Geometric stick-breaking processes for continuous-time Bayesian nonparametric modeling

RAMSES MENA, MATTEO RUGGIERO* AND STEPHEN G. WALKER
UNAM, University of Pavia and University of Kent

Abstract
We propose a new class of time dependent random probability measures and show how this can be used for Bayesian nonparametric inference in continuous time. By means of a nonparametric hierarchical model we define a random process with geometric stick-breaking representation and dependence structure induced via a one dimensional diffusion process of Wright-Fisher type. The sequence is shown to be a strongly stationary measure-valued process with continuous sample paths which, despite the simplicity of the weights structure, can be used for inferential purposes on the trajectory of a discretely observed continuous-time phenomenon. A simple estimation procedure is presented and illustrated with simulated and real financial data.

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 structures. A key point in Bayesian nonparametric methods is the construction of a random probability measure, the most well known of which is the Dirichlet process, introduced by Ferguson (1973). Let $X$ be a complete and separable metric space and $\mathcal{X}$ its Borel sigma-field, and define

$$P = \sum_{i=1}^{\infty} w_i \delta_{x_i}$$

where $\delta_x$ denotes a point mass at $x$, $(x_i)_{i \geq 1}$ is a vector of i.i.d. random variables from a non atomic probability measure $P_0$ on $(X, \mathcal{X})$, and $(w_i)_{i \geq 1}$ is a vector of random weights such that $\sum_{i \geq 1} w_i = 1$ almost surely. Let also $w_1$ be given by

$$w_1 = v_1, \quad w_i = v_i \prod_{j=1}^{i-1} (1 - v_j), \quad i \geq 2$$

where $v_j \sim\text{ind} \beta(a_j, b_j)$, which defines a so-called stick-breaking prior. See for example Ishwaran and James (2001). The Dirichlet process is obtained by (1)-(2) when $v_j$ are i.i.d. Beta$(1, c)$, for some $c > 0$. The above representation of the Dirichlet process is due to Sethuraman (1994), and such weights $(w_i)_{i \geq 1}$ are said to have one-parameter GEM$(c)$ distribution, named after Griffiths, Engen and McCloskey. The decreasingly ranked sequence of weights $(w(1), w(2), \ldots)$ is instead said to have Poisson-Dirichlet distribution with parameter $c$ (Kingman 1975).

*Also affiliated to Collegio Carlo Alberto, Moncalieri, Italy.
Considerable attention has been recently devoted, within the Bayesian nonparametric community, to the development of so-called dependent processes. The aim is to provide a family of random probability measures linked by a suitable dependence structure, e.g. by means of a set of covariates or a time parameter, and use it for drawing inferences on random phenomena in appropriate frameworks, usually with the aid of simulation techniques. The general idea for constructing dependent processes \( \{P_z, z \in Z\} \), where \( z \in Z \) can represent a set of covariates or a time index, is based on extensions of the stick-breaking structure in (1) of the type

\[
P_z = \sum_{i=1}^{\infty} w_i(z) \delta_{x_i(z)}
\]

where

\[
w_1(z) = v_1(z), \quad w_i(z) = v_i(z) \prod_{j=1}^{i-1} (1 - v_j(z)), \quad i \geq 2
\]

and the \( \{v_i(z), z \in Z\}_{i \geq 1} \) are independent processes.

When \( Z \equiv \mathbb{R}_+ \) and the dependence is indexed by time, then we are effectively dealing with probability-measure-valued processes. Apart from the theoretical developments offered by the probabilistic study of measure-valued processes, whose literature is certainly broad and well established (see for example Ethier and Kurtz (1986) and Dawson (1993)), their use for inference is a relatively young area and to date the most productive ideas have involved the Dirichlet process. Initiated with the papers by MacEachern (1999; 2000) who introduced the notion of dependent Dirichlet process, the current literature on the topic includes, among others: De Iorio et al. (2004) who proposed a model with an ANOVA-type dependence structure; Gelfand, Kottas, and MacEachern (2005) who apply the dependent Dirichlet process to spatial modeling by using a Gaussian process for the atoms; Griffin and Steel (2006) who let the dependence on the random masses be directed by a Poisson process; Caron et al. (2006) who model nonparametrically the noise in a dynamic linear model; Dunson and Park (2008) who construct an uncountable collection of dependent random probability measures based on a stick-breaking procedure with kernel-based weights; Rodriguez and Dunson (2010) who take \( v_i(t) = \Phi(u_i(t)) \) in (3), where \( \Phi(\cdot) \) is the standard normal cdf and \( \{u_i(t), t \geq 0\}_{i \geq 1} \) are independent Gaussian processes; Griffin and Steel (2010) who model the dependent process as a Markovian autoregression with updates incoming in a Feigin and Tweedie (1989) style. See also Dunson, Pillai, and Park (2007), Dunson, Xue, and Carin (2008), Rodriguez and Ter Horst (2008), Petrone, Guindani, and Gelfand (2009), Fuentes-García, Mena, and Walker (2009). A recent review on Bayesian nonparametric techniques can be found in the monograph edited by Hjort, Holmes, Müller, and Walker (2010). However, to the best of our knowledge, there is no current available statistical literature devoted to the study of continuously dependent random measures which both exploit the flexibility of a Bayesian nonparametric approach and enjoy desirable properties, besides Markovianity, such as reversibility or continuity of the sample paths for the constructed model. These features
are appealing in various modeling contexts and applications of stochastic processes such as finance or population genetics, where diffusion processes are typically used to model random phenomena evolving in time.

