Technical Report 1152, Department of Statistics, University of Wisconsin, Madison, Wisconsin, U.S.A. Version: 6/4/2009 (2009), *pp.* 1–16

Nonparametric Bayesian inference for the spectral density function of a random field

BY YANBING ZHENG

Department of Statistics, University of Kentucky, 817 Patterson Office Tower, Lexington, Kentucky 40506, U.S.A.

yanbing.zheng@uky.edu

JUN ZHU

Department of Statistics, University of Wisconsin, 1300 University Avenue, Madison, Wisconsin 53706, U.S.A.

jzhu@stat.wisc.edu

ANINDYA ROY

Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, Maryland 21250, U.S.A.

anindya@math.umbc.edu

SUMMARY

A powerful technique for inference concerning spatial dependence in a random field is to use spectral methods based on frequency domain analysis. Here we develop a nonparametric Bayesian approach to statistical inference for the spectral density of a random field. We construct a multi-dimensional Bernstein polynomial prior for the spectral density and establish its theoretical validity as a nonparametric prior. We devise a Markov chain Monte Carlo algorithm to simulate from the posterior of the spectral density. The posterior sampling enables us to obtain a smoothed estimate of the spectral density as well as credible bands at desired levels. Simulation shows that our proposed method is more robust than a parametric approach and for illustration, we analyze a soil data example.

Some key words: Bernstein polynomial prior; Dirichlet process; Markov chain Monte Carlo; Multi-dimensional Bernstein polynomial; Periodogram; Whittle's approximation.

1. INTRODUCTION

Random fields provide a flexible modeling framework for the analysis of spatially referenced data that arise in a wide variety of disciplines (Cressie, 1993; Stein, 1999; Schabenberger & Gotway, 2005). Much research effort has focused on properly modeling and inferring spatial dependence expressed as an autocovariance function in the spatial domain. A powerful alternative is to use spectral methods in the frequency domain. This paper aids further development of spectral methods by delineating a new nonparametric Bayesian approach to statistical inference for the spectral density of a spatial random field.

Let $\{X(s) : s \in D \subset \mathbb{R}^d\}$ denote a random field on the spatial domain D where d is a positive integer. We assume that the random field is second-order stationary with a constant mean function and an autocovariance function $C(h) = \mathrm{cov}\{X(s), X(s+h)\}$. As the Fourier transform of the autocovariance function, the spectral density is defined as $f^*(\omega^*) = (2\pi)^{-d} \int_{\mathbb{R}^d} \exp(-ih'\omega^*)C(h)dh$, where $\omega^* = (\omega_1^*, \ldots, \omega_d^*) \in \mathbb{R}^d$. If the spatial domain D is an integer lattice in $\mathbb{Z}^d$, the frequency ω^* is restricted to a finite-frequency band $(-\pi, \pi]^d$ and the spectral density of the random field can be written as,

$$f^*(\omega^*) = \frac{1}{(2\pi)^d} \sum_{h \in \mathbb{Z}^d} \exp(-ih'\omega^*)C(h). \quad (1)$$

Let D_n denote the spatial sampling locations in D and we suppose that D_n consists of an $n_1 \times \cdots \times n_d$ integer lattice in $\mathbb{Z}^d$. A periodogram provides a nonparametric estimate of the spectral density. It is defined as,

$$I(\omega^*) = \frac{1}{(2\pi)^d \prod_{i=1}^d n_i} \left| \sum_{s \in D_n} X(s) \exp(-is'\omega^*) \right|^2. \quad (2)$$

Under mild regularity conditions on the autocovariance function, the periodogram is an asymptotically unbiased estimator of the spectral density. However, it is inconsistent and needs to be adjusted by various smoothing techniques (Ripley, 1981; Heyde & Gay, 1993; Böhm et al., 2002; Schabenberger & Gotway, 2005; Robinson, 2007). Statistical inference of the spectral density is largely based on the asymptotic distribution of the smoothed periodogram.

Here we develop nonparametric Bayesian inference for the spectral density of a random field, as a novel alternative to traditional asymptotic inference. For time series, various nonparametric Bayesian methods were developed to estimate the spectral density (Choudhuri et al. (2004a) and the references therein). In particular, Choudhuri et al. (2004a) proposed a nonparametric Bayesian method to estimate the spectral density of a stationary time series, where the nonparametric prior on the spectral density was based on one-dimensional Bernstein polynomials (Petrone, 1999a,b). We construct a multi-dimensional Bernstein polynomial prior for the spectral density of a random field. We also establish its theoretical validity. Two- and multi-dimensional Bernstein polynomials have been developed for estimation of probability densities (Tenbusch, 1994) and as a prior for Bayesian inference (Epifani, 1999; Regazzini & Sazonov, 1999; Petrone, 2004; Kruijer & van der Vaart, 2008). In particular, Epifani (1999) studied Bernstein polynomials on hypercubes and their implication on nonparametric Bayesian inference, while Petrone (2004) developed general theoretical results for multi-dimensional Feller operators, of which multi-dimensional Bernstein polynomials are a special case. Equipped with the multi-dimensional Bernstein polynomial prior, we devise a Markov chain Monte Carlo algorithm to sample from the posterior of the spectral density. With the posterior, not only a smoothed estimate of the spectral density can be obtained, it also becomes straightforward to construct credible bands at desired levels and perform inference for other meaningful aspects of the spectral density.

2. MULTI-DIMENSIONAL BERNSTEIN POLYNOMIAL PRIOR

2·1. *Multi-Dimensional Bernstein Polynomials and the Bernstein Density*

Let $\Delta = [0,1]^d$ denote the unit cube in $\mathbb{R}^d$. Let $k = (k_1, \ldots, k_d)$ and $l = (l_1, \ldots, l_d)$ denote two vectors of nonnegative integers. Define the partial order $l \preceq k$ if $l_i \leq k_i$ for each i. Then define the multi-index sum as $\sum_{j=l}^{k} = \sum_{j_1=l_1}^{k_1} \cdots \sum_{j_d=l_d}^{k_d}$, where $j = (j_1, \ldots, j_d)$ and $l \preceq k$. Let $\mathrm{Cube}(j,k)$ denote the d-dimensional cube of the form $((j_1-1)/k_1, j_1/k_1] \times \cdots \times ((j_d-1)/k_d, j_d/k_d]$ with the convention that if $j_i = 0$, then the interval $((j_i-1)/k_i, j_i/k_i]$ is replaced by the point $\{0\}$. Let ${}^dk_1 = \prod_{i=1}^d k_i$, $\min\{k\} = \min\limits_{1\le i\le d} k_i$, and let j/k denote the vector $(j_1/k_1, \ldots, j_d/k_d)$. Also, let $\mathbf{0} = (0, \ldots, 0)$ and $\mathbf{1} = (1, \ldots, 1)$ denote the d-dimensional vectors of zeros and ones.

For a bounded function $G : \Delta \to \mathbb{R}$, a multi-dimensional Bernstein polynomial of order $k = (k_1, \ldots, k_d)$ associated with G is defined as,

$$B_{k,G}(\omega) = \sum_{j=\mathbf{0}}^{k} G(j/k) \prod_{i=1}^{d} p_{j_i k_i}(\omega_i), \tag{3}$$

where $\omega = (\omega_1, \ldots, \omega_d) \in \Delta$, $p_{j_i k_i}(\omega_i) = \binom{k_i}{j_i}\omega_i^{j_i}(1-\omega_i)^{k_i - j_i}$, for $0 \le \omega_i \le 1$, $j_i = 0, \ldots, k_i$, $k_i = 1, 2, \ldots$ and $i = 1, \ldots, d$. The order k can be interpreted as a multi-index parameter that controls the smoothness of $B_{k,G}$, with a smaller k_i associated with a smoother function along dimension i. When G is a distribution function with $G(0, 1, \ldots, 1) = \cdots = G(1, \ldots, 1, 0) = 0$, we will refer to the Bernstein polynomial as the Bernstein distribution function associated with G. As a special case of Theorem 1 of Petrone (2004), if G is bounded (or continuous), the Bernstein polynomial $B_{k,G}$ in (3) provides a good approximation of the function G. In our application the function G will always be a probability distribution function and hence, $B_{k,G}$ will be a probability distribution function on Δ as well.

