Exact simulation of the Wright-Fisher diffusion*

Paul A. Jenkins† & Dario Spanò
Department of Statistics, University of Warwick

Abstract

The Wright-Fisher family of diffusion processes is a class of evolutionary models widely used in population genetics, with applications also in finance and Bayesian statistics. Simulation and inference from these diffusions is therefore of widespread interest. However, simulating a Wright-Fisher diffusion is difficult because there is no known closed-form formula for its transition function. In this article we demonstrate that it is in fact possible to simulate *exactly* from the scalar Wright-Fisher diffusion with *general* drift, extending ideas based on retrospective simulation. Our key idea is to exploit an eigenfunction expansion representation of the transition function. This approach also yields methods for exact simulation from several processes related to the Wright-Fisher diffusion: (i) its moment dual, the ancestral process of an infinite-leaf Kingman coalescent tree; (ii) its infinite-dimensional counterpart, the Fleming-Viot process; and (iii) its bridges. Finally, we illustrate our method with an application to an evolutionary model for mutation and diploid selection. We believe our new perspective on diffusion simulation holds promise for other models admitting a transition eigenfunction expansion.

1 Introduction

Monte Carlo simulation of diffusion processes is of great interest, as it underlies methods of statistical inference from discrete observations in a variety of applications (e.g. Golightly and Wilkinson, 2006, 2008; Chib et al., 2010; Bladt and Sørensen, 2014; Bladt et al., 2014). Our interest in this paper is in the *Wright-Fisher* diffusion. This process is widely used for inference, especially in genetics, where it serves as a model for the evolution of the frequency $X_t \in [0,1]$ of a genetic variant, or *allele*, in a large randomly mating population. If there are two alternative alleles then the diffusion obeys a one-dimensional stochastic differential equation (SDE)

$$dX_t = \beta(X_t)dt + \sqrt{X_t(1-X_t)}dB_t, \quad X_0 = x_0,\ t \in [0,T]. \tag{1}$$

MSC 2000 subject classifications: Primary 65C05, 62M05; Secondary 60H35, 60J60, 92D15. *Keywords and phrases*: Monte Carlo methods, Simulation, Wright-Fisher diffusion, exact algorithm, Fleming-Viot process, diffusion bridge, retrospective simulation, Kingman's coalescent, population genetics.

†Supported in part by EPSRC Research Grant EP/L018497/1.

The drift coefficient, $\beta : [0, 1] \to \mathbb{R}$, can encompass a variety of evolutionary forces. For example,

$$\beta(x) = \frac{1}{2}[\theta_1(1 - x) - \theta_2 x] + \sigma x(1 - x)[x + h(1 - 2x)], \tag{2}$$

describes a process with recurrent mutation between the two alleles, governed by parameters θ_1 and θ_2, and with (diploid) natural selection causing fitness differences between individuals with different numbers of copies of the allele, governed by parameters σ and h. There is much interest among geneticists in inference from this and related diffusions (e.g. Williamson et al., 2005; Bollback et al., 2008; Gutenkunst et al., 2009; Malaspinas et al., 2012), and in the characteristics of the trajectories themselves (Schraiber et al., 2013; Zhao et al., 2013), as discretely observed data are becoming more and more available (for example, as genetic time series data for ancient human DNA and for viral evolution within hosts). Beyond genetics, Wright-Fisher diffusions have been considered for applications in several other fields. For example, in finance they have been used as models for time-evolving regime probability, discount coefficients or state price (e.g. Delbaen and Shirakawa, 2002; Gourieroux and Jasiak, 2006); they have been studied as hidden Markov signals in filtering problems (Chaleyat-Maurel and Genon-Catalot, 2009; Papaspiliopoulos and Ruggiero, 2014; Papaspiliopoulos et al., 2014); and in Bayesian statistics they have been proposed as models for time-evolving priors (Walker et al., 2007; Favaro et al., 2009; Griffiths and Spanò, 2013; Mena and Ruggiero, 2014).

Simulation from equation (1) is highly nontrivial because there is no known closed form expression for the transition function of the diffusion, even in the simple case $\beta(x) \equiv 0$. In the absence of a method of exact simulation, it is necessary to turn to approximate alternatives such as an Euler-Maruyama scheme (Schraiber et al., 2013), truncation of a spectral expansion of the transition function (Lukić et al., 2011; Song and Steinrücken, 2012; Steinrücken et al., 2013), or numerical solutions of the Kolmogorov equations (Williamson et al., 2005; Bollback et al., 2008; Gutenkunst et al., 2009). The error introduced by these methods can be difficult to quantify and must often be tested empirically. However, in another direction, a series of recent papers have shown that it is possible to simulate *exactly* from a certain class of diffusion processes (Beskos and Roberts, 2005; Beskos et al., 2006a, 2008; Pollock et al., 2014) using a retrospective rejection sampling algorithm with Brownian motion serving to construct its candidate paths. By "exact" we mean that samples from the finite-dimensional distributions of the target diffusion can be recovered (up to the precision of a computer) without any approximation error. When they are available, exactly simulated diffusions can even be embedded into a particle filter (Beskos et al., 2006b; Fearnhead et al., 2008, 2010) or within a Markov chain Monte Carlo framework (Sermaidis et al., 2013). The price for this is that such a diffusion must satisfy a number of regularity conditions, the most stringent perhaps that its law be absolutely continuous with respect to Brownian motion. The Wright-Fisher diffusion (1) fails in this regard, first because of its nonunit diffusion coefficient and second because of its finite boundaries. Although the first problem is easily solved via a Lamperti transformation [also known as Fisher's transformation when applied to (1)], it is not clear how to deal with the second. Beskos et al. (2008) point out that their exact algorithm can be adopted to the case of two finite entrance boundaries,

but this approach breaks down whenever the diffusion is proximate to the boundary (Jenkins, 2013); in any case, the boundaries of the Wright-Fisher diffusion can be exit, regular reflecting, or entrance, depending on $\beta(x)$. Another approach developed in Jenkins (2013) can handle the case of a single entrance boundary efficiently, but it also seems difficult to extend this to a second boundary or to boundaries that are not entrance.

In this paper we develop a novel and *exact* method for simulating the Wright-Fisher diffusion with *general* drift, as well as the corresponding bridges. The key approach, developed in Section 3, is to show it is possible to simulate directly (without rejection) and exactly from the neutral Wright-Fisher diffusion with mutation; that is from the solution to (1) but with drift

$$\alpha(x) = \frac{1}{2}[\theta_1(1-x) - \theta_2 x]. \tag{3}$$

We can then use samples from this process as candidates in a rejection algorithm for our target process with more general drift $\beta(x)$. This uses a retrospective approach similar to that of the exact algorithms based on Brownian motion; we develop these ideas in Section 5. The advantage of this approach is clear, since our candidate diffusion is now *constructed* with finite boundaries and is expected to mimic the target diffusion well provided $\alpha(x)$ and $\beta(x)$ are not too dissimilar. We illustrate the idea using the drift given in (2), which motivated our interest in this problem. In Section 6 we discuss practical considerations of the algorithm. In Section 7 we extend our approach to more than one—in fact infinitely many—dimensions. Section 8 contains the proofs of our results.

2 Overview of the exact algorithm

In this section we give a brief overview of the exact algorithm (EA) of Beskos and Roberts (2005), Beskos et al. (2006a, 2008), and Pollock et al. (2014), and we refer to those papers for further details. The EA returns a recipe for simulating the sample paths of a diffusion $X = (X_t)_{t\in[0,T]}$ satisfying the SDE

$$dX_t = \mu(X_t)dt + dB_t, \qquad X_0 = x_0, t \in [0,T], \tag{4}$$

with μ assumed to satisfy the requirements for (4) to admit a unique, weak solution. Denote the law of such a process, our target, by $\mathbb{Q}_{x_0}$. The idea of the EA is to use Brownian motion started at x_0, whose law will be denoted $\mathbb{W}_{x_0}$, as the candidate process in a rejection sampling algorithm. The goal is then to write down the rejection probability, which is possible under the following assumptions:

(A1) The Radon-Nikodým derivative of $\mathbb{Q}_{x_0}$ with respect to $\mathbb{W}_{x_0}$ exists and is given by Girsanov's formula,

$$\frac{d\mathbb{Q}_{x_0}}{d\mathbb{W}_{x_0}}(X) = \exp\left\{\int_0^T \mu(X_t)dX_t - \frac{1}{2}\int_0^T \mu^2(X_t)dt\right\}, \tag{5}$$

(A2) $\mu \in C^1$,

(A3) $\phi(x) := \frac{1}{2}[\mu^2(x) + \mu'(x)]$ is bounded below, by ϕ^- say.

(A4) $A(x) := \int_0^x \mu(z)dz$ is bounded above, by A^+ say.

