EXACT SIMULATION OF THE WRIGHT–FISHER DIFFUSION

BY PAUL A. JENKINS[1] AND DARIO SPANÒ

University of Warwick

The Wright–Fisher family of diffusion processes is a widely used class of evolutionary models. However, simulation 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 a broad class of Wright–Fisher diffusion processes and their bridges. For those diffusions corresponding to reversible, neutral evolution, our key idea is to exploit an eigenfunction expansion of the transition function; this approach even applies to its infinite-dimensional analogue, the Fleming–Viot process. We then develop an exact rejection algorithm for processes with more general drift functions, including those modelling natural selection, using ideas from retrospective simulation. Our approach also yields methods for exact simulation of the moment dual of the Wright–Fisher diffusion, the ancestral process of an infinite-leaf Kingman coalescent tree. 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., Bladt, Finch and Sørensen (2016), Bladt and Sørensen (2014), Chib, Pitt and Shephard (2010), Golightly and Wilkinson (2006, 2008)]. 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):

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

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

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

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., Bollback, York and Nielsen (2008), Gutenkunst et al. (2009), Malaspinas et al. (2012), Williamson et al. (2005)], and in the characteristics of the trajectories themselves [Schraiber, Griffiths and Evans (2013), Zhao et al. (2013)], as discretely observed data are becoming more and more available (e.g., 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 proposed in biophysics as a model for ion channel dynamics [Dangerfield, Kay and Burrage (2010), Dangerfield et al. (2012)]; they have been studied as hidden Markov signals in filtering problems [Chaleyat-Maurel and Genon-Catalot (2009), Papaspiliopoulos and Ruggiero (2014), Papaspiliopoulos, Ruggiero and Spanò (2014)]; and in Bayesian statistics they have been proposed as models for time-evolving priors [Favaro, Ruggiero and Walker (2009), Griffiths and Spanò (2013), Mena and Ruggiero (2016), Walker, Hatjispyros and Nicoleris (2007)].

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 $\gamma(x) \equiv 0$. In the absence of a method of exact simulation, it is necessary to turn to approximate alternatives such as an Euler–Maruyama scheme. Standard Euler-type methods fail here because simulated paths can easily leave the state space $[0, 1]$, and moreover standard assumptions for weak and strong convergence typically require that the diffusion coefficient be Lipschitz continuous [see Kloeden and Platen (1999)]. Consequently, a number of specialized time-discretization methods have been developed for the Wright–Fisher diffusion with various drifts; when the drift is of the form of $\beta(x)$, see Schurz (1996) for $\theta_1 = \theta_2 = \sigma = 0$, Dangerfield et al. (2012) for $\sigma = 0$, $\theta_1, \theta_2 > 0$, Schraiber, Griffiths and Evans (2013) for $\theta_1 = \theta_2 = 0$, $h = 1/2$ and Neuenkirch and Szpruch (2014) for $\sigma = 0$, $\theta_1, \theta_2 \geq 1$. Other approaches include truncating a spectral expansion of the transition function [Lukić, Hey and Chen (2011), Song and Steinrücken (2012), Steinrücken, Wang and Song (2013)] or numerical solutions of the Kolmogorov equations [Bollback, York and Nielsen (2008), Gutenkunst et al. (2009), Williamson et al. (2005)]. The error introduced by these methods can be difficult to quantify and must often be tested empirically.

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. 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. We build up our algorithm in stages, addressing how to simulate exactly from each of the following:

1. The neutral Wright–Fisher diffusion; that is, with drift

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

where $\theta_1, \theta_2 > 0$ (Section 2).

2. The neutral Wright–Fisher bridge (Section 3); informally, this is the process

$$(X_t)_{t \in [0,T]} \mid X_T = y.$$

3. The Wright–Fisher diffusion and its bridges with a very general class of drift functions (defined later), including drift $\beta(x)$ of the form (2) when $\theta_1, \theta_2 > 0$ (Section 5).

To achieve step 1, the key approach is to exploit an eigenfunction expansion representation of the transition function [see Griffiths and Spanó (2010), for review]. The expansion admits a probabilistic interpretation and, therefore, lends itself to simulation techniques, but these techniques are not straightforward to implement because the distributions involved are known only in infinite series form. Despite this hurdle, here we show that is possible to perform such simulation without error. The technique is very general and so we develop this section not just for the Wright–Fisher diffusion but for the *Fleming–Viot process*, its infinite-dimensional generalization.

To achieve step 2, we obtain a new probabilistic description of the transition function of a neutral Wright–Fisher *bridge*. This is again complicated by the appearance of distributions known only in infinite series form, but from which (we show) realizations can still be obtained by evaluating only a finite number of terms in the series.

Finally, we generalize these techniques to nonneutral processes in step 3. The eigenfunction expansion for a nonneutral process [Barbour, Ethier and Griffiths (2000)] is probabilistically intractable, so we take a different approach: we use the simulated neutral processes as candidates in a rejection algorithm. This uses a retrospective approach similar to that of the "exact algorithms" of Beskos and Roberts (2005), Beskos, Papaspiliopoulos and Roberts (2006, 2008) and Pollock, Johansen and Roberts (2016), which can return exact samples from a class of diffusions using *Brownian motion* as the candidate for rejection. We defer a detailed description to Section 5.1, but for now we note that a direct application of these techniques would require that the target diffusion 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, Papaspiliopoulos and Roberts (2008) point out that their exact algorithm can be adopted to the case of two finite entrance