A multi-dimensional Bernstein density, the derivative of the Bernstein distribution function, is defined as,

$$b_{k,G}(\omega) = \frac{\partial^d}{\partial\omega_1 \cdots \partial\omega_d} B_{k,G}(\omega) = \sum_{j=\mathbf{1}}^{k} u_{k,G}(j) \prod_{i=1}^{d} \beta(\omega_i; j_i, k_i - j_i + 1), \tag{4}$$

where $u_{k,G}(j)$ are mixing weights and $\beta(\cdot; a, b)$ denotes the probability density of a Beta distribution with parameters a and b. The mixing weights $u_{k,G}(j)$ are the probabilities of Cube(j, k) under G. It follows that $u_{k,G}(j) \ge 0$ and $\sum_{j=\mathbf{1}}^{k} u_{k,G}(j) = 1$ and thus the Bernstein density $b_{k,G}$ in (4) is a probability density on Δ. If G is a probability distribution function with a continuous probability density g, then it can be shown that g can be approximated uniformly by a sequence of Bernstein densities, $\{b_{k,G}(\omega)\}$, as shown in the following Theorem 1. For completeness, we include in this theorem properties relating to the Bernstein distribution function as well.

THEOREM 1. *If a function $G : \Delta \to \mathbb{R}$ is bounded on the unit cube Δ and let $B_{k,G}$ be the associated Bernstein polynomial defined in* (3)*. Then as $\min\{k\} \to \infty$, we have $B_{k,G}(\boldsymbol{\omega}) \longrightarrow G(\boldsymbol{\omega})$ at each point of continuity $\boldsymbol{\omega}$ of G. If G is continuous on Δ, then $B_{k,G}(\boldsymbol{\omega}) \longrightarrow G(\boldsymbol{\omega})$ uniformly on Δ. In addition, if G is a probability distribution function with a continuous density g on the unit cube Δ and $b_{k,G}$ is the associated Bernstein density function defined in* (4)*, then $b_{k,G}(\boldsymbol{\omega}) \longrightarrow g(\boldsymbol{\omega})$ uniformly on Δ.*

2·2. *Multi-Dimensional Bernstein Polynomial Prior*

Let Π denote the space of probability distribution functions on Δ and equip Π with its Borel σ-field $\mathcal{T}$ generated by the topology of weak convergence. We define a random multi-dimensional Bernstein distribution function as the random function of the form $B_{k,G}$ in (3), where the order k is an $\mathbb{N}^d$-valued random variable and given k, the coefficient is $G(j/k)$ with G being a random probability distribution function on Δ whose atoms are $\{\text{Cube}(j, k), j \preceq k\}$. The random multi-dimensional Bernstein distribution functions define a probability measure on the space $(\Pi, \mathcal{T})$ as to be established in Theorem 2 which we will refer to as a multi-dimensional Bernstein polynomial prior and denote by π. The parameter of π is the joint probability distribution of k and G and is denoted as $\mathcal{P}$.

Since any continuous density on Δ can be uniformly approximated by a sequence of Bernstein densities, we can define a measure on the space of continuous densities on Δ by defining a measure on the set of Bernstein densities. Let $\tilde{\Pi}$ denote the space of continuous probability densities on Δ and equip $\tilde{\Pi}$ with the Borel σ-field $\tilde{\mathcal{T}}$ generated by the topology of weak convergence. We define a random multi-dimensional Bernstein density as a random function of the form $b_{k,G}$ in (4), where the order k is an $\mathbb{N}^d$-valued random variable and given k, the mixing weights $u_k(j)$ are random with $u_k = (u_k(j) : j_i = 1, \ldots, k_i, i = 1, \ldots, d)$ belonging to the ${}^d k_1 - 1$ dimensional simplex,

$$\mathcal{S}_k = \left\{ u_k(j) : u_k(j) \geq 0, \sum_{j=\mathbf{1}}^{k} u_k(j) = 1 \right\}. \tag{5}$$

When induced by a random probability distribution function G, we denote the vector of mixing weights u_k as $u_{k,G}$. In order for the proposed multi-dimensional Bernstein polynomial prior π to be a valid nonparametric prior, a well-accepted criterion is that it should have full topological support, in the sense that it is able to reach the entire space of probability distribution functions Π. In Theorem 3, we will show that the multi-dimensional Bernstein polynomial prior developed here meets this criterion.

2·3. *Prior for the Spectral Density*

Although not crucial, we rescale the spectral density f^* in (1) to f on Δ such that, $f(\omega) = f^*(2\pi\omega - \pi)$, where $\omega \in \Delta$ and $2\pi\omega - \pi = (2\pi\omega_1 - \pi, \ldots, 2\pi\omega_d - \pi)$. We normalize f to

$$q(\omega) = f(\omega)/\tau, \tag{6}$$

where the normalizing constant $\tau = \int_\Delta f(\omega) d\omega$. Alternatively, as a reviewer suggested, a prior may be imposed directly on f^*. We suspect that this would require some form of rescaling of a bounded function G in order to apply a Dirichlet distribution, an issue that we will investigate further.

We impose a multi-dimensional Bernstein polynomial prior on the normalized spectral density q and independently impose a prior on τ. Following (4) and (6), we let,

$$f(\omega) = \tau \sum_{j=\mathbf{1}}^{k} u_{k,G}(j) \prod_{i=1}^{d} \beta(\omega_i; j_i, k_i - j_i + 1), \tag{7}$$

where G follows a Dirichlet process such that for any partition $\{A_1, \ldots, A_m\}$ of the sample space Δ, the m-dimensional random vector of $G(A_j)$ follows a Dirichlet distribution,

$$\{G(A_1), \ldots, G(A_m)\} \sim \text{Dirichlet}\{MG_0(A_1), \ldots, MG_0(A_m)\}, \tag{8}$$

where M is a weight parameter and G_0 is a base measure (Müller & Quintana, 2004). We will abbreviate (8) to $G \sim \mathcal{D}(M, G_0)$. A diffuse prior on G may be obtained by setting M to a small number and setting G_0 to a uniform distribution on Δ. Finally, we let the order k have a probability mass function $p(k) > 0$, for $k \in \mathbb{N}^d$ and let the normalizing constant τ have a probability density on $(0, \infty)$. We assume that G, k, and τ are *a priori* independent. The prior for f induced by the product of priors on G, k, and τ satisfies the conditions of Theorem 3 and thus has full topological support.

3. Posterior for the Spectral Density

3·1. *Dirichlet Process Representation*

We let the order of the Bernstein polynomials be restricted to the form $k = k\mathbf{1}$ so that the posterior computation simplifies considerably. We will refer to the common value k as the order of the Bernstein polynomial prior. Following Sethuraman (1994), we represent G as an infinite mixture of point masses at $Z_\ell = (Z_{\ell,1}, \ldots, Z_{\ell,d}) \sim G_0$ as,

$$G = \sum_{\ell=1}^{\infty} p_\ell \delta_{Z_\ell}, \tag{9}$$

where δ denotes a point mass function, $p_\ell = V_\ell \prod_{j<\ell}(1 - V_j)$, with $V_\ell \sim \text{Beta}(1, M)$, $Z_1, Z_2, \ldots$, $V_1, V_2, \ldots$ are independent of each other and are the parameters of G in the representation (9). To achieve a finite parameterization, we truncate the series in (9) at a large L and represent G by $G = \sum_{\ell=0}^{L} p_\ell \delta_{Z_\ell}$, where $p_0 = 1 - \sum_{\ell=1}^{L} p_\ell$, $Z_0 \sim G_0$ and is independent of the other parameters.