Using (A1–A4) and Itô's lemma, (5) can be re-expressed as

$$\frac{d\mathbb{Q}_{x_0}}{d\mathbb{W}_{x_0}}(X) \propto \exp\left\{A(X_T) - A^+\right\} \exp\left\{-\int_0^T [\phi(X_t) - \phi^-]dt\right\}. \tag{6}$$

Written in this form, the right hand side of (6) is less than or equal to 1, and therefore provides the required rejection probability. To make the accept/reject decision, it is necessary to construct an event occurring with probability (6). This is easy to achieve given a realized sample path $(X_t)_{t\in[0,T]} \sim \mathbb{W}_{x_0}$, but obtaining such a path would require an infinite amount of computation. Instead, note that the right hand term in (6) is the probability that all points in a Poisson point process $\mathbf{\Phi} = \{(t_j, \psi_j) : j = 0, 1, \ldots\}$ of unit rate on $[0, T] \times [0, \infty)$ lie in the epigraph of $t \mapsto [\phi(X_t) - \phi^-]$, and this event can be determined by simulating X only at times $t_1, t_2, \ldots$. Thus, the following algorithm returns a (random) collection of skeleton points from $X \sim \mathbb{Q}_{x_0}$:

Algorithm 1: Exact algorithm for simulating the path of a diffusion process with law $\mathbb{Q}_x$.

repeat
  Simulate $\mathbf{\Phi}$, a Poisson point process on $[0, T] \times [0, \infty)$.
  Simulate $U \sim$ Uniform$[0, 1]$.
  Given $\mathbf{\Phi} = \{(t_j, \psi_j) : j = 0, 1, \ldots\}$, simulate $B \sim \mathbb{W}_x$ at times $(t_j)_{j=0,1,\ldots}$ and at time T.
  if $\phi(B_{t_j}) \le \psi_j$ *for all* $j = 0, 1, \ldots$ *and* $U \le \exp\{A(B_T) - A^+\}$ **then**
    return $\{(t_j, B_{t_j}) : j = 0, 1 \ldots\} \cup \{(T, B_T)\}$.
  end
until *false*

If ϕ is bounded above, by K say, then Algorithm 1 can be implemented with finite computation by thinning $\mathbf{\Phi}$ to a Poisson Point process on $[0, T] \times [0, K]$; $|\mathbf{\Phi}|$ is then almost surely finite.

The output of Algorithm 1 is a set of skeleton points at a random, finite, collection of times. Once a skeleton has been accepted, further points can be filled in as required by sampling from Brownian bridges; no further reference to $\mathbb{Q}_x$ is necessary.

Remarks.

(i) Assumption (A4) is restrictive, and can in fact be removed by using a certain *biased* Brownian motion as an alternative candidate; this also improves the efficiency of the algorithm (Beskos and Roberts, 2005). We present the EA in the form above

since an analogue of (A4) *does* hold for the Wright-Fisher diffusion, and in any case a 'biased Wright-Fisher diffusion' is not available.

(ii) It is also in fact possible to implement Algorithm 1 with finite computation and without assuming an upper bound for ϕ (Beskos et al., 2006a, 2008; Pollock et al., 2014).

3 Simulating the neutral Wright-Fisher process

3.1 A transition density expansion

Recall that our goal is to develop an exact algorithm using the Wright-Fisher diffusion itself as a candidate process. In this section we demonstrate how simulation from a suitable candidate can be achieved. In particular, our strategy is first to simulate from the diffusion satisfying (1) with drift (3). Denote its law by $\mathbb{WF}_{\alpha,x}$, with transition density $f(x,\cdot\,;t)$. Throughout this paper we assume $\theta_1,\theta_2 > 0$; then $f(x,\cdot\,;t)$ is a probability density. We will exploit the following probabilistic representation of the transition function's eigenfunction expansion (Griffiths, 1979; Tavaré, 1984; Ethier and Griffiths, 1993; Griffiths and Spanò, 2010):

$$f(x,y;t) = \sum_{m=0}^{\infty} q_m^{\theta}(t) \sum_{l=0}^{m} \mathcal{B}_{m,x}(l)\mathcal{D}_{\theta_1+l,\theta_2+m-l}(y), \tag{7}$$

where

$$\mathcal{B}_{m,x}(l) = \binom{m}{l} x^l (1-x)^{m-l}, \qquad l = 0,1,\ldots,m,$$

is the probability mass function (PMF) of a binomial random variable,

$$\mathcal{D}_{\theta_1,\theta_2}(y) = \frac{1}{B(\theta_1,\theta_2)} y^{\theta_1-1}(1-y)^{\theta_2-1},$$

is the probability density function (PDF) of a beta random variable, $\theta = \theta_1 + \theta_2$, and $\{q_m^{\theta}(t) : m = 0,1,\ldots\}$ are the transition functions of a certain death process $A_{\infty}^{\theta}(t)$ with an entrance boundary at ∞. More precisely, let $\{A_n^{\theta}(t) : t \geq 0\}$ be a pure death process on $\mathbb{N}$ such that $A_n^{\theta}(0) = n$ almost surely and whose only transitions are $m \mapsto m-1$ at rate $m(m+\theta-1)/2$, for each $m = 1,2,\ldots,n$. Then $q_m^{\theta}(t) = \lim_{n\to\infty} \mathbb{P}(A_n^{\theta}(t) = m)$.

The representation (7) has a natural interpretation in terms of Kingman's coalescent, which is the moment dual to the Wright-Fisher diffusion. The ancestral process $A_{\infty}(t)$ represents the number of lineages surviving a time t back in an infinite-leaf coalescent tree, when lineages are lost both by coalescence and by mutation. For our purposes, the mixture expression (7) also provides an immediate method for *simulating* from $f(x,\cdot\,;t)$. We summarize this idea in Algorithm 2, which first appeared in Griffiths and Li (1983).

Algorithm 2: Simulating from the transition density $f(x, \cdot\,; t)$ of the neutral Wright-Fisher diffusion with mutation.

1 Simulate $A^\theta_\infty(t)$.
2 Given $A^\theta_\infty(t) = m$, simulate $L \sim \text{Binomial}(m, x)$.
3 Given $L = l$, simulate $Y \sim \text{Beta}(\theta_1 + l, \theta_2 + m - l)$.
4 **return** Y.

Steps 2 and 3 of Algorithm 2 are straightforward, but Step 1 requires the PMF $\{q^\theta_m(t) : m = 0, 1, \ldots\}$, which is not available in closed form. Griffiths and Li (1983) used a numerical approximation, but in the following subsection we show it is in fact possible to simulate from this distribution without error.

3.2 Simulating the ancestral process of Kingman's coalescent

Our goal in this subsection is to obtain *exact* samples from the discrete random variable with PMF $\{q^\theta_m(t) : m = 0, 1, \ldots\}$. Were each $q^\theta_m(t)$ available in closed form, then standard inversion sampling would return exact samples from this distribution (see for example Devroye, 1986, Ch. 2): for $U \sim \text{Uniform}[0, 1]$,

$$\inf\left\{M \in \mathbb{N} : \sum_{m=0}^{M} q^\theta_m(t) > U\right\}$$

is distributed according to $\{q^\theta_m(t) : m = 0, 1, \ldots\}$. However, $q^\theta_m(t)$ is known only as an infinite series (Griffiths, 1980; Tavaré, 1984):

$$q^\theta_m(t) = \sum_{k=m}^{\infty} (-1)^{k-m} a^\theta_{km} e^{-k(k+\theta-1)t/2}, \quad \text{where} \quad a^\theta_{km} = \frac{(\theta + 2k - 1)(\theta + m)_{(k-1)}}{m!(k-m)!}. \tag{8}$$

Here we have used the notation $a_{(x)} := \Gamma(a + x)/\Gamma(a)$ for $a > 0$ and $x \geq -1$.

Despite the apparently infinite amount of computation needed to evaluate (8), we now show that it is nonetheless possible to return exact samples from this distribution by a variant of the *alternating series method* (Devroye, 1986, Ch. 5), which we summarize for a discrete random variable X on $\mathbb{N}$ as follows. Suppose X has PMF $\{p_m : m = 0, 1, \ldots\}$ of the form

$$p_m = \sum_{k=0}^{\infty} (-1)^k b_k(m),$$

and such that

$$b_k(m) \downarrow 0 \text{ as } k \to \infty, \text{ for each } m. \tag{9}$$

Then for each $M, K \in \mathbb{N}$,

$$T^-_K(M) := \sum_{m=0}^{M} \sum_{k=0}^{2K+1} (-1)^k b_k(m) \leq \sum_{m=0}^{M} p_m \leq \sum_{m=0}^{M} \sum_{k=0}^{2K} (-1)^k b_k(m) =: T^+_K(M).$$

Furthermore, $T_K^-(M) \uparrow \mathbb{P}(X \le M)$ and $T_K^+(M) \downarrow \mathbb{P}(X \le M)$ as $K \to \infty$. Hence, for $U \sim \text{Uniform}[0,1]$ and $K_0(M) := \inf\left\{K \in \mathbb{N} : T_K^-(M) > U \text{ or } T_K^+(M) < U\right\}$,

$$\inf\left\{M \in \mathbb{N} : T_{K_0(M)}^-(M) > U\right\}$$

can be computed from finitely many terms and is exactly distributed according to $\{p_m : m = 0, 1, \ldots\}$.

