INTERNATIONAL CENTRE FOR ECONOMIC RESEARCH

ICER

WORKING PAPER SERIES

R. H. Mena, M. Ruggiero, S. Walker

GEOMETRIC STICK-BREAKING PROCESSES FOR CONTINUOUS-TIME NONPARAMETRIC MODELING

Working Paper No. 26/2009

APPLIED MATHEMATICS
WORKING PAPER SERIES

Geometric stick-breaking processes for continuous-time nonparametric modeling

Ramsés H. Mena, Matteo Ruggiero and Stephen Walker
Universidad Nacional Autónoma de México, University of Pavia and University of Kent

Abstract

This paper is concerned with the construction of a continuous parameter sequence of random probability measures and its application for modeling random phenomena evolving in continuous time. At each time point we have a random probability measure which is generated by a Bayesian nonparametric hierarchical model, and the dependence structure is induced through a Wright-Fisher diffusion with mutation. The sequence is shown to be a stationary and reversible diffusion taking values on the space of probability measures. A simple estimation procedure for discretely observed data is presented and illustrated with simulated and real data sets.

Keywords: Bayesian non-parametric inference, continuous time dependent random measure, Markov process, measure-valued process, stationary process, stick-breaking process

1. Introduction. The Bayesian nonparametric approach to statistical inference has become a useful methodology, not only in exchangeable contexts, but also as a modular component to describe phenomena with other kind of dependence structure. The key object in Bayesian nonparametric methods is the construction of a random distribution function, the most well known of which is the Dirichlet process, introduced in [6]. Let $(\mathcal{X}, \mathscr{B}(\mathcal{X}))$ be a measurable space, where $\mathcal{X}$ is a complete separable metric space and $\mathscr{B}(\mathcal{X})$ its Borel sigma-field, and define

$$\mu = \sum_{i\geq 1} w_i \, \delta_{x_i} \tag{1}$$

where δ_x denotes a point mass at x, $(x_i)_{i\geq 1}$ is a vector of i.i.d. samples from ν_0, assumed to be a non-atomic probability measure on $(\mathcal{X}, \mathscr{B}(\mathcal{X}))$, and $(w_i)_{i\geq 1}$ is a vector of weights. The Dirichlet process is obtained by (1) with one-parameter GEM(c) weights, that is when

$$w_i = v_i \prod_{j=i}^{i-1} (1 - v_j) \tag{2}$$

and $v_j \sim^{iid}$ Beta$(1, c)$ random variables, for some $c > 0$. The above representation of the Dirichlet process is due to [15].

Recent attention has focussed on the application of suitably defined measure-valued processes for nonparametric inference purposes. The idea is to provide a sequence of random probability measures linked by a suitable dependence structure, e.g. by means of a time parameter and an appropriate transition mechanism, and use it for drawing inferences on random phenomena, usually with the aid of simulation techniques. Apart from the theoretical developments offered by the probabilistic study of measure-valued processes, for example as in [3] and [5], whose literature is certainly broad and well established, on the statistical side this is a relatively young area. To date, the most productive ideas have involved the Dirichlet process. See, for example, [2], [4], [8], [10], [12] [14]. In particular, to the best

of our knowledge there is no current available statistical literature devoted to the study of continuously dependent random measures which exploits the flexibility of a Bayesian nonparametric approach and at the same time investigates desirable properties such as stationarity, reversibility, Markovianity, regularity of sample paths for the constructed model.

Here we propose the construction of a sequence of random probability measures which depends on a continuous parameter, has nice path properties and turns out to be a useful tool for nonparametric inference on continuous-time random phenomena despite its relative simplicity. The construction is suggested by a recently studied random probability measure (see [7]), which has proven to have some appealing features when used in Bayesian nonparametric mixture modeling and regression analysis contexts. Let $\mathscr{P}(\mathcal{X})$ be the set of probability measures on $(\mathcal{X}, \mathscr{B}(\mathcal{X}))$. Define $\mu \in \mathscr{P}(\mathcal{X})$ as

$$
\begin{aligned}
\mu &= \textstyle\sum_{i\geq 1} q_i \delta_{x_i} \\
q_i &= \lambda(1-\lambda)^{i-1} \qquad i \geq 1 \\
\lambda &\sim \text{Beta}(a, b) \\
x_i &\sim \nu_0 \qquad\qquad\quad i \geq 1
\end{aligned} \tag{3}
$$

where the non atomic probability measure ν_0 can be thought of as the prior guess for μ. At first sight the random probability measure provided by (3) can be misinterpreted as a special case of the Dirichlet process, since the weights of the former can be obtained by (2) by letting $(v_1, v_2, \ldots)$ be the realization of the same Beta random variable, so that ω_i, $i = 1, 2, \ldots$, are mixed geometric with Beta(a, b) as the mixing distribution. However this proves not to be the case. First because in the Dirichlet case the parameter a of the Beta distribution is constrained to be one. And more significantly because the Dirichlet process satisfies the distributional equation

$$
\mu \stackrel{d}{=} w_1 \delta_{x_1} + (1 - w_1)\mu \tag{4}
$$

obtained by eliciting the first term in the infinite sum (1) and exploiting the fact that with probability one

$$
\sum_{i=1}^{n} w_i = 1 - \prod_{i=1}^{n} (1 - w_i) \to 1
$$

as $n \to \infty$. See [15]. The same procedure applied to (3) yields

$$
\mu \stackrel{d}{=} \lambda \delta_{x_1} + (1 - \lambda)\mu. \tag{5}
$$

The crucial difference between these two cases is that in (4) μ is independent of (w_1, x_1), while in (4) μ is independent of x_1 but not of λ. Hence we are dealing with a different random probability measure. An interpretation of (3) as related to the Dirichlet process is nonetheless available. This is given by taking the expectation of (2), which yields

$$
\mathbb{E}(w_i) = \left(\frac{1}{1+c}\right)\left(\frac{c}{1+c}\right)^{i-1}
$$

and we let $\lambda = 1/(1 + c)$, c being the total mass of the Dirichlet process. Hence the random probability measure (3) can be thought of as obtained by removing a level of the hierarchy from the Dirichlet process model by replacing the random weights with their expected values.