The rescaled spectral density f in (7) becomes,

$$f(\omega) = \tau \sum_{j=\mathbf{1}}^{k\mathbf{1}} \left[\sum_{\ell=0}^{L} p_\ell \mathcal{I}\Big\{ (j_1 - 1)/k < Z_{\ell,1} \le j_1/k, \ldots, (j_d - 1)/k < Z_{\ell,d} \le j_d/k \Big\} \times \prod_{i=1}^{d} \beta(\omega_i; j_i, k - j_i + 1) \right]. \tag{10}$$

It suffices to consider a finite number of parameters $Z_0, Z_1, \ldots, Z_L, V_1, V_2, \ldots, V_L, k$, and τ.

3·2. *Posterior for the Spectral Density*

In practice, the periodogram is computed at a finite set of Fourier frequencies,

$$\Omega = \left\{ \omega^* = (\omega_1^*, \ldots, \omega_d^*) : \omega_i^* = -\frac{2\pi}{n_i}\left\lfloor \frac{n_i - 1}{2} \right\rfloor, \ldots, \frac{2\pi}{n_i}\left\lfloor \frac{n_i}{2} \right\rfloor, i = 1, \ldots, d \right\},$$

where $\lfloor \cdot \rfloor$ denotes the floor function. Then, the periodogram can be viewed as the Fourier transform of the sample autocovariance function,

$$I(\omega^*) = \frac{1}{(2\pi)^d} \sum_{h \in \mathcal{Z}_n} \hat{C}(h) \exp(-ih'\omega^*), \tag{11}$$

where $\omega^* \in \Omega$, $\mathcal{Z}_n = \{-n_1 + 1, \ldots, n_1 - 1\} \times \cdots \times \{-n_d + 1, \ldots, n_d - 1\}$ and $\hat{C}(h)$ is a sample autocovariance function.

For evaluating the negative log-likelihood function up to a constant, we use Whittle's approximation $1/2 \sum_{\omega^* \in \Omega} \{\log f^*(\omega^*) + I(\omega^*)/f^*(\omega^*)\}$, where the summation is over the set of Fourier frequencies in Ω, I is the periodogram (11), and f^* is the spectral density (Whittle, 1954). Thus the posterior of $Z_0, Z_1, \ldots, Z_L, V_1, V_2, \ldots, V_L, k, \tau$ is proportional to,

$$\prod_{\omega^* \in \Omega} \exp\left[-\frac{1}{2}\Big\{ \log f(\omega) + I(\omega^*)/f(\omega) \Big\} \right] \prod_{\ell=1}^{L} M(1 - v_\ell)^{M-1} \prod_{\ell=0}^{L} g_0(z_\ell) p(k) p(\tau), \tag{12}$$

where $f(\omega)$ is specified in (10) with $\omega = (\omega^* + \pi)/2\pi$ and g_0 is the probability density of G_0. A simulation study to be shown in Section 5·1 demonstrates the posterior consistency empirically. However, it is challenging to establish posterior consistency theoretically. We believe that the main difficulty is to develop a contiguity theory in the spirit of Choudhuri et al. (2004b) for time series, which we are currently investigating.

3·3. *Markov Chain Monte Carlo for Sampling from the Posterior*

To sample from the posterior (12), we implement a Gibbs sampler and update the parameters componentwise. For the normalizing constant τ, we assume a conjugate Inverse Gamma prior and sample τ according to its full conditional distribution. Moreover, we update V_ℓ and Z_ℓ by a Metropolis-Hastings algorithm. In particular, we let the proposal distribution of V_ℓ in the Metropolis-Hastings update have a uniform distribution on $[V_\ell - \epsilon_\ell, V_\ell + \epsilon_\ell]$, where ϵ_ℓ depends on ℓ in order to match the posterior variance of the corresponding V_ℓ. The proposal distribution of Z_ℓ is selected in a similar manner. Based on experimentation for the case $d = 2$, it appears that the choice of $\epsilon_\ell = 1/\{\ell + 2(n_1 n_2)^{1/2}\}$ works well. To determine a suitable L, we note that a larger L is better in terms of accuracy of the approximation, but increases the amount of computing time. Again, based on experimentation, we find it suitable to let $L = \max\{20, (n_1 n_2)^{1/3}\}$, where 20 is chosen to keep L sufficiently large for moderately large lattice sizes. The mean of the posterior, denoted as $\hat{f}$, is used as the smoothed estimate of the rescaled spectral density f and thus, the spectral density f^* is estimated by $\hat{f}^*(\omega^*) = \hat{f}\{(\omega^* + \pi)/2\pi\}$. The 2.5^{th} and 97.5^{th} percentiles of the posterior spectral densities are obtained in a similar manner, which are used to construct a 95% credible band of f^*.

3·4. *Alternative Priors*

As an alternative to the Dirichlet process prior for G, we investigate a more general family of stick-breaking random measures. The stick-breaking priors are of the form (9), but with more general $V_\ell \sim$ Beta(a_ℓ, b_ℓ) (Ishwaran & Zarepour, 2000; Ishwaran & James, 2001). In particular, we consider the two-parameter Poisson-Dirichlet process, also known as the Pitman-Yor process $\mathcal{PY}(a, b)$ defined as a two-parameter stick-breaking random measure with parameters $a_\ell = 1 - a$ and $b_\ell = b + \ell a$ with $0 \le a < 1, b > -a$ (Pitman & Yor, 1997). The higher-order terms in the infinite-mixture representation of the Pitman-Yor process are again truncated. With $a = 0$ and $b = \beta$, the Pitman-Yor process reduces to a Dirichlet process. Another example of the Pitman-Yor process is the stable-law process with parameters $a = \alpha$, $b = 0$, and $0 < \alpha < 1$. In general, with $a = \alpha$ and $b = \beta$, the prior of V_ℓ is Beta($1 - \alpha, \beta + \ell\alpha$). We impose diffuse priors on the parameters of the Pitman-Yor process and devise a Markov chain Monte Carlo algorithm to sample from the posterior of the $Z_0, Z_1, \ldots, Z_L$, $V_1, V_2, \ldots, V_L$, k, τ as before. In addition to a uniform distribution, we also consider a diffuse truncated normal distribution as the base measure G_0.

4. THEORETICAL PROPERTIES OF BERNSTEIN POLYNOMIAL PRIOR

The random Bernstein distribution function of the form (3) induces a probability measure π on the space of probability distribution functions.

THEOREM 2. *Let* Π *denote the space of probability distribution functions on the unit square* Δ *equipped with the Borel* σ*-field* $\mathcal{T}$ *generated by the topology of weak convergence. Then the Bernstein distribution functions induce a probability measure* π *on* $(\Pi, \mathcal{T})$.

The probability measure induced in Theorem 2 is the Bernstein polynomial prior π with parameter $\mathcal{P}$, where $\mathcal{P}$ is the joint distribution of k and G.

Next, we establish the validity of the Bernstein polynomial prior π. Let $\mathcal{G}_n = \{g_1, \ldots, g_n\}$ denote a collection of real-valued continuous functions on the unit cube Δ, where $n = 1, 2, \ldots$.

For any probability distribution function $Q \in \Pi$ and $\epsilon > 0$, define a weak neighborhood of Q as,

$$N_{\mathcal{G}_n,\epsilon}(Q) = \left\{ Q^* \in \Pi : \max_{g \in \mathcal{G}_n} \left| \int g dQ^* - \int g dQ \right| < \epsilon \right\}. \tag{13}$$

Let Π_1 denote the subclass of all continuous probability distribution functions on Δ. For any $\epsilon > 0$, define a strong neighborhood of $Q \in \Pi_1$ as,

$$N_\epsilon(Q) = \{Q^* \in \Pi_1 : \sup_{\omega \in \Delta} |Q^*(\omega) - Q(\omega)| < \epsilon\}. \tag{14}$$

THEOREM 3. *Let the prior on the order parameter k satisfy $p(k) > 0$ for all $k \in \mathbb{N}^d$ and given k, let the conditional prior $p(\cdot|k)$ on the mixing weights u_k have positive density with respect to the Lebesgue measure on $\mathbb{R}^{d k_1 - 1}$ for all $u_k \in \mathcal{S}_k$ and $\mathcal{S}_k$ is defined in* (5). *Then,*

(i) π has full topological support on $(\Pi, \mathcal{T})$, in the sense that for all choices of $n, \mathcal{G}_n, Q \in \Pi$ and $\epsilon > 0$, we have $\pi\{N_{\mathcal{G}_n,\epsilon}(Q)\} > 0$, where $N_{\mathcal{G}_n,\epsilon}(Q)$ is defined in (13).