In this paper we propose the construction of a continuous-time dependent process with a simple version of weights and nice sample-path properties, and show that its relative simplicity does not affect its efficiency for nonparametric inference purposes on continuous-time random phenomena. Instead of moving from the Dirichlet process and provide a dependent process with Dirichlet marginals, we start from a somewhat simpler object and provide its time dynamics. In particular it is possible to considerably simplify the stick-breaking structure of the weights of a Dirichlet process without compromising the support of the prior, thus providing, from certain aspects, a simpler but still very useful object. The construction is suggested by a recently studied random probability measure (see Fuentes-García et al. 2009; 2010), which has proven to have some appealing features when used in Bayesian nonparametric mixture modeling and regression analysis contexts.

Hence the purpose of the present paper can be seen as twofold. In the first part we introduce a dependent process based on the notion of geometric stick-breaking random probability measure, reviewed in Section 2. The dependent process can in probabilistic terms be seen as a purely atomic continuous-time measure-valued process, denoted \( \{P_t, t \geq 0\} \). The construction is formulated in a way such that the process \( P_t \) is marginally a random probability measure of geometric stick-breaking type. The dependent structure is induced on the sequence of random measures through the set of weights, by letting the only parameter that generates the sequence depend on time. In particular this will be a two-type Wright-Fisher diffusion, briefly reviewed in Section 3. This is a one-dimensional diffusion process with state space given by the interval \([0,1]\) and is thus a natural choice for letting the \([0,1]\)-valued parameter vary with time with a strong dependence structure. Wright-Fisher diffusions enjoy nice properties which are somewhat inherited by geometric stick-breaking processes, which are formally defined in Section 4 where their properties are investigated in detail. In particular, they are shown to be Markovian and to have continuous sample-paths on the set of probability measures \( \mathcal{P}(X) \), thus being measure-valued diffusions. They also turns out to be stationary and reversible processes, and the invariant measure is completely identified.

The second part of the paper deals with estimation. We apply the construction in a framework of Bayesian nonparametric inference for continuous time random phenomena whose data have been observed discretely. Given a set of available observations \((y_1, \ldots, y_n)\), recorded at times \((t_1, \ldots, t_n)\) respectively, we assume \(y_i\) is drawn from the time-dependent nonparametric mixture model

\[
(4) \quad f_{t_i}(y) = \int_X K(y|z)P_{t_i}(dz)
\]

where \(K\) is a density function on \(X\) conditional on the parameter \(z\) and \(P_t\) is a geometric stick-breaking process. In Section 5 we develop a simple estimation procedure for drawing inferences on the trajectory of such discretely observed continuous time phenomena, by means of Markov chain Monte Carlo techniques. In Section 6 we illustrate the estimation with
simulated data and a real data set regarding financial indices. The paper is concluded with a discussion and an appendix containing all proofs.

2 Geometric stick-breaking random probability measures

Let $\mathcal{P}(\mathcal{X})$ be the set of probability measures on $(\mathcal{X}, \mathcal{B})$. A geometric stick-breaking random probability measure $P \in \mathcal{P}(\mathcal{X})$ is defined as follows:

$$
P = \sum_{i=1}^{\infty} q_i \delta_{x_i}
$$

(5)

$$
q_i = \lambda (1 - \lambda)^{i-1}, \quad i \geq 1
$$

$$
\lambda \sim \text{Beta}(a, b)
$$

$$
x_i \overset{iid}{\sim} P_0, \quad i \geq 1.
$$

Here $P$ is an almost surely discrete random probability measure with locations $(x_i)_{i \geq 1}$, sampled from a non atomic probability measure $P_0$ on $\mathcal{X}$, and associated frequencies with geometric stick-breaking structure induced by a single observation $\lambda$ from a $\text{Beta}(a, b)$ distribution, with $a, b > 0$. As for the Dirichlet process, the probability measure $P_0$ can be thought of as the prior guess for $P$, since $\mathbb{E}[P] = P_0$.

At first sight the random probability measure provided by (5) can be misinterpreted as a special case of the Dirichlet process, since the weights of the former can be obtained from (2) by letting $(v_1, v_2, \ldots)$ be the realization of the same Beta random variable, so that $\omega_i, i = 1, 2, \ldots$, are mixed geometric with $\text{Beta}(a, b)$ as the mixing distribution. However this proves not to be the case. A first and immediate reason is that in the Dirichlet case the parameter $a$ of the Beta distribution is constrained to be one. Second, and more significantly, because the Dirichlet process satisfies the distributional equation

$$
P \overset{d}{=} w_1 \delta_{x_1} + (1 - w_1)P
$$

(6)

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 Sethuraman (1994). The same procedure applied to (5) yields

$$
P \overset{d}{=} \lambda \delta_{x_1} + (1 - \lambda)P.
$$

(7)

The crucial difference between these two cases is that in (6) $P$ is independent of $(w_1, x_1)$, while in (7) $P$ is independent of $x_1$ but not of $\lambda$. Hence (5) is not a Dirichlet-type random
probability measure. An interpretation of (5) as related to the Dirichlet process is nonetheless available. This is given by taking the expectation of (2), which yields

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

and letting $\lambda = 1/(1 + c)$, where $c$ is commonly referred to as the total mass of the Dirichlet process. Hence the random probability measure (5) can be thought of as obtained by removing a level of the hierarchy from the Dirichlet process model, replacing the random weights with their expected values.

The key issue that arises when considering whether to use a random probability measure of type (5) is the impact on inference of the strong constraint on the weights in terms of flexibility. The following proposition, whose proof can be found in Ongaro and Cattaneo (2004), justifies the use of the proposed model for inferential purposes.

**Proposition 2.1.** Let $\tilde{P}$ be the probability distribution induced on $\mathcal{P}(X)$ by random probability measures $P$ of type (5). Then the support of $\tilde{P}$ in the topology of weak convergence on $\mathcal{P}(X)$ is given by all probability measures $Q \in \mathcal{P}(X)$ such that the support of $Q$ is included in that of $P_0$.