boundaries, but this approach requires a further technical condition [(A3) below] which does not always hold here. When it does hold, this approach becomes arbitrarily inefficient when 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 the parameters of $\beta(x)$. But now we are armed with the ability to simulate the neutral Wright–Fisher process, which serves as a far more promising candidate than Brownian motion in a rejection algorithm; specifically, it is known that the law of a nonneutral process is absolutely continuous with respect to its neutral counterpart [Dawson (1978), Ethier and Kurtz (1993)]. We develop these ideas in full in Section 5. We also remark that a related approach is taken by Schraiber, Griffiths and Evans (2013), who were interested in simulating nonneutral Wright–Fisher bridges in the absence of mutation. In this context, one can condition each sample path to remain in (0, 1) (otherwise the path could be absorbed at 0 or 1 and could not terminate at a prespecified point), rendering the boundaries inaccessible. They show that the appropriate candidate in this case is a Bessel process of dimension 4, whose boundary at 0 is also of entrance type. However, their method is not exact in the sense given above, since the rejection probabilities are approximated via numerical integration. Furthermore, the Radon–Nikodým derivative of the Wright–Fisher process with respect to the Bessel(4) process is not bounded [another approach developed in Jenkins (2013) for a single entrance boundary suffers a similar limitation]. In any case, a direct comparison with our method is not possible since here we tackle $\theta_1, \theta_2 > 0$ rather than $\theta_1 = \theta_2 = 0$.

The remainder of the paper is structured as follows. In Section 4, we discuss practical considerations of the algorithm; Section 6 fills in one last gap by showing how to construct a nonneutral Wright–Fisher bridge; Section 7 discusses extensions of the algorithm and Section 8 contains the proofs of our results.

2. Simulating the neutral Wright–Fisher process. In this section, we demonstrate how exact simulation from the neutral Wright–Fisher diffusion can be achieved. To aid the exposition, we first focus on a one-dimensional process, and then later extend this idea to the Fleming–Viot process.

2.1. *A transition density expansion in one dimension.* Consider the diffusion satisfying (1) with drift (3). Denote its law by $\mathbb{WF}_{\alpha,x}$ and its 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 [Ethier and Griffiths (1993), Griffiths (1979), Griffiths and Spanó (2010), Tavaré (1984)]:

$$(4) \qquad f(x, y; t) = \sum_{m=0}^{\infty} q_m^{\theta}(t) \sum_{l=0}^{m} \mathcal{B}_{m,x}(l) \mathcal{D}_{\theta_1+l, \theta_2+m-l}(y),$$

Algorithm 1: Simulating from the transition density $f(x, \cdot; t)$ of the neutral Wright–Fisher diffusion with mutation

---

1 Simulate $A_\infty^\theta(t)$
2 Given $A_\infty^\theta(t) = m$, 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

---

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 (4) 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 (4) also provides an immediate method for *simulating* from $f(x, \cdot; t)$. We summarize this idea in Algorithm 1, which first appeared in Griffiths and Li (1983).

Steps 2 and 3 of Algorithm 1 are straightforward, but step 1 requires the PMF $\{q_m^\theta(t) : m = 0, 1, \ldots\}$, 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.

2.2. *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 $\{q_m^\theta(t) : m = 0, 1, \ldots\}$. Were each $q_m^\theta(t)$ available in closed form, then standard inversion sampling would return exact samples from this distribution [see, e.g., Devroye (1986), Chapter 2]: for $U \sim \text{Uniform}[0, 1]$,

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

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

$$q_m^\theta(t) = \sum_{k=m}^{\infty} (-1)^{k-m} a_{km}^\theta e^{-k(k+\theta-1)t/2},$$

(5)

$$\text{where } a_{km}^\theta = \frac{(\theta + 2k - 1)(\theta + m)_{(k-1)}}{m!(k - m)!}.$$

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 (5), 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), Chapter 4], 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

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

such that

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

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, $T_K^-(M) \uparrow \mathbb{P}(X \leq M)$ and $T_K^+(M) \downarrow \mathbb{P}(X \leq M)$ as $K \to \infty$. Hence, for $U \sim \text{Uniform}[0, 1]$ and

$$K_0(M) := \inf\{K \in \mathbb{N} : T_K^-(M) > U \text{ or } T_K^+(M) < U\},$$

the quantity $\inf\{M \in \mathbb{N} : T_{K_0(M)}^-(M) > U\}$ 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 (6) holds with the possible exception of the first few terms in m. For those exceptional terms, (6) 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* (5), *and let*

(7) $$C_m^{(t,\theta)} := \inf\{i \geq 0 : b_{i+m+1}^{(t,\theta)}(m) < b_{i+m}^{(t,\theta)}(m)\}.$$

Then:

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

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

    (iii) $C_m^{(t,\theta)} = 0$ *for all* $m > D_0^{(t,\theta)}$, *where for* $\varepsilon \in [0, 1)$,

$$(8) \quad D_\varepsilon^{(t,\theta)} := \inf\left\{ k \geq \left(\frac{1}{t} - \frac{\theta+1}{2}\right) \vee 0 : (\theta + 2k + 1)e^{-\frac{(2k+\theta)t}{2}} < 1 - \varepsilon \right\}.$$

(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 $m \leq D_0^{(t,\theta)}$, then precompute terms in $q_m^\theta(t)$ until the first time that the coefficients in (5) 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 $\mathbf{k} = (k_0, k_1, \ldots, k_M)$ and

$$(9) \quad S_{\mathbf{k}}^-(M) := \sum_{m=0}^{M} \sum_{i=0}^{2k_m+1} (-1)^i b_{m+i}^{(t,\theta)}(m), \qquad S_{\mathbf{k}}^+(M) := \sum_{m=0}^{M} \sum_{i=0}^{2k_m} (-1)^i b_{m+i}^{(t,\theta)}(m).$$

We summarize this procedure in Algorithm 2.