This approach can be applied—with some modification—to $\{q_m^\theta(t) : m = 0, 1, \ldots\}$ by the following proposition, which says that the required condition (9) holds with the possible exception of the first few terms in m. For those exceptional terms, (9) still holds beyond the first few terms in k, and there is an easy way to check when this condition has been reached.

Proposition 1. *Let* $b_k^{(t,\theta)}(m) = a_{km}^\theta e^{-k(k+\theta-1)t/2}$, *the relevant coefficient in* (8), *and let*

$$C_m^{(t,\theta)} := \inf\left\{i \ge 0 : b_{i+m+1}^{(t,\theta)}(m) < b_{i+m}^{(t,\theta)}(m)\right\}. \tag{10}$$

Then

(i) $C_m^{(t,\theta)} < \infty$, *for all* m.

(ii) $b_k^{(t,\theta)}(m) \downarrow 0$ *as* $k \to \infty$ *for all* $k \ge m + C_m^{(t,\theta)}$, *and*

(iii) $C_m^{(t,\theta)} = 0$ *for all* $m > D_0$, *where*

$$D_\epsilon := \inf\left\{k \ge \left(\frac{1}{t} - \frac{\theta+1}{2}\right) \vee 0 : (\theta + 2k + 1)e^{-\frac{(2k+\theta)t}{2}} < 1 - \epsilon\right\} \tag{11}$$

for $\epsilon \in [0, 1)$.

(The parameter ϵ is introduced for later use.) As a result of Proposition 1, we need only to make the following adjustment to the alternating series method: If $m \le D_0$ then precompute terms in $q_m^\theta(t)$ until the first time that the coefficients in (8) begin to decay; we know by Proposition 1(ii) that this decay then continues indefinitely. To allow for the number of computed coefficients to depend on m we introduce $\boldsymbol{k} = (k_0, k_1, \ldots, k_M)$ and

$$S_{\boldsymbol{k}}^-(M) := \sum_{m=0}^{M} \sum_{i=0}^{2k_m+1} (-1)^i b_{m+i}^{(t,\theta)}(m), \qquad S_{\boldsymbol{k}}^+(M) := \sum_{m=0}^{M} \sum_{i=0}^{2k_m} (-1)^i b_{m+i}^{(t,\theta)}(m). \tag{12}$$

We summarize this procedure in Algorithm 3.

Algorithm 3: Simulating from the ancestral process $A_\infty(t)$ of Kingman's coalescent with mutation.

Set $m \longleftarrow 0$, $k_0 \longleftarrow 0$, $\boldsymbol{k} \longleftarrow (k_0)$.
Simulate $U \sim \text{Uniform}[0,1]$.
repeat
  Set $k_m \longleftarrow \lceil C_m^{(t,\theta)}/2 \rceil$ [eq. (10)].
  while $S_{\boldsymbol{k}}^-(m) < U < S_{\boldsymbol{k}}^+(m)$ **do**
    Set $\boldsymbol{k} \longleftarrow \boldsymbol{k} + (1, 1, \ldots, 1)$
  end

  if $S_{\boldsymbol{k}}^-(m) > U$ **then**
    return m
  else if $S_{\boldsymbol{k}}^+(m) < U$ **then**
    Set $\boldsymbol{k} \longleftarrow (k_0, k_1, \ldots, k_m, 0)$.
    Set $m \longleftarrow m + 1$.
  end
until *false*

Note that computed coefficients a_{km}^θ in (8) can also be stored for future calls to this algorithm. Further performance considerations of this algorithm are discussed in Section 6.

4 Simulating a neutral Wright-Fisher bridge

4.1 A transition density expansion

Section 3 provides a method for returning exact samples from $f(x, \cdot\,; t)$ for any fixed $x \in [0,1]$ and $t > 0$. For the purpose of refining this path, and for later use in our rejection algorithm, we now detail how to sample a point from a neutral Wright-Fisher *bridge*. The density of a point $y \in (0,1)$ at time s in a Wright-Fisher bridge from x at time 0 to z at time t is given by (Fitzsimmons et al., 1992; Schraiber et al., 2013):

$$f_{z,t}(x, y; s) = \frac{f(x, y; s) f(y, z; t-s)}{f(x, z; t)}, \qquad 0 < s < t, \tag{13}$$

with $f(\cdot\,, \cdot\,; \cdot\,)$ as in (7). Motivated by (7), our goal is to facilitate easy simulation from $f_{z,t}(x, y; s)$ by putting it into mixture form. For the rest of this section we assume $0 < x, y, z < 1$.

Proposition 2. *The transition density of a Wright-Fisher bridge has expansion*

$$f_{z,t}(x, y; s) = \sum_{m=0}^{\infty} \sum_{k=0}^{\infty} \sum_{l=0}^{m} \sum_{j=0}^{k} p_{m,k,l,j}^{(x,z,s,t,\boldsymbol{\theta})} \mathcal{D}_{\theta_1+l+j, \theta_2+m+k-l-j}(y), \tag{14}$$

where

$$p_{m,k,l,j}^{(x,z,s,t,\boldsymbol{\theta})} = \mathcal{A}_{m,k,l,j}^{(x,z,\boldsymbol{\theta})} \frac{q_m^\theta(s) q_k^\theta(t-s)}{f(x,z;t)}, \tag{15}$$
$$\mathcal{A}_{m,k,l,j}^{(x,z,\boldsymbol{\theta})} = \mathcal{B}_{m,x}(l)\mathcal{D}_{\theta_1+j,\theta_2+k-j}(z)\binom{k}{j}\frac{B(\theta_1+l+j,\theta_2+m-l+k-j)}{B(\theta_1+l,\theta_2+m-l)},$$

for $0 \le l \le m$ *and* $0 \le j \le k$, *and* $p_{m,k,l,j}^{(x,z,s,t,\boldsymbol{\theta})} = 0$ *otherwise.*

By Proposition 2, we recognize equation (14) as a mixture of beta-distributed random variables, with mixture weights defining a PMF $\{p_{m,k,l,j}^{(x,z,s,t,\boldsymbol{\theta})} : m,k,l,j \in \mathbb{N}\}$ on $\mathbb{N}^4$. Thus, the following algorithm returns exact samples from $f_{z,t}(x,y;s)$.

Algorithm 4: Simulating from the transition density $f_{z,t}(x,y;s)$ of a bridge of the neutral Wright-Fisher diffusion with mutation.

1 Simulate $(M,K,L,J) \sim \{p_{m,k,l,j}^{(x,z,s,t,\boldsymbol{\theta})} : m,k,l,j \in \mathbb{N}\}$ [eq. (15)].
2 Given $(M,K,L,J) = (m,k,l,j)$, simulate $Y \sim \text{Beta}(\theta_1+l+j, \theta_2+m+k-l-j)$.
3 **return** Y

Again, while step 2 of Algorithm 4 is straightforward, Step 1 is complicated by the appearance of $q_m^\theta(s)q_k^\theta(t-s)/f(x,z;t)$ in (15); each term in this ratio is known only as an infinite series, as we have seen. We address this in the following subsection.

4.2 Simulating from the discrete random variable on $\mathbb{N}^4$ with PMF $\{p_{m,k,l,j}^{(x,z,s,t,\boldsymbol{\theta})} : m,k,l,j \in \mathbb{N}\}$

We will apply the alternating series approach of Section 3.2 separately to each of $q_m^\theta(s)$, $q_k^\theta(t-s)$, and $f(x,z;t)$, and then combine these to obtain monotonically converging upper and lower bounds on (15). The first two terms have been dealt with already in Section 3.2, so it remains to take a similar approach for $f(x,z;t)$. Note that this problem—the pointwise evaluation of $f(x,z;t)$—is separate from (and in this case, harder than) actually simulating from $f(x,\cdot\,;t)$.