An intuitive heuristic argument behind the result can be though in the context of mixture modeling. While the geometric weights would be practically useless if we could only have one of the $x_i$’s identifying a particular cluster location, the fact that they are infinite implies that we can have an unlimited number of them supporting a location. Hence, the weights for a particular cluster are obtained via a combination of the geometric weights and the number of $x_i$’s supporting that location.

### 3 Two-type Wright-Fisher diffusions

Before proceeding with the construction of the dependent nonparametric process, we briefly review the basic notions on one-dimensional Wright-Fisher diffusions, which will be used in the subsequent section. For more details we refer the reader to Karlin and Taylor (1981) or Ethier and Kurtz (1986).

Wright-Fisher diffusions arise in biology for approximating the time evolution of the frequencies of different species in a large population. The initial need and later fame for such type of processes is mainly due to the relative mathematical simplicity of computing quantities of interest through diffusions rather than through high-dimensional discrete time processes which describe individuals explicitly. In its simplest version a Wright-Fisher diffusion is a one dimensional diffusion with state space given by the interval $[0, 1]$, whose value at every instant is the relative frequency of one of the two species in the population. The basic evolutionary mechanisms that a Wright-Fisher diffusion can describe are known in the biological literature as mutation and random genetic drift, the latter being also referred to as resampling.
More formally, a one-dimensional Wright-Fisher diffusion is the solution of the stochastic differential equation on \([0, 1]\) given by
\[
\frac{dX_t}{dt} = \mu(X_t) \, dt + \sigma(X_t) \, dB_t
\]
where \(B_t\) is a standard Brownian motion and the coefficients are given by
\[
\mu(x) = \alpha(1 - x) - \beta x, \quad \sigma(x) = \sqrt{x(1 - x)}
\]
with \(\alpha, \beta > 0\). Letting \(x\) denote the frequency of species 1, \(\alpha\) and \(\beta\) are the intensity of a mutation from type 2 to type 1 and viceversa respectively. Note that 0 and 1 are reflecting barriers. Intuitively, when the process reaches any of the two, the diffusion coefficient drops to zero and the drift drives it instantaneously towards the interior of the state space. With null mutation rates \(\alpha\) and \(\beta\), the interval extremes 0 and 1 are absorbing barriers which determine, once touched, the so-called fixation of one of the two species and extinction of the other.

For non-null mutation rates, the diffusion driven by (8) is stationary and reversible with respect to a Beta\((2\alpha, 2\beta)\) distribution, with density with respect to the Lebesgue measure
\[
f(x) = \frac{\Gamma(2(\alpha + \beta))}{\Gamma(2\alpha)\Gamma(2\beta)} x^{2\alpha-1}(1 - x)^{2\beta-1}, \quad x \in [0, 1].
\]
It is therefore intuitive to understand why a two-type Wright-Fisher diffusion is a natural choice for modeling the time-dependence of the parameter \(\lambda\) which generates the sequence of geometric weights in (5). The formalization of this intuition is the object of the following section.

### 4 Geometric stick-breaking processes

Let \(\mathcal{P}(\mathcal{X})\) be the set of Borel probability measures on the measurable space \((\mathcal{X}, \mathcal{F})\).

**Definition 4.1.** Let \(\{\lambda_t, t \geq 0\}\) be a two-type Wright-Fisher diffusion with mutation, driven by the stochastic differential equation
\[
\frac{d\lambda_t}{dt} = \left[\frac{a}{2}(1 - \lambda_t) - \frac{b}{2}\lambda_t\right] \, dt + \sqrt{\lambda_t(1 - \lambda_t)} \, dB_t
\]
where \(B_t\) is a standard Brownian motion, and let \(P_0\) be a non atomic probability measure on \((\mathcal{X}, \mathcal{F})\). A geometric stick-breaking process with parameters \(a, b > 0\) and \(P_0\) is a random process \(\{P_t, t \geq 0\}\) taking values in \(\mathcal{P}(\mathcal{X})\) defined at each \(t \geq 0\) by
\[
P_t = \sum_{i=1}^{\infty} q_{i,t} \, \delta_{x_i}
\]
(10)
\[
q_{i,t} = \lambda_t(1 - \lambda_t)^{i-1}, \quad i \geq 1
\]
\[
x_i \overset{iid}{\sim} P_0, \quad i \geq 1
\]
and denoted \(GSB(a, b, P_0)\).
As anticipated, a geometric stick-breaking process is a time-dependent random probability measure with series representation whose locations are i.i.d. and all distinct and whose weights are obtained by sequentially breaking a unit-length stick geometrically. Moreover the weights are themselves random processes driven by a unique one-dimensional diffusion on $[0,1]$ of Wright-Fisher type. Before illustrating how an object of this type can be used for Bayesian nonparametric inference in continuous-time, in the remainder of this section we investigate the properties of such process. As can be inferred intuitively from inspection of (10), the GSB process inherits nice properties from the Wright-Fisher diffusion.

Assume for simplicity that the complete and separable space $X$ is compact, for we are going to consider continuous functions on $X$. The compactness requirement is not restrictive as we could also take $X$ to be locally compact and consider continuous functions on $X$ which vanish at infinity. Let $\mathcal{P}_g(X) \subset \mathcal{P}(X)$ be the set of purely atomic probability measures with geometric weights as in (5), and denote $C_{\mathcal{P}_g(X)}([0,\infty))$ the space of continuous functions from $[0,\infty)$ to $\mathcal{P}_g(X)$. The following propositions are essentially proved by using the infinitesimal generator of the GSB($a,b,P_0$) process, whose identification is deferred to the Appendix due to its technical nature. See Proposition A.1.

The first proposition states that \{\Pt, t \geq 0\} is a purely atomic measure-valued diffusion.