Of course, the *false* condition in line 15 of Algorithm 2 is never met, but Proposition 1 guarantees that the algorithm still halts in finite time. Further performance

---

Algorithm 2: Simulating from the ancestral process $A_\infty(t)$ of Kingman's coalescent with mutation

---
1 Set $m \longleftarrow 0, k_0 \longleftarrow 0, \mathbf{k} \longleftarrow (k_0)$

2 Simulate $U \sim \text{Uniform}[0, 1]$

3 repeat

4   Set $k_m \longleftarrow \lceil C_m^{(t,\theta)}/2 \rceil$ [eq. (7)]

5   **while** $S_{\mathbf{k}}^-(m) < U < S_{\mathbf{k}}^+(m)$ **do**

6   Set $\mathbf{k} \longleftarrow \mathbf{k} + (1, 1, \ldots, 1)$

7   **end**

8   ;

9   **if** $S_{\mathbf{k}}^-(m) > U$ **then**

10   **return** m

11   **else if** $S_{\mathbf{k}}^+(m) < U$ **then**

12   Set $\mathbf{k} \longleftarrow (k_0, k_1, \ldots, k_m, 0)$

13   Set $m \longleftarrow m + 1$

14   **end**

15 until *false*;

considerations of this algorithm are discussed in Section 4. Also note that computed coefficients a_{km}^θ in (5) can also be stored for future calls to this algorithm.

2.3. *A transition density expansion in higher dimensions.* It is worth pointing out that an interesting by-product of Proposition 1 (and of Algorithm 2) 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. 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:

$$p(\mu, d\nu; t) = \sum_{m=0}^\infty q_m^\theta(t) \int_{E^m} \mu^{\otimes m}(d\xi_1, \ldots, d\xi_m) \Pi_{\theta+m, \frac{m}{\theta+m}\eta_m + \frac{\theta}{\theta+m}P_0}(d\nu),$$

(10)
$$t \geq 0, \mu, \nu \in \mathcal{M}_1(E),$$

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 (10) is thus the following modification of Algorithm 1.

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 2 provides a way of filling the only missing gap (step 1 of Algorithm 3) to make the transition function (10) 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 1 is viable for exact simulation of neutral $(d-1)$-dimensional extensions of the Wright–Fisher diffusion (1) with drift (3).

Algorithm 3: 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

---

3. Simulating a neutral Wright–Fisher bridge. In this section, we demonstrate how exact simulation from the neutral Wright–Fisher diffusion *bridges* can be achieved, via a new probabilistic description of its transition density. For the remainder of the paper, we return to processes of dimension one.

3.1. *A transition density expansion.* Section 2 provides a method for returning exact samples from $f(x, \cdot; t)$ for any fixed $x \in [0, 1]$ and $t > 0$. 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, Pitman and Yor (1993), Schraiber, Griffiths and Evans (2013):

$$(11) \qquad f_{z,t}(x, y; s) = \frac{f(x, y; s) f(y, z; t - s)}{f(x, z; t)}, \qquad 0 < s < t,$$

with $f(\cdot, \cdot; \cdot)$ as in (4). Motivated by (4), 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*

$$(12) \qquad f_{z,t}(x, y; s) = \sum_{m=0}^\infty \sum_{k=0}^\infty \sum_{l=0}^m \sum_{j=0}^k p_{m,k,l,j}^{(x,z,s,t,\boldsymbol{\theta})} \mathcal{D}_{\theta_1+l+j, \theta_2+m+k-l-j}(y),$$

where

$$(13) \quad p_{m,k,l,j}^{(x,z,s,t,\boldsymbol{\theta})} = \mathcal{B}_{m,x}(l) \mathcal{D}_{\theta_1+j, \theta_2+k-j}(z) \mathcal{DM}_{\theta_1+l, \theta_2+m-l; k}(j) \frac{q_m^\theta(s) q_k^\theta(t - s)}{f(x, z; t)},$$

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, where

$$\mathcal{DM}_{a,b;k}(j) = \binom{k}{j} \frac{B(a + j, b + k - j)}{B(a, b)}$$

is the PMF of a beta-binomial random variable on $\{0, 1, \ldots, k\}$.

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. (13)]
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

By Proposition 2, we recognize equation (12) 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 $f_{z,t}(x, y; s)$.

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 (13); each term in this ratio is known only as an infinite series, as we have seen. We address this in the following subsection.

3.2. *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 2.2 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 (13). The first two terms have been dealt with already in Section 2.2, 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 (4) and (5) to write

$$f(x, z; t) = \sum_{m=0}^{\infty} \sum_{k=m}^{\infty} (-1)^{k-m} c_{k,m}^{(x,z,t,\boldsymbol{\theta})}$$

(14)

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

and $L_m \sim \text{Binomial}(m, x)$. Our strategy is to group the triangular array of coefficients $(c_{k,m}^{(x,z,t,\boldsymbol{\theta})})_{k \geq m}$ in such a way that, with the exception of the first few terms, they exhibit a property analogous to (6). We will compare terms in the sequence $(d_i)_{i=0,1,\dots}$ of antidiagonals, defined by

$$\text{(15)} \quad d_{2m} = \sum_{j=0}^{m} c_{m+j,m-j}^{(x,z,t,\boldsymbol{\theta})}, \qquad d_{2m+1} = \sum_{j=0}^{m} c_{m+1+j,m-j}^{(x,z,t,\boldsymbol{\theta})}, \qquad m = 0, 1, \dots$$