To employ the alternating series approach, use (7) and (8) to write

$$f(x,z;t) = \sum_{m=0}^{\infty}\sum_{k=m}^{\infty}(-1)^{k-m}c_{k,m}^{(x,z,t,\boldsymbol{\theta})}, \tag{16}$$
$$\text{where} \quad c_{k,m}^{(x,z,t,\boldsymbol{\theta})} = a_{km}^\theta e^{-k(k+\theta-1)t/2}\mathbb{E}[\mathcal{D}_{\theta_1+L_m,\theta_2+m-L_m}(z)],$$

and $L_m \sim \text{Binomial}(m,x)$. Our strategy is to group the triangular array of coefficients $(c_{k,m}^{(x,z,t,\boldsymbol{\theta})})_{k\ge m}$ in such a way that, with the exception of the first few terms, they exhibit a property analogous to (9). We will compare terms in the sequence $(d_i)_{i=0,1,\ldots}$ of

antidiagonals, defined by

$$d_{2m} = \sum_{j=0}^{m} c_{m+j,m-j}, \qquad d_{2m+1} = \sum_{j=0}^{m} c_{m+1+j,m-j}, \qquad m = 0, 1, \ldots, \tag{17}$$

(see Figure 1), and dropping the superscript for convenience. Notice that the coefficients within each entry of this sequence are all multiplied by the same sign in (16), so that $f(x, z; t) = d_0 - d_1 + d_2 - d_3 + \ldots$ will be our alternating sequence. The main complication in this approach is to find *explicitly* the first i for which the coefficients (d_i) begin to decrease in magnitude. To this end we define

$$E := \inf \left\{ m \geq 0 : j \geq C_{m-j}^{(t,\theta)} \text{ for all } j = 0, \ldots, m \right\}, \tag{18}$$

which simply provides the first entry in (d_{2m}) for which every member of the corresponding antidiagonal is decaying as a function of its first index. We now need the following lemma.

Lemma 1. *Let* $L_m \sim$ *Binomial*(m, x) *and*

$$K := \frac{x}{z} + \frac{1-x}{1-z}. \tag{19}$$

Then for all $m \in \mathbb{N}$,

$$\mathbb{E}[\mathcal{D}_{\theta_1 + L_{m+1}, \theta_2 + m + 1 - L_{m+1}}(z)] < K \mathbb{E}[\mathcal{D}_{\theta_1 + L_m, \theta_2 + m - L_m}(z)]. \tag{20}$$

Using Lemma 1 we are in a position to obtain the required analogue of property (9).

Proposition 3. *Let* D_ϵ, $(d_i)_{i=0,1,\ldots}$, E, *and* K *be defined as in* (11), (17), (18), *and* (19), *respectively, and* $\epsilon \in (0, 1)$. *Then*

$$d_{2m+2} < d_{2m+1} < d_{2m}$$

for all $m \geq E \vee D_\epsilon \vee 2K/\epsilon$.

We can now combines Propositions 1 and 3 in order to construct a sequence amenable to simulation from $\{p_{m,k,l,j}^{(x,z,s,t,\boldsymbol{\theta})} : m, k, l, j \in \mathbb{N}\}$ [eq. (15)], via the alternating series method.

Proposition 4. *Define*

$$F_{m,k,l,j} := C_m^{(t,\theta)} \vee C_k^{(t-s,\theta)} \vee E \vee D_\epsilon \vee 2K/\epsilon, \tag{21}$$

and

$$e_{m,k,l,j}(v) := \mathcal{A}_{m,k,l,j}^{(x,z,\boldsymbol{\theta})} \left[\sum_{i=0}^{v} (-1)^i b_{m+i}^{(s,\theta)}(m) \right] \left[\sum_{i=0}^{v} (-1)^i b_{k+i}^{(t-s,\theta)}(k) \right] \Bigg/ \sum_{i=0}^{v+1} (-1)^i d_i \, .$$

Then for $2v \geq F_{m,k,l,j}$,

$$e_{m,k,l,j}(2v+1) < e_{m,k,l,j}(2v+3) < p_{m,k,l,j}^{(x,z,s,t,\boldsymbol{\theta})} < e_{m,k,l,j}(2v+2) < e_{m,k,l,j}(2v).$$

In other words, for sufficiently large v the odd and even terms in the sequence $(e_{m,k,l,j}(v))_{v=0}^{\infty}$ provide monotonically converging lower and upper bounds on $p_{m,k,l,j}^{(x,z,s,t,\boldsymbol{\theta})}$, respectively, and "sufficiently large" can be verified explicitly.

The above results are summarized in Algorithm 5. To explore $\mathbb{N}^4$ we introduce for convenience a bijective pairing function $\Sigma : \mathbb{N} \to \mathbb{N}^4$, such that $\Sigma(n) = (m, k, l, j)$. As in Algorithm 3, we also introduce $\boldsymbol{v} = (v_0, v_1, \ldots, v_n)$ and

$$V_{\boldsymbol{k}}^{-}(N) := \sum_{n=0}^{N} \sum_{i=0}^{2v_n+1} (-1)^i e_{\Sigma(n)}(v), \qquad V_{\boldsymbol{k}}^{+}(N) := \sum_{n=0}^{N} \sum_{i=0}^{2v_n} (-1)^i e_{\Sigma(n)}(v).$$

Algorithm 5: Simulating from the discrete random variable on $\mathbb{N}^4$ with PMF $\{p_{m,k,l,j}^{(x,z,s,t,\boldsymbol{\theta})} : m, k, l, j \in \mathbb{N}\}$.

Set $n \longleftarrow 0$, $v_0 \longleftarrow 0$, $\boldsymbol{v} \longleftarrow (v_0)$.
Simulate $U \sim \text{Uniform}[0, 1]$.
repeat
  Set $v_n \longleftarrow \lceil F_{\Sigma(n)}/2 \rceil$ [eq. (21)].
  while $V_{\boldsymbol{v}}^{-}(n) < U < V_{\boldsymbol{v}}^{+}(n)$ **do**
    Set $\boldsymbol{v} \longleftarrow \boldsymbol{v} + (1, 1, \ldots, 1)$
  end

  if $V_{\boldsymbol{v}}^{-}(n) > U$ **then**
    return $\Sigma(n)$
  else if $V_{\boldsymbol{v}}^{+}(n) < U$ **then**
    Set $\boldsymbol{v} \longleftarrow (v_0, v_1, \ldots, v_n, 0)$.
    Set $n \longleftarrow n + 1$.
  end
until *false*

5 Simulating a Wright-Fisher coordinate with general drift

5.1 Exact algorithm

In this section we develop an exact rejection algorithm for simulating from the Wright-Fisher diffusion (1) with general drift, whose law we denote $\mathbb{WF}_{\beta,x_0}$. We use $\mathbb{WF}_{\alpha,x_0}$ [eq. (3)] to define our candidate process and employ the retrospective sampling techniques outlined in Section 2 to make the accept/reject decision. To that end we impose the following assumptions on $\beta(x)$:

(WF1) The Radon-Nikodým derivative of $\mathbb{WF}_{\beta,x_0}$ with respect to $\mathbb{WF}_{\alpha,x_0}$ exists and is

given by Girsanov's formula,

$$\frac{d\mathbb{WF}_{\beta,x_0}}{d\mathbb{WF}_{\alpha,x_0}}(X) = \exp\left\{\int_0^T \frac{\beta(X_t)-\alpha(X_t)}{X_t(1-X_t)}dX_t - \frac{1}{2}\int_0^T \frac{\beta^2(X_t)-\alpha^2(X_t)}{X_t(1-X_t)}dt\right\}, \tag{22}$$

(WF2) β is continuously differentiable on $(0,1)$.

(WF3) $\widetilde{\phi}(x)$ is bounded on $(0,1)$: $\widetilde{\phi}^- \le \widetilde{\phi}(x) \le \widetilde{\phi}^+$, where

$$\widetilde{\phi}(x) := \frac{1}{2}\left[\frac{\beta^2(x)-\alpha^2(x)}{x(1-x)} + \beta'(x) - \alpha'(x) - [\beta(x)-\alpha(x)]\frac{1-2x}{x(1-x)}\right].$$

(WF4) $\widetilde{A}(x) := \int_0^x \frac{\beta(z)-\alpha(z)}{z(1-z)}dz$ is bounded above, by $\widetilde{A}^+$ say.

Specific conditions on α and β to satisfy (WF1) are detailed e.g. in Theorem 8.6.8 in Øksendal (2003). Following Section 2, we apply Itô's lemma to $\widetilde{A}(x)$ to re-express (22) as

$$\frac{d\mathbb{WF}_{\beta,x_0}}{d\mathbb{WF}_{\alpha,x_0}}(X) \propto \exp\left\{\widetilde{A}(X_T) - \widetilde{A}^+\right\}\exp\left\{-\int_0^T [\widetilde{\phi}(X_t) - \widetilde{\phi}^-]dt\right\}. \tag{23}$$

We recognize the rightmost term in (23) as the probability that no points in a Poisson point process on $[0,T]\times[0,\widetilde{\phi}^+]$ lie in the epigraph of $t \mapsto \widetilde{\phi}(X_t) - \widetilde{\phi}^-$. Hence, the following algorithm returns exact samples from $\mathbb{WF}_{\beta,x_0}$.

