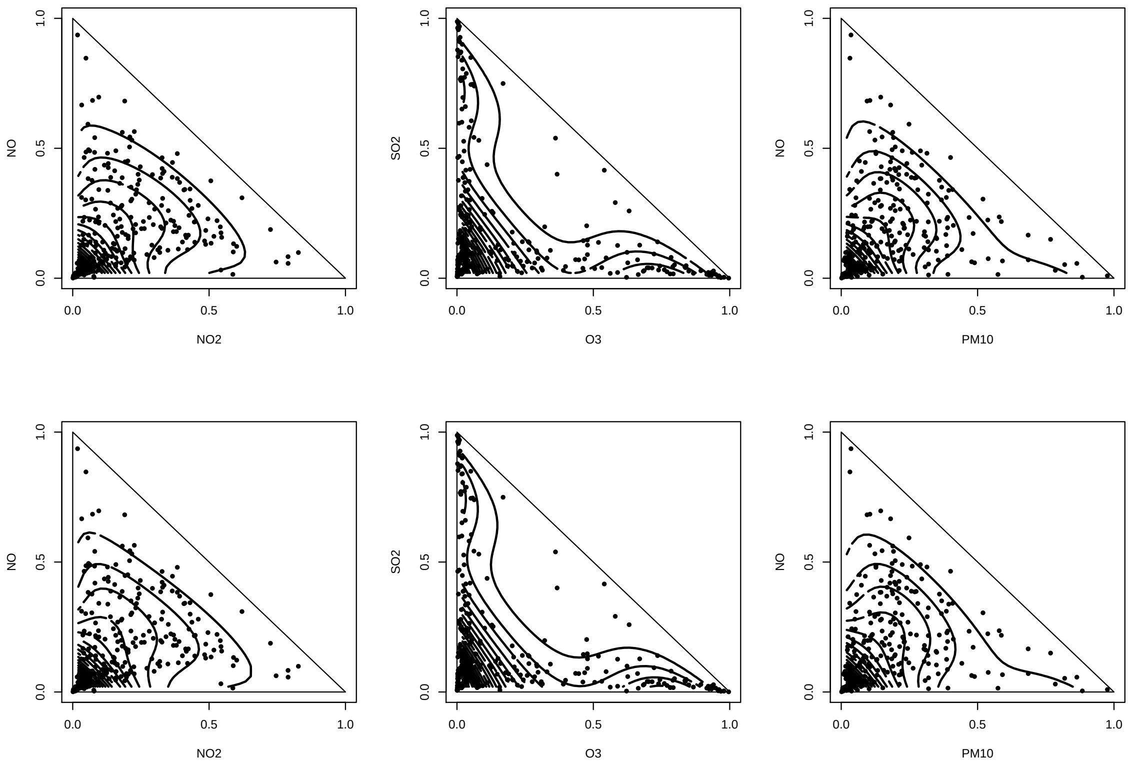

Figure 1: Leeds air quality data; top row is $J = 11$ yielding $m = 210$ basis functions and bottom row is $J = 12$ yielding $m = 330$.

in the estimates, confirming that picking m 'large enough' will adequately model the data. Figure 1 can be compared to Fig. 5 in Boldi and Davison (2007) and Fig. 6 in Sabourin and Naveau (2014); all methods provide somewhat similar inferences, although the $k = 10$ in Boldi and Davison (2007) shows rather concentrated Dirichlet mixands.

Beyond this real data example, several simulated datasets were considered of varying complexity and dimension. When data were generated from a Bernstein polynomial parameters for both models (Sections 2 and 4) were consistent and asymptotically unbiased. For data generated otherwise, e.g. as a mean $p^{-1}\mathbf{1}_p$ mixture of Dirichlets, the Bernstein polynomial model estimated the true density very well. Little posterior sensitivity was noted for increasing m after a certain point, however c does play a small role in how 'bumpy' the estimates are with smaller c allowing for more heterogenious estimates. Note that $c \to \infty$ forces the Bernstein polynomial to be uniform over S_p.