(see Figure 2), and dropping the superscript for convenience. Notice that the coefficients within each entry of this sequence are all multiplied by the same sign in (14), so that $f(x, z; t) = d_0 - d_1 + d_2 - d_3 + \cdots$ 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

$$(16) \qquad E^{(t,\theta)} := \inf\{m \geq 0 : 2j \geq C_{m-j}^{(t,\theta)} \text{ for all } j = 0, \dots, m\},$$

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 $L_m \sim \text{Binomial}(m, x)$ and*

$$(17) \qquad K^{(x,z)} := \frac{x}{z} + \frac{1-x}{1-z}.$$

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

$$(18) \qquad \mathbb{E}\big[\mathcal{D}_{\theta_1 + L_{m+1}, \theta_2 + m + 1 - L_{m+1}}(z)\big] < K^{(x,z)}\mathbb{E}\big[\mathcal{D}_{\theta_1 + L_m, \theta_2 + m - L_m}(z)\big].$$

Using Lemma 1, we are in a position to obtain the required analogue of property (6).

PROPOSITION 3. *Let $D_\varepsilon^{(t,\theta)}$, $(d_i)_{i=0,1,\dots}$, $E^{(t,\theta)}$, and $K^{(x,z)}$ be defined as in (8), (15), (16) and (17), respectively, and $\varepsilon \in (0, 1)$. Then*

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

for all $m \geq E^{(t,\theta)} \vee D_\varepsilon^{(t,\theta)} \vee 2K^{(x,z)}/\varepsilon$.

We can now combine Propositions 1 and 3 in order to construct a sequence amenable to simulation from $\{p_{m,k,l,j}^{(x,z,s,t,\boldsymbol{\theta})} : m, k, l, j \in \mathbb{N}\}$ [equation (13)], via the alternating series method.

PROPOSITION 4. *Define*

$$(19) \qquad F_{m,k,l,j}^{(s,t,\theta,x,z)} := C_m^{(s,\theta)} \vee C_k^{(t-s,\theta)} \vee E^{(t,\theta)} \vee D_\varepsilon^{(t,\theta)} \vee 2K^{(x,z)}/\varepsilon$$

and

$$e_{m,k,l,j}(v) := \mathcal{B}_{m,x}(l)\mathcal{D}_{\theta_1+j,\theta_2+k-j}(z)\mathcal{DM}_{\theta_1+l,\theta_2+m-l;k}(j)$$
$$\times \left[\sum_{i=0}^{v}(-1)^i b_{m+i}^{(s,\theta)}(m)\right]\left[\sum_{i=0}^{v}(-1)^i b_{k+i}^{(t-s,\theta)}(k)\right] \Big/ \sum_{i=0}^{v+1}(-1)^i d_i.$$

Then for $2v \geq F_{m,k,l,j}^{(s,t,\theta,x,z)}$,

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

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

---

1 Set $n \longleftarrow 0$, $v_0 \longleftarrow 0$, $\mathbf{v} \longleftarrow (v_0)$
2 Simulate $U \sim \text{Uniform}[0, 1]$
3 repeat
4   Set $v_n \longleftarrow \lceil F_{\Sigma(n)}/2 \rceil$ [eq. (19)]
5   **while** $V_{\mathbf{v}}^-(n) < U < V_{\mathbf{v}}^+(n)$ **do**
6 | Set $\mathbf{v} \longleftarrow \mathbf{v} + (1, 1, \ldots, 1)$
7   **end**
8   ;
9   **if** $V_{\mathbf{v}}^-(n) > U$ **then**
10 | **return** $\Sigma(n)$
11   **else if** $V_{\mathbf{v}}^+(n) < U$ **then**
12 | Set $\mathbf{v} \longleftarrow (v_0, v_1, \ldots, v_n, 0)$
13 | Set $n \longleftarrow n + 1$
14   **end**
15 until *false*;

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 2, we also introduce $\mathbf{v} = (v_0, v_1, \ldots, v_N)$ and

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

4. Performance of algorithms for neutral processes. There are several easy improvements to the underlying Algorithm 2. 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 $\mu^{(t,\theta)}$ in Theorem 1 below. It may also be possible to improve on Algorithm 2 by allowing different $q_m^\theta(t)$ to be refined at different rates, that is, 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 2; these in turn depend on

the quantities $D_0^{(t,\theta)}$ and $C_m^{(t,\theta)}$ (recall Proposition 1). These quantities are in general manageably small, suggesting that the number of coefficients that need to be computed in Algorithms 2 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. The following result makes precise the complexity of Algorithm 2.

PROPOSITION 5. *As* $t \to 0$:

  (i) $C_m^{(t,\theta)} = O(t^{-1})$;
  (ii) $\max_m C_m^{(t,\theta)} = O(t^{-1} \log(t^{-1}))$;
  (iii) $D_0^{(t,\theta)} = o(t^{-(1+\kappa)})$, *for any* $\kappa > 0$.

Let $N^{(t,\theta)}$ *denote the total number of coefficients that must be computed in an implementation of Algorithm* 2. *Then* $\mathbb{E}[N^{(t,\theta)}] < \infty$, *and in particular*:

  (iv) $\mathbb{E}[N^{(t,\theta)}] = o(t^{-(1+\kappa)})$, *for any* $\kappa > 0$.

The growth in Algorithm 2 as $t \to 0$ is closely related to the well-known numerical instability of (5) as $t \to 0$ [Griffiths (1984)], which afflicts any method based on the expansion (4) [or an expansion using a basis of orthogonal polynomials, which is equivalent to (5); Griffiths and Spanó (2010)]. In any practical implementation of our algorithm, we are obliged to use an approximation should the separation between two points be very small, and we found Algorithm 2 to fail for $t < 0.05$ or so. One option is to revert to the Euler–Maruyama approximation for small t. Alternatively, there has been much previous work in coalescent theory on approximating the distribution (5) [e.g., Griffiths (1984, 2006), Jewett and Rosenberg (2014)]; by inserting those approximations into Algorithm 2 they readily define new algorithms for approximate simulation of the dual *diffusion*. We consider the following result, due to Griffiths (1984), Theorem 4.