(ii) π has full topological support on $(\Pi_1, \mathcal{T}_1)$, in the sense that for all choices of $\epsilon > 0$ and $Q \in \Pi_1$, we have $\pi\{N_\epsilon(Q)\} > 0$, where $N_\epsilon(Q)$ is defined in (14).

Theorem 3(i) states that if the probability distribution of k is positive over the entire set $\mathbb{N}^d$ and for any $k \in \mathbb{N}^d$, the conditional distribution of u_k given k is nonzero at every point in $\mathcal{S}_k$, then the Bernstein polynomial prior π has full topological support on Π in the sense that every weak neighborhood in $\mathcal{T}$ has positive π-measure. Under the same condition, Theorem 3(ii) states that any neighborhood of a given continuous probability distribution function has positive π-probability, where the neighborhood is defined in terms of a Kolmogorov-Smirnov distance. The two parts of the theorem ensure that every probability distribution on the unit cube Δ is in the topology of weak convergence of the Bernstein polynomial prior and moreover, every continuous probability distribution is in the topology of uniform convergence of the Bernstein polynomial prior, provided that the distributions $p(k)$ and $p(u_k|k)$ have full support.

5. NUMERICAL EXAMPLES

5·1. *Simulation Study*

For the simulation study, we focus on the case $d = 2$ and vary the $r \times r$ sampling grid by letting $r = 5, 10$, or 20. For each grid size, we simulate data from a stationary and isotropic Gaussian process. The mean of the Gaussian process is set to 0. We consider the Matérn class of autocovariance functions $C(h) = \sigma(h/2\rho)^\nu 2K_\nu(h/\rho)/\Gamma(\nu)$, where $h = ||h||$ denotes the distance of a spatial lag h, σ is a variance parameter, ρ is a range parameter, and ν is a smoothing parameter (Cressie, 1993). We let $\sigma = 1$, but vary the range parameter ρ and the smoothing parameter ν. We let $\rho = 1$ or 3, corresponding to a shorter and a longer range of dependence. We also let $\nu = 1/2$ or ∞, corresponding to the exponential model and the Gaussian model. We simulate 100 data sets for each combination of r, ρ, and ν.

For each simulated data set, we apply the nonparametric Bayesian method developed above to estimate the spectral density. For the Gibbs sampler with Metropolis-Hastings updates, we let the order of the Bernstein polynomials have a probability mass $p(k) = c\exp(-0.05k^2)$. We set the weight parameter to $M = 1$ and G_0 to a uniform distribution in the Dirichlet process for specifying the prior of G. The burn-in length is 1,000 and the Monte Carlo sample size after burn-in is 6,000. Let $\hat{f}^*$ denote the posterior mean of the spectral density. To calibrate the accuracy of $\hat{f}^*$ as an estimate of the true spectral density f^*, we define an L_1-error as $\|\hat{f}^* - f^*\|_1 =$

Table 1. *Median and interquartile range of the L_1-errors in the estimation of spectral density functions*

Grid Size	Method	Exponential		Gaussian	
		$\rho = 1$	$\rho = 3$	$\rho = 1$	$\rho = 3$
5×5	NBE	0.42 (0.35, 0.50)	0.58 (0.52, 0.67)	0.31 (0.28, 0.36)	0.57 (0.53, 0.66)
	PE	0.16 (0.10, 0.29)	0.20 (0.13, 0.29)	0.33 (0.28, 0.42)	0.46 (0.42, 0.51)
10×10	NBE	0.33 (0.28, 0.38)	0.48 (0.45, 0.52)	0.18 (0.15, 0.23)	0.49 (0.47, 0.51)
	PE	0.07 (0.04, 0.11)	0.23 (0.17, 0.28)	0.25 (0.22, 0.31)	0.51 (0.47, 0.58)
20×20	NBE	0.29 (0.27, 0.31)	0.44 (0.43, 0.46)	0.13 (0.12, 0.13)	0.32 (0.28, 0.35)
	PE	0.04 (0.02, 0.06)	0.11 (0.06. 0.17)	0.21 (0.20, 0.23)	0.51 (0.49, 0.54)

NBE, nonparametric Bayesian; PE, parametric Bayesian

$\int_{\omega^* \in (-\pi,\pi]\times[0,\pi]} |\hat{f}^*(\omega^*) - f^*(\omega^*)| d\omega^*$. Note that, due to the symmetry of the spectral density f^*, we only focus on the L_1-error of $\hat{f}^*$ on $(-\pi, \pi] \times [0, \pi]$.

For comparison, we devise a parametric approach for estimating the spectral density. Regardless of the data generating mechanism, we assume that the random field is a Gaussian process with an exponential model as the autocovariance function $C(h) = \sigma \exp(-h/\rho)$, where σ is a variance parameter and ρ is a range parameter. We apply a fully Bayesian approach to estimate the model parameters σ and ρ and thus the posterior of the spectral density. In particular, we develop another Markov chain Monte Carlo algorithm to draw Monte Carlo samples of the model parameters from the posterior distribution of σ and ρ. We then obtain the posterior of the spectral density according to the parametric form and let the posterior mean be the estimate of the true spectral density f^*. The L_1-error of the spectral density estimate is computed in the same way. The burn-in length is set to 1,000 and the Monte Carlo sample size after burn-in is 6,000.

Table 1 gives the median and interquartile range of the L_1-errors based on 100 simulations for the three grid sizes and two spatial models. When the true autocovariance function follows an exponential model, the median L_1-error of the nonparametric Bayesian estimate becomes smaller and thus the estimation becomes more accurate with increasing grid size. The corresponding median L_1-errors of the parametric Bayesian estimate, in contrast, are smaller, because the correct exponential model is used to fit the data. There tends to be more variation in the L_1-error with the parametric Bayesian approach. When the true autocovariance function follows a Gaussian model, again as the grid size increases, estimation of the spectral density becomes more accurate. This is in contrast to the corresponding median L_1-error of the parametric Bayesian estimate, when the model is misspecified as an exponential model. Except the case of $r = 5$ and $\rho = 3$, the L_1-error is smaller with the nonparametric Bayesian estimate. Again there is more variation in the L_1-error using the parametric Bayesian approach.

5·2. *Soil Data Example*

For illustration, we analyze a soil property data set featured in the geoR package of R (R Development Core Team, 2008). In the example, soil samples were collected on a 25×10 regular grid and soil chemical properties were measured. Here we focus on the catium exchange capability and apply the nonparametric Bayesian method to obtain the posterior of the spectral density. Figure 1(a) and (b) show the posterior means and 95% credible bands of the spectral density on the $\log_{10}$ scale at $\omega_1^* = 0$ and $\omega_2^* = 0$ respectively. Figure 1(c) and (d) show the surface and contour plot of the posterior estimates on the $\log_{10}$ scale. These figures reveal a larger peak at $\omega^* = (0, 0)$ and two smaller ones at frequencies around $(-2.5, 0)$ and $(2.5, 0)$. Furthermore, we perform sensitivity analysis of the prior selection and the main features of the posterior estimates