Algorithm 6: Exact algorithm for simulating the path of a diffusion process with law $\mathbb{WF}_{\beta,x}$.

repeat
  Simulate $\mathbf{\Phi}$, a Poisson point process on $[0,T]\times[0,\widetilde{\phi}^+]$.
  Simulate $U \sim$ Uniform$[0,1]$.
  Given $\mathbf{\Phi} = \{(t_j,\psi_j) : j = 0,1,\ldots,J\}$, simulate $X \sim \mathbb{WF}_{\alpha,x}$ at times $(t_j)_{j=0,1,\ldots,J}$ and at time T.
  if $\phi(X_{t_j}) \le \psi_j$ *for all* $j = 0,1,\ldots,J$ *and* $U \le \exp\{\widetilde{A}(X_T) - \widetilde{A}^+\}$ **then**
    return $\{(t_j, X_{t_j}) : j = 0,1\ \ldots, J\} \cup \{(T, X_T)\}$.
  end
until *false*

Step 4 is achieved via Algorithm 2. Once a skeleton is accepted, further points can be filled in via Algorithm 4.

5.2 Example: Wright-Fisher diffusion with diploid selection

When $\beta(x)$ is given by (2) and $\alpha(x)$ by (3), that (WF1) holds follows by e.g. Ethier and Kurtz (1993, Theorem 3.3, in the particular case where the type space E consists of two

points) or Dawson (1978, Theorem 5.1). We also find

$$\widetilde{\phi}(x) = \frac{\sigma}{2}x(1-x)\left[\sigma\Big(x+h(1-2x)\Big)^2 + 1 - 2h\right] + \frac{\sigma}{2}\Big(x+h(1-2x)\Big)[\theta_1(1-x) - \theta_2 x], \tag{24}$$

So in this example $\widetilde{\phi}(x)$ is a (quartic) polynomial; (WF2) and (WF3) are easily satisfied. Bounds on $\widetilde{\phi}(x)$ can be found by numerically maximizing and minimizing this function on $[0,1]$ to a given tolerance (this does not void the exactness of the algorithm). In the haploid case ($h = 1/2$), $\widetilde{\phi}(x)$ is quadratic and analytic bounds are available.

It is also easy to see that (WF4) holds, since

$$\widetilde{A}(x) := \sigma x\left[h + \frac{x}{2}(1-2h)\right].$$

We implemented our exact algorithm on this model, and investigated its performance by considering several combinations of parameters; results are shown in Table 1. For moderate selection ($|\sigma| = 1$) the algorithm is extremely efficient, with only slightly more than one candidate needed per acceptance. Furthermore, most simulations resulted in zero Poisson points. These results are quite robust to the length of the path t, the initial frequency x, and the sign of the selection parameter. For stronger selection ($\sigma = 10$) we observe some deterioration in efficiency because of the greater mismatch between candidate and target paths—to the extent that simulation of paths of length $t = 5.0$ became prohibitive. Nonetheless, it is still feasible to simulate a collection of shorter paths in a few seconds (and to string these together to construct longer paths, if necessary).

6 Performance issues

There are several easy improvements to the underlying Algorithm 3. For example, we are free to vary the order of inspection of each m by any finite permutation of $\mathbb{N}$, and we found a dramatic improvement by radiating outwards from (an approximation of) the mode of $\{q_m^\theta(t) : m = 0, 1, \ldots\}$ than to start at $m = 0$ and work upwards. Our approximation used Theorem 1 below. It may also be possible to improve on Algorithm 3 by allowing different $q_m^\theta(t)$ to be refined at different rates, i.e. by using a vector other than $(1, 1, \ldots, 1)$ in Step 6; we do not explore that here.

A crucial quantity governing the efficiency of our algorithms is the number of coefficients $b_k^{(t,\theta)}(m)$ we must compute in Algorithm 3; these in turn depend on the quantities D_0 and C_m (recall Proposition 1). The dependence of these quantities on θ and t is summarized in Table 2. It is evident from the table that these quantities are in general manageably small, suggesting that the number of coefficients that need to be computed in Algorithms 3 and 5 (line 4 in each) should not be excessive.

One exception to this observation is when t is very small, for which the number of relevant coefficients grows quickly. This is closely related to the well known numerical

Table 1: Performance of Algorithm 6 applied to the Wright-Fisher diffusion with symmetric mutation and additive selection. Each row (except total running time) reports averages across $1,000$ paths initiated at $X_0 = x$ and run for time t, with mutation parameter θ and selection parameter σ (and $h = 0.5$). Reported are the total numbers (per *accepted* path) of: attempts, Poisson points simulated, skeleton points simulated (including the endpoint), random variables generated, the number of coefficients $b_k^{(t,\theta)}(m)$ needed, and the number of times the simulation resorted to the approximation (25).

$\sigma = -1$

t	x	Attempts	Poisson points	Coefficients	Random variables	G1984	Total time (s)
0.1	0.5	1.3	0.0	118.4	7.5	0.0	2.0
0.1	0.25	1.1	0.0	105.6	6.6	0.0	2.0
0.1	0.01	1.0	0.0	91.1	6.0	0.0	2.0
0.5	0.5	1.2	0.0	18.6	7.3	0.0	1.0
0.5	0.25	1.1	0.0	17.1	6.8	0.0	1.0
0.5	0.01	1.0	0.0	15.0	6.1	0.0	0.0
5.0	0.5	1.3	0.2	5.6	8.8	0.0	0.0
5.0	0.25	1.2	0.2	6.3	7.8	0.0	0.0
5.0	0.01	1.0	0.2	4.9	7.2	0.0	0.0

$\sigma = 1$

0.1	0.5	1.3	0.0	116.1	7.4	0.0	2.0
0.1	0.25	1.4	0.0	132.9	8.0	0.0	2.0
0.1	0.01	1.6	0.0	141.2	8.9	0.0	3.0
0.5	0.5	1.2	0.0	18.6	7.2	0.0	0.0
0.5	0.25	1.5	0.0	21.4	8.4	0.0	0.0
0.5	0.01	1.6	0.0	23.8	9.0	0.0	0.0
5.0	0.5	1.3	0.2	5.4	8.5	0.0	1.0
5.0	0.25	1.5	0.2	7.3	9.5	0.0	0.0
5.0	0.01	1.6	0.3	7.2	10.6	0.0	1.0

$\sigma = 10$

0.1	0.5	11.9	3.8	1003.1	63.3	1.9	18.0
0.1	0.25	41.3	13.0	3675.5	223.0	7.9	67.0
0.1	0.01	146.5	46.2	15006.7	861.0	46.2	264.0
0.5	0.5	13.4	20.9	650.0	110.5	2.5	12.0
0.5	0.25	43.7	68.6	2721.4	398.3	10.8	46.0
0.5	0.01	144.1	227.7	16489.8	1808.9	69.3	297.0

D_0					
	θ				
t	0.01	0.1	1.0	10.0	100.0
0.01	729	729	728	724	679
0.1	46	46	45	41	0
0.5	5	5	5	0	0
1.0	2	2	1	0	0
10.0	0	0	0	0	0

$\max_m C_m$					
	θ				
t	0.01	0.1	1.0	10.0	100.0
0.01	56	56	56	56	56
0.1	5	5	5	5	0
0.5	1	1	1	0	0
1.0	1	1	0	0	0
10.0	0	0	0	0	0

Table 2: Values of D_0 [eq. (11)] and $\max_m C_m$ [eq. (10)] for various choices of t and θ.

instability of (8) as $t \to 0$ (Griffiths, 1984), which afflicts any method based on the expansion (7) (or an expansion using a basis of orthogonal polynomials, which is equivalent to (8); Griffiths and Spanò, 2010). Since we are unable to guarantee any lower bound on the separation between simulated Poisson points, in any practical implementation of our algorithm we are obliged to use an approximation should this separation be very small. Fortunately, there has been much previous work in coalescent theory on approximating the distribution (8) (e.g. Griffiths, 1984, 2006; Jewett and Rosenberg, 2014); by inserting those approximations here they readily define new algorithms for approximate simulation of the dual *diffusion*. Griffiths (1984, Theorem 4) found the following result; we applied it whenever $t \leq 0.05$, which we found to be useful.