THEOREM 1 (Griffiths (1984)). *Suppose* $\beta = \frac{1}{2}(\theta - 1)t$, *and let*

$$\mu^{(t,\theta)} = \frac{2\eta}{t},$$

$$(\sigma^{(t,\theta)})^2 = \begin{cases} \dfrac{2\eta}{t}(\eta + \beta)^2\Big(1 + \dfrac{\eta}{\eta + \beta} - 2\eta\Big)\beta^{-2}, & \beta \neq 0, \\ \dfrac{2}{3t}, & \beta = 0, \end{cases}$$

where

$$\eta = \begin{cases} \dfrac{\beta}{e^\beta - 1}, & \beta \neq 0, \\ 1, & \beta = 0. \end{cases}$$

Then $\mathbb{P}((A_\infty^\theta(t) - \mu^{(t,\theta)})(\sigma^{(t,\theta)})^{-1} \leq x) \to \Phi(x)$ *as* $t \to 0$, *where* $\Phi(\cdot)$ *is the cumulative distribution function* (*CDF*) *of a standard normal random variable*.

[The statement in Griffiths (1984) is missing the factor β^{-2}.] To apply this approximation in practice when t is small, we replace line 1 in Algorithm 1 with

1′ . Simulate $A_\infty(t) \sim N(\mu^{(t,\theta)}, (\sigma^{(t,\theta)})^2)$ and round it to the nearest nonnegative integer.

4.1. *Comparison with Euler–Maruyama simulation.* To check the correctness of our algorithm and to compare its performance to the Euler–Maruyama simulation, we performed the following experiment. We fixed $\theta_1 = \theta_2 = 1/2$, explored various fixed values of x and t, and for each parameter combination drew 10,000 samples from $f(x, \cdot; t)$ using Algorithm 1. To quantify whether the resulting sample was consistent with the true distribution $f(x, \cdot; t)$, we performed a one-sample Kolmogorov–Smirnov (K–S) test. We then performed the same experiment instead using Euler simulation with various stepsizes δ. For this purpose, we used the Balanced Implicit Split Step (BISS) algorithm of Dangerfield et al. (2012), an Euler-type algorithm with some advanced modifications that guarantee each sample path stays within [0, 1], and which is state-of-the-art for $\theta_1, \theta_2 > 0$. (Their algorithm has an additional tuning parameter ε; we followed their recommendation and set $\varepsilon = \delta/4$.)

To obtain an accurate expression for the CDF of the true distribution for use within the K–S statistic, we exploited the fact that, in the special case $\theta_1 = \theta_2 = 1/2$, a Lamperti transformation of (1) (or conversion to Stratonovich form) leads to

$$X_t = \frac{1}{2}(1 - \cos B_t),$$

where $(B_t)_{t \geq 0}$ is a Brownian motion commenced from $\arccos(1 - 2x)$ and reflecting at 0 and π. A rapidly converging series expression for the CDF of B_t (and hence of X_t) is available [Linetsky (2005), equation (26)]; the first 1000 terms in the series suggested convergence to machine precision and were used as the reference CDF.

As is evident from Table 1, exact simulation strongly outperforms the BISS algorithm over almost all the parameter combinations investigated. Over a timescale of $t \gtrsim 0.1$, errors in the Euler-type method accumulate sufficiently fast that the resulting samples are easy to reject in a K–S test. Even reducing the stepsize so that its running time is several orders of magnitude greater than the exact method provides only a modest improvement to the quality of the sample. Note also that the performance of the BISS algorithm deteriorates for paths started close to the boundary (compare $x = 0.01$ with $x = 0.5$), whereas the exact method is indifferent to starting position. One reason p-values are persistently small for the BISS algorithm is that sample paths are constrained by construction to stay inside

TABLE 1

Comparison of exact simulation methods for the neutral Wright–Fisher diffusion. BISS: Algorithm of Dangerfield et al. (2012) with stepsize δ. Exact: Algorithm 1. Exact′: Algorithm 1 with the approximation of Griffiths (1984) described on page 1491. Tabulated are the computing time needed to simulate 10,000 sample paths and the p-value of a K–S test applied to the resulting collection of endpoints. Paths are initiated at $X_0 = x$ and run for length t. Mutation parameters are $\theta_1 = \theta_2 = 1/2$

