

# Latent Stick-Breaking Processes

Abel RODRÍGUEZ, David B. DUNSON, and Alan E. GELFAND

---

We develop a model for stochastic processes with random marginal distributions. Our model relies on a stick-breaking construction for the marginal distribution of the process, and introduces dependence across locations by using a latent Gaussian copula model as the mechanism for selecting the atoms. The resulting latent stick-breaking process (LaSBP) induces a random partition of the index space, with points closer in space having a higher probability of being in the same cluster. We develop an efficient and straightforward Markov chain Monte Carlo (MCMC) algorithm for computation and discuss applications in financial econometrics and ecology. This article has supplementary material online.

KEY WORDS: Nonparametric Bayes; Option pricing; Point-referenced counts; Random probability measure; Random stochastic processes.

---

## 1. INTRODUCTION

Recent literature in Bayesian nonparametrics repeatedly focused on models for collections of distributions based on extensions of the Dirichlet process and other stick-breaking priors (Ferguson 1973; Sethuraman 1994; Ishwaran and James 2001; Ongaro and Cattaneo 2004). These extensions are generated by introducing dependence in the weights and/or atoms of the stick-breaking construction (MacEachern 2000; DeIorio et al. 2004; Gelfand, Kottas, and MacEachern 2005; Teh et al. 2006; Griffin and Steel 2006b; Duan, Guindani, and Gelfand 2007; Dunson and Park 2008; Rodriguez, Dunson, and Gelfand 2008), or by considering mixtures of independent samples from one or more Dirichlet processes (Müller, Quintana, and Rosner 2004; Griffin and Steel 2006a; Dunson, Pillai, and Park 2007). In contrast, this article focuses on models for prior distributions on stochastic processes on an index space  $D \subset \mathbb{R}^d$  with rich but constant marginal distributions. We no longer focus on building different distributions for each possible value of the index space, but instead construct a stochastic process where observations at different locations are dependent, but have a common (albeit unknown) marginal distribution. We call the resulting construction a latent stick-breaking process (LaSBP). As the name indicates, we rely on a stick-breaking construction to represent the unknown marginal distribution, while introducing an underlying (latent) Gaussian process on  $D$  to drive the selection of the atoms at each location  $s \in D$ , as suggested in MacEachern (2007).

As a motivation, consider the problem of estimating species abundance from point counts at pre-specified locations. It is clear that abundance changes spatially with the specific characteristics of micro-habitats, and modeling these changes is key to understanding population intensities and dynamics. However, biologists are interested in not only the spatial distribution of a species, but also its local and global abundance. When studying abundance of European Starlings (a nonnative bird species)

in North Carolina (see Section 7.2), one may be interested in the marginal distribution of the number of individual birds observed at randomly selected locations. In this context, the underlying Gaussian process introduced previously can be interpreted as a latent, location-specific habitat quality score, which changes spatially making higher rates more likely in certain regions. Similarly, we can consider the price of a financial asset to be a random variable whose volatility changes in time according to an underlying score reflecting market sentiments, with the common marginal distribution providing a summary of the return levels that can be expected from the asset (see Section 7.1). Current nonparametric Bayes models for related problems focus on priors for dependent collections of conditional distributions, and therefore, assume that the marginal distribution of the process changes at each location. Instead, we are interested in models that assume a constant marginal distribution across the index space, but allow samples from it to be dependent.

Due to the discrete nature of its marginal distribution, LaSBP models induce a random segmentation of the index space  $D$ . This segmentation is spatially coherent, in the sense that, the closer two locations are in space, the larger the probability of them being assigned to the same cluster. Typically, the segmentation has an interesting interpretation in specific applications. For example, the partitions can be identified with micro-habitats in species distribution applications. Nonetheless, smooth interpolations and predictions can be recovered through model averaging. The literature in signal and image segmentation is enormous, some interesting reviews include Pal and Pal (1993), Pham, Xu, and Prince (2000), and Suri, Wilson, and Laxminarayan (2005). Mixture models are the most relevant for our discussion. In contrast to our method, many of them do not incorporate information on spatial location in the prior for allocating observations to mixture components. One interesting exception is the model developed in Figueiredo (2005) and Figueiredo, Cheng, and Murino (2007) for image segmentation, which uses a latent conditionally autoregressive (CAR) structure to define spatially dependent weights, in a similar spirit to Rodriguez, Dunson, and Taylor (2009).

Since the selection process in the LaSBP inherits the characteristics of the underlying Gaussian process, our models can

---

Abel Rodriguez is Assistant Professor, Department of Applied Mathematics and Statistics, University of California, Mailstop SOE2, Santa Cruz, CA 95064 (E-mail: [abel@soe.ucsc.edu](mailto:abel@soe.ucsc.edu)). David B. Dunson is Professor (E-mail: [dunson@stat.duke.edu](mailto:dunson@stat.duke.edu)) and Alan E. Gelfand is Professor (E-mail: [alan@stat.duke.edu](mailto:alan@stat.duke.edu)), Department of Statistical Sciences, Duke University, Box 90251, Durham, NC 27708. The work of the first author was supported in part by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences and NSF DMS 0915272. The work of the third author was supported in part by NSF DMS 0504953. The authors thank the associate editor and two anonymous referees, whose comments greatly contributed to improve the quality of the manuscript.

easily accommodate a rich variety of correlation structures, including nonstationary processes. This is especially interesting because (1) typical change-point models (like those used in finance) assume a stationary Markovian structure to simplify computation, and (2) nonstationary dependence structures are hard to specify in other nonparametric models like the order-dependent Dirichlet process (Griffin and Steel 2006b). Also, the model does not require replicates at each location, giving it an advantage over Gelfand, Kottas, and MacEachern (2005) for the purpose of our applications.

In addition, the use of order constraints in the definition of the LaSBP allows us to generate skewed mixing distributions. For scale mixtures (as in Section 7.1), this feature yields models that favor heavier tails for the marginal distribution of the observations than those obtained by dependent Dirichlet process (DDP) mixtures with the same baseline measure. For location mixtures (as in Section 7.2), the LaSBP favors distributions that have a heavy right or left tail. This is appealing in many applications, but can be avoided through suitable adjustment of the distributions of the stick-breaking weights if desired.

The LaSBP can also be formulated as a Gaussian copula process with a common and unknown marginal distribution, which is assigned a stick-breaking prior. The literature on copula modeling is extensive; Nelsen (1999), among others, provided an introduction. Within a quasi-Bayesian setting, Hoff (2007) recently developed a multivariate nonparametric model for mixed discrete and continuous data that uses Gaussian copulas and marginal set likelihoods. Although similar in spirit to the LaSBP, this model avoids the specification of the marginal distributions by using a pseudo-likelihood, and is formulated as a procedure to obtain a finite-dimensional distribution and not a process. Also, Kottas, Müller, and Quintana (2005) formulated a model for ordinal data using a nonparametric specification for the copula. Again, the goals of this model are different from ours, as they focus on contingency tables and build a mixture model for the copula describing the correlation across entries in the table.

The LaSBP is related to the hybrid Dirichlet process (hDP) models in Petrone, Guindani, and Gelfand (2009), who also proposed the use of Gaussian copula processes to build dependence across locations. However, our construction is conceptually simpler because it uses scalar atoms rather than atoms that are process realizations. In addition, while we focus on problems where prediction and interpolation are the main goal, the hDP is designed for local surface clustering. Although the hDP is, in principle, a more flexible model (as it allows for more general atoms than the LaSBP), the underlying latent field lacks interpretability and can produce multimodal predictive distributions as an artifact (see Section 3 and Figure 2 in it). This is a direct consequence of the use of process realizations in the definition of the hDP, as an adjacent switch of labels does not imply any closeness for the adjacent atoms.

For some applications (e.g., the stochastic volatility model in Section 7.1), the infinite hidden Markov model (iHMM) (Beal, Ghahramani, and Rasmussen 2001; Fox et al. 2008; van Gael et al. 2008), which is closely connected to the hierarchical Dirichlet process (Teh et al. 2006), is a viable alternative to the LaSBP. However, the iHMM assumes by construction that the index space  $D$  for the stochastic process is discrete and unidimensional, which prevents its use in other interesting applications

such as the one presented in Section 7.2. In addition, the iHMM relies on a Markovian assumption, which only allows a very restrictive local dependence structure in allocation to mixture components, while the LaSBP is much more flexible in using a latent Gaussian process to drive the allocation.

This article is organized as follows: Section 2 reviews the literature on stick-breaking processes and sets the stage for the LaSBP model. Sections 3 and 4 define the latent stick-breaking process and discusses some of its properties, with Section 5 describing an efficient Markov chain Monte Carlo (MCMC) algorithm for posterior computation. Section 6 discusses extensions of the model to multivariate problems. Section 7 presents two applications in financial econometrics and ecology. Finally, in Section 8 we present a short discussion and some future directions.

## 2. STICK-BREAKING PRIORS

Let  $G$  be a random probability measure on  $(\Theta, \mathcal{B})$ , with  $\mathcal{B}$  the Borel sets of  $\Theta$ . Then  $G$  has a stick-breaking prior if

$$G(\cdot) = \sum_{l=1}^L w_l \delta_{\theta_l^*}(\cdot),$$

with  $\theta_l^* \stackrel{\text{iid}}{\sim} H$  for some baseline probability measure  $H$  on  $\Theta$  and  $w_l = v_l \prod_{k=1}^{l-1} (1 - v_k)$  where  $v_l \sim \text{beta}(a_l, b_l)$  for  $l < L$  and  $v_l = 1$  for  $l = L$ . The number of atoms  $L$  can be finite (either known or unknown) or infinite. The class of stick-breaking priors includes some well-known processes. For example, taking  $L = \infty$ ,  $a_l = a$  and  $b_l = b + l(1 - a)$  for  $0 \leq a < 1$  and  $b > -(1 - a)$  yields the two-parameter Poisson-Dirichlet Process (Pitman 1996), with the choice  $a = 1$  resulting in the Dirichlet process with baseline measure  $H$  and precision parameter  $b$  (Sethuraman 1994).

For  $L < \infty$ , setting  $v_L = 1$  ensures that  $\sum_{l=1}^L w_l = 1$  with probability one. If  $L = \infty$ , Ishwaran and James (2001) showed that the weights add up to 1 (and therefore the process yields a well-defined probability distribution) almost surely if and only if  $\sum_{l=1}^{\infty} E(\log(1 - v_k)) = -\infty$ . For any measurable set  $B \in \mathcal{B}$ ,  $G(B)$  is a random variable whose moments are given by

$$E(G(B)) = H(B),$$

$$\text{Var}(G(B)) = H(B)(1 - H(B))$$

$$\begin{aligned} & \times \sum_{l=1}^L \left[ \frac{a_l(1 + a_l)(a_l + b_l)}{(a_l + b_l)^2(a_l + b_l + 1)} \right. \\ & \left. \times \prod_{k=1}^{l-1} \frac{b_k(1 + b_k)(a_k + b_k)}{(a_k + b_k)^2(a_k + b_k + 1)} \right]. \end{aligned}$$

Therefore,  $H$  describes the expected shape of the distributions, while the sequences  $\{a_l\}_{l=1}^L$  and  $\{b_l\}_{l=1}^L$  control the variance of the sampled distributions around  $H$ .

To increase the flexibility of the model, most applications use convolutions of the form  $\int K(\cdot|\theta)G(d\theta)$ , where  $G$  follows a stick-breaking prior and  $K$  is a kernel, to model unknown probability distributions. Under fairly general conditions on the kernel  $K$ , mixtures of this type yield models that are dense (in the  $L^1$  sense) in the space of absolutely continuous distributions (Lo 1984).