Theorem 1 (Griffiths (1984))**.** *$q_m^\theta(t)$ has an asymptotic Normal(μ, σ^2) distribution as $t \to 0$ while $\beta = \frac{1}{2}(\theta - 1)t$ is bounded above, where*

$$\mu = \frac{2\eta}{t}, \quad \sigma^2 = \begin{cases} \frac{2\eta}{t}(\eta+\beta)^2\left(1 + \frac{\eta}{\eta+\beta} - 2\eta\right), & \beta \neq 0, \\ \frac{2}{3t}, & \beta = 0, \end{cases} \quad \text{and} \quad \eta = \begin{cases} \frac{\beta}{e^\beta - 1}, & \beta \neq 0, \\ 1, & \beta = 0. \end{cases} \tag{25}$$

7 Simulating neutral Wright-Fisher transition functions in higher dimensions

It is worth pointing out that an interesting by-product of Proposition 1 (and of Algorithm 3) is the possibility of simulating exactly from the transition function of the (parent-independent) neutral Wright-Fisher diffusion *in any dimension*, even in infinite dimensions, *as long as one is not concerned to condition the process on a given terminal point*. Wright-Fisher diffusions in d dimensions can be seen as d-dimensional projections of a so-called neutral (parent-independent) Fleming-Viot measure-valued diffusion $\mu = (\mu_t : t \geq 0)$ with state space $\mathcal{M}_1(E)$, the set of all the probability measures on a given (Polish) type space E, equipped with the Borel sigma-algebra induced by the weak convergence topology. Given a total mutation parameter θ and a mutation distribution $P_0 \in \mathcal{M}_1(E)$, the process μ is reversible with stationary distribution given by

the Dirichlet process with parameter (θ, P_0), here denoted with Π_{θ,P_0}, characterized by Dirichlet finite-dimensional distributions:

$$\Pi_{\theta,P_0}\left(\bigcap_{i=1}^{d}\{\mu(A_i)\in dx_i,\}\right)\propto\left[\prod_{i=1}^{d}x_i^{\theta P_0(A_i)-1}dx_i\right]\mathbb{I}_{\Delta_{(d-1)}}(x_1,\ldots,x_d)$$

for any d and every measurable partition $A_1,\ldots,A_d$ of E, where $\Delta_{(d-1)}=\{(x_1,\ldots,x_d)\in[0,1]^d:\sum_1^d x_i=1\}$.

The transition function describing the evolution of μ admits a probabilistic series expansion as mixture of (posterior) Dirichlet processes:

$$p(\mu,d\nu;t)=\sum_m q_m^\theta(t)\int_{E^n}\mu^{\otimes m}(d\xi_1,\ldots,d\xi_m)\Pi_{\theta+m,\frac{m}{\theta+m}\eta_n+\frac{\theta}{\theta+m}P_0}(d\nu),$$
$$t\geq 0,\mu,\nu\in\mathcal{M}_1(E),\quad(26)$$

where $\mu^{\otimes n}$ denotes the n-fold μ-product measure and $\eta_m:=m^{-1}\sum_{i=1}^m\delta_{\xi_i}$ (see Ethier and Griffiths, 1993). The coefficients of the series expansion are given by *i.i.d.* samples (the ξ-random variables) from the starting measure, μ, of random size given by the coalescent lines-of-descent counting process $A_\infty^\theta(t)$ with distribution $q_m^\theta(t)$. An algorithm for simulating from the transition function (26) is thus the following modification of Algorithm 2.

Algorithm 7: Simulating from the transition density $p(\mu,\cdot;t)$ of the neutral Fleming-Viot process with parent-independent mutation.

1 Simulate $A_\infty^\theta(t)$.
2 Given $A_\infty^\theta(t)=m$, simulate $\xi_1,\ldots,\xi_m\overset{iid}{\sim}\mu$.
3 Given $m^{-1}\sum_{i=1}^m\delta_{\xi_i}=\eta_m$, simulate $\nu\sim\Pi_{\theta+m,\frac{m}{\theta+m}\eta_n+\frac{\theta}{\theta+m}P_0}$.
4 **return** ν.

Notice that step 3 requires sampling a (potentially infinite-dimensional) random measure distributed according to a Dirichlet process. Techniques for exact sampling from a Dirichlet process have been available in the literature [e.g. Papaspiliopoulos and Roberts (2008) and Walker (2007)] for quite some time. Hence Algorithm 3 provides a way of filling the only missing gap (Step 1 of Algorithm 7) to make the transition function (26) viable for exact simulation. When E consists of d points ($d\in\mathbb{N}$) the process reduces to the $(d-1)$-dimensional Wright-Fisher diffusion, thus Algorithm 2 is viable for exact simulation of neutral $(d-1)$-dimensional extensions of the Wright-Fisher diffusion (1) with drift (3). The problem of sampling from $(d-1)$-dimensional ($2<d\leq\infty$) Wright-Fisher bridges is an interesting open problem currently under study.

In a wider perspective, we believe that the approach proposed here might serve as a template for developing new techniques for sampling exactly from diffusion processes by

means of non-Brownian bridges whose transition function admits a transparent transition function expansion.

8 Proofs

Proof of Proposition 1. First suppose $m > 0$.

(i) Note that

$$\frac{b_{k+1}^{(t,\theta)}(m)}{b_k^{(t,\theta)}(m)} = \frac{\theta+m+k-1}{k-m+1} \cdot \frac{\theta+2k+1}{\theta+2k-1} e^{-\frac{(2k+\theta)t}{2}} =: f_m^\theta(k) e^{-\frac{(2k+\theta)t}{2}}, \tag{27}$$

say. Treat $f_m^\theta(k)$ as having domain $\mathbb{R}$; it then suffices to show that $(f_m^\theta)'(k) < 0$ for all sufficiently large k. Then the right hand side of (27) is subsequently decreasing in k monotonically to 0. Part (i) follows for the finite k $(= m + i)$ for which the right hand side of (27) drops below 1. Routine calculations show that $(f_m^\theta)'(k) < 0$ for all $k > (\sqrt{2(m-1)+\theta} - \theta)/2$.

(ii) Note that

$$\frac{\sqrt{2(m-1)+\theta} - \theta}{2} < \sqrt{\frac{m-1}{2}} + \frac{\sqrt{\theta} - \theta}{2} < \sqrt{\frac{m-1}{2}} + \frac{1}{8} < m,$$

so in fact $(f_m^\theta)'(k) < 0$ for *all* $k \geq m$, and thus as soon as $b_{k+1}^{(t,\theta)}(m) < b_k^{(t,\theta)}(m)$ for some k, it must also hold for all subsequent k.

(iii) The right hand side of (27) can be made independent of m by noting that

$$f_m^\theta(k) < f_k^\theta(k) = \theta + 2k + 1. \tag{28}$$

Thus for $C_m^{(t,\theta)} = 0$ to hold we need m to exceed the upper of the two solutions on $\mathbb{R}$ of

$$(\theta + 2k + 1)e^{-(2k+\theta)t/2} = 1.$$

The definition of D_0 is one way to express this condition, since the mode of $(\theta + 2k + 1)e^{-(2k+\theta)t/2}$ which separates the two solutions is at $k = \left(\frac{1}{t} - \frac{\theta+1}{2}\right)$.

Finally, consider the special case $m = 0$. If $\theta > 1$ then similar arguments as in (i–ii) above continue to hold. However, if $\theta \leq 1$ then in fact $(f_0^\theta)'(k) > 0$ for all k, with $f_0^\theta(k)$ continuous on $k \geq 1$. But then $f_0^\theta(k) < f_0^\theta(\infty) = 1$ for $k \geq 1$, so $f_0^\theta(k)e^{-(2k+\theta)t/2} < 1$ for all $k \geq 1$ and hence (i–ii) still hold, with $C_0^{(t,\theta)} \leq 1$. □

Proof of Proposition 2. This follows by substituting (7) into (13), multiplying by $B(\theta_1 + l + j, \theta_2 + m - l + k - j)/B(\theta_1 + l + j, \theta_2 + m - l + k - j)$, and rearranging. □

Proof of Lemma 1. First suppose $l \leq \lfloor mz \rfloor$. Then, using $\Gamma(x+1) = x\Gamma(x)$,

$$\begin{aligned}
&\mathbb{P}(L_{m+1} = l)\mathcal{D}_{\theta_1+l,\theta_2+m+1-l}(z) \\
&\quad = \left[\frac{m+1}{m+1-l}(1-x)\frac{\theta+m}{\theta_2+m-l}(1-z)\right]\mathbb{P}(L_m = l)\mathcal{D}_{\theta_1+l,\theta_2+m-l}(z), \\
&\quad \leq \left[\frac{m+1}{m+1-mz}(1-x)\frac{\theta+m}{\theta_2+m-mz}(1-z)\right]\mathbb{P}(L_m = l)\mathcal{D}_{\theta_1+l,\theta_2+m-l}(z), \\
&\quad \leq \left[\frac{1-x}{1-z}\right]\mathbb{P}(L_m = l)\mathcal{D}_{\theta_1+l,\theta_2+m-l}(z),
\end{aligned}$$

maximizing the term in square brackets first in l and then in m, noting for the last inequality that this term is increasing in m. Hence, summing over $l = 0, \ldots, \lfloor mz \rfloor$,

$$\sum_{l=0}^{\lfloor mz \rfloor}\mathbb{P}(L_{m+1} = l)\mathcal{D}_{\theta_1+l,\theta_2+m+1-l}(z) < \frac{1-x}{1-z}\sum_{l=0}^{\lfloor mz \rfloor}\mathbb{P}(L_m = l)\mathcal{D}_{\theta_1+l,\theta_2+m-l}(z). \tag{29}$$