The purpose of the present paper is twofold. Firstly we introduce a purely atomic continuous-time measure-valued Markov process $\{\mu(t,\omega), t \geq 0\}$, defined on an appropriate probability space $(\Omega, \mathscr{F}, \mathbb{P})$, such that at each fixed $\omega \in \Omega$, $\mu(t,\omega)$ is a continuous function from $[0,\infty)$ to $\mathscr{P}(\mathcal{X})$, and at each $t \geq 0$, $\mu(t,\omega)$ is a random probability measure of the type (3). We proceed by inducing a dependent structure on a sequence of random measures analogous to (3) by letting λ depend on time and be a two-type Wright-Fisher diffusion process with selectively neutral mutation and random genetic drift (see, for example, [5], Chapter 10). We investigate the properties of the resulting process which turns out to be a stationary and reversible diffusion taking values in $\mathscr{P}(\mathcal{X})$. We characterise the process in terms of its generator and we identify the invariant measure. Secondly, we apply the construction in a framework of non parametric Bayesian inference on some continuous time random phenomena whose data have been observed discretely. Given a set of available observations $(y_1, \ldots, y_n)$, recorded at instants $(t_1, \ldots, t_n)$ respectively, we assume y_i is drawn from the nonparamteric mixture model

$$f_{t_i}(y) = \int_{\mathcal{X}} f_{t_i}(y|z)\mu_{t_i}(\mathrm{d}z)$$

where $\mu_t := \mu(t,\omega)$ is, as above mentioned, of type (3). We develop a simple estimation procedure for drawing inferences on the trajectory of continuous time phenomena which are observed discretely, by means of Markov chain Monte Carlo techniques, and illustrate it with simulated and real data sets.

2. Geometric stick-breaking processes. In this section we define a dependent nonparametric model by introducing a continuous parameter probability-measure-valued process in a Bayesian framework which provides (3) at every instant. Let $\mathscr{P}(\mathcal{X})$ be the set of Borel probability measures on $(\mathcal{X}, \mathscr{B}(\mathcal{X}))$.

Definition 1. Let $\{\mu(t,\omega), t \geq 0\}$ be a $\mathscr{P}(\mathcal{X})$-valued stochastic process defined on an appropriate probability space $(\Omega, \mathscr{F}, \mathbb{P})$ such that at each $t \geq 0$

$$\begin{aligned} \mu(t,\omega) &= \textstyle\sum_{i\geq 1} q_i(t,\omega)\delta_{x_i} \\ q_i(t,\omega) &= \lambda(t,\omega)(1-\lambda(t,\omega))^{i-1} \qquad i \geq 1 \\ x_i &\sim \nu_0 \qquad\qquad\qquad\qquad\quad i \geq 1 \end{aligned} \tag{6}$$

where ν_0 is a non-atomic distribution on $(\mathcal{X}, \mathscr{B}(\mathcal{X}))$ and $\{\lambda(t,\omega), t \geq 0\}$ is a two-type Wright-Fisher diffusion with mutation, with paths in the set $C_{[0,1]}([0,\infty))$ of continuous functions from $[0,\infty)$ to $[0,1]$ and infinitesimal generator

$$\mathcal{A} = \left[\frac{a}{2}(1-\lambda) - \frac{b}{2}\lambda\right]\frac{\mathrm{d}}{\mathrm{d}\lambda} + \frac{1}{2}\lambda(1-\lambda)\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}. \tag{7}$$

The domain of $\mathcal{A}$ is taken to be $\mathscr{D}(\mathcal{A}) = C^2([0,1])$. Such process $\{\mu(t,\omega), t \geq 0\}$ will be referred to as *geometric stick-breaking process* with parameters (a, b, ν_0) and denoted $\mathrm{GSB}(a, b, \nu_0)$.

In the following we will make use of the shorter notation $\{\mu_t\}_{t\geq 0} = \{\mu(t,\omega), t \geq 0\}$ and $\{\lambda_t\}_{t\geq 0} = \{\lambda(t,\omega), t \geq 0\}$. The GSB process describes the evolution of a random probability measures of type (3) whose weights evolve smoothly in time. It is well-known that a Wright-Fisher diffusion process with generator (7) is reversible and stationary with respect to a

Beta(a,b) distribution. We will show that also $\{\mu_t\}_{t\geq 0}$ is reversible and stationary, and identify the invariant measure. First though we show that it is a measure-valued diffusion.