One way to extend the stick-breaking priors to *collections* of distributions on a space  $D$  is to replace the random variables involved in the definition of the random distribution with stochastic processes on  $D$ . This is the strategy behind the dependent Dirichlet process (DP) (MacEachern 2000; DeIorio et al. 2004; Gelfand, Kottas, and MacEachern 2005; Griffin and Steel 2006b). Inherently, all of these approaches (including those that allow the weights to depend on the location) assume that, conditional on the stick-breaking parameters, the probability of assigning observations to atoms is known. For example, for the spatial Dirichlet process (SDP) (Gelfand, Kottas, and MacEachern 2005),

$$\Pr(\theta(\mathbf{s}) = \theta_i^*(\mathbf{s}), \theta(\mathbf{s}') = \theta_j^*(\mathbf{s}') | \{\theta_l^*(\mathbf{s})\}_{l=1}^L, \{w_l(\mathbf{s})\}_{l=1}^L) = \begin{cases} 1, & i=j, \\ 0, & i \neq j, \end{cases}$$

for all  $\mathbf{s}, \mathbf{s}' \in D$  and all  $i, j \in \{1, \dots, L\}$ . Therefore, conditional on the atoms and weights, the selection mechanism is constant for all locations  $\mathbf{s} \in S$ . In the sequel, we explore an alternative approach, which constructs a process where the selection probabilities are not conditionally independent across locations while the atoms and weights are.

### 3. UNIVARIATE LATENT STICK-BREAKING PROCESS

Our goal is to construct a stochastic process  $\theta = \{\theta(\mathbf{s}) : \mathbf{s} \in D \subset \mathbb{R}^d\}$  where  $\theta(\mathbf{s}) \in \Theta \subset \mathbb{R}$  for all  $\mathbf{s}$  such that, marginally,  $\theta(\mathbf{s})|G \sim G$  for all  $\mathbf{s} \in D$  and some unknown  $G$ , but  $\text{Cov}(\theta(\mathbf{s}), \theta(\mathbf{s}')|G) \neq 0$ . To do this, consider the general approach motivated by copula models: Define a stochastic process  $U = \{U(\mathbf{s}) : \mathbf{s} \in D\}$  with uniform marginals, then define  $\theta(\mathbf{s}) = G^-(U(\mathbf{s}))$ , where  $G$  is a random discrete probability distribution and  $G^-$  is its *generalized* inverse. Then clearly  $\theta(\mathbf{s})$  is distributed according to  $G$  for any given  $\mathbf{s}$  and  $\theta(\mathbf{s})$  and  $\theta(\mathbf{s}')$  are correlated. The question of model development becomes how to choose the process  $U$  and a prior for  $G$ . In the sequel, we do this through an independent triplet,

$$\{z(\mathbf{s}) : \mathbf{s} \in D\}, \{v_l\}_{l=1}^L, \{\theta_l^*\}_{l=1}^L. \tag{1}$$

Here, the *latent* Gaussian process  $z$  is such that  $z(\mathbf{s}) \sim N(0, 1)$  for all  $\mathbf{s} \in D$  and  $\text{Cor}(z(\mathbf{s}), z(\mathbf{s}')) = \gamma(\mathbf{s}, \mathbf{s}')$ , then  $U(\mathbf{s}) = \Phi(z(\mathbf{s}))$  and the sequences  $\{v_l\}_{l=1}^L$  and  $\{\theta_l^*\}_{l=1}^L$  are used to define a random distribution

$$G(\cdot) = \sum_{l=1}^L w_l \delta_{\theta_l^*}(\cdot),$$

where the probability weights are defined as  $w_l = v_l \prod_{k < l} (1 - v_k)$  and satisfy  $\sum_l w_l = 1$  almost surely. The sequences of *stick breaking ratios*  $\{v_l\}_{l=1}^L$  is such that  $v_l \sim \text{beta}(a_l, b_l)$  for  $l < L$  and  $v_L = 1$ . The sequence of *atoms*  $\{\theta_l^*\}_{l=1}^L$  is constructed by imposing an order constraint on a sample from a baseline measure  $H$  on  $\mathbb{R}$  using the following mechanism: Sample  $\theta_1^* \sim H$  and for  $l > 1$  draw  $\theta_l^* \sim H_l$ , where  $H_l$  is defined as the restriction of  $H$  to the set  $S_l = \{\theta : \theta > \theta_{l-1}^*\}$ , i.e.,  $H_l(B) = H(B \cap S_l) / H(S_l)$  for any measurable set  $B \in \mathcal{B}$ . Referring back to our example in ecology, we can think of  $z(\mathbf{s})$  as the quality of the habitat at location  $\mathbf{s}$ ,  $\{\theta_l^*\}_{l=1}^L$  as the (ordered) set of possible abundance rates that can potentially be observed,

and  $\{v_l\}_{l=1}^L$  as the marginal prior probability of observing the different rates.

Now, for any finite set of locations  $\mathbf{s}_1, \dots, \mathbf{s}_n \in D$ , and any  $n$  let

$$\begin{aligned} \Pr(\theta(\mathbf{s}_1) = \theta_{l_1}^*, \dots, \theta(\mathbf{s}_n) = \theta_{l_n}^* | \{\theta_l^*\}_{l=1}^L, \{v_l\}_{l=1}^L) \\ = \Pr(z(\mathbf{s}_1) \in [\Phi^{-1}(\pi_{l_1-1}), \Phi^{-1}(\pi_{l_1})], \dots, \\ z(\mathbf{s}_n) \in [\Phi^{-1}(\pi_{l_n-1}), \Phi^{-1}(\pi_{l_n})]), \end{aligned} \tag{2}$$

where  $\Phi(\cdot)$  denotes the cumulative standard normal distribution function and  $\pi_l = 1 - \prod_{k \leq l} (1 - v_k)$  is the proportion of the unit stick assigned to the first  $l$  atoms, with  $\pi_0 = 0$ . We call the prior distribution induced by the hierarchical model above on the law  $P$  of the process a latent stick-breaking process (LaSBP), and write  $\theta|P \sim P$  with  $P \sim \text{LaSBP}_L(\{a_l\}_{l=1}^L, \{b_l\}_{l=1}^L, H, \gamma)$ . The definition in Equation (2) makes clear that we are modeling a joint distribution, which encourages closeness among  $\theta(\mathbf{s})$  and  $\theta(\mathbf{s}')$ . This is very different from the effect of DDP models, which encourage closeness among the distributions at  $\mathbf{s}$  and  $\mathbf{s}'$ , but not necessarily between  $\theta(\mathbf{s})$  and  $\theta(\mathbf{s}')$ .

*Theorem 1.* For every  $n > 1$ ,  $P \sim \text{LaSBP}_L(\{a_l\}_{l=1}^L, \{b_l\}_{l=1}^L, H, \gamma)$  satisfies Kolmogorov's conditions almost surely,

(1) symmetry:

$$\begin{aligned} P_{\mathbf{s}_1, \dots, \mathbf{s}_n}(\theta(\mathbf{s}_1) \in A_1, \dots, \theta(\mathbf{s}_n) \in A_n) \\ = P_{\mathbf{s}_{v_1}, \dots, \mathbf{s}_{v_n}}(\theta(\mathbf{s}_{v_1}) \in A_{v_1}, \dots, \theta(\mathbf{s}_{v_n}) \in A_{v_n}) \end{aligned}$$

for any permutation  $\mathbf{v} = (v_1, \dots, v_n)$  of the set  $\{1, \dots, n\}$ ;

(2) consistency:

$$\begin{aligned} P_{\mathbf{s}_1, \dots, \mathbf{s}_{n+1}}(\theta(\mathbf{s}_1) \in A_1, \dots, \theta(\mathbf{s}_n) \in A_n, \theta(\mathbf{s}_{n+1}) \in (-\infty, \infty)) \\ = P_{\mathbf{s}_1, \dots, \mathbf{s}_n}(\theta(\mathbf{s}_1) \in A, \dots, \theta(\mathbf{s}_n) \in A_n), \end{aligned}$$

where  $A_1, \dots, A_n$  are measurable subsets of  $\Theta$ .

The proof is in the [Appendix](#). Note that for any single given location  $P_{\mathbf{s}}(\cdot) = \sum_{l=1}^L w_l \delta_{\theta_l^*}(\cdot) = G(\cdot)$ , and therefore, the marginal distribution of the process is constant. However, draws at two different locations  $\mathbf{s}$  and  $\mathbf{s}'$  are correlated because  $z(\mathbf{s})$  and  $z(\mathbf{s}')$  are correlated. For example, predictions on the value  $\theta(\mathbf{s}')$  for an unobserved location  $\mathbf{s}'$  depend on the predictions for  $z(\mathbf{s}')$ , in particular

$$\begin{aligned} \Pr(\theta(\mathbf{s}') = \theta_l^* | \theta(\mathbf{s}_1), \dots, \theta(\mathbf{s}_n), \{z(\mathbf{s}_i)\}_{i=1}^n, \{\theta_l^*\}_{l=1}^L, \{v_l\}_{l=1}^L) \\ = \Pr(z(\mathbf{s}') \in [\Phi^{-1}(\pi_{l-1}), \Phi^{-1}(\pi_l)] | z(\mathbf{s}_1), \dots, z(\mathbf{s}_n)) \\ = \int_{\Phi^{-1}(\pi_{l-1})}^{\Phi^{-1}(\pi_l)} \phi(t | \Psi'_{\mathbf{s}', \mathbf{s}} \Psi_{\mathbf{s}, \mathbf{s}}^{-1} \mathbf{z}; 1 - \Psi'_{\mathbf{s}', \mathbf{s}} \Psi_{\mathbf{s}, \mathbf{s}}^{-1} \Psi_{\mathbf{s}, \mathbf{s}}) dt, \end{aligned}$$

where  $\phi(\cdot | a; b)$  denotes the density of the univariate Gaussian distribution with mean  $a$  and variance  $b$ ,  $\Psi_{\mathbf{s}', \mathbf{s}}$  is a column vector with entries  $[\Psi_{\mathbf{s}', \mathbf{s}}]_i = \gamma(\mathbf{s}_i, \mathbf{s}')$ ,  $\Psi_{\mathbf{s}, \mathbf{s}}$  is a  $n \times n$  matrix with entries  $[\Psi_{\mathbf{s}, \mathbf{s}}]_{ij} = \gamma(\mathbf{s}_i, \mathbf{s}_j)$ , and  $\mathbf{z} = (z(\mathbf{s}_1), \dots, z(\mathbf{s}_n))'$ .

An alternative but equivalent construction for the LaSBP can be obtained by introducing the labeling process  $\xi = \{\xi(\mathbf{s}) : \mathbf{s} \in$

$D$ ) with  $\xi(\mathbf{s}) \in \{1, \dots, L\}$  such that  $\xi|\{v_l\}_{l=1}^L \sim Q$  and  $Q$  is characterized by the finite-dimensional distributions

$$\begin{aligned} Q_{s_1, \dots, s_n}(l_1, \dots, l_n) &= \Pr(\xi(\mathbf{s}_1) = l_1, \dots, \xi(\mathbf{s}_n) = l_n | \{v_l\}_{l=1}^L) \\ &= \Pr\left(z(\mathbf{s}_1) \in [\Phi^{-1}(\pi_{l_1-1}), \Phi^{-1}(\pi_{l_1})], \dots, \right. \\ &\quad \left. z(\mathbf{s}_n) \in [\Phi^{-1}(\pi_{l_n-1}), \Phi^{-1}(\pi_{l_n})]\right). \end{aligned}$$

For example,  $Q_s(k) = \Pr(\xi(\mathbf{s}) = k | \{v_l\}_{l=1}^L) = w_k$  and

$$\begin{aligned} Q_{s, s'}(l, k) &= \Phi_{\gamma(s, s')}(\Phi^{-1}(\pi_l), \Phi^{-1}(\pi_k)) \\ &\quad - \Phi_{\gamma(s, s')}(\Phi^{-1}(\pi_{l-1}), \Phi^{-1}(\pi_k)) \\ &\quad - \Phi_{\gamma(s, s')}(\Phi^{-1}(\pi_l), \Phi^{-1}(\pi_{k-1})) \\ &\quad + \Phi_{\gamma(s, s')}(\Phi^{-1}(\pi_{l-1}), \Phi^{-1}(\pi_{k-1})), \end{aligned}$$

where  $\Phi_r(\cdot, \cdot)$  denotes the cumulative distribution of the standard bivariate normal with correlation  $r$ . Conditionally on the labeling process we can define  $\theta(\mathbf{s})|\xi(\mathbf{s}), \{\theta_l^*\}_{l=1}^L = \theta_{\xi(\mathbf{s})}^*$ . Integrating out  $\xi$  we get  $\theta|P \sim P$  and  $P \sim \text{LaSBP}(\{a_l\}_{l=1}^L, \{b_l\}_{l=1}^L, H, \gamma)$  as before, however, by explicitly introducing the latent labeling process  $\xi$  we immediately realize that the LaSBP induces a segmentation of the  $\theta$  surface by discretizing a Gaussian process.