**Proposition 4.2.** Let $X$ be a compact complete and separable space, and let \{\Pt, t \geq 0\} be a GSB($a,b,P_0$) process on $\mathcal{P}_g(X)$. Then \{\Pt, t \geq 0\} is a Feller process with sample paths in $C_{\mathcal{P}_g(X)}([0,\infty))$.

Hence the realizations of a GSB process are continuous functions from $[0,\infty)$ to the set of probability measures. The continuity of paths of the one-dimensional diffusion \{\lambda_t, t \geq 0\} has been somewhat transferred to the process of measures \{\Pt, t \geq 0\}. Again, this is quite intuitive since the locations $(x_i)_{i \geq 1}$ are random but fixed across time. An immediate observation would be that letting the $x_i$’s also vary leads to a more flexible model. While this is certainly true, we have two reasons for keeping the $x_i$’s fixed. On the probabilistic side this would most likely tear apart the nice properties this process enjoys. But more importantly, in view of Proposition 2.1 and as the illustration in the following sections will show, for inferential purposes this is not even needed.

We now turn to stationarity. Denoted $x = (x_i)_{i \geq 1}$, in order to emphasize the dependence on $\lambda$ define, conditionally on $x$, the continuous map $\phi_x : [0,1] \to \mathcal{P}_g(X)$ by

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

Recall that the vector of weights associated to the locations $(x_1, x_2, \ldots)$ is always disposed in decreasing order. Hence, defining $g_x(P) = P(\{x_1\}) = \lambda$, we can set $\phi_x^{-1} = g_x$. Denote also by $B(a,b)$ the Beta distribution with parameters $a,b$ and let $\tilde{B}(a,b) = B(a,b) \circ \phi_x^{-1}$.

**Proposition 4.3.** Let \{\Pt, t \geq 0\}_{t \geq 0} be a GSB($a,b,P_0$) process. Then \{\Pt, t \geq 0\}_{t \geq 0} is reversible with respect to $\tilde{B}(a,b)$. 

Recall that reversibility implies that the time-reversal process has the same finite-dimensional distributions of the original process. An immediate consequence of time reversibility is given by the following corollary, stated without proof.

**Corollary 4.4.** Let \( \{P_t, t \geq 0\}_{t \geq 0} \) be a \( \text{GSB}(a,b,P_0) \) process. Then \( \{P_t, t \geq 0\}_{t \geq 0} \) has invariant measure given by \( \tilde{\mathcal{B}}(a,b) \).

Thus a GSB process exhibits smooth realizations and is strongly stationary, with its invariant measure completely identified. Although the infinitesimal generator (A.21) characterizes all the dependence properties inherent to the GSB processes, in the following proposition we depict an explicit expression for the autocorrelation of \( P_t \). To this end, denote \( d_\theta^n(t) = \mathbb{P}\{D_t = n\} \), where \( \{D_t, t \geq 0\} \) is a pure death process such that \( \mathbb{P}\{D_0 = \infty\} = 1 \), and death events occur at rate \( n(\theta + n - 1)/2 \). Denote also by

\[(12) \quad (x)_y = x(x+1)\ldots(x+y-1)\]

the Pochhammer symbol. Tavaré (1984), for example, computed that

\[d_0(t) = 1 - \sum_{m=1}^{\infty} (-1)^{m-1} \frac{(\theta + 2m - 1)\theta(m-1)}{m!} e^{-m(\theta+m-1)t/2}\]

and for \( n \in \mathbb{N} \)

\[d_\theta^n(t) = \sum_{m=n}^{\infty} (-1)^{m-n} \binom{m}{n} \frac{(\theta + 2m - 1)(\theta + n)(m-1)}{m!} e^{-m(\theta+m-1)t/2}.

**Proposition 4.5.** Let \( \{P_t, t \geq 0\}_{t \geq 0} \) be a \( \text{GSB}(a,b,P_0) \) process. For \( t,s > 0 \) and \( A \in \mathcal{X} \)

\[\text{Corr}(P_t(A), P_{t+s}(A)) = \frac{k_s}{\sum_{i \geq 1} \kappa_{i,1,i}}\]

where

\[k_s = \sum_{n=0}^{\infty} d^{a+b}_n(s) \sum_{k=1}^{n} \binom{n}{k} (a+k) \sum_{i=1}^{\infty} \frac{(b+n-k)(i-1)}{(a+b+n)i} \kappa_{n,k,i},\]

and

\[\kappa_{h,j,l} = \mathbb{E}(\lambda_t^{j+1}(1-\lambda_t)^{h+j+l-1})\]

Note that the stationarity of the Wright-Fisher diffusion implies that \( \kappa_{h,j,l} \) does not depend on \( t \).

The next section shows 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.
5 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

\[
  f(t, y) = \int_X K(y|z)P_t(dz) = \sum_{l=1}^{\infty} \lambda_t (1 - \lambda_t)^{l-1} K(y | \theta_l)
\]

where \(P_t\) is of type (10), \(K(\cdot | \theta)\) is a well defined density function for all \(\theta\), \(\{\lambda_t, t \geq 0\}\) 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 \(P_0\). The non parametric mixture model (13) can alternatively be written in hierarchical form

\[
  y_i | t_i, x_i \sim K(\cdot | \theta_i)
\]

\[
  \{\theta_i\} \sim P_t
\]

\[
  P_t \sim GSB(a, b, P_0).
\]

where for simplicity we have set \(x_i = x_{t_i}\).

In order to estimate the model we undertake a Gibbs sampler algorithm scheme borrowing some slice sampler techniques to overcome infinite summations in both the random probability measure and the series representation for the Wright-Fisher transition density.