Assume for simplicity that $\mathcal{X}$ is compact (otherwise we could take $\mathcal{X}$ to be locally compact and consider functions on $\mathcal{X}$ which vanish at infinity), and let $\mathscr{P}_g(\mathcal{X}) \subset \mathscr{P}(\mathcal{X})$ be the set of purely atomic probability measures with geometric weights as in (3). Given $x := (x_i)_{i\geq 1}$, in order to emphasize the dependence on λ, define the continuous map $\phi_x : [0,1] \to \mathscr{P}_g(\mathcal{X})$ as

$$\phi_x(\lambda) = \sum_{i\geq 1} \lambda(1-\lambda)^{i-1}\delta_{x_i}. \tag{8}$$

For $m \in \mathbb{N}$ and $f \in C(\mathcal{X}^m)$, the space of continuous functions on $\mathcal{X}^m$, let

$$\varphi_m(\mu) = \varphi_m(\phi_x(\lambda)) := \langle f, \mu^m \rangle \tag{9}$$

where $\langle f, \mu\rangle = \int f \mathrm{d}\mu$ and $\mu^m = \mu \times \ldots \times \mu$ is an m-fold product measure. Let $C_P(\mathscr{P}_g(\mathcal{X}))$ be the subalgebra of $C(\mathscr{P}_g(\mathcal{X}))$ given by the linear span of monomials of type (9) when $\mu \in \mathscr{P}_g(\mathcal{X})$.

Proposition 1. Let $\{\mu_t\}_{t\geq 0}$ be a GSB(a,b,ν_0) process. Then $\{\mu_t\}_{t\geq 0}$ has infinitesimal operator

$$\begin{aligned}\mathcal{B}\varphi_m(\mu) =& \left(\frac{a}{2}(1-\lambda) - \frac{b}{2}\lambda\right) \sum_{i_1,\ldots,i_m\geq 1} f(x_{i_1},\ldots,x_{i_m}) \frac{\partial}{\partial\lambda} h(\lambda; m, i_1, \ldots, i_m) \\ &+ \frac{1}{2}\lambda(1-\lambda) \sum_{i_1,\ldots,i_m\geq 1} f(x_{i_1},\ldots,x_{i_m}) \frac{\partial^2}{\partial\lambda^2} h(\lambda; m, i_1, \ldots, i_m)\end{aligned}$$

with domain

$$\mathscr{D}(\mathcal{B}) = \left\{\varphi \in C(\mathscr{P}_g(\mathcal{X})) :\ \varphi = \varphi_m(\mu) = \langle f, \mu^m\rangle,\ f \in C(\mathcal{X}^m),\ m \in \mathbb{N}\right\}$$

and where

$$h(\lambda; m, i_1, \ldots, i_m) = \lambda_t^m (1-\lambda_t)^{\sum_{j=1}^m i_j - m}.$$

The subalgebra of $C(\mathscr{P}_g(\mathcal{X}))$ given by the linear span of monomials of type (9) when $\mu \in \mathscr{P}_g(\mathcal{X})$ is a core for $\mathcal{B}$.

Proof. From (9) we can write

$$\varphi_m(\mu_t) = \sum_{i_1\geq 1} \ldots \sum_{i_m \geq 1} \lambda_t^m (1-\lambda_t)^{i_1+\ldots+i_m-m} f(x_{i_1},\ldots,x_{i_m}). \tag{10}$$

It follows that the operator semigroup $\{T(t)\}_{t\geq 0}$ for the process $\{\mu_t\}_{t\geq 0}$ is such that for every $t \geq 0$

$$T(t)\varphi_m(\mu_0) = \sum_{i_1,\ldots,i_m\geq 1} \int_{[0,1]} \lambda_t^m (1-\lambda_t)^{\sum_{j=1}^m i_j - m} p(\mathrm{d}\lambda_t|\lambda_0) f(x_{i_1},\ldots,x_{i_m}) \tag{11}$$

where $\mu_0 = \mu(0)$ is the initial value of the process and $p(\mathrm{d}\lambda_t|\lambda_0)$ is the transition function for $\{\lambda_t\}_{t\geq 0}$. From the fact that the two-type Wright-Fisher process is a well-defined diffusion on $[0,1]$, with infinitesimal operator (7) defined on $C^2([0,1])$, it can be easily seen that

$$t^{-1}[T(t)\varphi_m(\mu_0) - \varphi_m(\mu_0)]$$

converges strongly, as $t \downarrow 0$, to

$$\sum_{i_1,\dots,i_m \geq 1} f(x_{i_1},\dots,x_{i_m}) \mathcal{A} h(\lambda; m, i_1, \dots, i_m) \tag{12}$$

where

$$h(\lambda; m, i_1, \dots, i_m) = \lambda_t^m (1-\lambda_t)^{\sum_{j=1}^m i_j - m}.$$

In view of (7), from (12) the generator for $\{\mu_t\}_{t\geq 0}$ in terms of $\{\lambda_t\}_{t\geq 0}$ can be written

$$\begin{aligned}\mathcal{B}\varphi_m(\mu) =& \Big(\frac{a}{2}(1-\lambda) - \frac{b}{2}\lambda\Big) \sum_{i_1,\dots,i_m \geq 1} h_1(\lambda; m, i_1, \dots, i_m) f(x_{i_1},\dots,x_{i_m}) \\ &+ \frac{1}{2}\lambda(1-\lambda) \sum_{i_1,\dots,i_m \geq 1} h_2(\lambda; m, i_1, \dots, i_m) f(x_{i_1},\dots,x_{i_m})\end{aligned}$$

where $h_1(\lambda; m, i_1, \dots, i_m)$ and $h_2(\lambda; m, i_1, \dots, i_m)$ are the first and second derivatives of $h(\lambda; m, i_1, \dots, i_m)$ with respect to λ. From [3], Lemma 2.1.2, it follows that $C_P(\mathscr{P}_g(\mathcal{X}))$ is dense in $C(\mathscr{P}_g(\mathcal{X}))$, so that $C_P(\mathscr{P}_g(\mathcal{X}))$ is a core for $\mathcal{B}$, and we can take the domain of $\mathcal{B}$ to be $C(\mathscr{P}_g(\mathcal{X}))$. □

The following proposition states that $\{\mu_t\}_{t\geq 0}$ is a purely atomic measure-valued Feller diffusion.