To gain some intuition into this construction, we show in Figure 1 three random realizations of the  $\theta$  surface on  $D = [0, 1]^2$

associated with three different marginal distributions  $P$  sharing the same underlying latent process  $z$ , which was generated using an exponential correlation function. For this simulation, we took  $L = 100$  and  $a_l = 1, b_l = \alpha$ , for three different values of  $\alpha$ . The latent process was sampled at 3600 locations over an evenly spaced  $60 \times 60$  grid, and we used a standard Gaussian baseline measure  $H$  to generate the atoms of the process. The resulting surfaces are tiled, segmenting the space according to the level of the underlying Gaussian process. The parameter  $\alpha$  influences the roughness of the segmentation (essentially, the number of different levels that the random surface takes), the baseline measure  $H$  controls the height of the tiles, and the correlation function  $\gamma$  influences the relative size of the different tiles.

Figure 1 also provides insight into the order constraint we introduced in the definition of the process. The order constraint allows us to link the behavior of the atoms of the stick-breaking construction to the behavior of the latent process. In particular, it enforces the internal consistency condition  $z(\mathbf{s}) > z(\mathbf{s}') \Rightarrow \theta(\mathbf{s}) \geq \theta(\mathbf{s}')$ , which simplifies interpretation. In the species abundance application, this means that higher habitat quality scores should imply a larger number of sightings. In a similar spirit, a more nervous market should be subject to larger price changes. Without the order constraint forcing adjacent atoms to have similar values, interpretability is lost and label switching can unexpectedly lead to very wiggly estimates of the underlying latent field  $z$  and multimodal predictive distributions at new locations as an artifact.

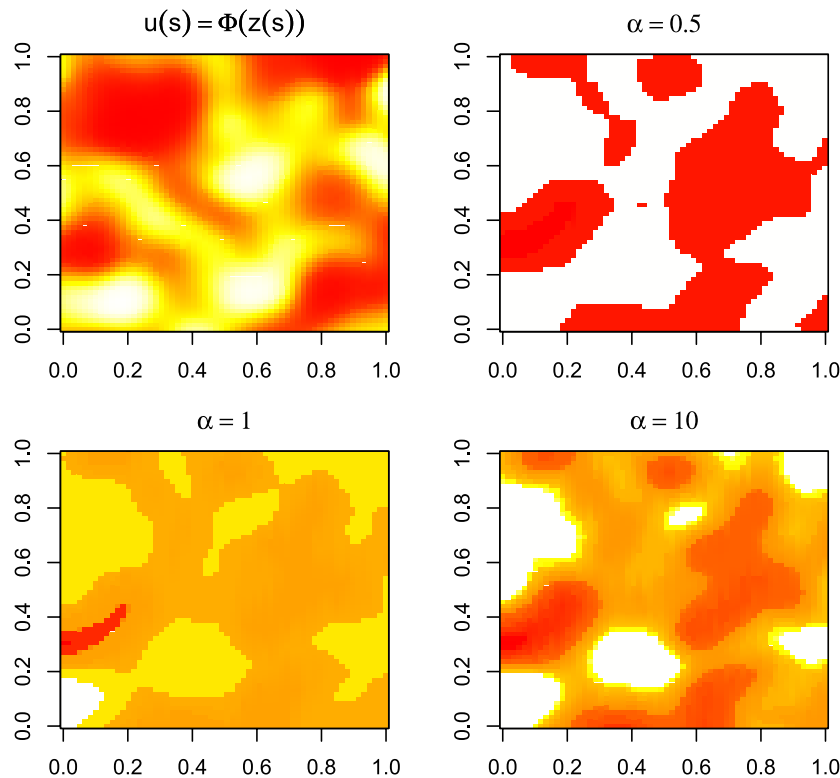


Figure 1. Realizations from a random law  $P$  drawn from a LaSBP process on  $[0, 1]^2$  with a standard Gaussian baseline measure. We illustrate the effect of different concentrations while keeping the underlying Gaussian process (shown in the upper left panel) constant to simplify interpretation. In the color scale, white corresponds to large values, yellow to medium values, and red to low values of the process. A color version of this figure is available in the electronic version of this article.



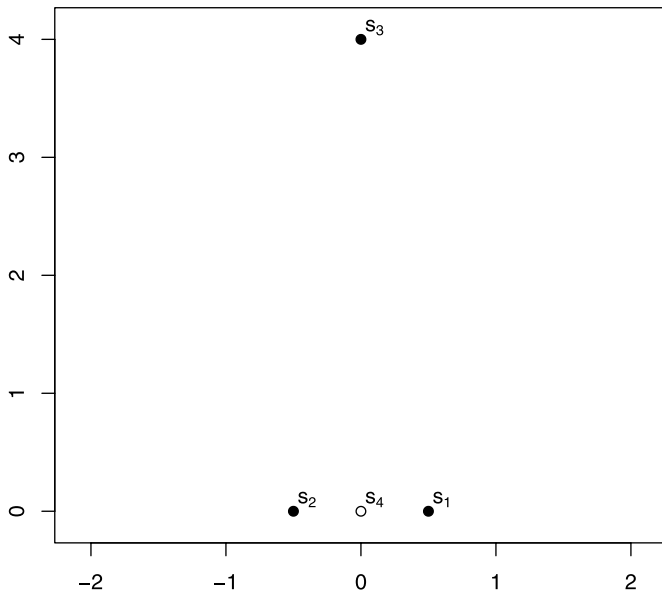


Figure 2. Locations in our toy example. Black points correspond to locations where data are observed, the white point is a location where prediction is desired.

As an illustration of this phenomenon, consider a very simple setup where  $L = 3$  and the process is observed at three locations (labeled  $s_1$ ,  $s_2$ , and  $s_3$  in Figure 2). Let the stick-breaking weights be such that  $\Phi^{-1}(\pi_1) = -1$  and  $\Phi^{-1}(\pi_2) = 1$  and the underlying  $z$  take the values  $z(s_1) = -2$ ,  $z(s_2) = 2$ , and  $z(s_3) = 0$ . This implies that the component membership indicators are  $\xi(s_1) = 1$ ,  $\xi(s_2) = 3$ , and  $\xi(s_3) = 2$ . Also, let the atoms take the value  $\theta_1^* = -5$ ,  $\theta_2^* = 5$ , and  $\theta_3^* = 0$  (this is a valid combination of values if the order constraint is removed).

Consider now the value of the underlying process and the corresponding predictions at location  $s_4$ . Since the Gaussian process is continuous, predictions for  $z(s_4)$  will be close to 0, making  $\xi(s_4) = 2$  (and a prediction close to 5) highly likely, even though  $s_3$  is far away from  $s_1$ ,  $s_2$ , and  $s_4$ . This is undesirable if we are interested in convex interpolations, as in that case we will expect  $\theta(s_4) < 0$ . This feature can be particularly acute with small samples or low spatial coverage, as there may be little information to distinguish between multiple latent surfaces consistent with different label permutations.

In most applications, the LaSBP will not be used to model the observable  $Y(\mathbf{s})$  directly, but will be used to construct LaSBP mixtures. Briefly, we will assume that  $Y(\mathbf{s})|\theta(\mathbf{s}) \sim f(\cdot|\theta(\mathbf{s}))$ , where  $\theta(\mathbf{s})$  is a scalar, and  $f(\cdot|\theta(\mathbf{s}))$  is a density function, e.g.,  $N(\mu, 1/\theta(\mathbf{s}))$  (as in Section 7.1) or  $\text{Poi}(\theta(\mathbf{s}))$  (as in Section 7.2), and let  $\theta|P \sim P$  where  $P \sim \text{LaSBP}_L(\{a_l\}_{l=1}^L, \{b_l\}_{l=1}^L, H, \gamma)$ . This implies that  $Y(\mathbf{s})$  and  $Y(\mathbf{s}')$  are correlated since  $\theta(\mathbf{s})$  and  $\theta(\mathbf{s}')$  are correlated, and that marginally,  $Y(\mathbf{s})|G \sim \sum_{l=1}^L w_l f(\cdot|\theta_l^*)$ , which defines the same, possibly multimodal, marginal for any  $\mathbf{s}$ .

#### 4. PROPERTIES

First, consider the expectation of  $\theta(\mathbf{s})$ . If  $\psi$  is a measurable function on  $(\Theta, \mathcal{B})$ ,

$$\mathbf{E}(\psi(\theta(\mathbf{s}))) = \sum_{l=1}^L \mathbf{E}(w_l) \mathbf{E}(\psi(\theta_l^*))$$

$$\begin{aligned} &= \sum_{l=1}^L \frac{a_l}{a_l + b_l} \left[ \prod_{k<l} \frac{b_k}{a_k + b_k} \right] \mathbf{E}(\psi(\theta_l^*)) \\ &> \mathbf{E}_H(\psi(\theta_1^*)). \end{aligned}$$

Therefore, due to the order constraint, some care must be exercised when interpreting the baseline measure in our model. In particular, the average level of the process will depend not only on the expected value of the baseline measure  $H$ , but also on the stick-breaking parameters  $\{a_l\}_{l=1}^L$  and  $\{b_l\}_{l=1}^L$ . To understand the consequences of this behavior, consider the case  $a_l = 1$  and  $b_l = b$ . In that case, if  $b \rightarrow \infty$  then  $\mathbf{E}(\theta(\mathbf{s})) \rightarrow \infty$ . In turn, for a finite  $b$ , the mean value of the process is strictly greater than the mean of the baseline measure.

*Example 1.* Consider  $\theta|P \sim P$  and  $P \sim \text{LaSBP}_L(a, b, H, \gamma)$ , where  $H$  is an exponential baseline measure with mean  $\lambda$ . This coincides with the illustrations we present in Section 7. In this case,  $\theta_l^*|\theta_{l-1}^*$  follows a shifted exponential distribution and  $\mathbf{E}(\theta_l^*|\theta_{l-1}^*) = \lambda + \theta_{l-1}^*$ . This implies that  $\mathbf{E}(\theta_l^*) = l\lambda$  and

$$\begin{aligned} \mathbf{E}(\theta(\mathbf{s})) &= \sum_{l=1}^L \lambda l \frac{a}{a+b} \left( \frac{b}{a+b} \right)^{l-1} \\ &= \lambda \frac{a+b}{a} \left[ 1 - \left( \frac{b}{a+b} \right)^{L+1} \right. \\ &\quad \left. - (L+1) \left( \frac{a}{a+b} \right) \left( \frac{b}{a+b} \right)^L \right]. \end{aligned}$$

Note that

$$\lim_{L \rightarrow \infty} \mathbf{E}[\theta(\mathbf{s})] = \lambda \frac{a+b}{a}.$$

Therefore, the process is well defined (in the sense that it has a finite first moment) even with an infinite number of atoms as long as  $b$  is finite. A similar argument yields

$$\lim_{L \rightarrow \infty} \mathbf{E}[(\theta(\mathbf{s}))^2] = 2\lambda^2 \left( \frac{a+b}{a} \right)^2$$

and

$$\lim_{L \rightarrow \infty} \text{Var}[\theta(\mathbf{s})] = \lambda^2 \left( \frac{a+b}{a} \right)^2.$$

Example 1 illustrates that in many cases we can interpret finite LaSBP's as truncations of infinite processes. In this situation, the theory of truncations developed in Ishwaran and James (2001, 2003) or Rodríguez, Dunson, and Gelfand (2008) can be directly modified for the LaSBP. The following theorem gives sufficient conditions for the limiting process to have finite expectation.