By a similar argument, for $l \geq \lfloor mz \rfloor$:

$$\mathbb{P}(L_{m+1} = l+1)\mathcal{D}_{\theta_1+(l+1),\theta_2+m+1-(l+1)}(z) \leq \frac{x}{z}\mathbb{P}(L_m = l)\mathcal{D}_{\theta_1+l,\theta_2+m-l}(z),$$

(this time it is crucial we compare $L_{m+1} = l+1$ with $L_m = l$, rather than with $L_m = l+1$), and hence

$$\sum_{l=\lfloor mz \rfloor+1}^{m+1}\mathbb{P}(L_{m+1} = l)\mathcal{D}_{\theta_1+l,\theta_2+m+1-l}(z) < \frac{x}{z}\sum_{l=\lfloor mz \rfloor}^{m}\mathbb{P}(L_m = l)\mathcal{D}_{\theta_1+l,\theta_2+m-l}(z). \tag{30}$$

Finally, sum (29) and (30) to yield (20), noting that the overlap on the right-hand side at $l = \lfloor mz \rfloor$ necessitates the given definition of K (instead of the simpler bound $\frac{1-x}{1-z} \vee \frac{x}{z}$). □

Proof of Proposition 3. First, since $m \geq E$ we know $j \geq C_{m-j}^{(t,\theta)}$ for $j = 0, \ldots, m$ and hence by Proposition 1 that $b_{m+j+1}^{(t,\theta)}(m-j) < b_{m+j}^{(t,\theta)}(m-j)$. Now multiply this inequality by $\mathbb{E}[\mathcal{D}_{\theta_1+L_{m-j},\theta_2+m-j-L_{m-j}}(z)]$ to yield

$$c_{m+j+1,m-j}^{(x,z,t,\boldsymbol{\theta})} < c_{m+j,m-j}^{(x,z,t,\boldsymbol{\theta})}. \tag{31}$$

Thus, summing over $j = 0, 1, \ldots, m$,

$$\sum_{j=0}^{m} c_{m+1+j,m-j}^{(x,z,t,\boldsymbol{\theta})} < \sum_{j=0}^{m} c_{m+j,m-j}^{(x,z,t,\boldsymbol{\theta})}, \tag{32}$$

which says precisely that $d_{2m+1} < d_{2m}$. We also need to show that $d_{2m+2} < d_{2m+1}$, but this case is more subtle since the left hand side is a sum over one more term than the right. Instead, we will increment the first index in (31) and then sum over $j = 1, \ldots, m$:

$$\sum_{j=1}^{m} c_{m+2+j,m-j}^{(x,z,t,\boldsymbol{\theta})} < \sum_{j=1}^{m} c_{m+1+j,m-j}^{(x,z,t,\boldsymbol{\theta})}. \tag{33}$$

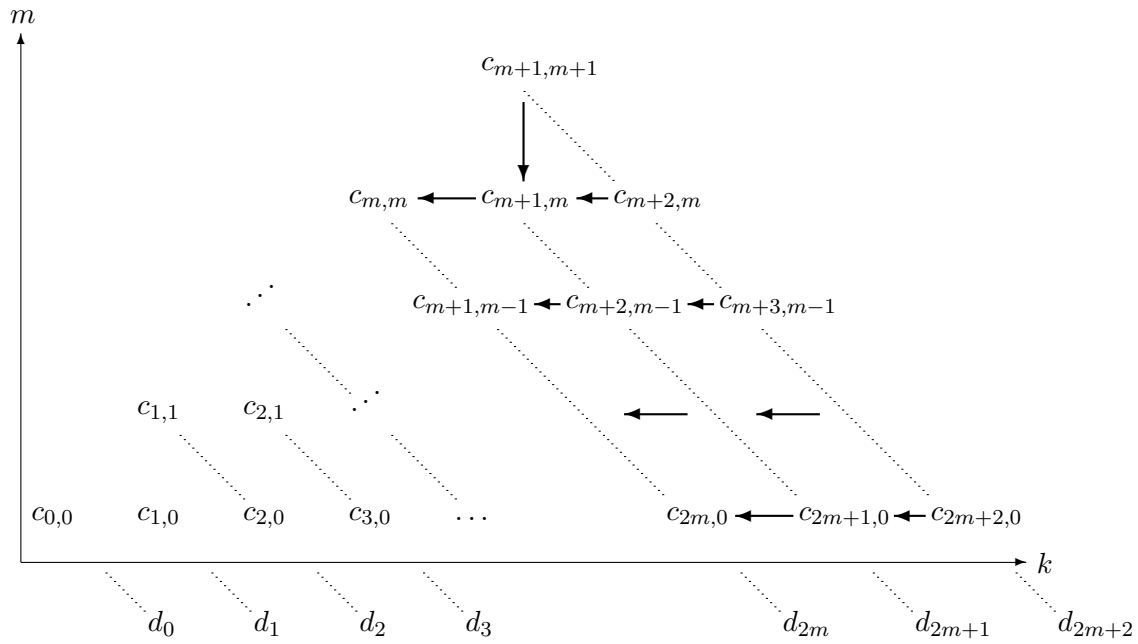

Figure 1: Computation of $c_{k,m}^{(x,z,t,\boldsymbol{\theta})}$. The sum of each antidiagonals (dashed) defines the sequence $(d_i)_{i=0,1,\ldots}$. To show that $d_{i+1} < d_i$ terms are paired off as shown by the arrows; that is, the coefficient at the head of a set of arrows is greater in magnitude than the sum of the terms at its tails.

It then suffices to show

$$c_{m+1,m+1}^{(x,z,t,\boldsymbol{\theta})} + c_{m+2,m}^{(x,z,t,\boldsymbol{\theta})} < c_{m+1,m}^{(x,z,t,\boldsymbol{\theta})}, \tag{34}$$

for if we sum (33) and (34) we obtain $d_{2m+2} < d_{2m+1}$ as required (Figure 1). To show (34), first note that

$$\frac{c_{k+1,m}^{(x,z,t,\boldsymbol{\theta})}}{c_{k,m}^{(x,z,t,\boldsymbol{\theta})}} = \frac{b_{k+1}^{(t,\theta)}(m)}{b_k^{(t,\theta)}(m)} = f_m^{\theta}(k) e^{-\frac{(2k+\theta)t}{2}} \leq (\theta + 2k + 1) e^{-\frac{(2k+\theta)t}{2}},$$

with $f_m^{\theta}(k)$ defined as in (27) and the inequality following from (28). Hence, choosing $k = m + 1$ and noting that $m \geq C_{\epsilon}^{(t,\theta)}$,

$$c_{m+2,m}^{(x,z,t,\boldsymbol{\theta})} \leq (\theta + 2k + 1) e^{-\frac{(2k+\theta)t}{2}} c_{m+1,m}^{(x,z,t,\boldsymbol{\theta})} < (1 - \epsilon) c_{m+1,m}^{(x,z,t,\boldsymbol{\theta})}. \tag{35}$$

Second, note that

$$\begin{aligned}\frac{c^{(x,z,t,\boldsymbol{\theta})}_{m+1,m+1}}{c^{(x,z,t,\boldsymbol{\theta})}_{m+1,m}} &= \frac{1}{m+1}\frac{\theta+2m}{\theta+m}\frac{\mathbb{E}[\mathcal{D}_{\theta_1+L_{m+1},\theta_2+m+1-L_{m+1}}(z)]}{\mathbb{E}[\mathcal{D}_{\theta_1+L_m,\theta_2+m-L_m}(z)]},\\ &\leq \frac{1}{m+1}\cdot 2\frac{\mathbb{E}[\mathcal{D}_{\theta_1+L_{m+1},\theta_2+m+1-L_{m+1}}(z)]}{\mathbb{E}[\mathcal{D}_{\theta_1+L_m,\theta_2+m-L_m}(z)]},\\ &< \epsilon, \end{aligned} \tag{36}$$

using Lemma 1 and $m+1 \geq 2K/\epsilon$ for the last inequality. Rearrange (36) and sum with (35) to get (34). □

Proof of Proposition 4. This follows immediately from Propositions 1 and 3. It can also be viewed as an application of Proposition 1 of Beskos et al. (2008) to a function $g(u_1, u_2, u_3) \propto u_1 u_2/u_3$. □

Acknowledgements

We thank Bob Griffiths, Jere Koskela, Omiros Papaspiliopoulos, and Murray Pollock for helpful discussions.