Proposition 2. Let $\mathcal{X}$ be a compact complete separable metric space, and let $\{\mu_t\}_{t\geq 0}$ be a GSB(a, b, ν_0) process on $\mathscr{P}_g(\mathcal{X})$. Then $\{\mu_t\}_{t\geq 0}$ is a Feller process with sample paths in $C_{\mathscr{P}_g(\mathcal{X})}([0,\infty))$.

Proof. Recall (8) and note that the vector of weights associated to the locations $(x_1, x_2, \dots)$ is always disposed in decreasing order. Hence, letting $g_x(\mu) := \mu(\{x_1\}) = \lambda$, we can define $\phi_x^{-1} := g_x$. Denoting $\{S(t)\}_{t\geq 0}$ the Feller semigroup on $C([0,1])$ corresponding to $\{\lambda_t\}_{t\geq 0}$, we can define a strongly continuous, positive, conservative, contraction semigroup $\{T(t)\}_{t\geq 0}$ on $C(\mathscr{P}_g(\mathcal{X}))$ by

$$T(t)\varphi = [S(t)(\varphi \circ \phi_x)] \circ \phi_x^{-1}.$$

Then Theorem 4.2.7 of [5] implies that for every $\nu \in \mathscr{P}_g(\mathcal{X})$, there exists a Markov process $\{\mu_t\}_{t\geq 0}$ corresponding to $\{T(t)\}$ with initial distribution ν and sample paths in $D_{\mathscr{P}_g(\mathcal{X})}([0,\infty))$, the space of *càdlàg* functions from $[0,\infty)$ to $\mathscr{P}_g(\mathcal{X})$.

Denote now with $P(t, \mu, \mathrm{d}\nu)$ the transition function corresponding to the semigroup $\{T(t)\}_{t\geq 0}$, that is

$$T(t)\varphi(\mu) = \int_{\mathscr{P}(\mathcal{X})} \varphi(\nu) P(t, \mu, \mathrm{d}\nu).$$

In order to show that the sample paths of $\{\mu_t\}_{t\geq 0}$ are continuous in $\mathscr{P}_g(\mathcal{X})$, it suffices to show that for every $\mu \in \mathscr{P}_g(\mathcal{X})$ and every $\varepsilon > 0$ we have

$$t^{-1} P(t, \mu, B(\mu, \varepsilon)^c) \to 0 \qquad \text{as } t \to 0 \tag{13}$$

where $B(\mu, \varepsilon)^c$ is the complement of an ε-neighborhood of μ in a topology which makes $\mathscr{P}_g(\mathcal{X})$ locally compact and separable. See [5], Lemma 4.2.9. For our purposes it is enough to show that the sample paths are continuous in $\mathscr{P}_{g,x}(\mathcal{X})$, defined to be the restriction of

$\mathscr{P}_g(\mathcal{X})$ to the set of purely atomic probability measures with geometric weights and a given set of locations $(x_i)_{i\geq 1}$, for every initial distribution $\mu \in \mathscr{P}_{g,x}(\mathcal{X})$. Thus it suffices to show that (13) holds for every $\mu \in \mathscr{P}_{g,x}(\mathcal{X})$ and every $\varepsilon > 0$, where $B(\mu,\varepsilon)^c = B_W(\mu,\varepsilon)^c$ is the complement of, say, an ε-neighborhood of μ in the weak topology. Observe now that $\mathscr{P}_{g,x}(\mathcal{X})$ is locally compact since, from the continuity of (8), for every $\mu \in \mathscr{P}_{g,x}(\mathcal{X})$ one can find a compact neighborhood of μ in $\mathscr{P}_{g,x}(\mathcal{X})$ by letting λ vary smoothly. Also $\mathscr{P}_{g,x}(\mathcal{X})$ is separable, since the set $\{\nu \in \mathscr{P}_{g,x}(\mathcal{X}) : \lambda \in \mathbb{Q} \cap [0,1]\}$ is countable and dense in $\mathscr{P}_{g,x}(\mathcal{X})$. The key now is the fact that in (11) the transition function providing $\{T(t)\}_{t\geq 0}$ is expressed in terms of $p(\lambda_t|\lambda_0)$. That is, if $\mu = \sum_{i\geq 1} \lambda_0(1-\lambda_0)^{i-1}\delta_{x_i}$ and $\nu = \sum_{i\geq 1} \lambda_t(1-\lambda_t)^{i-1}\delta_{y_i}$, then $P(t,\mu,\mathrm{d}\nu) = p(\mathrm{d}\lambda_t|\lambda_0)\prod_{i=1}^{\infty}\delta_{x_i}(\mathrm{d}y_i)$. Hence for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$P(t,\mu,B_W(\mu,\varepsilon)^c) = p(B(\lambda_0,\delta)^c|\lambda_0)\prod_{i=1}^{\infty}\delta_{x_i}(\mathrm{d}y_i)$$

Since $\{\lambda_t\}_{t\geq 0}$ is a diffusion, we have

$$t^{-1}p(B(\lambda_0,\delta)^c|\lambda_0) \to 0, \qquad t \to 0.$$

Since also the product is bounded by one, (13) follows. □

We now turn to stationarity. Denote $B_{a,b} = \mathrm{Beta}(a,b)$ and $\tilde{B}_{a,b} = B_{a,b}\circ\phi_x^{-1}$, with ϕ_x defined as in (8). Then we have the following.