It is convenient to start by considering the part of the model related to the Wright-Fisher diffusion \(\{\lambda_t, t \geq 0\}\). Assume we have observations \((t_i, s_i)_{i=1}^n\), where \(s_i\) is a latent random variable that indicates which \(K(\cdot | \theta)\) the observation \(y_i\) comes from. The model is then written as follows:

\[
  s_i | \lambda_t \sim \text{Geom}(\lambda_t)
\]

with \(\lambda_i = \lambda_{t_i}\) and

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

where

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

\((a + b)_m\) reads as (12) and \(\tau_i = t_i - t_{i-1}\). Also,

\[
  D(\lambda_i | m, \lambda_{i-1}) = \sum_{k=0}^{m} \text{Beta}(\lambda_i | a + k, b + m - k) \text{Bin}(k | m, \lambda_{i-1}).
\]
Here, equation (15) corresponds to a representation of the transition density of a general class of Beta-Binomial diffusion processes, which are the solution of the SDE

\[ d\lambda_t = \left[ \frac{c}{a + b - 1} (a - (a + b)\lambda) \right] dt + \sqrt{\frac{2c}{a + b - 1} \lambda(1 - \lambda)} \, dB_t \]

and include the Wright-Fisher diffusion as a special case. See Mena and Walker (2009) 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 by letting \( c = (a + b - 1)/2 \) leads back to the Wright-Fisher model driven by (9).

To simplify and accommodate a Gibbs sampler, and more specifically to avoid the infinite summations needed for (15), we need to introduce a further set of 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

\[ 1(u_i < g(d_i)) \frac{p_2(d_i)}{g(d_i)} \text{Beta}(\lambda_i|a + k_i, b + d_i - k_i) \text{Bin}(k_i|d_i, \lambda_{i-1}), \]

where \( g \) is a decreasing function with known inverse. See Kalli et al. (2010) for more discussion and use of this slice idea. Integrating out the latent variables clearly yields \( p(\lambda_i|\lambda_{i-1}) \). Hence, the likelihood function with the complete data is

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

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

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

Here \( C \) is the normalizing which does not depend 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 (ARS) of Gilks and Wild (1992). The full conditional for \( b \) follows similarly and is also log-concave. The full conditional for \( c \) can be written as

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

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

\[
\pi(k_i|\ldots) \propto \left( \frac{d_i}{k_i} \right) 1(k_i \in \{0, 1, \ldots, d_i\}) \frac{\lambda_i \lambda_{i-1}}{\Gamma(a + k_i) \Gamma(b + d_i - k_i)} \left\{ \frac{\lambda_i \lambda_{i-1}}{(1 - \lambda_i)(1 - \lambda_{i-1})} \right\}^{k_i}.
\]
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 a uniform distribution on $(0, g(d_i))$, where $g$ is chosen at convenience, for example $g(d) = e^{-d}$ or $g(d) = d^{-2}$, so that $g^{-1}$ is known. The benefit of this is apparent when we consider the full conditional for $d_i$. This is indeed given by

$$
\pi(d_i \mid \ldots) \propto \frac{p_i(d_i)}{g(d_i)} \left\{ \Gamma(a + b + d_i) \right\} \left\{ \Gamma(b + d_i - k_i) \right\} \left\{ (1 - \lambda_{i-1})(1 - \lambda_i) \right\}^{d_i} \mathbf{1}(k_i \leq d_i \leq g^{-1}(u_i))
$$

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

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

$$(17) \quad \pi(\lambda_i \mid \ldots) = \text{Beta}(1 + a + k_i + k_{i+1}, s_i - 1 + b + d_i + d_{i+1} - k_i - k_{i+1}),$$

whereas

$$(18) \quad \pi(\lambda_0 \mid \ldots) = \text{Beta}(a + k_1, b + d_1 - k_1)$$

and

$$(19) \quad \pi(\lambda_n \mid \ldots) = \text{Beta}(1 + a + k_n, s_n - 1 + b + d_n - k_n).$$

Note 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 (15). This deals with the part of the model related to the Wright-Fisher process.

For the remaining part of the model, which for a given observation is given by

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

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

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

Namely, we slice the infinite summation again. 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 $\theta_l$’s. Now,

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

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

$$\pi(\theta_l \mid \ldots) \propto \prod_{s_i = l} K(y_i \mid \theta_l) g_0(\theta_l).$$
where $g_0$ is the density corresponding to $P_0$. Notice that in principle we would need to sample an infinite number of $\theta_l$'s. However, due to the auxiliary variable $v_i$, the Gibbs sampler algorithm only needs to consider those corresponding to the choices of the $s_i$'s. Hence, we only need to sample \((\theta_l)_{l=1}^M\) where $M = \max_i M_i$ and \(\{1, \ldots, M_i\} = \{l : \psi_l > v_i\}\).

We thus have all the full conditional distributions required to implement the Gibbs sampler needed for the estimation of model (13) given a discretely observed trajectory. The following algorithm summarizes the procedure.

### Algorithm

1. Select $g(\cdot)$ and $\psi(\cdot)$ functions, e.g. $g(x) = \psi(x) = e^{-x}$

2. Set initial values for:
   - Wright-Fisher diffusion parameters \((a_0, b_0, c_0)\)
   - Parameters in the Kernel $K$ and possibly in $P_0$, e.g. $\theta^0$
   - Latent variables needed to overcome infinite summations, \((u^0_i, s^0_i, k^0_i, d^0_i)_{i=1}^n\). For these an initial value for the augmented random probability measure is also needed, e.g. $M^0 = 20$
   - Use these values to initiate $\lambda^0 = (\lambda^0_i)_{i=0}^n$

then for $j = 1, \ldots, I$

3. Update $v^j = (v^j_i)_{i=0}^n$, i.e. $v^j_i \sim U[0, \psi_{s^{j-1}}]$, and compute
   \[ M^j = \max M^j_i \text{ with } \{1, \ldots, M^j_i\} = \{l : \psi(l) > v^j_i\} \]

4. Update $\lambda^j = (\lambda^j_i)_{i=0}^n$, $\theta^j = (\theta^j_l)_{l=1}^M$, and $(u^j_i, s^j_i, k^j_i, d^j_i)_{i=1}^n$ using the corresponding full conditionals

5. Update $(a_j, b_j, c_j)$, e.g. via ARS algorithm

The $I$ iterations can then be used to build a Monte Carlo estimator $f_t$ of any desired functional.