References

Beskos, A. and Roberts, G. O. (2005). Exact simulation of diffusions. *Annals of Applied Probability*, **15**, 2422–2444.

Beskos, A., Papaspiliopoulos, O., and Roberts, G. O. (2006a). Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli*, **12**, 1077–1098.

Beskos, A., Papaspiliopoulos, O., Roberts, G. O., and Fearnhead, P. (2006b). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *Journal of the Royal Statistical Society: Series B*, **68**, 333–382.

Beskos, A., Papaspiliopoulos, O., and Roberts, G. O. (2008). A factorisation of diffusion measure and finite sample path constructions. *Methodology and Computing in Applied Probability*, **10**, 85–104.

Bladt, M. and Sørensen, M. (2014). Simple simulation of diffusion bridges with application to likelihood inference for diffusions. *Bernoulli*, **20**, 645–675.

Bladt, M., Finch, S., and Sørensen, M. (2014). Simulation of multivariate diffusion bridges. *arXiv:1405.7728*.

Bollback, J. P., York, T. L., and Nielsen, R. (2008). Estimation of $2N_e s$ from temporal allele frequency data. *Genetics*, **179**, 497–502.

Chaleyat-Maurel, M. and Genon-Catalot, V. (2009). Filtering the Wright-Fisher diffusion. *ESAIM: Probability and Statistics*, **13**, 197–217.

Chib, S., Pitt, M., and Shephard, N. (2010). Likelihood based inference for diffusion driven state space models. *Working paper*.

Dawson, D. (1978). Geostochastic calculus. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, **6**, 143–168.

Delbaen, F. and Shirakawa, H. (2002). An interest rate model with upper and lower bounds. *Asia-Pacific Financial Markets*, **9**, 191–209.

Devroye, L. *Non-uniform random variate generation*. Springer-Verlag, New York, 1986.

Ethier, S. N. and Griffiths, R. C. (1993). The transition function of a Fleming-Viot process. *Annals of Probability*, **21**, 1571–1590.

Ethier, S. N. and Kurtz, T. G. (1993). Fleming-Viot processes in population genetics. *SIAM Journal of Control and Optimization*, **31**, 345–386.

Favaro, S., Ruggiero, M., and Walker, S. G. (2009). On a Gibbs sampler based random process in Bayesian nonparametrics. *Electronic Journal of Statistics*, **3**, 1556–1566.

Fearnhead, P., Papaspiliopoulos, O., and Roberts, G. O. (2008). Particle filters for partially observed diffusions. *Journal of the Royal Statistical Society: Series B*, **40**, 755–777.

Fearnhead, P., Papaspiliopoulos, O., Roberts, G. O., and Stuart, A. (2010). Random-weight particle filtering of continuous time processes. *Journal of the Royal Statistical Society: Series B*, **72**, 497–512.

Fitzsimmons, P., Pitman, J., and Yor, M. Markovian bridges: construction, Palm interpretation, and splicing. In Çinlar, E., Chung, K. L., Sharpe, M. J., Bass, R. F., and Burdzy, K., editors, *Seminar on stochastic processes*, volume 33 of *Progress in probability*, pages 101–134. Birkhäuser, Boston, 1992.

Golightly, A. and Wilkinson, D. J. (2006). Bayesian sequential inference for nonlinear multivariate diffusions. *Statistics and Computing*, **16**, 323–338.

Golightly, A. and Wilkinson, D. J. (2008). Bayesian sequential inference for nonlinear multivariate diffusions observed with error. *Computational Statistics & Data Analysis*, **52**, 1674–1693.

Gourieroux, C. and Jasiak, J. (2006). Multivariate Jacobi process with application to smooth transitions. *Journal of Econometrics*, **131**, 475–505.

Griffiths, R. C. (1979). A transition density expansion for a multi-allele diffusion model. *Advances in Applied Probability*, **11**, 310–325.

Griffiths, R. C. (1980). Lines of descent in the diffusion approximation of neutral Wright-Fisher models. *Theoretical Population Biology*, **17**, 37–50.

Griffiths, R. C. (1984). Asymptotic line-of-descent distributions. *Journal of Mathematical Biology*, **21**, 67–75.

Griffiths, R. C. (2006). Coalescent lineage distributions. *Advances in Applied Probability*, **38**, 405–429.

Griffiths, R. C. and Li, W.-H. (1983). Simulating allele frequencies in a population and the genetic differentiation of populations under mutation pressure. *Theoretical Population Biology*, **23**, 19–33.

Griffiths, R. C. and Spanò, D. Diffusion processes and coalescent trees. In Bingham, N. H. and Goldie, C. M., editors, *Probability and Mathematical Genetics, Papers in Honour of Sir John Kingman*, volume 378 of *LMS Lecture Note Series*, chapter 15, pages 358–375. Cambridge University Press, 2010.

Griffiths, R. C. and Spanò, D. (2013). Orthogonal polynomial kernels and canonical correlations for Dirichlet measures. *Bernoulli*, **19**, 548–598.

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., and Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, **5**, e1000695.

Jenkins, P. A. (2013). Exact simulation of the sample paths of a diffusion with a finite entrance boundary. *arXiv:1311.5777*.

Jewett, E. M. and Rosenberg, N. A. (2014). Theory and applications of a deterministic approximation to the coalescent model. *Theoretical Population Biology*, **93**, 14–29.

Lukić, S., Hey, J., and Chen, K. (2011). Non-equilibrium allele frequency spectra via spectral methods. *Theoretical Population Biology*, **79**, 203–219.

Malaspinas, A.-S., Malaspinas, O., Evans, S. N., and Slatkin, M. (2012). Estimating allele age and selection coefficient from time-serial data. *Genetics*, **192**, 599–607.

Mena, R. and Ruggiero, M. (2014). Dynamic density estimation with diffusive Dirichlet mixtures. *arXiv:1410.2477*.

Øksendal, B. *Stochastic differential equations*. Springer-Verlag, 5th printing of 6th edition, 2003.

Papaspiliopoulos, O. and Roberts, G. O. (2008). Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika*, **95**, 169–186.

Papaspiliopoulos, O. and Ruggiero, M. (2014). Optimal filtering and the dual process. *Bernoulli*, **20**, 1999–2019.

Papaspiliopoulos, O., Ruggiero, M., and Spanò, D. (2014). Filtering hidden Markov measures. *arXiv:1411.4944*.

Pollock, M., Johansen, A. M., and Roberts, G. O. (2014). On the exact and ϵ-strong simulation of (jump) diffusions. *Bernoulli*. In press. *arXiv:1304.1377*.

Schraiber, J., Griffiths, R. C., and Evans, S. N. (2013). Analysis and rejection sampling of Wright-Fisher diffusion bridges. *Theoretical Population Biology*, **89**, 64–74.

Sermaidis, G., Papaspiliopoulos, O., Roberts, G. O., Beskos, A., and Fearnhead, P. (2013). Markov chain Monte Carlo for exact inference for diffusions. *Scandinavian Journal of Statistics*, **40**, 294–321.

Song, Y. S. and Steinrücken, M. (2012). A simple method for finding explicit analytic transition densities of diffusion processes with general diploid selection. *Genetics*, **190**, 1117–1129.

Steinrücken, M., Wang, Y. X. R., and Song, Y. S. (2013). An explicit transition density expansion for a multi-allelic Wright-Fisher diffusion with general diploid selection. *Theoretical Population Biology*, **83**, 1–14.

Tavaré, S. (1984). Line-of-descent and genealogical processes, and their applications in population genetics models. *Theoretical Population Biology*, **26**, 119–164.

Walker, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics—Simulation and Computation*, **36**, 45–54.

Walker, S. G., Hatjispyros, S. J., and Nicoleris, T. (2007). A Fleming-Viot process and Bayesian nonparametrics. *Annals of Applied Probability*, **17**, 67–80.

Williamson, S. H., Hernandez, R., Fledel-Alon, A., Zhu, L., and Bustamante, C. D. (2005). Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 7882–7887.

Zhao, L., Lascoux, M., Overall, A. D. J., and Waxman, D. (2013). The characteristic trajectory of a fixing allele: a consequence of fictitious selection that arises from conditioning. *Genetics*, **195**, 993–1006.

Paul A. Jenkins
Department of Statistics
University of Warwick
Coventry CV4 7AL, U.K.
E-mail: p.jenkins@warwick.ac.uk

Dario Spanò
Department of Statistics
University of Warwick
Coventry CV4 7AL, U.K.
E-mail: d.spano@warwick.ac.uk