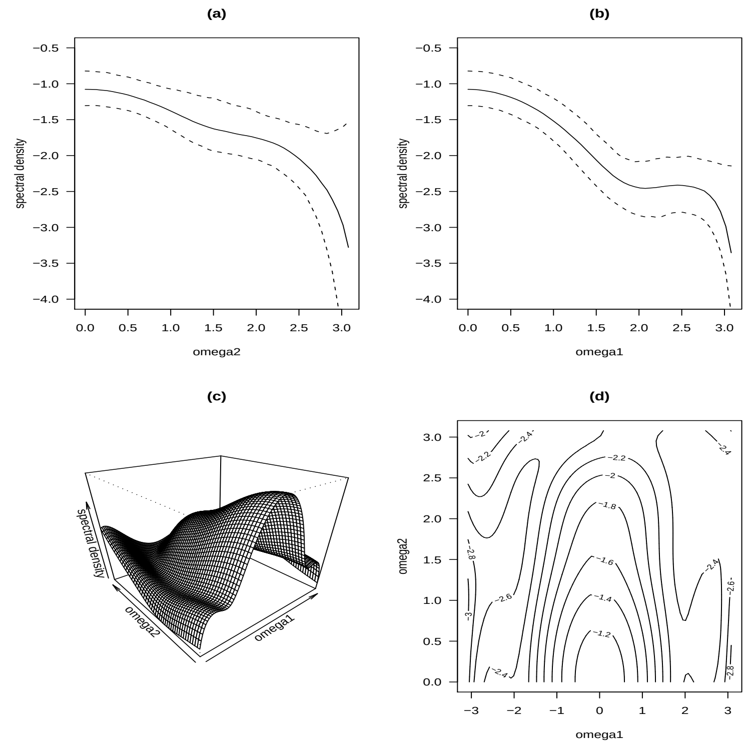

Fig. 1. Spectral density function estimate on the $\log_{10}$ scale using nonparametric Bayesian (a) $\omega_1^* = 0$ (b) $\omega_2^* = 0$ (both with a 95% credible band); (c) surface plot; (d) contour plot.

are similar under different Pitman-Yor process priors and base measures. Figure 2 shows the contour plots of the posterior mean on the $\log_{10}$ scale for the three forms of the Pitman-Yor process combined with two different base measures. We add that the smoothness of the posterior do not differ much among the various stick-breaking priors and the base measures.

6. Shape Restriction for Bernstein Polynomials

For density estimation problems, it is important that the Bernstein polynomial density can be restricted to an admissible class of densities with certain geometric/shape properties of the target density. For spectral density problems, such shape restrictions are not common. However, in certain circumstances, prior information may be available on the geometric shape of the spectral density. For example, smoothness of the random field may be imposed by specifying decreasing spectral densities with more contribution from lower frequencies. Therefore, it is desirable that the Bernstein polynomial prior capture such shape restrictions. It is well-known that the one-dimensional Bernstein polynomials retain shape properties such as monotonicity and concavity of the function generating the coefficients (Lorentz, 1986). We investigate similar properties in the multi-dimensional case with the objective of specifying shape restrictive priors.

The derivatives of the Bernstein polynomial are given by,

$$\frac{\partial^{|\alpha|}}{\partial\omega_1^{\alpha_1}\cdots\partial\omega_d^{\alpha_d}} B_{k,G}(\omega) = \Big\{ \big(\prod_{i=1}^{d} \frac{k_i!}{(k_i-\alpha_i)!} \Big\} \sum_{j=\mathbf{0}}^{k-\alpha} \nabla^{\alpha} G(j/k) \prod_{i=1}^{d} p_{j_i,k_i-\alpha_i}(\omega_i), \tag{15}$$

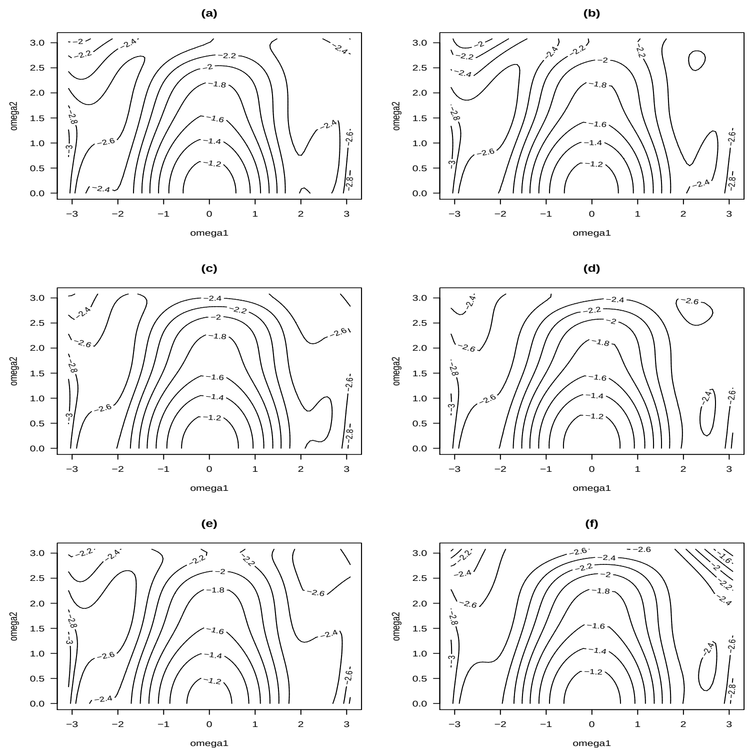

Fig. 2. Contour plots of spectral density estimate on the $\log_{10}$ scale using non-parametric Bayesian under different prior of the distribution G and the base measure G_0. (a) G = Dirichlet process, G_0 = uniform; (b) G = stable law process, G_0 = uniform; (c) G = Pitman-Yor process, G_0 = uniform; (d) G = Dirichlet process, G_0 = normal; (e) G = stable law process, G_0 = normal; (f) G = Pitman-Yor process, G_0 = normal.

where ∇ is the forward difference operator and $\alpha = (\alpha_1, \ldots, \alpha_d)$ is a multi-index with $|\alpha| = \alpha_1 + \cdots + \alpha_d$. Let d_i denote the vector with all ones except a two at the i^{th} place. Then the Bernstein density is monotonically increasing along the i^{th} coordinate if $\partial^{|\alpha|}/\partial\omega_1^{\alpha_1} \cdots \partial\omega_d^{\alpha_d} B_{k,G}(\omega) \geq 0$ for $\alpha = d_i$. Clearly, from (15), a sufficient condition for the Bernstein density to be increasing along the i^{th} coordinate is that $\nabla^{d_i} G(j/k) \geq 0$. Following arguments in Chang et al. (2007), the set of Bernstein polynomials with coefficients monotonically increasing along the i^{th} coordinate will be dense in the set of continuous functions on the unit cube which are monotonically increasing along the i^{th} coordinate, equipped with the topology of uniform convergence. Prior positivity of weak neighborhoods of functions increasing along the i^{th} coordinate can be proved in the same manner as in Theorem 3.

Convexity of the coefficients does not guarantee global convexity of the resulting polynomial. When the Bernstein polynomials are defined on the simplex, Chang & Davis (1984), Chang & Feng (1984), Sauer (1991) among others gave sufficient conditions for the Hessian of the Bernstein polynomial to be semi-positive definite in order to have convexity. Sauer (1991) defined two alternative notions of convexity that are inherited from the coefficients of the Bernstein polynomial. However, when the functions are defined on the unit cube, such results are not available. Even though exact convexity is not retained, convexity of the coefficients does guarantee asymptotic convexity of the polynomials. By results of Butzer (1953), if the density associated with the coefficients admits continuous cross-partial derivatives up to the second order, then the cross-partial derivatives (up to the second order) of the Bernstein density will converge uniformly to the corresponding derivatives of the target density g. Hence, if g is convex, one can find a K such that if $\min\{k_i\} > K$ then the Hessian of the associated Bernstein polynomial $B_{k,G}$ will be nonnegative. If a prior on g is restricted to the set of convex probability densities and $\min\{k_i\}$ is large, then the corresponding random Bernstein density will be convex with large probability.