*Theorem 2.* Let  $\theta|P \sim P$  and  $P \sim \text{LaSBP}_L(a, b, H, \gamma)$ , where  $H$  is such that, for large  $l$ ,

$$\mathbf{E}(\theta_l^*|\theta_{l-1}^*) \leq \mu + \lambda \theta_{l-1}^*$$

for some real constants  $\mu$  and  $\lambda$  such that  $0 < \lambda \leq 1$ . Then

$$\lim_{L \rightarrow \infty} \mathbf{E}(\theta(\mathbf{s})) < \infty.$$

*Corollary 1.* If  $\mathbf{E}(\theta_l^*) < c_1 l^p + c_2 \lambda^l$  for some real constants  $c_1$ ,  $c_2$ , and  $\lambda$  with  $0 < \lambda \leq 1$  and integer constant  $p$ , then  $\lim_{L \rightarrow \infty} \mathbf{E}(\theta(\mathbf{s})) < \infty$ .

*Corollary 2.* If  $H$  has a compact support then  $\lim_{L \rightarrow \infty} \mathbf{E}(\theta(\mathbf{s})) < \infty$ .

The details of the proof are shown in the [Appendix](#), and it can be easily extended to provide conditions for finite higher-order moments. As the following two examples show, the condition in Theorem 2 is generally satisfied when  $H$  corresponds to a distribution with light tails.

*Example 2.* Let  $\theta|P \sim P$  and  $P \sim \text{LaSBP}_L(a, b, H, \gamma)$  where  $H$  is a Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$ . In this case

$$\mathbf{E}(\theta_l^*|\theta_{l-1}^*) = \mu + \sigma \frac{\phi((\theta_{l-1}^* - \mu)/\sigma)}{1 - \Phi((\theta_{l-1}^* - \mu)/\sigma)},$$

where  $\phi$  denotes the density of the standard normal distribution. Using some well-known properties ([Abramowitz and Stegun 1965](#), p. 298),

$$\begin{aligned} & \sigma \frac{\phi((\theta_{l-1}^* - \mu)/\sigma)}{1 - \Phi((\theta_{l-1}^* - \mu)/\sigma)} \\ & \leq \frac{1}{\sqrt{2\pi}} [(\theta_{l-1}^* - \mu) + \sqrt{(\theta_{l-1}^* - \mu)^2 + 2\sigma^2}] \\ & \approx \frac{2}{\sqrt{2\pi}} (\theta_{l-1}^* - \mu) \end{aligned}$$

for large  $\theta_{l-1}^*$ . Since  $2/\sqrt{2\pi} < 1$ , Theorem 1 applies and  $\lim_{L \rightarrow \infty} \mathbf{E}(\theta(\mathbf{s})) < \infty$ .

*Example 3.* Consider now a Pareto baseline measure with density

$$p(\theta) = \frac{\alpha}{\eta} \left(\frac{\eta}{\theta}\right)^{\alpha+1}, \quad \theta \geq \eta.$$

The conditional density reduces to

$$p(\theta_l^*|\theta_{l-1}^*) = \frac{\alpha}{\theta_{l-1}^*} \left(\frac{\theta_{l-1}^*}{\theta_l^*}\right)^{\alpha+1}, \quad \theta_l^* \geq \theta_{l-1}^*,$$

and  $\mathbf{E}(\theta_l^*|\theta_{l-1}^*) = \frac{\alpha}{\alpha-1} \theta_{l-1}^*$ . Therefore  $\lim_{L \rightarrow \infty} \mathbf{E}(\theta(\mathbf{s})) < \infty$  iff  $b/(a+b) < (\alpha-1)/\alpha$ .

In addition to the mean of the process, it is important to understand the covariance structure induced by our construction. The covariance of the  $\theta$  process is given by

$$\begin{aligned} \text{Cov}(\theta(\mathbf{s}), \theta(\mathbf{s}')) &= \sum_{l=1}^L \sum_{k=1}^L [\mathbf{E}(Q_{s,s'}(l, k))\mathbf{E}(\theta_l^*\theta_k^*) \\ &\quad - \mathbf{E}(Q_s(l))\mathbf{E}(Q_{s'}(k))\mathbf{E}(\theta_l^*)\mathbf{E}(\theta_k^*)]. \end{aligned} \quad (3)$$

Although we do not have a closed-form expression for the covariance function, it is clear from Equation (3) that the process on  $\theta$  is isotropic, stationary or nonstationary if  $z$  has these properties. Additional properties for the selection process  $U$  can be seen in the [Appendix](#).

Realizations from  $P$  are *discontinuous* almost surely, which is a consequence of the stick-breaking construction. However, note that

$$\begin{aligned} Q_{s,s'}(l, k) &\rightarrow Q_s(l)Q_{s'}(k) \quad \text{as } \gamma(\mathbf{s}, \mathbf{s}') \rightarrow 0, \\ Q_{s,s'}(l, k) &\rightarrow \begin{cases} Q_s(l), & l = k, \\ 0, & \text{otherwise} \end{cases} \quad \text{as } \gamma(\mathbf{s}, \mathbf{s}') \rightarrow 1. \end{aligned}$$

In particular, this implies that  $\Pr(\theta(\mathbf{s}) = \theta(\mathbf{s}')) \rightarrow 1$  as  $\mathbf{s}' \rightarrow \mathbf{s}$  almost surely, and the selection mechanism is spatially coherent. This result also shows that, as the range of the latent Gaussian process goes to zero, samples at different locations become iid observations from  $G$ . On the other hand, as the range grows, the random surfaces drawn from  $P$  are constant and the corresponding LaSBP mixture reduces to a parametric model with a prior  $H$  on the unknown parameter.

It is important to note that  $\theta(\mathbf{s})$  and  $\theta(\mathbf{s}')$  are never independent. Indeed, even if  $\gamma(\mathbf{s}, \mathbf{s}') = 0$ , implying that  $\Pr(\xi(\mathbf{s}) = l, \xi(\mathbf{s}') = k|\{v_r\}_{r=1}^L) = \Pr(\xi(\mathbf{s}) = l|\{v_r\}_{r=1}^L)\Pr(\xi(\mathbf{s}') = k|\{v_r\}_{r=1}^L) = Q_s(l)Q_{s'}(k)$ , it is clear from Equation (3) that, in general,  $\text{Cov}(\theta(\mathbf{s}), \theta(\mathbf{s}')) \neq 0$  because  $\text{Cov}(\mathbf{E}(\theta(\mathbf{s})|G), \mathbf{E}(\theta(\mathbf{s}')|G)) \neq 0$ . It is also interesting to formally consider the effect of the stick-breaking distribution on the surface. Assume for simplicity that  $a_l = a$  and  $b_l = b$  for all  $l$ , which we will do in most practical applications. Note that as  $\frac{a}{a+b} \rightarrow 1$  we have  $w_l \rightarrow 1$  and realizations from  $P$  will correspond to flat random surfaces,  $\theta(\mathbf{s}) = \theta_1^*$  for all  $s \in D$ .

Finally, we also note that, due to the order constraint on the atoms, the LaSBP with constant  $a$  and  $b$  tends to generate skewed marginal distributions as the process will also encourage larger weights for smaller values of  $\theta$ . This provides our model with a particular flavor that differentiates it from the regular Dirichlet process and the Dirichlet-multinomial model underlying the hDP in [Petrone, Guindani, and Gelfand \(2009\)](#). More precisely, the weights arising from a stick-breaking construction are stochastically ordered. Under the order constraint for the  $\theta_l^*$ 's, this implies that smaller  $\theta_l^*$ 's receive stochastically larger weights. For scale LaSBP mixtures (as in Sections 7.1 and 7.2), this feature yields models that favor heavier tails for the marginal distribution of the observations than those obtained by dependent Dirichlet process (DDP) mixtures. For location mixtures, the LaSBP favors distributions that have a heavy right or left tail, which in our experience, can be appealing property in many applications. Indeed, there is often prior information available that one tail of the distribution is heavier and is more subject to secondary modes, and our specification allows a natural incorporation of such prior information. In the applications we considered, this was the case. To provide a few additional examples: (1) In studies of premature delivery, the natural response is the gestational age at delivery. The right tail tends to be very stable across populations and is very light, while the left tail (corresponding to premature deliveries) is quite heavy and varies substantially across ethnic groups and with other factors; (2) in studies of the level of a pollutant, the left tail (corresponding to unusually low pollution levels) tends to be very light relative to the right tail (corresponding to high pollution levels); and (3) similar behavior is observed typically for continuous adverse health responses, such as the frequency of DNA strand breaks in cells. In applications where this property not appealing, it can be mitigated through appropriate choices of the nonconstant sequences  $\{a_l\}_{l=1}^L$  and  $\{b_l\}_{l=1}^L$  controlling the stick breaking weights.

### 5. INFERENCE FOR MIXTURES OF LASBP

For inference purposes, we develop an MCMC algorithm. We assume  $a_l = a$ ,  $b_l = b$ ,  $L < \infty$  and use a blocked sampler ([Ishwaran and James 2001](#)). If  $L = \infty$ , a retrospective sampler

(Roberts and Papaspiliopoulos 2008) can be easily obtained as an extension. To facilitate computation, the latent processes,  $z$  and  $\xi$  are sampled explicitly. Assuming that  $n$  data points  $y(\mathbf{s}_1), \dots, y(\mathbf{s}_n)$  were collected, one for each location  $\mathbf{s}_1, \dots, \mathbf{s}_n$ , and using the familiar bracket notation, the joint distribution is given in this case by

$$\prod_{i=1}^n [y(\mathbf{s}_i)|\theta^*, \xi(\mathbf{s}_i)] \times \prod_{i=1}^n [\xi(\mathbf{s}_i)|z(\mathbf{s}_i), \mathbf{v}] \times [z(\mathbf{s}_1), \dots, z(\mathbf{s}_n)|\lambda] \times \prod_{l=1}^{\infty} [v_l|\alpha] \times \prod_{l=1}^{\infty} [\theta_l] \times [\alpha] \times [\lambda].$$

After setting up initial values for all parameters in the model, the algorithm proceeds by sequentially updating the parameters according to the following steps:

(1) Jointly update the latent processes  $\xi$  and  $z$ , one location at a time, by first sampling  $\xi(\mathbf{s}_i)$  from a discrete distribution such that

$$\Pr(\xi(\mathbf{s}_i) = l) \propto p(y(\mathbf{s}_i)|\theta_l) \Pr(z(\mathbf{s}_i) \in [\Phi^{-1}(\pi_{l-1}), \Phi^{-1}(\pi_l)] | z(\mathbf{s}_1), \dots, z(\mathbf{s}_{i-1}), z(\mathbf{s}_{i+1}), \dots, z(\mathbf{s}_n)),$$

where the prior probability of component  $l$  can be obtained by a univariate integration, and then sampling  $z(\mathbf{s}_i)$  from the restricted univariate normal distribution defined by

$$[z(\mathbf{s}_i)|z(\mathbf{s}_1), \dots, z(\mathbf{s}_{i-1}), z(\mathbf{s}_{i+1}), \dots, z(\mathbf{s}_n)] \mathbf{1}_{\Omega_l},$$

where  $\Omega_l = \{z(\mathbf{s}_i) : z(\mathbf{s}_i) \in [\Phi^{-1}(\pi_{\xi(\mathbf{s}_i)-1}), \Phi^{-1}(\pi_{\xi(\mathbf{s}_i)})]\}$ . Note that, if  $\mathbf{s}_i = \mathbf{s}_j$  for some  $j$ , then the prior probability for observation  $i$  being assigned to component  $\xi(\mathbf{s}_j)$  is one, and therefore,  $\xi(\mathbf{s}_i) = \xi(\mathbf{s}_j)$  and  $z(\mathbf{s}_i) = z(\mathbf{s}_j)$ , as expected.