|		$x = 0.01$				$x = 0.5$	
Method	**δ**	**Time (s)**	**p-value**	**Method**	**δ**	**Time (s)**	**p-value**
				$t = 0.01$			
BISS	10^{-3}	0.03	$<10^{-100}$	BISS	10^{-3}	0.02	8.0×10^{-3}
	10^{-4}	0.05	$<10^{-100}$		10^{-4}	0.05	0.36
	10^{-5}	0.30	$<10^{-100}$		10^{-5}	0.30	0.22
	10^{-6}	2.83	$<10^{-100}$		10^{-6}	2.78	0.30
	10^{-7}	27.61	$<10^{-100}$		10^{-7}	27.44	0.78
Exact′	—	0.19	0.94	Exact′	—	0.17	0.18
				$t = 0.05$			
BISS	10^{-3}	0.04	$<10^{-100}$	BISS	10^{-3}	0.04	9.0×10^{-4}
	10^{-4}	0.16	$<10^{-100}$		10^{-4}	0.16	0.35
	10^{-5}	1.40	$<10^{-100}$		10^{-5}	1.39	0.53
	10^{-6}	13.77	$<10^{-100}$		10^{-6}	13.66	0.02
	10^{-7}	138.08	$<10^{-100}$		10^{-7}	137.06	0.72
Exact	—	0.35	0.09	Exact	—	0.34	0.64
Exact′	—	0.17	0.30	Exact′	—	0.17	0.97
				$t = 0.5$			
BISS	10^{-3}	0.16	$<10^{-100}$	BISS	10^{-3}	0.16	1.2×10^{-28}
	10^{-4}	1.43	$<10^{-100}$		10^{-4}	1.43	9.0×10^{-18}
	10^{-5}	14.07	$<10^{-100}$		10^{-5}	14.08	9.4×10^{-17}
	10^{-6}	137.82	$<10^{-100}$		10^{-6}	138.23	6.3×10^{-13}
	10^{-7}	1378.33	$<10^{-100}$		10^{-7}	1368.33	8.1×10^{-13}
Exact	—	0.19	0.16	Exact	—	0.19	0.81
Exact′	—	0.17	0.21	Exact′	—	0.17	0.60
				$t = 5$			
BISS	10^{-2}	0.17	$<10^{-100}$	BISS	10^{-2}	0.16	$<10^{-100}$
	10^{-3}	1.43	$<10^{-100}$		10^{-3}	1.43	$<10^{-100}$
	10^{-4}	14.01	$<10^{-100}$		10^{-4}	14.04	$<10^{-100}$
	10^{-5}	137.95	$<10^{-100}$		10^{-5}	138.09	4.7×10^{-100}
	10^{-6}	1375.75	$<10^{-100}$		10^{-6}	1378.54	1.6×10^{-100}
Exact	—	0.18	0.58	Exact	—	0.18	0.88
Exact′	—	0.17	1.2×10^{-47}	Exact′	—	0.17	0.49
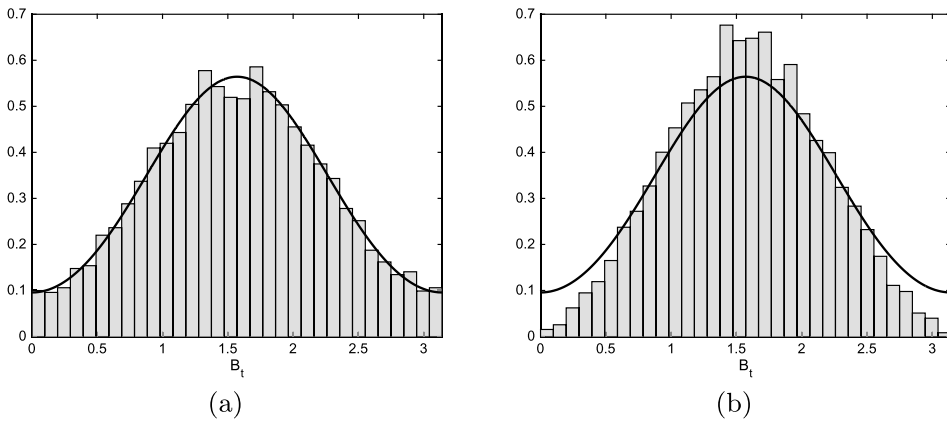
FIG. 1.   *Histogram of* 10,000 *simulated points using* (a) *Algorithm* 1 *and* (b) *the Euler-type BISS algorithm of* Dangerfield et al. (2012). *Parameters are* $\theta_1 = \theta_2 = 1/2$, $x = 0.5$, $t = 0.5$, $\delta = 10^{-6}$. *For visual clarity, plotted are samples of the driving reflecting Brownian motion* $B_t = \arccos(1 - 2X_t)$ *rather than* X_t (*since the density function of* B_t—*shown as a solid line—is bounded and can be calculated*; *see main text*).

$[\delta/4, 1 - \delta/4]$, yet the narrow region close to the boundaries is precisely where we expect to find much of the probability mass for many choices of parameters. An example of how this affects the resulting transition density is given in Figure 1. By contrast, in no application of the K–S test to samples from our method would we have rejected at level 0.05 the hypothesis that they were drawn from the true distribution. Only over short timescales and away from the boundaries (Table 1, $t \leq 0.05$ and $x = 0.5$) do the two methods seem comparable. At $t = 0.05$, the same computational investment as in the exact method but applied to the BISS method buys a stepsize of about $\delta = 10^{-4}$, which is adequate in the interior of the state space ($x = 0.5$) but not near the boundaries ($x = 0.01$).

When $t \geq 0.05$, it is also possible to compare the exact method both with and without the approximation of Griffiths (1984) (see page 1491); the two versions exhibit similar running times and generate *bona fide* samples from the true distribution (according to a K–S test) for moderate t. However, the approximate version deteriorates (as reported by the K–S p-value) for large t (see Table 1, $t = 5$), away from its asymptotic regime. Thus, a suitable rule-of-thumb is to use the exact algorithm for $t \geq 0.05$ and its approximate version for $t < 0.05$, with the two methods in good agreement in their region of overlap in t.

We investigated several other choices of x and t with predictable results (not shown); for example, performance of the exact method seems unrelated to starting position x, while the BISS method deteriorates even further as $x \to 0$.

5. Simulating the nonneutral Wright–Fisher process.   In this section, we develop an exact rejection algorithm for simulating from the Wright–Fisher diffu-

sion (1) with general drift. We make use of retrospective sampling techniques for the exact simulation of diffusions, which we first summarize.