For specification of prior it is necessary to specify a random measure on the restricted space of continuous convex densities on the unit cube. For the one-dimensional case, a sufficient condition for the Bernstein density to be convex is that $\nabla^2 u_{k,G}(j) \geq 0$ for each j. In the multi-dimensional case, an obvious class of functions for which the Bernstein density is convex is where separation of variables occurs. That is, the target density is of the form $g(\omega) = \sum_{i=1}^{d} w_i g_i(\omega_i)$, where $w_i \geq 0$ and g_i are convex. However, from a practical point of view, this is too restrictive in the sense that such Bernstein polynomials can only model convex functions in which separation of variables occurs. This would require that all cross-partial derivatives vanish everywhere. In the two-dimensional case, the requirement that cross-partial derivatives vanish everywhere can be relaxed. We present a set of sufficient conditions for convexity of the Bernstein density in the bivariate case in the following proposition. It will be convenient to denote the coefficients $u_{k,G}(j)$ as $c_j = c_{j_1,\ldots,j_d}$. We define $c_{j_1,j_2} = 0$ if either $j_1 < 0$ or $j_2 < 0$.

PROPOSITION 1. *Let $B_{k,G}(\omega)$ be the Bernstein distribution defined in* (3) *and let the corresponding Bernstein density be $b_{k,G}(\omega)$ defined in* (4). *Then $b_{k,G}(\omega)$ is convex if:*
(i) $\nabla^{(2,\,0)} c_{j_1,j_2} \geq 0$ *and* $\nabla^{(0,\,2)} c_{j_1,j_2} \geq 0$ *for all* j_1 *and* j_2.
(ii) $\nabla^{(1,\,1)} c_{j_1,j_2} = 0$ *if either* $j_1 = 0$ *or* $j_2 = 0$.
(iii) $Q_{j_i,j_2,l_1,l_2} \geq 0$ *for* $j_1, j_2, l_1, l_2 \geq 0$ *where,*

$$Q_{j_i,j_2,l_1,l_2} = j_1 l_2 \nabla^{(2,\,0)} c_{j_1-1,j_2+1} \nabla^{(0,\,2)} c_{l_1+1,l_2-1} - (j_2+1)(l_1+1) \nabla^{(1,\,1)} c_{j_1,j_2} \nabla^{(1,\,1)} c_{l_1,l_2}.$$

Condition (i) guarantees that the Bernstein density is convex along each direction. Condition (ii) imposes the condition of vanishing cross-partial derivatives along the boundary of the unit square. Condition (iii) is a type of local convexity condition at any interior point. Clearly, if the coefficients are convex along each direction and the cross-partial differences vanish everywhere then conditions (i)-(iii) are satisfied. Thus, the conditions of the proposition are weaker than the condition of separation of variables. It is not immediately clear how to sample coefficients from the set of coefficients that satisfy conditions (i)-(iii). Further simplification of the conditions to make them amenable to prior specification is a topic of future research.

7. FURTHER DISCUSSION

In summary, we have demonstrated that the multi-dimensional Bernstein polynomial prior provides a way to smooth a Dirichlet process and a convenient parameterization that enables nonparametric Bayesian inference for the spectral density of a random field. Although uncommon for spectral density estimation, shape restriction is of great interest in general density estimation problems. We have attempted to address shape restriction, but acknowledge that further work will be needed to make it truly usable for practical problems. In addition, parameterization other than that by multi-dimensional Bernstein polynomials may be worth investigating. For example, for spectral density that is continuous and bounded away from zero, logistic density may offer an alternative parameterization with $f(\omega) = \exp\{g(\omega)\}$, where $g \in \mathcal{L}^2(\Delta)$ (Lenk, 1988). One possibility is to represent g as the realization of a Gaussian process, which naturally induces a prior. Another option may be to represent g by a spline basis and impose Gaussian prior on the coefficients of the representation. We will investigate such alternative approaches in our future research as well.

APPENDIX 1: PROOF OF THEOREM 1

Proof. For any $1 \leq i \leq d$, let $p_{j_i k_i}$ denote the probability mass function of Binomial(k_i, ω_i). By the variance formula, we have $\sum_{j_1=0}^{k_i}(j_i - k_i\omega_i)^2 p_{j_i k_i} = k\omega_i(1-\omega_i)$. Therefore,

$$\sum_{j_i:|j_i/k_i(\omega_i)-\omega_i|\geq\delta} p_{j_i k_i}(\omega_i) \leq \sum_{j_i:|j_i/k_i-\omega_i|\geq\delta} (j_i/k_i-\omega_i)^2 p_{j_i k_i}(\omega_i)/\delta^2 \leq \omega_i(1-\omega_i)/(k_i\delta^2)$$
$$\leq 1/(4\delta^2 k_i) \leq 1/(4\delta^2 \min\{k\}),$$

where the last inequality holds since $\omega_i(1-\omega_i) \leq 1/4$ on $[0,1]$.

By assumption, the function G is bounded, say $|G(\boldsymbol{\omega})| \leq M$ on the unit cube Δ. Let $\boldsymbol{\omega}$ be a point of continuity. Then for a given $\epsilon > 0$, we can find a $\delta > 0$ such that if $|\omega_i^+ - \omega_i| < \delta,\ i = 1, 2, \ldots, d$, implies that $|G(\boldsymbol{\omega}^+) - G(\boldsymbol{\omega})| < \epsilon$. Let $B_i^0 = \{j_i : |j_i/k_i - \omega_i| < \delta\}$ and $B_i^1 = \{j_i : |j_i/k_i - \omega_i| \geq \delta\}$. Let ϵ_i be either 0 or 1 and let $\prod B_i^{\epsilon_i} = B_1^{\epsilon_1} \times \cdots \times B_d^{\epsilon_d}$ be a rectangular subset of the d-dimensional lattice $\{1, \ldots, k_1\} \times \cdots \times \{1, \ldots, k_d\}$. Then we have,

$$\begin{aligned}
\Big|G(\boldsymbol{\omega}) - B_{k,G}(\boldsymbol{\omega})\Big| &= \left|\sum_{j=\mathbf{0}}^{k}\Big[G(j/k) - G(\boldsymbol{\omega})\Big]\prod_{i=1}^{d} p_{j_i k_i}(\omega_i)\right| \\
&\leq \sum_{j\in\prod B_i^0}\Big|G(j/k) - G(\boldsymbol{\omega})\Big|\prod_{i=1}^{d} p_{j_i k_i}(\omega_i) \\
&\quad + \sum_{j\in\prod B_i^{\epsilon_i}\,:\,\prod\epsilon_i\neq 0}\Big|G(j/k) - G(\boldsymbol{\omega})\Big|\prod_{i=1}^{d} p_{j_i k_i}(\omega_i) \\
&\leq \epsilon + 2M \sum_{j\in\prod B_i^{\epsilon_i}\,:\,\prod\epsilon_i\neq 0}\prod_{i=1}^{d} p_{j_i k_i}(\omega_i) \\
&\leq \epsilon + 2M\{[1 + (4\delta^2\min\{k\})^{-1}]^d - 1\}. \qquad (16)
\end{aligned}$$

When $\min\{k\}$ is sufficiently large, we have $|G(\boldsymbol{\omega}) - B_{k,G}(\boldsymbol{\omega})| < 2\epsilon$. If G is continuous then it is uniformly continuous on the unit square Δ, and for any $\epsilon > 0$, we can find a δ such that for two points $\boldsymbol{\omega}$ and $\boldsymbol{\omega}^+$, if $|\omega_i^+ - \omega_i| < \delta,\ i = 1, 2, \ldots, d$, then $|G(\boldsymbol{\omega}^+) - G(\boldsymbol{\omega})| < \epsilon$. Thus, (16) holds for some δ that is independent of $\boldsymbol{\omega}$ and thus $B_{k,G}(\boldsymbol{\omega})$ converges to $G(\boldsymbol{\omega})$ uniformly.