(2) Sample the stick-breaking ratios one at a time from their full conditional

$$[v_l|\mathbf{v}_{(-l)}] \propto v_l^{a-1} (1 - v_l)^{b-1} \mathbf{1}_{A_l},$$

where  $A_l = \{v_l : q_l^l \leq v_l < q_l^u\}$  and

$$q_l^l = \max_{\{i:\xi(\mathbf{s}_i) \geq l\}} \left\{ 1 - \frac{1 - \Phi(z(\mathbf{s}_i))}{\prod_{k \leq \xi(\mathbf{s}_i), k \neq l} (1 - v_k)} \right\},$$

$$q_l^u = \min_{\{i:\xi(\mathbf{s}_i) \geq l+1\}} \left\{ 1 - \frac{1 - \Phi(z(\mathbf{s}_i))}{\prod_{k \leq \xi(\mathbf{s}_i)-1, k \neq l} (1 - v_k)} \right\}.$$

Note that  $v_l$  depends on  $\mathbf{v}_{(-l)}$  only through the constraint on the support. Also,  $v_l$  depends on  $z$  only through  $\{z(\mathbf{s}_i) : \xi(\mathbf{s}_i) \geq l\}$ . Therefore, for  $l > l^* = \max_i \{\xi(\mathbf{s}_i)\}$ ,  $v_l$  is conditionally independent from  $\mathbf{v}_{(-l)}$ .

(3) Sample the atoms  $\{\theta_l\}$ . The set of full conditionals is given by

$$p(\theta_1 | \dots) \propto h(\theta_1) \prod_{\{i:\xi(\mathbf{s}_i)=1\}} p(\mathbf{y}(\mathbf{s}_i)|\theta_1)$$

and

$$p(\theta_l | \dots) \propto h(\theta_l|\theta_{l-1}) \prod_{\{i:\xi(\mathbf{s}_i)=l\}} p(\mathbf{y}(\mathbf{s}_i)|\theta_l), \quad l > 1,$$

where  $h_l$  is the density associated with the measure  $H_l$ .

(4) The prior parameters on the stick-breaking ratios  $a$  and  $b$  can be jointly sampled using a random-walk Metropolis–Hastings algorithm.

(5) The parameters of the underlying Gaussian process can be sampled conditional on the current realization  $z(\mathbf{s}_1), \dots, z(\mathbf{s}_n)$  using a random-walk Metropolis–Hastings algorithm, as is customary in Gaussian process models.

(6) Since the conditional moves in step (1) can raise some concerns about mixing rates, we suggest to additionally sample the whole Gaussian process jointly. This can be easily done by noting that the joint full conditional is given by

$$[z(\mathbf{s}_1), \dots, z(\mathbf{s}_n) | \dots] \dots \sim \mathbf{N}(\mathbf{0}, \Psi) \mathbf{1}_{\Omega},$$

where  $[\Psi]_{ij} = \gamma(\mathbf{s}_i, \mathbf{s}_j)$ ,  $\Omega = \bigcap_{i=1}^n \Omega_i$  and  $\Omega_i = \{z(\mathbf{s}_i) : \Phi^{-1}(\pi_{\xi(\mathbf{s}_i)-1}) \leq z(\mathbf{s}_i) < \Phi^{-1}(\pi_{\xi(\mathbf{s}_i)})\}$  as before. Due to the form of the restrictions, this is a relatively straightforward step.

The supplemental material available at <http://pubs.amstat.org/loi/jasa> discusses some details related to efficient computational implementation for this model.

## 6. MULTIVARIATE LATENT STICK-BREAKING PROCESSES

The order constraint in the definition of the LaSBP makes it hard to directly extend our formulation to multivariate atoms. One alternative we explore here is to use a latent multivariate Gaussian process to induce dependence in the selection mechanism of multiple stick-breaking representations. As an example, consider a two-dimensional process  $\theta = \{(\theta_1(\mathbf{s}), \theta_2(\mathbf{s}))' : \mathbf{s} \in D\}$  with marginal distributions  $G_1$  and  $G_2$ , which have stick-breaking representations  $G_i(\cdot) = \sum_{l=1}^L w_{il} \delta_{\theta_{il}^*}(\cdot)$  with  $\theta_{il}^* \leq \theta_{i,l+1}^*$  for  $i = 1, 2$  as defined in Section 3 for the univariate LaSBP, with components independent for  $i = 1, 2$ . Also, let  $(z_1(\mathbf{s}), z_2(\mathbf{s}))' = \mathbf{z}(\mathbf{s}) = \mathbf{A}\mathbf{x}(\mathbf{s})$ , where  $x_i$  is a standardized Gaussian process with correlation function  $\gamma_i(\mathbf{s}, \mathbf{s}')$  and  $\mathbf{A}$  is a matrix such that

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ \eta & \sqrt{1 - \eta^2} \end{pmatrix} \Rightarrow \mathbf{A}'\mathbf{A} = \begin{pmatrix} 1 & \eta \\ \eta & 1 \end{pmatrix}.$$

Letting  $\mathbf{a}_j$  be the  $j$ th column of  $\mathbf{A}$ , the corresponding cross-covariance matrix for  $\mathbf{z}$  is given by  $\Sigma_{\mathbf{z}(\mathbf{s}), \mathbf{z}(\mathbf{s}')} = \gamma_1(\mathbf{s}, \mathbf{s}') \mathbf{a}_1 \mathbf{a}_1' + \gamma_1(\mathbf{s}, \mathbf{s}') \mathbf{a}_2 \mathbf{a}_2'$ . In particular, this implies that  $\text{Var}(z_i(\mathbf{s})) = 1$  for  $i = 1, 2$  and  $\text{Cov}(z_1(\mathbf{s}), z_2(\mathbf{s})) = \text{Cor}(z_1(\mathbf{s}), z_2(\mathbf{s})) = \eta$ . As in Section 3, we can define finite-dimensional distributions

$$P_{\mathbf{s}_1, \dots, \mathbf{s}_n}(\theta_1(\mathbf{s}_1) = \theta_{1l_1}^*, \dots, \theta_1(\mathbf{s}_n) = \theta_{1l_n}^*, \theta_2(\mathbf{s}_1) = \theta_{2l_{n+1}}^*, \dots, \theta_2(\mathbf{s}_n) = \theta_{2l_{2n}}^* | \{\theta_{1l}, \theta_{2l}\}_{l=1}^L, \{v_{1l}, v_{2l}\}_{l=1}^L) = \Pr(z_1(\mathbf{s}_1) \in [\Phi^{-1}(\pi_{1,l_1-1}), \Phi^{-1}(\pi_{1,l_1})], \dots, z_1(\mathbf{s}_n) \in [\Phi^{-1}(\pi_{1,l_n-1}), \Phi^{-1}(\pi_{1,l_n})], z_2(\mathbf{s}_1) \in [\Phi^{-1}(\pi_{2,l_{n+1}-1}), \Phi^{-1}(\pi_{2,l_{n+1}})], \dots, z_2(\mathbf{s}_n) \in [\Phi^{-1}(\pi_{2,l_{2n}-1}), \Phi^{-1}(\pi_{2,l_{2n}})]).$$

Choosing  $\eta = 0$  implies that  $\theta_1(\mathbf{s})$  is independent from  $\theta_2(\mathbf{s})$ . More generally, we can calculate the cross-covariance function

for this process by noting that

$$\text{Cov}(\theta_i(\mathbf{s}), \theta_j(\mathbf{s}')) = \sum_{l=1}^L \sum_{k=1}^L \mathbb{E}[Q_{s,s'}^{ij}(l, k)] \mathbb{E}(\theta_{il}^*) \mathbb{E}(\theta_{jk}^*) - \mathbb{E}[Q_s^i(l)] \mathbb{E}[Q_s^j(k)] \mathbb{E}(\theta_{il}^*) \mathbb{E}(\theta_{jk}^*), \quad (4)$$

where as in Section 4,  $Q_{s,s'}^{ij}(l, k) = \Pr(\Phi(z_i(\mathbf{s})) \in [\pi_{i,l-1}, \pi_{i,l}], \Phi(z_j(\mathbf{s}')) \in [\pi_{j,k-1}, \pi_{j,k}] | \{v_{1l}, v_{2l}\}_{l=1}^L)$  and  $Q_s^i(l) = \Pr(\Phi(z_i(\mathbf{s})) \in [\pi_{i,l-1}, \pi_{i,l}] | \{v_{1l}\}_{l=1}^L) = w_{il}$ . As expected, Equation (4) reduces to Equation (3) when  $i = j$ .

The use of dependent Gaussian processes typically provides an interesting interpretation for the model. For example, in the context of the stochastic volatility model described in Section 7.1, the dependence between latent Gaussian processes measures the correlation between market sentiments across different assets. Additionally, sampling for this multivariate extension can proceed along the lines described in Section 5, with an additional step to sample the parameters in  $\mathbf{A}$  conditionally on all other parameters.

The multivariate extension we describe in this section is really just a straightforward modification of the univariate case in that the order restrictions are placed on the atoms in each marginal in the same manner as in the univariate case. Hence, the assumptions are not more stringent than in the univariate case, so our previous comments comparing the LaSBP with the hDP in the univariate case still apply. As we discussed in Section 4, it is very often the case that such order restrictions are justified. This is also true in the multivariate case. For example, instead of one pollutant, we may measure multiple pollutants, all of which should have a heavier right tail in their concentration distribution a priori. A posteriori these assumptions are not strictly enforced; if there is sufficient evidence in the data to refute our a priori (informed) guess of the shape of the distribution, the posterior estimate will be appropriately adapted.

## 7. ILLUSTRATIONS

### 7.1 Modeling Stochastic Volatility

In this section we consider an application of the LaSBP to the modeling of stock-market returns. This example is interesting as it allows us to compare the LaSBP with other more standard models in a real life example.

The dataset under consideration consists of the weekly returns of the S&P500 index covering the 10-year period between April 21, 1997 and April 9, 2007, for a total of 520 observations. Figure 3 shows the evolution of these returns. The series does not exhibit any long-term trend, but different levels of volatility can be clearly seen. In particular, two slightly different regimes are apparent; before May 2003, periods of high volatility are relatively frequent, while after May 2003, we can appreciate longer low-volatility periods. Indeed, it is well known that financial time series typically exhibit heavy tails and that periods of low/high volatility tend to cluster together.

Similarly to other approaches in the literature (Bollerslev 1986; Jacquier, Polson, and Rossi 1994), we model  $r(t)$ , the log-return of the S&P500 at time  $t$ , as conditionally following a normal distribution, i.e.,  $r(t) | \mu, \theta(t) \sim \mathbf{N}(\mu, \theta(t)^{-1})$ , where  $\theta | P \sim P$  and the law  $P$  follows a LaSBP with an exponential baseline measure with mean  $1/\phi_0$  and a decreasing (rather than

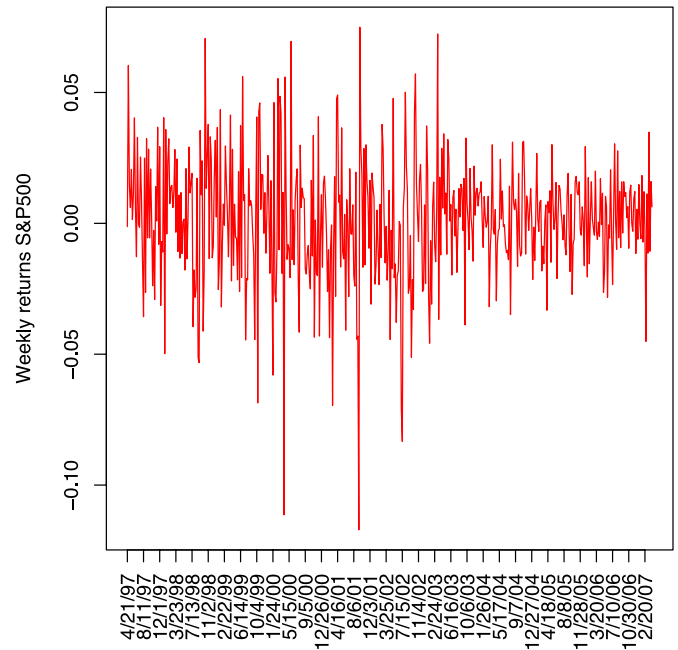


Figure 3. Weekly returns on the S&P500 index between April 21, 1997 and April 9, 2007. A color version of this figure is available in the electronic version of this article.

increasing) order constraint. Also, we take  $a_l = 1$  and  $b_l = \alpha$  for all  $l$ . We noted before that this choice places stochastically larger probabilities on larger variances. A prior with this structure is well motivated in this application, because low volatility (high precision) is the norm, while high volatility (low precision) is associated with sporadic shocks to the system. To demonstrate the methodology we use an exponential correlation function with range parameter  $\lambda$  for the latent process, i.e.,  $\gamma(t, t') = \exp\{-\frac{|t-t'|}{\lambda}\}$ ; however, implementations using more general correlation functions (like Matérn, power exponential or some nonstationary family) is just as straightforward. This model results in a marginal distribution for the returns that is a scale mixture of Gaussian distributions, yielding a rich model on the class of unimodal distributions that can accommodate heavy tails. Also, the structure of the LaSBP naturally induces volatility clustering, while potentially allowing for richer dependence structures than Markov-switching models.