Proposition 3. Let $\{\mu_t\}_{t\geq 0}$ be a $\mathrm{GSB}(a,b,\nu_0)$ process. Then $\{\mu_t\}_{t\geq 0}$ is reversible with respect to $\tilde{B}_{a,b}$.

Proof. Given $(x_i)_{i\geq 1}$, it is enough to show that for every $t \geq 0$ and every $f,g \in C(\mathscr{P}_g(\mathcal{X}))$ we have $\int_{\mathscr{P}_g(\mathcal{X})} fT(t)g\,\mathrm{d}\tilde{B}_{a,b} = \int_{\mathscr{P}_g(\mathcal{X})} gT(t)f\mathrm{d}\tilde{B}_{a,b}$. Let $\{S(t)\}_{t\geq 0}$ and $\{T(t)\}_{t\geq 0}$ be as in Proposition 2, and note that $f\circ\phi_x$ and $g\circ\phi_x$ are in $C([0,1])$. Then we have

$$\begin{aligned}\int_{\mathscr{P}_g(\mathcal{X})} fT(t)g\,\mathrm{d}\tilde{B}_{a,b} &= \int_{\mathscr{P}_g(\mathcal{X})} f[S(t)(g\circ\phi_x)]\circ\phi_x^{-1}\mathrm{d}(B_{a,b}\circ\phi_x^{-1}) \\ &= \int_{[0,1]}(f\circ\phi_x)[S(t)(g\circ\phi_x)]\mathrm{d}B_{a,b}\end{aligned}$$

from which the result is implied by the reversibility of $\{\lambda_t\}_{t\geq 0}$ with respect to $B_{a,b}$. □

Corollary 1. Let $\{\mu_t\}_{t\geq 0}$ be a $\mathrm{GSB}(a,b,\nu_0)$ process. Then $\{\mu_t\}_{t\geq 0}$ has invariant law $\tilde{B}_{a,b}$.

Proof. This is an immediate consequence of the previous proposition. □

Hence $\{\mu_t\}_{t\geq 0}$ is a continuous-time purely atomic measure-valued diffusion with continuous sample paths, which is also stationary and reversible. The next section show how this model can be used in a Bayesian nonparametric setting for inference purposes on some continuous-time random phenomenon which has been observed discretely.

3. Estimation. Assume we are interested in a continuous time process which is observed discretely, and a set of observations $(y_i)_{i=1}^n$ recorded at times $(t_i)_{i=1}^n$ is available. We wish to model the dynamics driving such a trajectory by assuming that at time t the corresponding observation follows a random distribution with density given by

$$f_t(y) = \int_{\mathcal{X}} f_t(y|z)\mu_t(\mathrm{d}z) = \sum_{l=1}^{\infty} \lambda_t \, (1-\lambda_t)^{l-1} \, \mathsf{K}(y \mid \theta_l) \tag{14}$$

where μ_t is of type (3), $\mathsf{K}(\cdot \mid \theta)$ is a well defined density function for all θ, $\{\lambda_t\}_{t\geq 0}$ is a two-type Wright-Fisher diffusion on $[0,1]$ and $(\theta_l)_{l=1}^{\infty}$ are i.i.d. from a non atomic probability measure ν_0. The non parametric mixture model (14) can alternatively be written in hierarchical form

$$\begin{aligned} y_i \mid t_i, x_i &\sim \mathsf{K}(\cdot \mid x_i) \\ \{x_i\} &\sim \mu_t \\ \mu_t &\sim \mathrm{GSB}(a, b, \nu_0). \end{aligned} \tag{15}$$

where for simplicity we have set $x_i := x_{t_i}$.

In order to estimate the model we undertake a Gibbs sampler algorithm scheme. For this purpose, it is convenient to start by considering the Wright-Fisher diffusion process $\{\lambda_t\}_{t\geq 0}$. Assume we have observations $(t_i, s_i)_{i=1}^n$, where s_i is a random variable that indicates to us which $\mathsf{K}(\cdot \mid \theta_l)$ the observation y_i comes from. The model is then written as follows:

$$s_i|\lambda_i \sim \mathrm{Geom}(\lambda_i)$$

with $\lambda_i := \lambda_{t_i}$ and

$$p(\lambda_i|\lambda_{i-1}) = \sum_{m=0}^{\infty} p_i(m) \, D(\lambda_i|m, \lambda_{i-1}) \tag{16}$$

where

$$p_i(m) = \frac{(a+b)_m \, e^{-m\,c\,\tau_i}}{m!}(1-e^{-c\,\tau_i})^{a+b}$$

and $\tau_i := t_i - t_{i-1}$. Also,

$$D(\lambda_i|m, \lambda_{i-1}) = \sum_{k=0}^{m} \mathrm{Beta}(\lambda_i|a+k, b+m-k) \, \mathrm{Bin}(k|m, \lambda_{i-1}).$$

Here, equation (16) corresponds to a representation of the transition density of a general class of Beta-Binomial diffusion process, which includes the Wright-Fisher diffusion with (7) as a particular case, and is characterized by the infinitesimal generator

$$\mathcal{A} = \left[\frac{c}{a+b-1}(a-(a+b)\lambda)\right]\frac{\mathrm{d}}{\mathrm{d}\lambda} + \frac{c}{a+b-1}\lambda(1-\lambda)\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}. \tag{17}$$

Such a generator corresponds to a strictly stationary reversible diffusion process with $\mathrm{Beta}(a,b)$ invariant distribution. See [13] for details. We describe the algorithm in this, slightly more general, setting since the inclusion of the parameter c provides a clearer and more general

interpretation for the dependence structure in the diffusion model. A simple reparameterization, i.e. $c := (a+b-1)/2$, leads us back to the Wright-Fisher model with generator (7).