### 6 Illustration

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 series consists of 251 daily observations (corresponding to 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 on 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 Ait-Sahalia (1996)). Adopting a nonparametric approach based on measure-valued processes provides enough flexibility to avoid such committing assumptions.

In order to mimic some of the models typically used in financial applications, we assume the following specifications needed for model (13):

\[
\theta = (m, v), \quad K(y \mid \theta) = N(y \mid m, v^{-1}),
\]

\[
g_0(\theta) = N(m; \omega, \gamma v^{-1}) \text{Ga}(v; \alpha, \beta)
\]

where \(\gamma, \alpha, \beta > 0, \omega \in \mathbb{R},\) with \(g_0\) defined on \(X = \mathbb{R} \times \mathbb{R}_+\). In this way the possible distributions describing the phenomenon are not limited to a single parametric family. The well-known result that mixtures of a sufficiently large number of normal densities can be used to approximate any smooth density accurately makes the time-dependent nonparametric mixture (4) an appropriate and flexible tool to be used in this framework.

In general, with the MC algorithm described in Section 5, we are able to draw inferences on the trajectory of the whole density \(f_t\) process or the distribution of any functional process \(\int h(y) f_t(y) dy\). In particular, it is of interest the distribution of the mean functional \(\bar{\eta}_t := \int y f_t(y) dy\), namely the evolution of the mean, which imitate that of one-dimensional diffusion process. Figure 1 shows the MCMC estimate of \(\bar{\eta}_t\), together with its 95% high posterior density (HPD) intervals, for the BM simulated dataset. Analogously, Figure 2 shows the MCMC estimate (heat contours) for the density process, \(\hat{f}_t\), and the corresponding mean of the functional \(\bar{\eta}_t\) (solid line) for the S&P 500 data set (points). For both data sets the choice of hyper-parameters was \(\omega = 0, \gamma = 1000\) and \((\alpha, \beta) = (0, 10, 1)\). This choice was made so the inverse gamma distribution corresponding to \(g_0\) has finite expectation and in a way that \(\gamma(\beta/(\alpha - 1))\) is large enough to account for a large variance in the Gaussian kernel. For visualization purposes a smaller portion of this latter estimate is displayed in Figure 3.

For the BM dataset estimations, 10000 effective iterations after 2000 iterations of burn in were used. In the case of the S&P 500 dataset, a higher number of iterations, 100000, after a 20000 of burn in (thinned each 10), were required to attain a satisfactory convergence of the sampler. Figure 4 shows the Markov chains corresponding to parameters \((a, b)\) and the corresponding posterior densities, the results corresponding to \(c\) are proportional to those corresponding to \(b\) for the S&P 500 data set. A standard convergence analysis was performed, in particular the Gelman and Rubin (1992) visual test and the Raftery and Lewis (1992) diagnosis test were satisfactory.

It is apparent that the approach, here undertaken, is able to capture the dependence
induced by these datasets. Furthermore the model adapts well to drastic changes like that observed in the S&P 500 index.

7 Discussion

The paper proposes a new class of dependent processes which exhibit nice sample path properties and can be used to model and estimate continuous time phenomena in a Bayesian nonparametric setting. 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. At first sight the model does not seem suitable to be used for the estimation 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. In fact the support of the prior is the same even with the apparent more general stick breaking weights (2). This is due to the infinite collection of \( \theta \)'s being able to control the weights at locations by collecting sufficient quantities of them in the neighborhood of a particular point.

Acknowledgements

The authors wish to thank three anonymous referees for their comments which substantially improved the manuscript. R.H. Mena and S.G. Walker are grateful for the support of the International Centre for Economic Research, Turin and CONACyT, grant J50160. M. Ruggiero is partially supported by MIUR.
Appendix

The following technical result identifies the infinitesimal generator of the process. For $m \in \mathbb{N}$ and $f \in C(\mathbb{X}^m)$, the space of continuous functions on $\mathbb{X}^m$, let

\begin{equation}
\varphi_m(P) = \varphi_m(\phi_x(\lambda)) = \langle f, P^m \rangle
\end{equation}

where $\langle f, P \rangle = \int f \, dP$ and $P^m = P \times \ldots \times P$ is an $m$-fold product measure. Let $C_P(\mathcal{P}_g(\mathbb{X}))$ be the subalgebra of $C(\mathcal{P}_g(\mathbb{X}))$ given by the linear span of monomials of type (A.20) when $P \in \mathcal{P}_g(\mathbb{X})$.

Proposition 7.1. Let $\{P_t, t \geq 0\}$ be a GSB$(a, b, P_0)$ process. Then $\{P_t, t \geq 0\}$ has infinitesimal generator

$$
\mathcal{B}\varphi_m(P) = \left(\frac{a}{2}(1 - \lambda) - \frac{b}{2}\lambda\right) \sum_{i_1, \ldots, i_m \geq 1} f(x_{i_1}, \ldots, x_{i_m}) \frac{\partial}{\partial \lambda} h(\lambda; m, i_1, \ldots, i_m)
$$
Figure 4: MCMC iterations and posterior densities for parameters \((a, b)\). The estimates are based on 10000 effective iterations, drawn from 100000 iterations thinned each 10 iterations, of the Gibbs sampler algorithm after 20000 iterations of burn in.