We complete the model by specifying priors for the different parameters. The mean of the returns is given a normal distribution with mean  $\mu_0$  and variance  $\tau^2$ , the spatial range is given a  $\text{Gam}(a_\lambda, b_\lambda)$  distribution, the precision  $\alpha$  is assigned a  $\text{Gam}(a_\alpha, b_\alpha)$  prior and  $\phi_0$  is given a  $\text{Gam}(\delta, \phi_{00})$  prior. Since the historical annual volatility of the S&P500 is traditionally estimated to be in the range of 12%–15%, we take  $\tau^2 = 0.15^2/52$  and  $\mathbb{E}(1/\phi_0) = 0.12^2/52$ . The prior mean for the average return is taken to be  $\mu_0 = 0$ , and the degrees of freedom are set to  $\delta = 2$ . We picked  $a_\alpha = b_\alpha = 1$  so that the prior for the precision parameter  $\alpha$  is centered around 1 and has variance 1. The prior for the temporal correlation is set as  $a_\lambda = 1$  and  $b_\lambda = 1/5$ , implying that the a priori range of the latent process driving the correlation is approximately 15 weeks. Forty atoms ( $L = 40$ ) were used in the stick-breaking construction. A small sensitivity analysis was conducted by changing the hyperpriors on  $\sigma_0^2$ ,



$\alpha$ , and  $\lambda$ , as well as  $L$ , and results seemed mostly robust to the selection. All results are based on 50,000 iterations obtained after a burn-in period of 5000 samples. We resampled the latent process every 10 iterations to improve mixing. Visual inspection of the trace plots does not reveal any obvious mixing or convergence problems. Sensitivity analysis was performed by considering values for  $E(1/\phi_0)$  in the range between  $0.05^2/52$  and  $0.4^2/52$ , leading to essentially the same results.

The model identifies between 2 and 11 volatility regimes, with models having between 3 and 6 components receiving the most weight (estimated posterior probabilities 0.27, 0.33, 0.20, and 0.10, respectively). The mean of the precision  $\alpha$  is estimated to be 0.700 [median 0.650, symmetric 90% credible interval (0.260; 1.312)]. The posterior distribution of the correlation parameter  $\lambda$  has a mean of 110.72 [90% credible interval (34.05; 241.53)], showing strong evidence of long-range dependence in the sample. This feature, which is well known in financial time series, is hard to capture with stochastic volatility models based on low-order autoregressive processes.

The posterior mean weekly return is estimated to be 0.00194 [credible interval (0.00050; 0.00336)], a reasonable value given historical evidence. Figure 4 shows the posterior mean for the volatility path estimated by the LaSBP model, along with the sample standard deviation and the volatility path obtained from the “standard” stochastic volatility described in Jacquier, Polson, and Rossi (1994). Note that, although under the LaSBP each realization of the volatility path is piecewise constant, by taking an average over multiple realizations of the posterior distribution we obtain a smoother path. Also, although both models provide similar volatility estimates, there are some interesting differences. For example, the low volatility period between October 2003 and December 2006 is estimated by the

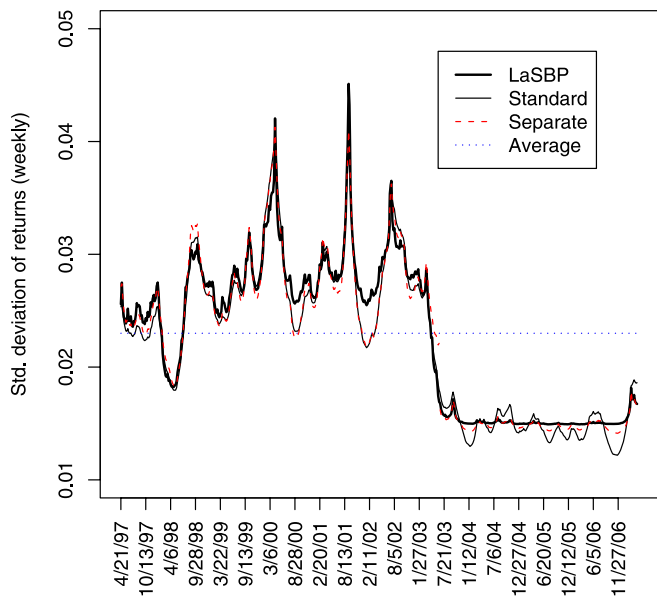


Figure 4. Smoothed weekly volatilities for the S&P500 using the LaSBP stochastic volatility model, a standard stochastic volatility model fitted to the whole series, and separate standard stochastic volatility models for the periods before and after 06/16/2003. The flat dotted line corresponds to the empirical standard deviation of the sample. A color version of this figure is available in the electronic version of this article.

LaSBP analysis to be a period with essentially constant volatility, while the regular volatility model implies frequent fluctuations. Nonetheless, both models rapidly adapt to the rise in volatility in early 2007. Also the three high volatility peaks in 2000, 2001, and 2002 are more pronounced under our model.

To better understand the behavior of these models, we also fitted separate standard stochastic volatility models for the periods before and after 06/16/2003 (labeled “separate” in Figure 4). Note that the separate stochastic volatility models fit a smoother volatility estimate than the standard model to the second period. This suggests that lack of smoothness in the second period for the standard stochastic volatility model is induced by the lack of flexibility in the model.

Unlike standard volatility models, the LaSBP automatically produces a nonparametric estimate of the marginal distribution of returns on S&P500, which is shown in Figure 5. As expected from historical evidence, the tails of the estimated marginal distribution are heavier than the tails of a normal distribution with equivalent mean and variance.

### 7.2 Understanding Species Abundance: A Model for Point-Referenced Counts

This section concentrates on models for spatial count data, in the spirit of Richardson and Green (2001). The Christmas Bird Count (CBC) is an annual census of early-winter bird populations conducted by over 50,000 observers each year between December 14th and January 5th. The primary objective of the CBC is “to monitor the status and distribution of bird populations across the Western Hemisphere.” Parties of volunteers follow specified routes through a designated 15-mile diameter circle, counting every bird they see or hear. The parties are organized by compilers who are also responsible for reporting total counts to the organization that sponsors the activity, the

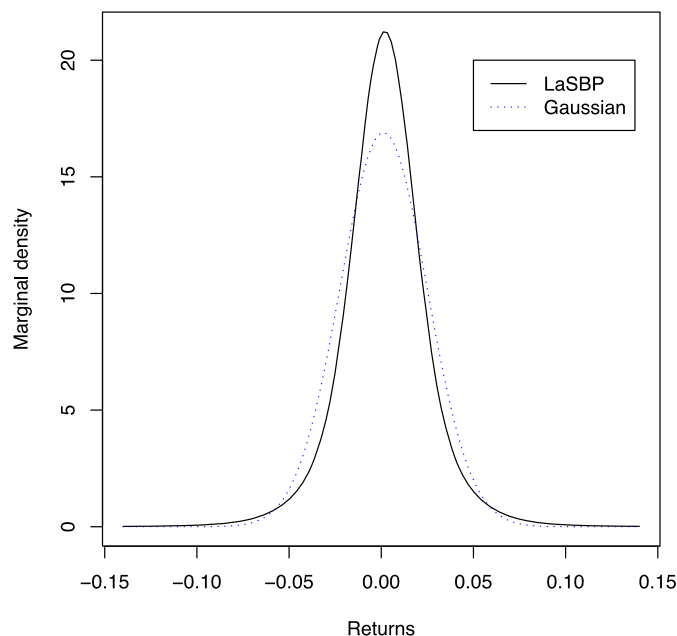


Figure 5. Marginal distribution of returns from the S&P500, compared with an empirical fit to the Gaussian distribution. A color version of this figure is available in the electronic version of this article.

Audubon Society. Data and additional details about this survey are available at <http://www.audubon.org/bird/cbc/index.html>.

In this illustration, we concentrate on jointly modeling the abundance of two bird species in North Carolina during the 2006–2007 winter season: *Sturnus vulgaris*, also known as European Starling, and *Zenaida macroura*, also known as Mourning Dove. In this period, forty-four circles were active within the state. Since the diameter of the circles is very small compared to the size of the region under study, we treat the data as point referenced to the center of the circle. We elaborate on the model described in Section 6 to construct a bivariate LaSBP mixture of Poisson kernels.

Specifically, letting  $y_i(\mathbf{s})$  stand for the number of birds observed at location  $\mathbf{s}$  (expressed in latitude and longitude) belonging to species  $i$ , we assume  $y_i(\mathbf{s}) \sim \text{Poi}(h(\mathbf{s})\theta_i(\mathbf{s}))$ , where  $h(\mathbf{s})$  represents the number of man-hours invested at location  $\mathbf{s}$ . Next, we assume that  $(\theta_1, \theta_2) | P$  and  $P$  follows a bivariate LaSBP with exponential baseline measures  $H_{10}$  and  $H_{20}$  with means  $\theta_{10}$  and  $\theta_{20}$ , stick-breaking weights  $w_{il} = v_{il} \prod_{k < l} (1 - v_{ik})$  with  $v_{il} \sim \text{beta}(1, \alpha_i)$ , and exponential correlation functions  $\gamma_1(\cdot, \cdot)$  and  $\gamma_2(\cdot, \cdot)$  with range  $\rho_1$  and  $\rho_2$ , respectively.

We allow for a hierarchical structure on the mean of the baseline distributions  $H_{10}$  and  $H_{20}$ , whose inverse is assumed to follow a gamma distribution,  $1/\theta_{i0} \sim \text{Gam}(2, 2\theta_{i00})$ . Based on historical data, we set  $\theta_{100} = 10$  and  $\theta_{200} = 4$ . The correlation  $\eta$  and range parameters  $\rho_1$  and  $\rho_2$  defining the latent Gaussian processes are assumed unknown, and priors  $\eta \sim \text{Uni}[-1, 1]$ ,  $\rho_1 \sim \text{Gam}(1, 1/2)$  and  $\rho_2 \sim \text{Gam}(1, 1/2)$  are, respectively, assigned. Precision parameters are assigned independent priors  $\alpha_i \sim \text{Gam}(1, 1)$ . Results are based on 100,000 iterations obtained after a burn-in period of 20,000 samples, and the latent process was resampled every 10 iterations. Visual inspection of trace plots reveals no obvious convergence problems, and autocorrelations seem within reasonable levels.