5.1. *Overview of the exact algorithm.* Here we give a brief overview of the exact algorithm (EA) of Beskos and Roberts (2005), Beskos, Papaspiliopoulos and Roberts (2006, 2008) and Pollock, Johansen and Roberts (2016), and we refer the reader 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

$$(20) \qquad dX_t = \mu(X_t)\,dt + dB_t, \qquad X_0 = x_0, t \in [0,T],$$

with μ assumed to satisfy the requirements for (20) 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,

$$(21) \qquad \frac{d\mathbb{Q}_{x_0}}{d\mathbb{W}_{x_0}}(X) = \exp\left\{ \int_0^T \mu(X_t)\,dX_t - \frac{1}{2}\int_0^T \mu^2(X_t)\,dt \right\};$$

(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, (21) can be reexpressed as

$$(22) \qquad \frac{d\mathbb{Q}_{x_0}}{d\mathbb{W}_{x_0}}(X) \propto \exp\{A(X_T) - A^+\} \exp\left\{ -\int_0^T [\phi(X_t) - \phi^-]\,dt \right\}.$$

Written in this form, the right-hand side of (22) 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 (22). 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 (22) is the probability that all points in a Poisson point process $\boldsymbol{\Phi} = \{(t_j, \psi_j) : j = 0, 1, \ldots\}$ 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 $X \sim \mathbb{Q}_{x_0}$.

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. If ϕ is bounded above, by ϕ^+ say, then Algorithm 6 can be implemented with finite computation by thinning $\boldsymbol{\Phi}$ to a Poisson Point process on $[0, T] \times [0, \phi^+ - \phi^-]$;

Algorithm 6: Exact algorithm for simulating the path of a diffusion process with law $\mathbb{Q}_x$

---

1 repeat

2   Simulate $\boldsymbol{\Phi}$, a Poisson point process on $[0, T] \times [0, \infty)$

3   Simulate $U \sim \text{Uniform}[0, 1]$

4   Given $\boldsymbol{\Phi} = \{(t_j, \psi_j) : j = 0, 1, \ldots\}$, simulate $B \sim \mathbb{W}_x$ at times $(t_j)_{j=0,1,\ldots}$ and at time T.

5   **if** $\phi(B_{t_j}) - \phi^- \leq \psi_j$ *for all* $j = 0, 1, \ldots$ *and* $U \leq \exp\{A(B_T) - A^+\}$ **then**

6   **return** $\{(t_j, B_{t_j}) : j = 0, 1 \ldots\} \cup \{(T, B_T)\}$

7   **end**

8 until *false*;

$|\boldsymbol{\Phi}|$ is then almost surely finite. (However, this requirement on ϕ can be relaxed [Beskos, Papaspiliopoulos and Roberts (2006, 2008)].)

We remark that 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.

5.2. *Exact algorithm for the Wright–Fisher diffusion.* As noted earlier, the requirements (A1–A4) need not hold when our target is the Wright–Fisher diffusion. However, related techniques can be used if the candidate process is chosen to be another Wright–Fisher diffusion, and in particular one with the same mutation parameters. Denote the target law $\mathbb{WF}_{\gamma,x_0}$ (with drift γ), and the candidate law $\mathbb{WF}_{\alpha,x_0}$ [with drift (3)]. We will write $\gamma \in \mathcal{WF}$ if there exists a drift α of the form (3) such that the following hold:

(WF1) The Radon–Nikodým derivative of $\mathbb{WF}_{\gamma,x_0}$ with respect to $\mathbb{WF}_{\alpha,x_0}$ exists and is given by Girsanov's formula,

(23)
$$\frac{d\mathbb{WF}_{\gamma,x_0}}{d\mathbb{WF}_{\alpha,x_0}}(X) = \exp\left\{\int_0^T \frac{\gamma(X_t) - \alpha(X_t)}{X_t(1 - X_t)} \, dX_t \right.$$
$$\left. - \frac{1}{2}\int_0^T \frac{\gamma^2(X_t) - \alpha^2(X_t)}{X_t(1 - X_t)} \, dt\right\}.$$

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

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

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

Algorithm 7: Exact algorithm for simulating the path of a diffusion process
with law $\mathbb{WF}_{\gamma,x}$

---

1 repeat
2   Simulate $\mathbf{\Phi}$, a Poisson point process on $[0, T] \times [0, \widetilde{\phi}^+ - \widetilde{\phi}^-]$
3   Simulate $U \sim \text{Uniform}[0, 1]$
4   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
5   **if** $\widetilde{\phi}(X_{t_j}) - \widetilde{\phi}^- \le \psi_j$ *for all* $j = 0, 1, \ldots, J$ *and* $U \le \exp\{\widetilde{A}(X_T) - \widetilde{A}^+\}$
      then
6 $\mid$ **return** $\{(t_j, X_{t_j}) : j = 0, 1 \ldots, J\} \cup \{(T, X_T)\}$
7   **end**
8 until *false*;

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

Specific conditions on α and γ to satisfy (WF1) are detailed, for example, in Theorem 8.6.8 in Øksendal (2003). [This theorem imposes some unduly restrictive conditions to ensure that the SDE has a unique weak solution, but this can be established for (1) by other means; Pardoux (2009), Chapter 4.] Following Section 5.1, we apply Itô's lemma to $\widetilde{A}(x)$ to reexpress (23) as

$$(24) \qquad \frac{d\mathbb{WF}_{\gamma,x_0}}{d\mathbb{WF}_{\alpha,x_0}}(X) \propto \exp\{\widetilde{A}(X_T) - \widetilde{A}^+\} \exp\left\{ -\int_0^T [\widetilde{\phi}(X_t) - \widetilde{\phi}^-] \, dt \right\}.$$

We recognize the rightmost term in (24) as the probability that no points in a Poisson point process on $[0, T] \times [0, \widetilde{\phi}^+ - \widetilde{\phi}^-]$ lie in the epigraph of $t \mapsto \widetilde{\phi}(X_t) - \widetilde{\phi}^-$. Hence, Algorithm 7 returns exact samples from $\mathbb{WF}_{\gamma,x_0}$. Step 4 of Algorithm 7 is achieved via Algorithm 1. Once a skeleton is accepted, further points can be filled in via Algorithm 4.