\[
(A.21) \quad + \frac{1}{2} \lambda(1 - \lambda) \sum_{i_1, \ldots, i_m \geq 1} f(x_{i_1}, \ldots, x_{i_m}) \frac{\partial^2}{\partial \lambda^2} h(\lambda; m, i_1, \ldots, i_m)
\]

with domain

\[
\mathcal{D}(\mathcal{B}) = \left\{ \varphi \in C(\mathcal{P}(\mathbb{X})) : \varphi = \varphi_m(P) = (f, P^m), \; f \in C(\mathbb{X}^m), \; m \in \mathbb{N} \right\}
\]
and where
\[ h(\lambda; m, i_1, \ldots, i_m) = \lambda_i^m (1 - \lambda_i) \sum_{j=1}^{m} i_j - m. \]

The subalgebra of \( C(\mathcal{P}_g(\mathbb{X})) \) given by the linear span of monomials of type (A.20) when \( P \in \mathcal{P}_g(\mathbb{X}) \) is a core for \( \mathcal{B} \).

**Proof.** From (A.20) we can write
\[
A \varphi_m(P_t) = \sum_{i_1 \geq 1} \sum_{i_m \geq 1} \lambda_i^m (1 - \lambda_i)^{i_1 + \ldots + i_m - m} f(x_{i_1}, \ldots, x_{i_m}).
\]

It follows that the operator semigroup \( \{T(t), t \geq 0\} \) for the process \( \{P_t, t \geq 0\} \) is such that for every \( t \geq 0 \)
\[
T(t)\varphi_m(P_0) = \sum_{i_1, \ldots, i_m \geq 1} \int_{[0,1]} \lambda_i^m (1 - \lambda_i)^{i_1 + \ldots + i_m - m} p(d\lambda_i|\lambda_0) f(x_{i_1}, \ldots, x_{i_m})
\]

where \( P_0 = P(0) \) is the initial value of the process and \( p(d\lambda_i|\lambda_0) \) 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
\[
A = \left[ a \frac{1}{2} (1 - \lambda) - b \frac{1}{2} \right] \frac{d}{d\lambda} + \frac{1}{2} \lambda (1 - \lambda) \frac{d^2}{d\lambda^2}.
\]

defined on \( C^2([0,1]) \), it can be easily seen that
\[
t^{-1} [T(t)\varphi_m(P_0) - \varphi_m(P_0)]
\]
converges strongly, as \( t \downarrow 0 \), to
\[
\sum_{i_1, \ldots, i_m \geq 1} f(x_{i_1}, \ldots, x_{i_m}) A h(\lambda; m, i_1, \ldots, i_m)
\]

where
\[ h(\lambda; m, i_1, \ldots, i_m) = \lambda_i^m (1 - \lambda_i) \sum_{j=1}^{m} i_j - m. \]

In view of (A.24), from (A.25) the generator for \( \{P_t, t \geq 0\} \) in terms of \( \{\lambda_t, t \geq 0\} \) can be written
\[
B \varphi_m(P) = \left( a \frac{1}{2} (1 - \lambda) - b \frac{1}{2} \right) \sum_{i_1, \ldots, i_m \geq 1} h_1(\lambda; m, i_1, \ldots, i_m) f(x_{i_1}, \ldots, x_{i_m})
\]
\[
+ \frac{1}{2} \lambda (1 - \lambda) \sum_{i_1, \ldots, i_m \geq 1} h_2(\lambda; m, i_1, \ldots, i_m) f(x_{i_1}, \ldots, x_{i_m})
\]

where \( h_1(\lambda; m, i_1, \ldots, i_m) \) and \( h_2(\lambda; m, i_1, \ldots, i_m) \) are the first and second derivatives of \( h(\lambda; m, i_1, \ldots, i_m) \) with respect to \( \lambda \). From Dawson (1993), Lemma 2.1.2, it follows that \( C_P(\mathcal{P}_g(\mathbb{X})) \) is dense in \( C(\mathcal{P}_g(\mathbb{X})) \), so that \( C_P(\mathcal{P}_g(\mathbb{X})) \) is a core for \( \mathcal{B} \), and we can take the domain of \( \mathcal{B} \) to be \( C(\mathcal{P}_g(\mathbb{X})) \). \( \square \)
Proof of Proposition 4.2
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 \( C(\mathcal{P}_g(X)) \) by

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

Then Theorem 4.2.7 of Ethier and Kurtz (1986) implies that for every \( Q \in \mathcal{P}_g(X) \), there exists a Markov process \( \{P_t, t \geq 0\} \) corresponding to \( \{T(t), t \geq 0\} \) with initial distribution \( Q \) and sample paths in \( D_{\mathcal{P}_g(X)}([0,\infty)) \), the space of càdlàg functions from \([0,\infty)\) to \( \mathcal{P}_g(X) \).

Denote now with \( P(t,P,dQ) \) the transition function corresponding to the semigroup \( \{T(t), t \geq 0\} \), that is

\[
T(t)\varphi(P) = \int_{\mathcal{P}(X)} \varphi(Q) P(t,P,dQ).
\]