To simplify and accommodate a Gibbs sampler, in particular to avoid the infinite summations needed for (16), we need to introduce latent variables $(u_i, d_i, k_i)_{i=1}^n$ whereby the joint density $p(\lambda_i, u_i, k_i, d_i|\lambda_{i-1})$ is given by

$$\mathbf{1}(u_i < g(d_i))\,\frac{p_i(d_i)}{g(d_i)}\,\mathrm{Beta}(\lambda_i|a+k_i, b+d_i-k_i)\,\mathrm{Bin}(k_i|d_i,\lambda_{i-1}),$$

where g is a decreasing function with known inverse. Integrating out the latent variables clearly yields $p(\lambda_i|\lambda_{i-1})$. Hence, the likelihood function with the complete data is

$$l(a,b,c) = \mathrm{Beta}(\lambda_0|a,b)\prod_{i=1}^n p(\lambda_i,u_i,k_i,d_i|\lambda_{i-1})\,\lambda_i(1-\lambda_i)^{s_i-1}.$$

We now concentrate on establishing the full conditional distributions and start with (a,b,c) assuming independent standard exponential distributions as priors. We therefore consider $\pi(a|b,c,\ldots) \propto l(a,b,c)\,e^{-a}$ for which

$$\log\pi(a|b,c,\ldots) = \sum_{i=1}^n \Big\{2\log\Gamma(a+b+d_i) - \log\Gamma(a+k_i) + a\log(1-e^{-c\,\tau_i}) + a\log\lambda_i\Big\} + a\log\lambda_0 - \log\Gamma(a) - (n-1)\log\Gamma(a+b) - a + C.$$

Here C is the normalising constant not depending on a. It is not difficult to see that $\log\pi(a|b,c,\ldots)$ is concave and hence we can implement the adaptive rejection sampler of [9]. The full conditional for b follows similarly and is also log-concave. The full conditional for c can be written as

$$\log\pi(c|a,b,\ldots) = \sum_{i=1}^n \big\{(a+b)\log(1-e^{-c\,\tau_i}) - d_i\,c\,\tau_i\big\} - c + C,$$

again C does not depend on c. This can also be shown to be concave and so the adaptive rejection sampler also applies. The full conditional distribution for k_i is given by

$$\pi(k_i|\ldots) \propto \binom{d_i}{k_i}\frac{\mathbf{1}(k_i\in\{0,1,\ldots,d_i\})}{\Gamma(a+k_i)\Gamma(b+d_i-k_i)}\left\{\frac{\lambda_i\lambda_{i-1}}{(1-\lambda_i)(1-\lambda_{i-1})}\right\}^{k_i}$$

which is clearly easy to sample since k_i can only take a finite number of values.

The full conditional for u_i is simply uniform distribution on $(0, g(d_i))$, where we choose g for simplicity; for example, $g(d) = e^{-d}$ or $g(d) = d^{-2}$, so that g^{-1} is known and the benefit of this is apparent when we consider the full conditional for d_i; which is next.

The full conditional for d_i is given by

$$\pi(d_i|\ldots) \propto \frac{p_i(d_i)}{g(d_i)}\binom{d_i}{k_i} \times \frac{\Gamma(a+b+d_i)}{\Gamma(b+d_i-k_i)}\{(1-\lambda_{i-1})(1-\lambda_i)\}^{d_i}\,\mathbf{1}(k_i\le d_i\le g^{-1}(u_i))$$

which by virtue of u_i is restricted to a finite set.

The full conditional for λ_i, for $i \neq 0, n$, is given by

$$\pi(\lambda_i|\ldots) = \text{Beta}(1 + a + k_i + k_{i+1}, s_i - 1 + b + d_i + d_{i+1} - k_i - k_{i+1}),$$

whereas

$$\pi(\lambda_0|\ldots) = \text{Beta}(a + k_1, b + d_1 - k_1)$$

and

$$\pi(\lambda_n|\ldots) = \text{Beta}(1 + a + k_n, s_n - 1 + b + d_n - k_n).$$

Let us notice that the augmentation mechanism employed here can also be useful to estimate other one-dimensional diffusion processes where the corresponding transition densities have infinite series representations similar to (16).

This deals with a part of our overall model, namely the Wright-Fisher part. For the remaining part of the model, which for a given observation is given by

$$y_i|t_i, \lambda_i, \theta \sim \sum_{l=1}^{\infty} \lambda_i(1 - \lambda_i)^{l-1}\mathsf{K}(y_i|\theta_l),$$

we introduce two latent variables (s_i, v_i) and a deterministic decreasing sequence of numbers (ψ_l) for which $\{l : \psi_l > v\}$ is a known set, such that

$$y_i, v_i, s_i|\lambda_i, \theta \sim \psi_{s_i}^{-1}\mathbf{1}(v_i < \psi_{s_i})\,\lambda_i(1 - \lambda_i)^{s_i-1}\,\mathsf{K}(y_i|\theta_{s_i}).$$

In order to complete the Gibbs sampler for the model we need to describe how to sample the s_i from their full conditional and also the θ_s's. Now,

$$\pi(s_i|\ldots) \propto \psi_{s_i}^{-1}\lambda_i(1 - \lambda_i)^{s_i-1}\,\mathsf{K}(y_i|\theta_{s_i})\,\mathbf{1}(s_i \in \{l : \psi_l > v_i\})$$

and clearly the full conditional for v_i is uniform distribution on $(0, \psi_{s_i})$. Since $\{l : \psi_l > v_i\}$ is a finite set this, once more, is easy to sample.