If G admits a continuous density g, then the coefficients $u_k(j)$ in $b_{k,G}$ are given by $u_{k,G}(j) = \int_{\text{Cube}(j,k)} g(\boldsymbol{\eta})d\boldsymbol{\eta}$. Since g is continuous on Δ, we can find M_g such that $g(\boldsymbol{\eta}) < M_g$ for all $\boldsymbol{\eta} \in \Delta$. Since g is uniformly continuous on Δ, we can find a δ_g such that for two points $\boldsymbol{\omega}$ and $\boldsymbol{\omega}^+$, if $|\omega_i^+ - \omega_i| < \delta_g,\ i = 1, 2, \ldots, d$, then $|g(\boldsymbol{\omega}^+) - g(\boldsymbol{\omega})| < \epsilon$. Also, note that ${}^d k_1 =$

$1/Vol\{\text{Cube}(j,k)\}$ and $\beta(\omega_i; j_i, k_i - j_i + 1) = k_i p_{j_i-1,k_i-1}(\omega_i)$. Hence,

$$\Big|b_{k,G}(\boldsymbol{\omega}) - g(\boldsymbol{\omega})\Big| = \left|\sum_{j=\mathbf{1}}^{k}\Big[u_{k,G}(j) - g(\boldsymbol{\omega})Vol\{\text{Cube}(j,k)\}\Big]\ {}^{d}k_1 \prod_{i=1}^{d} p_{j_i-1,k_i-1}(\omega_i)\right|$$

$$\leq \sum_{j=\mathbf{1}}^{k}\Big[\int_{\text{Cube}(j,k)} |g(\boldsymbol{\eta}) - g(\boldsymbol{\omega})| d\boldsymbol{\eta}\Big]\ {}^{d}k_1 \prod_{i=1}^{d} p_{j_i-1,k_i-1}(\omega_i) \tag{17}$$

Changing the variable j to $j - \mathbf{1}$, and replacing δ with δ_g in the definition of $B_i^{\epsilon_i}$, we have

$$\begin{aligned}\Big|b_{k,G}(\boldsymbol{\omega}) - g(\boldsymbol{\omega})\Big| &\leq \sum_{j\in\prod B_i^0}\Big|\int_{\text{Cube}(j+\mathbf{1},k)} |g(\boldsymbol{\eta}) - g(\boldsymbol{\omega})| d\boldsymbol{\eta}\Big|\ {}^{d}k_1 \prod_{i=1}^{d} p_{j_i,k_i-1}(\omega_i)\\ &\quad + \sum_{j\in\prod B_i^{\epsilon_i}\,:\,\prod \epsilon_i\neq 0}\Big|\int_{\text{Cube}(j+\mathbf{1},k)} |g(\boldsymbol{\eta}) - g(\boldsymbol{\omega})| d\boldsymbol{\eta}\Big|\ {}^{d}k_1 \prod_{i=1}^{d} p_{j_i,k_i-1}(\omega_i)\\ &\leq \epsilon + 2M_g \sum_{j\in\prod B_i^{\epsilon_i}\,:\,\prod \epsilon_i\neq 0} \prod_{i=1}^{d} p_{j_i,k_i-1}(\omega_i)\\ &\leq \epsilon + 2M\{[1 + (8\delta_g^2 \min\{k\})^{-1}]^d - 1\}. \end{aligned} \tag{18}$$

Because δ_g is independent of ϵ and $\boldsymbol{\omega}$, we have the result. □

APPENDIX 2: PROOF OF THEOREM 2

Proof. We extend the proof of a one-dimensional case in Petrone (1999a). Equip the space Π of probability distribution functions on the unit square Δ with a Borel σ-field $\mathcal{T}$ generated by the topology of weak convergence. Let $\mathbb{N}^d$, the set of positive integer lattice in d-dimension, be equipped with the power set $\mathcal{P}(\mathbb{N}^d)$. Let $\Omega = \mathbb{N}^d \times \Pi$, $\mathcal{B}(\Omega)$ be the product σ-field $\mathcal{P}(\mathbb{N}^d) \times \mathcal{T}$, and P be a probability measure on $(\Omega, \mathcal{B}(\Omega))$. For each $k \in \mathbb{N}^d$ and $G \in \Pi$, define an operator from Ω to Π as,

$$B_{k,G}(\omega) = \sum_{j=\mathbf{0}}^{k} G(j(\omega)/k) \prod_{i=1}^{d} p_{j_i k_i}(\omega_i)$$

where $j(\omega) = (j_1(\omega), \ldots, j_d(\omega))$ and $j_i(\omega) = 0$ if $\omega_i < 0$, $j_i(\omega) = j_i$ if $0 \leq \omega_i \leq 1$ and $j_i(\omega) = k_i$ if $\omega_i > 1$. For fixed k and G, $B_{k,G}(\cdot)$ is a probability distribution function in Π.

A mapping H from $(\Omega, \mathcal{B}, P)$ to Π is a random distribution function, if and only if for each fixed k and ω, the real function $H_{k,\cdot}(\omega)$ on Ω is a Π-valued measurable function (Billingsley, 1999). Furthermore, if a random distribution function H is a measurable map from $(\Omega, \mathcal{B}, P)$ to Π, the distribution of H is a prior probability measure on $(\Pi, \mathcal{T})$. In this case, $B_{k,G}(\omega)$ is a random Bernstein polynomial, because for each k and ω, $B_{k,\cdot}(\omega)$ is a random variable in $(\Pi, \mathcal{T})$. Therefore, $B_{k,G}(\omega)$ induces a prior π on $(\Pi, \mathcal{T})$. □

APPENDIX 3: PROOF OF THEOREM 3

Proof. Let $\mathcal{G}_n = \{g_1, \ldots, g_n\}$ denote a collection of real-valued continuous functions on the unit cube Δ, where $n = 1, 2, \ldots$. Let $Q \in \Pi$ be a probability distribution function in Δ and let $\epsilon > 0$. As in Petrone (1999a), to show that π has full topological support on $(\Pi, \mathcal{B})$, it suffices to show that $\pi\{N_{\mathcal{G}_n,\epsilon}(Q)\} > 0$ for all choices of $n, \mathcal{G}_n, Q$ and ϵ where $N_{\mathcal{G}_n,\epsilon}(Q)$ is defined in (13). Fix $n, \mathcal{G}_n, Q$, and ϵ. For any $k \in \mathbb{N}^d$, let $\mathcal{B}_k$ denote the set of Bernstein probability distributions in Π that have order k. We can find k^* such that $B_{k^*,Q} \in N_{\mathcal{G}_n,\epsilon/2}(Q)$. Then $\pi\{N_{\mathcal{G}_n,\epsilon}(Q)\} > \pi\{\mathcal{B}_{k^*} \cap N_{\mathcal{G}_n,\epsilon/2}(Q)\}$. Let $B_{k^*} \in \mathcal{B}_{k^*}$ be any Bernstein distribution with order k^*. Then by definition,

$$b_{k^*}(\omega) = \sum_{j=\mathbf{1}}^{k^*} u_{k^*}(j) \prod_{i=1}^{d} \beta(\omega_i; j_i, k_i^* - j_i + 1),$$

for some $u_{k^*} = (u_{k^*}(j), j_i = 1, \ldots, k_i^*, i = 1, \ldots, d) \in \mathcal{S}_{k^*}$. Let $R = \max_{g \in \mathcal{G}_n} \sup_{\omega \in \Delta} g(\omega)$. Then, it is easy to show that,

$$\max_{g \in \mathcal{G}_n} \left| \int g dB_{k^*} - \int g dB_{k^*,Q} \right| \leq R({}^d k_1^*) \|u_{k^*} - u_{k^*,Q}\|_\infty,$$

where $\|u_{k^*} - u_{k^*,Q}\|_\infty = \max_j |u_{k^*}(j) - u_{k^*,Q}(j)|$. Therefore,

$$\pi\{N_{\mathcal{G}_n,\epsilon}(Q)\} > \pi\{\mathcal{B}_{k^*} \cap N_{\mathcal{G}_n,\epsilon/2}(Q)\} > p(k^*) p(\{u_{k^*} : \|u_{k^*} - u_{k^*,Q}\|_\infty < \epsilon/(2R({}^d k_1^*))\} \mid k^*).$$

Because the set $\{u_{k^*} \in \mathcal{S}_{k^*} : \|u_{k^*} - u_{k^*,Q}\|_\infty < \epsilon/(2R({}^d k_1^*))\}$ has positive Lebesgue measure and the conditional prior $p(\cdot|k^*)$ has positive Lebesgue density over $\mathcal{S}_{k^*}$, we have part (i).