Figure 6 shows the posterior mean surfaces corresponding to the per-man-hour sighting rates for each of the two species. As expected, rates for both species tend to be higher in the coastal plains region; the European Starling and the Mourning Dove tend to be associated with man-altered environments, preferring open country and grasslands, and avoiding woodlands, deserts, and swamps. The latent processes appear to be positively correlated, the posterior expectation for  $\eta$  is 0.26 and (0.12; 0.44) is a symmetric, 90% credible interval. However, the difference in the level of spatial dependence between species is striking; while suitable habitats for the Mourning Dove seem to cover large geographical regions, European Starlings seem to concentrate on small microhabitats.

Figure 7 shows estimates of the marginal distribution of the number of sightings in 1 man-hour for both species under consideration. Both distributions are bimodal and heavy tailed. The smaller mode, which can be interpreted as the baseline rate of sightings, falls around four or six birds per man per hour for the European Starling and around two to four birds for the Mourning Dove. The higher mode, which can be interpreted as corresponding to specially suited habitats, is around 45 sightings per man-hour for the European Starling and around 23 sightings for the Mourning Dove.

We wish like to emphasize that parametric alternatives to our proposed model like a Poisson log-linear model do not allow for a multimodal marginal distribution like the one shown previously. On the other hand, nonparametric alternatives like

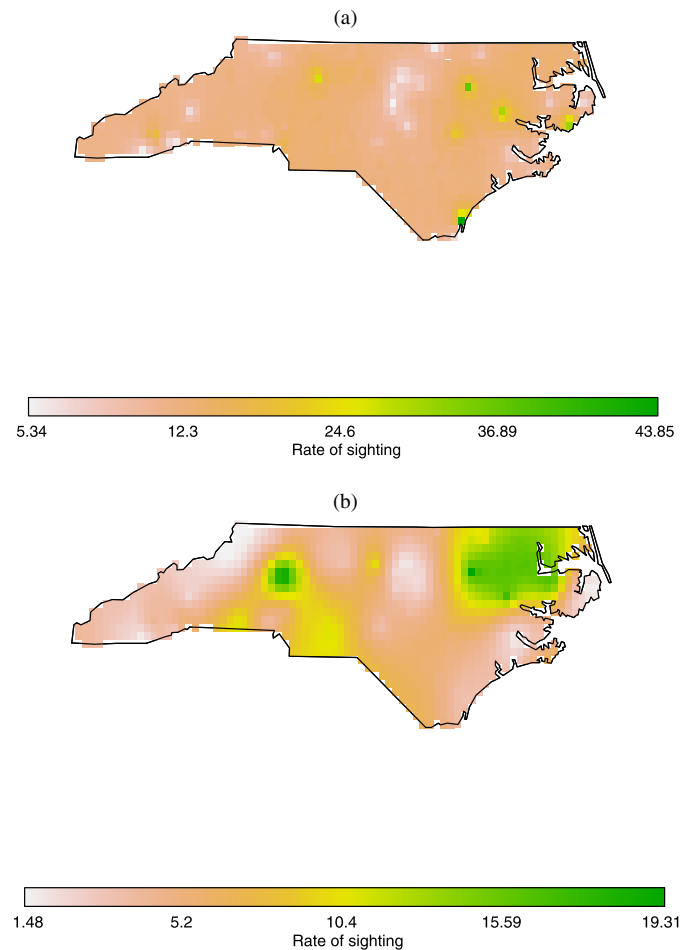


Figure 6. Estimates of the spatial abundance (a) the European Starling and (b) Mourning Dove in North Carolina. A color version of this figure is available in the electronic version of this article.

the SDP (Gelfand, Kottas, and MacEachern 2005) and the order dependent DP (Griffin and Steel 2006b) allow for flexible location-specific distributions for the number of sightings, but again do not provide a common estimate like the one in Figure 7.

## 8. DISCUSSION

We develop and illustrate a novel mechanism to generate stochastic processes with random marginal distributions. This is done by representing the common marginal distribution using a stick-breaking construction and inducing dependence in the selection mechanism of the atoms. Implementation of these models is straightforward using standard MCMC algorithms. One important difference between our model and other hybridization schemes is its ability to provide reasonable predictions and interpolations. Indeed, we show that some sort of identifiability constraint on the atoms is necessary to facilitate interpretation and avoid multimodality in the posterior distribution due to label switching.

LaSBP's are very flexible. Unlike other nonparametric models in the literature, implementation of anisotropic and nonstationary models is straightforward, as is the modeling of discrete or continuous data. In particular, we are currently exploring hierarchical specifications for the latent process that also allow

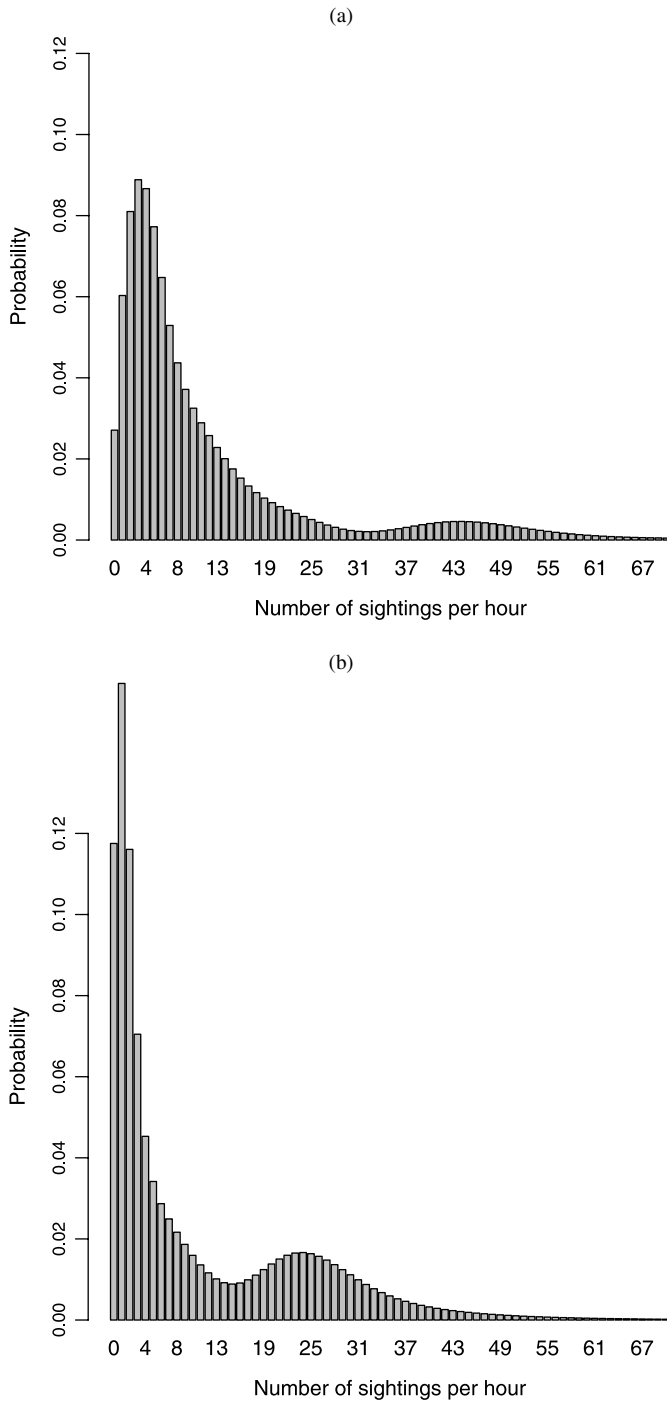


Figure 7. Marginal distribution of sightings for (a) the European Starling and (b) the Mourning Dove in North Carolina.

efficient high-dimensional specifications where the correlation between processes varies in space.

An alternative construction for the LaSBP can be obtained by using the order statistics from a sample of size  $L$  from the baseline measure as the atoms of  $G$ . This approach, which was suggested by one of the referees, will be explored elsewhere.

## APPENDIX

### A.1 Proof of Theorem 1

This is a simple consequence of the properties of a Gaussian process, and can be extended for any other consistent selection process

(in the Kolmogorov sense). Note that, from Equation (2), the finite-dimensional distributions can be represented as mixtures. Therefore, it is enough to prove that the process satisfies Kolmogorov's conditions conditional on  $\{v_l\}_{l=1}^L$  and  $\{\theta_l^*\}_{l=1}^L$ , as mixtures of consistent processes are consistent. Symmetry is immediate,

$$\begin{aligned}
 G(\theta(\mathbf{s}_{v_1}) = \theta_{v_1}^*, \dots, \theta(\mathbf{s}_{v_n}) = \theta_{v_n}^* | \{\theta_l^*\}_{l=1}^L, \{v_l\}_{l=1}^L) \\
 &= \Pr(z(\mathbf{s}_{v_1}) \in [\Phi^{-1}(\pi_{l_{v_1-1}}), \Phi^{-1}(\pi_{l_{v_1}})), \dots, \\
 &\quad z(\mathbf{s}_{v_1}) \in [\Phi^{-1}(\pi_{l_{v_n-1}}), \Phi^{-1}(\pi_{l_{v_n}}))] \\
 &= \Pr(z(\mathbf{s}_1) \in [\Phi^{-1}(\pi_{l_1-1}), \Phi^{-1}(\pi_{l_1})], \dots, \\
 &\quad z(\mathbf{s}_n) \in [\Phi^{-1}(\pi_{l_n-1}), \Phi^{-1}(\pi_{l_n})]) \\
 &= G(\theta(\mathbf{s}_1) = \theta_1^*, \dots, \theta(\mathbf{s}_n) = \theta_n^* | \{\theta_l^*\}_{l=1}^L, \{v_l\}_{l=1}^L).
 \end{aligned}$$

For consistency

$$\begin{aligned}
 \sum_{u \in \{\theta_l^*\}_{l=1}^L} G(\theta(\mathbf{s}_1) = \theta_1^*, \dots, \theta(\mathbf{s}_n) = \theta_n^*, \theta(\mathbf{s}_{n+1}) = u | \\
 \{\theta_l^*\}_{l=1}^L, \{v_l\}_{l=1}^L) \\
 &= \sum_{l_{n+1}=1}^L \Pr(z(\mathbf{s}_1) \in [\Phi^{-1}(\pi_{l_1-1}), \Phi^{-1}(\pi_{l_1})], \dots, \\
 &\quad z(\mathbf{s}_{n+1}) \in [\Phi^{-1}(\pi_{l_{n+1-1}}), \Phi^{-1}(\pi_{l_{n+1}})]) \\
 &= \Pr(z(\mathbf{s}_1) \in [\Phi^{-1}(\pi_{l_1-1}), \Phi^{-1}(\pi_{l_1})], \dots, z(\mathbf{s}_{n+1}) \in (-\infty, \infty)) \\
 &= \Pr(z(\mathbf{s}_1) \in [\Phi^{-1}(\pi_{l_1-1}), \Phi^{-1}(\pi_{l_1})], \dots, \\
 &\quad z(\mathbf{s}_n) \in [\Phi^{-1}(\pi_{l_n-1}), \Phi^{-1}(\pi_{l_n})]) \\
 &= G(\theta(\mathbf{s}_1) = \theta_1^*, \dots, \theta(\mathbf{s}_n) = \theta_n^* | \{\theta_l^*\}_{l=1}^L, \{v_l\}_{l=1}^L).
 \end{aligned}$$

### A.2 Proof of Theorem 2

If  $\mathbb{E}(\theta_l | \theta_{l-1}) \leq \mu + \lambda \theta_{l-1}$  then solving the recursion

$$\mathbb{E}(\theta_l) \leq \mu \sum_{k=0}^{l-1} \lambda^k + \lambda^l \mathbb{E}(\theta_1) = \frac{\mu}{1-\lambda} + \lambda^l (\mathbb{E}(\theta_1) - \mu)$$

and

$$\begin{aligned}
 \mathbb{E}(\theta(\mathbf{s})) &= \sum_{l=1}^{\infty} \mathbb{E}(\theta_l) \frac{a}{a+b} \left( \frac{b}{a+b} \right)^{l-1} \\
 &\leq \sum_{l=1}^{\infty} \left[ \frac{\mu}{1-\lambda} + \lambda^l (\mathbb{E}(\theta_1) - \mu) \right] \frac{a}{a+b} \left( \frac{b}{a+b} \right)^{l-1} \\
 &= \frac{\mu}{1-\lambda} + \frac{a\lambda(\mathbb{E}(\theta) - \mu)}{a+b(1-\lambda)},
 \end{aligned}$$

only if  $\lambda b/(a+b) < 1$ . This result can be easily extended to the case where  $\mathbb{E}(\theta_l) \leq c_1 \lambda^p + c_2 \lambda^l$  by noting that the expected value of the first term corresponds to the value of the polylogarithm function of order  $p$  (Abramowitz and Stegun 1965), which is finite.