Finally, we sample the θ_l's from

$$\pi(\theta_l|\ldots) \propto \prod_{s_i=l} \mathsf{K}(y_i|\theta_l)\,g_0(\theta_l)$$

where g_0 is the density corresponding to ν_0. Notice that in principle we would need to sample an infinite number of θ_l's, however due to the auxiliary variable v_i the Gibbs sampler algorithm only needs to consider those corresponding for the choices of s_i's. Hence, we only need to sample $(\theta_l)_{l=1}^{M}$ where $M = \max_i M_i$ and $\{1, \ldots, M_i\} = \{l : \psi_l > v_i\}$.

Hence we have all the full conditional distributions required to implement the Gibbs sampler needed for the estimation of model (14) given a discretely observed trajectory.

4. Ilustration. In this section we illustrate how the modeling scheme described above is able to capture the dynamics of continuous time phenomena. For this purpose, we will consider data coming from two time series. The first consists of 50 observations simulated from a standard Brownian motion (BM); the second of 251 daily observations (a financial year) coming from the adjusted close quotations of the S&P 500 index during the period 03.03.2008 to 27.02.2009 (the data set can be found at `http://finance.yahoo.com`).

Modelling these type of data sets is central for some applications in mathematical finance. Depending of the particular application, e.g. interest rates or asset modelling, these data are typically modeled through parametric diffusion processes. However, one could argue to what extent such restrictive assumptions are justified. For example, in the case of interest rates one could choose among many existing models, such as the Cox-Ingersoll-Ross (CIR) diffusion, the Brennan-Schwartz diffusion or the Duffie-Kan diffusion (see [1]). Adopting a nonparametric approach based on measure-valued processes provides enough flexibility to avoid such restriction.

In order to mimic some of the models typically used in financial applications, we assume the following specifications needed for model (14):

$$\theta = (m, v), \qquad \mathsf{K}(y \mid \theta) = \mathsf{N}(y \mid m, v^{-1}),$$

$$g_0(\theta) = \mathrm{N}(m; \mu, \gamma v^{-1})\,\mathrm{Ga}(v; \alpha, \beta)$$

where $\gamma, \alpha, \beta > 0$. Note that g_0 is defined on $\mathcal{X} = \mathbb{R} \times \mathbb{R}_+$. In this way the possible distributions describing the phenomenon are not limited to a single parametric family.

In general, with the MC algorithm described in Section 3, we are able make inferences about the trajectory of the whole density f_t process or the distribution of any functional process $\eta_t := \int h(y) f_t(y) \mathrm{d}y$. In particular, it is of interest the distribution of the mean functional $\bar{\eta}_t := \int y f_t(y) \mathrm{d}y$, namely the evolution of the mean, which imitate that of one-dimensional diffusions. Figure 1 shows the MC estimator of $\bar{\eta}_t$ together with its 95% high posterior density (HPD) intervals for the BM simulated dataset. Analogously, Figure 2 shows the MC estimator and the 99% HPD intervals for the estimations corresponding to the S&P 500 dataset. It is evident that the approach undertaken is able to capture the dependence induced by these datasets, and in particular drastic changes like the one observed in the S&P 500 index.

5. Discussion. We have constructed a stationary, Markovian and reversible measure-valued diffusion process which we have then used to model and estimate continuous time phenomena. The most striking feature of the constructed random process, in view of inferential applications, is the simplicity of the weights structure, which are decreasingly ordered. A consequence is that at first sight the model does not seem suitable to be used for estimating of a trajectory, but it turns out that this is the case. The key interpretation here is that a simple, and somewhat constrained, weights structure for the sequence of random probability measures, which is allowed to change over time, is sufficient for modeling purposes in this framework. The infinite location parameters at the model's disposal compensate to guarantee the necessary level of flexibility.

Acknowledgements. R.H. Mena is grateful for the support from the International Centre for Economic Research (ICER), Turin, Italy. R.H. Mena and S.G. Walker also acknowledge partial support from CONACyT, grant J50160.

References.

[1] Aït-Sahalia, Y. (1996). Nonparametric pricing of interest rate derivative securities. *Econometrica* **64**, 527–560.