6 Final remarks

We develop a multivariate Bernstein polynomial-based model for the angular measure of a multivariate extreme value distribution, which allows for a generalization that places mass at the boundaries of the simplex. `FORTRAN 90` programs for fitting the mean-constrained Bernstein polynomials proposed here are given in the Supplementary Materials; they are presented 'as is' for others to modify and use freely.

In some settings of applied interest the main concern may not be about a single angular measure, but rather on a family of angular densities $\{h_1, \ldots, h_K\}$, and in the latter case a main concern is how to borrow strength instead of fitting each h_k separately. While not explored here, such borrowing of strength is straightforward from a Bayesian perspective by adding another level to the hierarchy, without the need of solving sophisticated constrained optimization problems as in de Carvalho and Davison (2014).

While the focus of the paper has been on imposing the moment constraint of centering the (angular) density on the barycenter, $p^{-1}\mathbf{1}_p$, the approach in Section 2 can be readily extended to impose other types of moment constraints, and thus our methods can be readily adapted for compositional data analysis (Aitchison, 1986), for contexts where marginal moments from census data may be available, or for settings where other population level information is available (see, for instance, Oguz-Alper and Berger, 2016).

Acknowledgments. We thank the Co-Editor and the two referees, whose comments and suggestions led to an improved version of the manuscript.

References

Aitchison, J. 1986. The Statistical Analysis of Compositional Data. Chapman & Hall/CRC, London.

Barrientos, A.F., Jara, A., Quintana, F. 2015. Bayesian density estimation for compositional data using random Bernstein polynomials. J. Statist. Plann. Infer. 166, 116–125.

Boldi, M.-O., Davison, A.C. 2007. A mixture model for multivariate extremes. J. R. Statist. Soc. Ser. B 69(2), 217–229.

Chen, Y., Hanson, T., Zhang, J. 2014. Accelerated hazards model based on parametric families generalized with Bernstein polynomials. Biometrics 70(1), 192–201.

Coles, S. 2001. An Introduction to Statistical Modeling of Extreme Values. Springer, New York.

de Carvalho, M., Davison, A.C. 2014. Spectral density ratio models for multivariate extremes. J. Amer. Statist. Assoc. 109(506), 764–776.

Guillotte, S., Perron, F., Segers, J. 2011. Non-parametric Bayesian inference on bivariate extremes. J. R. Statist. Soc. Ser. B 73(3), 377–406.

Haario, H., Saksman, E., Tamminen, J. 2005. Componentwise adaptation for high dimensional MCMC. Comput. Statist. 20, 265–273

Jones, G., Johnson, W., Hanson, T., Christensen, R. 2010. Identifiability of models for multiple diagnostic testing in the absence of a gold standard. Biometrics 66(3), 855–863.

Lorentz, G.G. 1986. Bernstein Polynomials. 2nd ed. AMS Chelsea Publishing, Providence.

Marcon, G., Padoan, S.A., Antoniano-Villalobos, I. 2016. Bayesian inference for the extremal dependence. Elect. J. Statist. in press.

Nijenhuis, A., Wilf, H.S. 1978. Combinatorial Algorithms for Computers and Calculators. 2nd ed. Academic Press, New York.

Oguz-Alper, M., Berger, Y.G. 2016. Modelling complex survey data with population level information: An empirical likelihood approach. Biometrika 103(2), 447–459.

Petrone, S. 1999a. Bayesian density estimation using Bernstein polynomials. Canadian J. Statist. 27(1), 105–126.

Petrone, S. 1999b. Random Bernstein polynomials. Scandinavian J. Statist. 26(3), 373–393.

Pickands, J. 1981. Multivariate extreme value distributions. In: Proc. 43th Sess. Int. Statist. Inst.

Sabourin, A., Naveau, P. 2014. Bayesian Dirichlet mixture model for multivariate extremes: A re-parameterization. Comput. Statist. Data Anal. 71, 542–567.

Sauer, T. 1999. Multivariate Bernstein polynomials, convexity and related shape properties. In: J.M. Peña Ed., Shape Preserving Representations in Computer Aided Design. pp. 133–147. Nova Science Publishers.

Zheng, Y., Zhu, J., Roy, A. 2010. Nonparametric Bayesian inference for the spectral density function of a random field. Biometrika 97(1), 238–245