### A.3 Properties of the Selection Process

Let  $\{z(\mathbf{s}) : \mathbf{s} \in D\}$  be a standard Gaussian process with correlation function  $\gamma(\mathbf{s}, \mathbf{s}')$  and define  $U(\mathbf{s}) = \Phi(z(\mathbf{s}))$ . We call  $U(\mathbf{s})$  the selection process, as it controls which atom is associated with each location. This section describes some of the properties of  $U(\mathbf{s})$ .

By construction, the marginals of  $U(\mathbf{s})$  are uniform on  $[0, 1]$ , therefore,  $E(U(\mathbf{s})) = 1/2$  for all  $\mathbf{s} \in D$ . Equivalently, we can obtain this result through computing the expectation,

$$E(U(\mathbf{s})) = \int_{-\infty}^{\infty} \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\} dx dz.$$

If we define random variables  $X$  and  $Z$  such that  $(X, Z)' \sim N(\mathbf{0}, \mathbf{I})$ , it turns out that

$$\int_{-\infty}^{\infty} \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\} dx dz = \Pr(Z - X > 0) = 1/2,$$

since  $Z - X \sim N(0, 2)$ . We can use a similar approach to compute  $E(U(\mathbf{s})U(\mathbf{s}'))$ . Indeed,

$$E(U(\mathbf{s})U(\mathbf{s}')) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{z_1} \int_{-\infty}^{z_2} \frac{1}{2\pi} \exp\left\{-\frac{x_1^2 + x_2^2}{2}\right\} \times \frac{1}{2\pi} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2} \mathbf{z}' \Sigma^{-1} \mathbf{z}\right\} dx dz,$$

where  $\mathbf{z} = (z_1, z_2)'$  and

$$\Sigma = \begin{pmatrix} 1 & \gamma(\mathbf{s}, \mathbf{s}') \\ \gamma(\mathbf{s}, \mathbf{s}') & 1 \end{pmatrix}.$$

Define

$$\begin{pmatrix} X_1 \\ X_2 \\ Z_1 \\ Z_2 \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \gamma(\mathbf{s}, \mathbf{s}') \\ 0 & 0 & \gamma(\mathbf{s}, \mathbf{s}') & 1 \end{pmatrix} \right),$$

then

$$E(U(\mathbf{s})U(\mathbf{s}')) = \Pr(Z_1 - X_1 > 0, Z_2 - X_2 > 0),$$

where

$$(Z_1 - X_1 > 0, Z_2 - X_2 > 0) \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & \gamma(\mathbf{s}, \mathbf{s}') \\ \gamma(\mathbf{s}, \mathbf{s}') & 2 \end{pmatrix} \right).$$

Therefore,

$$\text{Cov}(U(\mathbf{s}), U(\mathbf{s}')) = \Phi_{\Psi(\mathbf{s}, \mathbf{s}')} (0, 0) - \frac{1}{4},$$

where  $\Phi_{\Sigma}$  denotes the cumulative distribution of a bivariate normal random variable with covariance matrix  $\Sigma$  and  $\Psi(\mathbf{s}, \mathbf{s}') = \begin{pmatrix} 2 & \gamma(\mathbf{s}, \mathbf{s}') \\ \gamma(\mathbf{s}, \mathbf{s}') & 2 \end{pmatrix}$ . Clearly, if  $\gamma(\mathbf{s}, \mathbf{s}') = 0$  then  $E(U(\mathbf{s})U(\mathbf{s}')) = \Pr(Z_1 - X_1 > 0, Z_2 - X_2 > 0) = \Pr(Z_1 - X_1 > 0)\Pr(Z_2 - X_2 > 0) = E(U(\mathbf{s}))E(U(\mathbf{s}'))$ , and therefore,  $\text{Cov}(U(\mathbf{s}), U(\mathbf{s}')) = 0$ . Higher-order moments of the process can be similarly computed, resulting in expressions where the  $i$ th moment depends on the cumulative distribution of an  $i$ th variate normal random variable.

### SUPPLEMENTAL MATERIALS

**Short note:** Discusses details pertaining to the efficient implementation of the Markov chain Monte Carlo sampler discussed in Section 5, including the efficient generation of pseudo-random numbers from truncated multivariate and conditional normal distributions. (supplemental\_material.pdf)

[Received April 2008. Revised November 2009.]

### REFERENCES

Abramowitz, M., and Stegun, I. A. (eds.) (1965), *Handbook of Mathematical Functions* (1st ed.), New York: Dover. [652,657]

Beal, M. J., Ghahramani, Z., and Rasmussen, C. E. (2001), "The Infinite Hidden Markov Model," in *Proceedings of Fourteenth Annual Conference on Neural Information Processing Systems*, Vancouver, BC, Canada. [648]

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics*, 31, 307–327. [654]

DeIorio, M., Müller, P., Rosner, G. L., and MacEachern, S. N. (2004), "An ANOVA Model for Dependent Random Measures," *Journal of the American Statistical Association*, 99, 205–215. [647,649]

Duan, J. A., Guindani, M., and Gelfand, A. E. (2007), "Generalized Spatial Dirichlet Process Models," *Biometrika*, 94, 809–825. [647]

Dunson, D. B., and Park, J.-H. (2008), "Kernel Stick-Breaking Processes," *Biometrika*, 95, 307–323. [647]

Dunson, D. B., Pillai, N., and Park, J.-H. (2007), "Bayesian Density Regression," *Journal of the Royal Statistical Society, Ser. B*, 69, 163–183. [647]

Ferguson, T. S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, 1, 209–230. [647]

Figueiredo, M. A. T. (2005), "Bayesian Image Segmentation Using Gaussian Field Priors," in *Energy Minimization Methods in Computer Vision and Pattern Recognition*, eds. A. Rangarajan, B. Vemuri, and A. L. Yuille, Berlin, Heidelberg: Springer-Verlag, pp. 74–89. [647]

Figueiredo, M. A., Cheng, D. S., and Murino, V. (2007), "Clustering Under Prior Knowledge With Application to Image Segmentation," in *Advances in Neural Information Processing Systems 19*, eds. B. Schölkopf, J. Platt, and T. Hoffman, Cambridge, MA: MIT Press, pp. 401–408. [647]

Fox, E., Sudderth, E., Jordan, M. I., and Willsky, A. (2008), "An HDP-HMM for Systems With State Persistence," in *Proceedings of the 25th International Conference on Machine Learning (ICML)*, Helsinki, Finland. [648]

Gelfand, A. E., Kottas, A., and MacEachern, S. N. (2005), "Bayesian Nonparametric Spatial Modeling With Dirichlet Process Mixing," *Journal of the American Statistical Association*, 100, 1021–1035. [647–649,656]

Griffin, J. E., and Steel, M. F. J. (2006a), "Nonparametric Inference in Time Series Problems," technical report, University of Warwick, Dept. of Statistics. [647]

——— (2006b), "Order-Based Dependent Dirichlet Processes," *Journal of the American Statistical Association*, 101, 179–194. [647–649,656]

Hoff, P. D. (2007), "Extending the Rank Likelihood for Semiparametric Copula Estimation," *The Annals of Applied Statistics*, 1, 265–283. [648]

Ishwaran, H., and James, L. F. (2001), "Gibbs Sampling Methods for Stick-Breaking Priors," *Journal of the American Statistical Association*, 96, 161–173. [647,648,651,652]

——— (2003), "Some Further Developments for Stick-Breaking Priors: Finite and Infinite Clustering and Classification," *Sankhya*, 65 577–592. [651]

Jacquier, E., Polson, N. G., and Rossi, P. E. (1994), "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics*, 12, 371–389. [654,655]

Kottas, A., Müller, P., and Quintana, F. (2005), "Nonparametric Bayesian Modeling for Multivariate Ordinal Data," *Journal of Computational and Graphical Statistics*, 14, 610–625. [648]

Lo, A. Y. (1984), "On a Class of Bayesian Nonparametric Estimates: I. Density Estimates," *The Annals of Statistics*, 12, 351–357. [648]

MacEachern, S. N. (2000), "Dependent Dirichlet Processes," technical report, Ohio State University, Dept. of Statistics. [647,649]

——— (2007), Discussion of "Bayesian Nonparametric Modelling for Spatial Data Using Dirichlet Processes" by A. E. Gelfand, M. Guindani, and S. Petrone, in *Bayesian Statistics 8*, eds. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West, New York: Oxford University Press, pp. 196–198. [647]

Müller, P., Quintana, F., and Rosner, G. (2004), "Hierarchical Meta-Analysis Over Related Non-Parametric Bayesian Models," *Journal of the Royal Statistical Society, Ser. B*, 66, 735–749. [647]

Nelsen, R. B. (1999), *An Introduction to Copulas*, New York: Springer. [648]

Ongaro, A., and Cattaneo, C. (2004), "Discrete Random Probability Measures: A General Framework for Nonparametric Bayesian Inference," *Statistics and Probability Letters*, 67, 33–45. [647]

Pal, N. R., and Pal, S. K. (1993), "A Review on Image Sementation Techniques," *Pattern Recognition*, 26, 1277–1294. [647]

Petrone, S., Guindani, M., and Gelfand, A. E. (2009), "Hybrid Dirichlet Mixture Models for Functional Data," *Journal of the Royal Statistical Society, Ser. B*, 71, 755–782. [648,652]

Pham, D. L., Xu, C., and Prince, J. L. (2000), "A Survey of Current Methods in Medical Image Segmentation," *Review of Biomedical Engineering*, 2, 315–337. [647]

Pitman, J. (1996), "Some Developments of the Blackwell–MacQueen Urn Scheme," in *Statistics, Probability and Game Theory. Papers in Honor of David Blackwell*, eds. T. S. Ferguson, L. S. Shapeley, and J. B. MacQueen, Hayward, CA: IMS, pp. 245–268. [648]



- Richardson, S., and Green, P. J. (2001), "Hidden Markov Models and Disease Mapping," *Journal of the American Statistical Association*, 97, 1055–1070. [655]
- Roberts, G., and Papaspiliopoulos, O. (2008), "Retrospective Markov Chain Monte Carlo Methods for Dirichlet Process Hierarchical Models," *Biometrika*, 95, 169–186. [653]
- Rodríguez, A., Dunson, D. B., and Gelfand, A. E. (2008), "The Nested Dirichlet Process" (with discussion), *Journal of the American Statistical Association*, 103, 1131–1154. [647,651]
- Rodríguez, A., Dunson, D. B., and Taylor, J. (2009), "Bayesian Hierarchically Weighted Finite Mixtures Models for Samples of Distributions," *Biostatistics*, 10, 155–171. [647]
- Sethuraman, J. (1994), "A Constructive Definition of Dirichlet Priors," *Statistica Sinica*, 4, 639–650. [647,648]
- Suri, J. S., Wilson, D., and Laxminarayan, S. (eds.) (2005), *Handbook of Biomedical Image Analysis. Volume 2: Segmentation Models. Part B*, New York: Springer. [647]
- Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006), "Sharing Clusters Among Related Groups: Hierarchical Dirichlet Processes," *Journal of the American Statistical Association*, 101, 1566–1581. [647,648]
- van Gael, J., Saatchi, Y., Teh, Y.-W., and Ghahramani, Z. (2008), "Beam Sampling for the Infinite Hidden Markov Model," in *Proceedings of the 25th International Conference on Machine Learning (ICML)*, Helsinki, Finland. [648]