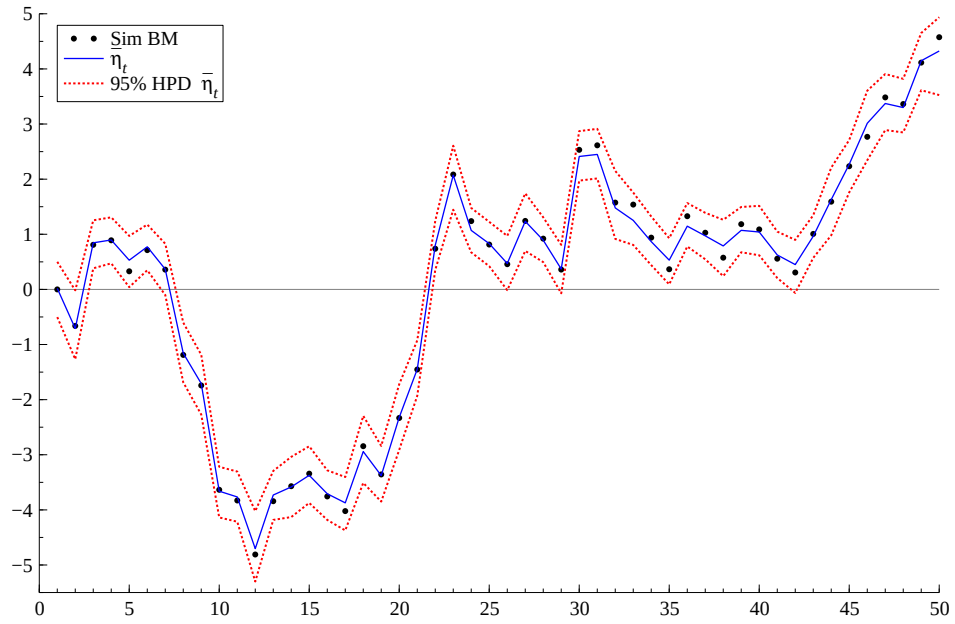

Figure 1: MC estimator for $\bar{\eta}_t$ (solid) and corresponding 95% highest posterior density intervals (dotted) for the Brownian motion simulated data set (dots). The estimates are based on 1000 iterations of the Gibbs sampler algorithm after 500 iterations of burn in.

[2] Caron, F., Davy, M., Doucet, A., Duflos, E. and Vanheeghe, P. (2006). Bayesian inference for dynamic models with Dirichlet process mixtures. In *9th IEEE International Conference on Information Fusion*, Florence, Italie.

[3] Dawson, D.A. (1993). *Measure-valued Markov processes*. Ecole d'Eté de Probabilités de Saint Flour XXI. Lecture Notes in Mathematics **1541**. Springer, Berlin.

[4] Dunson, D.B. and Park, J.H. (2008). Kernel stick-breaking processes. *Biometrika* **95**, 307–323.

[5] Ethier, S.N. and Kurtz, T.G. (1986). *Markov processes: Characterization and Convergence*. John Wiley & Sons Inc., New York.

[6] Ferguson, F. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1**, 209–230.

[7] Fuentes-García, R., Mena, R.H. and Walker, S.G. (2009). A new Bayesian nonparametric mixture model. *Submitted*.

[8] Fuentes-García, R., Mena, R.H. and Walker, S.G. (2009). A nonparametric dependent process for Bayesian regression. *Statist. Probab. Lett.* **79**, 1112–1119.

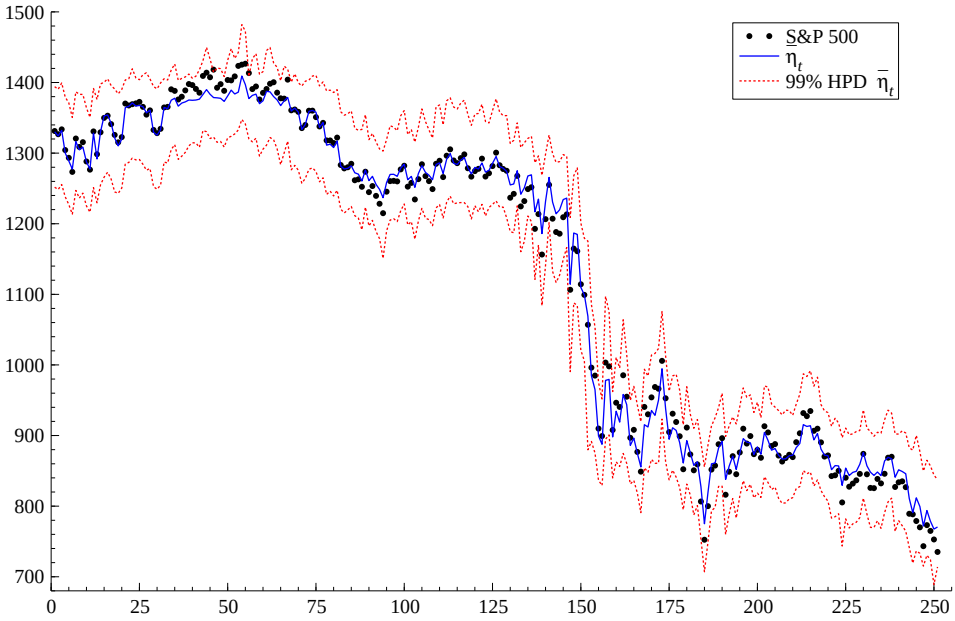

Figure 2: MC estimator for $\bar{\eta}_t$ (solid) and corresponding 99% highest posterior density intervals (dotted) for the S&P 500 data set (dots). The estimates are based on 10000 iterations of the Gibbs sampler algorithm after 2000 iterations of burn in.

[9] Gilks, W.R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Appl. Statist.* **41**, 337–348.

[10] Griffin, J.E. and Steel, M.F.J. (2006). Order-based dependent Dirichlet processes. *J. Amer. Statist. Assoc.* **101**, 179–194.

[11] Ishwaran, H. and James, L.F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.* **96**, 161–173.

[12] MacEachern, S.N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the section on Bayesian Statistical Science*, 50–55, Alexandria, VA, American Statistical Association.

[13] Mena, R.H. and Walker, S.G. (2009). On a construction of Markov models in continuous time. *Submitted*.

[14] Rodriguez, A. and Ter Horst, E. (2008). Bayesian dynamic density estimation. *Bayesian Analysis* **3**, 339–366.

[15] Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4**, 639–650, 1994.