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

\[
(A.26) \quad t^{-1}P(t,P,B(P,\varepsilon)^c) \to 0 \quad \text{as } t \to 0
\]

where \( B(P,\varepsilon)^c \) is the complement of an \( \varepsilon \)-neighborhood of \( P \) in a topology which makes \( \mathcal{P}_g(X) \) locally compact and separable. See Ethier and Kurtz (1986), Lemma 4.2.9. For our purposes it is enough to show that the sample paths are continuous in \( \mathcal{P}_{g,x}(X) \), defined to be the restriction of \( \mathcal{P}_g(X) \) 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 \( P \in \mathcal{P}_{g,x}(X) \). Thus it suffices to show that \((A.26)\) holds for every \( P \in \mathcal{P}_{g,x}(X) \) and every \( \varepsilon > 0 \), where \( B(P,\varepsilon)^c = B_{W}(P,\varepsilon)^c \) is the complement of, say, an \( \varepsilon \)-neighborhood of \( P \) in the weak topology. Observe now that \( \mathcal{P}_{g,x}(X) \) is locally compact since, from the continuity of \((11)\), for every \( P \in \mathcal{P}_{g,x}(X) \) one can find a compact neighborhood of \( P \) in \( \mathcal{P}_{g,x}(X) \) by letting \( \lambda \) vary smoothly. Also \( \mathcal{P}_{g,x}(X) \) is separable, since the set \( \{Q \in \mathcal{P}_{g,x}(X) : \lambda \in \mathbb{Q} \cap [0,1] \} \) is countable and dense in \( \mathcal{P}_{g,x}(X) \). The key now is the fact that in \((A.23)\) the transition function providing \( \{T(t), t \geq 0\} \) is expressed in terms of \( p(\lambda \varepsilon | \lambda_0) \). That is, if \( P = \sum_{i \geq 1} \lambda_0(1 - \lambda_0)^{i-1}\delta_{x_i} \) and \( Q = \sum_{i \geq 1} \lambda_t(1 - \lambda_t)^{i-1}\delta_{y_i} \), then \( P(t,P,dQ) = p(\lambda \varepsilon | \lambda_0) \prod_{i=1}^{\infty} \delta_{x_i}(dy_i) \). Hence for every \( \varepsilon > 0 \) there exists a \( \delta > 0 \) such that

\[
P(t,P,B_{W}(P,\varepsilon)^c) = p(B(\lambda_0, \delta)^c | \lambda_0) \prod_{i=1}^{\infty} \delta_{x_i}(dy_i)
\]

Since \( \{\lambda_t, t \geq 0\} \) is a diffusion, we have

\[
t^{-1}p(B(\lambda_0, \delta)^c | \lambda_0) \to 0, \quad t \to 0.
\]

Since also the product is bounded by one, \((A.26)\) follows. \(\square\)

Proof of Proposition 4.3
Given \((x_i)_{i \geq 1}\), from a result in Fukushima and Stroock (1986) it is enough to show that for
every $t \geq 0$ and every $f, g \in C(\mathcal{P}_g(\mathcal{X}))$ the following holds

$$\int_{\mathcal{P}_g(\mathcal{X})} fT(t)g \, d\tilde{B}(a, b) = \int_{\mathcal{P}_g(\mathcal{X})} gT(t)f \, d\tilde{B}(a, b).$$

Let $\{S(t), t \geq 0\}$ and $\{T(t), t \geq 0\}$ be as in Proposition 4.2, and note that $f \circ \phi_x$ and $g \circ \phi_x$ are in $C([0,1])$. Then we have

$$\int_{\mathcal{P}_g(\mathcal{X})} fT(t)g \, d\tilde{B}(a, b) = \int_{\mathcal{P}_g(\mathcal{X})} f[S(t)(g \circ \phi_x)] \circ \phi_x^{-1} \, d(B(a, b) \circ \phi_x^{-1})$$

$$= \int_{[0,1]} (f \circ \phi_x)[S(t)(g \circ \phi_x)] \, dB(a, b)$$

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

**Proof of Proposition 4.5**

We have

$$\mathbb{E}(P_t(A)P_{t+s}(A)) = \mathbb{E}\left(\sum_{i \geq 1} q_{i,t} \delta_{x_i}(A) \sum_{j \geq 1} q_{j,t+s} \delta_{x_j}(A)\right)$$

$$= \mathbb{E}\left(\sum_{i \geq 1} q_{i,t}q_{i,t+s} \delta_{x_i}(A) + \sum_{i \geq 1} \sum_{j \geq 1, j \neq i} q_{i,t}q_{j,t+s} \delta_{x_i}(A) \delta_{x_j}(A)\right)$$

$$= k_{t,s}P_0(A) + (1 - k_{t,s})P_0^2(A)$$

where

$$k_{t,s} = \mathbb{E}\left(\sum_{i \geq 1} q_{i,t}q_{i,t+s}\right) = \sum_{i \geq 1} \mathbb{E}(q_{i,t}q_{i,t+s})$$

and $1 - k_{t,s}$ is obtained by subtraction, since

$$\sum_{i \geq 1} q_{i,t} \sum_{j \geq 1} q_{j,t+s} = \sum_{i \geq 1} q_{i,t}q_{i,t+s} + \sum_{i \geq 1} \sum_{j \geq 1, j \neq i} q_{i,t}q_{j,t+s} = 1,$$

from which

$$\text{Cov}(P_t(A), P_{t+s}(A)) = k_{t,s}P_0(A)(1 - P_0(A)).$$

Analogously one finds that

$$\text{Var}(P_t(A)) = P_0(A)(1 - P_0(A)) \sum_{i \geq 1} \mathbb{E}(\lambda_t^2(1 - \lambda_t)^{2(i-1)}).$$

Note that $\text{Var}(P_t(A))$ does not depend on $t$ due to the stationarity of the Wright-Fisher diffusion. Furthermore we have

$$\mathbb{E}(q_{i,t}q_{i,t+s}) = \mathbb{E}\left[\mathbb{E}(q_{i,t}q_{i,t+s} | \lambda_t)\right]$$
where $p(d\lambda_t|\lambda_0)$ is the transition function of the one-dimensional Wright-Fisher process. Using Corollary 1.9 in Ethier and Griffiths (1993), this equals

$$
\mathbb{E}\left[ q_{i,t} \int_0^1 \lambda_{t+s}(1 - \lambda_{t+s})^{i-1} p(d\lambda_{t+s} | \lambda_t) \right]
$$

which does not depend on $t$ due to the stationarity of the Wright-Fisher diffusion, so that $k_{t,s} = k_s$. Summing over $i$ gives the result. □

References