Now, we prove part (ii). Let $d(Q^*, Q) = \sup_\omega |Q^*(\omega) - Q(\omega)|$. Then similar to the proof of part (i), we have,

$$\pi\left(\{Q^* \in \Pi_1 : d(Q^*, Q) < \epsilon\}\right) = \pi\left(\left\{B_{k,Q^*}, k \in \mathbb{N}^d, Q^* \in \Pi_1 : d(B_{k,Q^*}, Q)| < \epsilon\right\}\right).$$

Now $d(B_{k,Q^*}, Q) \leq d(B_{k,Q^*}, B_{k,Q}) + d(B_{k,Q}, Q)$. Since Q is a continuous probability distribution function, $B_{k,Q}$ converges to Q uniformly over the unit square Δ. Therefore $d(B_{k,Q}, Q)$ can be made arbitrarily small by choosing $\min\{k\}$ large enough. Furthermore, for any $\omega \in \Delta$,

$$|B_{k,Q^*}(\omega) - B_{k,Q}(\omega)| = \sum_{j=\mathbf{0}}^{k} \left| Q^*(j/k) - Q(j/k) \right| \prod_{i=1}^{d} p_{j_i k_i}(\omega_i)$$
$$\leq \max_{j \preceq k} \left| Q^*(j/k) - Q(j/k) \right| \leq {}^d k_1 \|u_{k,Q^*} - u_{k,Q}\|_\infty.$$

Using the argument given in the proof of part (i), we have part (ii). □

APPENDIX 4: PROOF OF PROPOSITION 1

Proof. The determinant of the Hessian matrix H is,

$$|H| = \Big\{ k_1 k_2 (k_1-1)(k_2-1) \sum_{j_1=0}^{k_1-2} \sum_{j_2=0}^{k_2} \sum_{l_1=0}^{k_1} \sum_{l_2=0}^{k_2-2} \nabla^{(2,\,0)} c_{j_1,j_2} \nabla^{(0,\,2)} c_{l_1,l_2} \, p_{j_1(k_1-2)}(\omega_1) p_{j_2 k_2}(\omega_2) p_{l_1 k_1}(\omega_1) p_{l_2(k_2-2)}(\omega_2) \Big\} - \Big\{ k_1^2 k_2^2 \sum_{j_1=0}^{k_1-1} \sum_{j_2=0}^{k_2-1} \sum_{l_1=0}^{k_1-1} \sum_{l_2=0}^{k_2-1} \nabla^{(1,\,1)} c_{j_1,j_2} \nabla^{(1,\,1)} c_{l_1,l_2} \, p_{j_1(k_1-1)}(\omega_1) p_{j_2 k_2-1}(\omega_2) p_{l_1 k_1-1}(\omega_1) p_{l_2(k_2-1)}(\omega_2) \Big\}.$$

Thus, using condition (ii) and after some algebra the determinant of the Hessian is equal to $T_1 + T_2$ where,

$$T_1 = \Big\{ k_1 k_2 (k_1-1)(k_2-1) \sum_{j_1=0}^{k_1-2} \sum_{j_2=0}^{k_2} \sum_{l_1=0}^{k_1} \sum_{l_2=0}^{k_2-2} \nabla^{(2,\,0)} c_{j_i,j_2} \nabla^{(0,\,2)} c_{l_i,l_2} \, p_{j_1(k_1-2)}(\omega_1) p_{j_2 k_2}(\omega_2) p_{l_1 k_1}(\omega_1) p_{l_2(k_2-2)}(\omega_2) \Big\},$$

with the restriction that in the sum either j_2 is zero and/or l_1 is zero. The term T_2 is,

$$T_2 = k_1^2 k_2^2 \sum_{j_1=0}^{k_1-1} \sum_{j_2=0}^{k_2-1} \sum_{l_1=0}^{k_1-1} \sum_{l_2=0}^{k_2-1} (j_2+1)^{-1} (l_1+1)^{-1} Q_{j_1,j_2,l_1,l_2} \, p_{j_1(k_1-1)}(\omega_1) p_{j_2 k_2-1}(\omega_2) p_{l_1 k_1-1}(\omega_1) p_{l_2(k_2-1)}(\omega_2).$$

By condition (i) $T_1 \geq 0$ and by condition (iii) $T_2 \geq 0$. By condition (i) the diagonal entries of H are non-negative. Hence, H is semi-positive definite. □

REFERENCES

BILLINGSLEY, P. (1999). *Convergence of Probability Measures.* New York: Wiley, 2nd ed.

BÖHM, S., HEINRICH, L. & SCHMIDT, V. (2002). Kernel estimation of the spectral density of stationary random closed sets. *Australian and New Zealand Journal of Statistics* **46**, 41–51.

BUTZER, P. L. (1953). On two-dimensional Bernstein polynomials. *Canadian Journal of Mathematics* **5**, 107–113.

CHANG, G. & DAVIS, P. J. (1984). The convexity of Bernstein polynomials over triangles. *Journal of Approximation Theory* **40**, 11–28.

CHANG, G. & FENG, Y. Y. (1984). An improved condition for the convexity of Bernstein Bezier patches. *Computer Aided Geometric Design* **1**, 279–283.

CHANG, I. S., CHIEN, L. C., CHAO, A. H., WEN, C. & WU, Y. J. (2007). Shape restricted regression with random Bernstein polynomials. *Complex Datasets and Inverse Problems: Tomography, Networks and Beyond, IMS Lecture Notes* **54**, 187–202.

CHOUDHURI, N., GHOSAL, S. & ROY, A. (2004a). Bayesian estimation of the spectral density of a time series. *Journal of the American Statistical Association* **99**, 1050–1059.

CHOUDHURI, N., GHOSAL, S. & ROY, A. (2004b). Contiguity of the Whittle measure for a Gaussian time series. *Biometrika* **91**, 211–218.

CRESSIE, N. (1993). *Statistics for Spatial Data.* New York: Wiley, revised ed.

EPIFANI, I. (1999). *Some Results on the Probability Law of Random Bernstein Polynomials and their Applications in Bayesian Statistics (in Italian).* PhD Dissertation, University of Trento, Trento.

HEYDE, C. C. & GAY, R. (1993). Smoothed periodogram asymptotics and estimation for processes and fields with possible long-range dependence. *Stochastic Processes and their Applications* **45**, 169–182.

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 in hierarchical models. *Biometrika* **87**, 371–390.

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

LENK, P. J. (1988). The logistic normal distribution for Bayesian, nonparametric, predictive densities. *Journal of the American Statistical Association* **83**, 509–516.

LORENTZ, G. G. (1986). *Bernstein Polynomials*. New York: Chelsea.

MÜLLER, P. & QUINTANA, F. (2004). Nonparametric Bayesian data analysis. *Statistical Science* **19**, 95–110.

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

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

PETRONE, S. (2004). On the role of mixtures in Bayesian nonparametrics. *Atti della XLII Riunione Scientifica della Societa Italiana di Statistica* **1**, 257–268.

PITMAN, J. & YOR, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Annals of Probability* **25**, 855–900.

R DEVELOPMENT CORE TEAM (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

REGAZZINI, E. & SAZONOV, V. (1999). Approximation of laws of multinomial parameters by mixtures of Dirichlet distributions with applications to Bayesian inference. *Acta Applicandae Mathematicae* **58**, 247–264.

RIPLEY, B. D. (1981). *Spatial Statistics*. New York: Wiley.

ROBINSON, P. M. (2007). Nonparametric spectrum estimation for spatial data. *Journal of Statistical Inference and Planning (Special Issue in honor of Puri, M. L.)* **137**, 1024–1034.

SAUER, T. (1991). Multivariate Bernstein polynomials and convexity. *Computer Aided Geometric Design* **8**, 465–478.

SCHABENBERGER, O. & GOTWAY, C. A. (2005). *Statistical Methods for Spatial Data Analysis*. Boca Raton: Chapman & Hall.

SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4**, 639–650.

STEIN, M. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer.

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

WHITTLE, P. (1954). On stationary processes in the plane. *Biometrika* **41**, 431–449.