5.3. *A class of drifts for which exact simulation is possible.* The class $\mathcal{WF}$ defines the set of drifts for which exact simulation is possible. Here, we show that $\mathcal{WF}$ contains all the most popular population genetics diffusion processes with mutation and selection (including frequency-dependent selection), whose drift admits the general form

$$(25) \qquad\qquad \gamma(x) = \alpha(x) + x(1 - x)\eta(x),$$

where α is as in (3) for some $\theta_1, \theta_2 > 0$ and η is a reasonably regular function of x. The case of diploid selection (2) is recovered by setting $\eta(x) \propto (x + h(1 - 2x))$. For general η the properties of diffusion processes with drift (25) and of the corresponding genealogies are studied, among others, by Coop and Griffiths (2004).

PROPOSITION 6. *Any drift of the form* (25), *for α as in* (3) *for some $\theta_1, \theta_2 > 0$ and for η continuously differentiable in* (0, 1), *is a member of $\mathcal{WF}$*.

It is worth noting that the additive structure in the drift γ (i.e., with a selection component added to the mutation component α) is a widely accepted and theoretically justified property of all population genetics diffusion models [see, e.g., the discussion in Karlin and Taylor (1981), pages 186–187].

5.3.1. *Example*: *Wright–Fisher diffusion with diploid selection.* By Proposition 6, the drift β [equation (2)] satisfies $\beta \in \mathcal{WF}$. In fact, in the haploid case ($h = 1/2$), there is much simplification: $\widetilde{\phi}$ is a quadratic polynomial for which analytic bounds are available, and $\widetilde{A}(x) = \sigma x/2$. We implemented our exact algorithm on this model, and investigated its performance by considering several combinations of parameters; results are shown in Table 2. For moderate selection

TABLE 2
Performance of Algorithm 7 applied to the Wright–Fisher diffusion with symmetric mutation and additive selection. Each row reports means (per accepted path) across a simulation to generate 1000 accepted paths. Paths were initiated at $X_0 = x$ and run for time t, with mutation parameters $\theta_1 = \theta_2 = 0.01$ and selection parameter σ (and $h = 0.5$). Reported are the total numbers (per accepted path) of: attempts, Poisson points simulated, coefficients $b_k^{(t,\theta)}(m)$ needed, random variables generated (i.e., the aggregate of all draws from underlying constituent distributions: uniforms, betas and so on), the number of times the simulation resorted to the approximation of Theorem 1, and the running time in milliseconds

t	x	Attempts	Poisson points	Coefficients	Random variables	G1984	Time (ms)
				$\sigma = 1$			
0.1	0.5	1.27	0.00	116.07	7.37	0.00	0.007
0.1	0.25	1.40	0.00	132.87	8.01	0.00	0.008
0.1	0.01	1.58	0.01	141.34	8.91	0.01	0.009
0.5	0.5	1.23	0.03	18.65	7.22	0.00	0.003
0.5	0.25	1.48	0.03	21.60	8.45	0.01	0.003
0.5	0.01	1.58	0.03	23.84	9.02	0.01	0.003
5.0	0.5	1.29	0.24	5.23	8.58	0.00	0.002
5.0	0.25	1.49	0.27	6.18	9.64	0.01	0.002
5.0	0.01	1.67	0.28	7.36	10.79	0.01	0.002
				$\sigma = 10$			
0.1	0.5	11.83	3.72	995.92	62.68	1.83	0.062
0.1	0.25	41.69	12.97	3714.82	224.95	7.86	0.225
0.1	0.01	145.73	45.96	14,937.52	856.21	45.88	0.879
0.5	0.5	13.16	20.96	641.52	109.00	2.47	0.054
0.5	0.25	43.82	69.05	2729.80	399.95	10.89	0.214
0.5	0.01	149.21	235.34	17,044.43	1869.54	71.54	1.185

($|\sigma| = 1$), 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).

To make this observation more precise, Beskos, Papaspiliopoulos and Roberts (2006), Proposition 3, obtained an explicit upper bound on the expected number of Poisson points required of Algorithm 6, and hence on the computational complexity of the algorithm. A careful reading of their result shows that it relies on the existence of the bounds $\phi^{\pm}$ but does not depend on the laws of the diffusions involved, and carries over easily to the Wright–Fisher case. We therefore omit a proof of the following.

PROPOSITION 7. *Let $N^{(T,\theta)}$ denote the number of Poisson points required until the first accepted path. Then*

$$\mathbb{E}[N^{(t,\theta)}] \leq (\widetilde{\phi}^+ - \widetilde{\phi}^-)T e^{(\widetilde{\phi}^+ - \widetilde{\phi}^-)T + \widetilde{A}^+}.$$

An immediate consequence of Proposition 7 is that the complexity of simulating a path of length KT is $O((\widetilde{\phi}^+ - \widetilde{\phi}^-)KT e^{(\widetilde{\phi}^+ - \widetilde{\phi}^-)KT + \widetilde{A}^+})$ as $KT \to \infty$. However, superior performance can be achieved by splitting the problem into K separate simulations; the complexity is then $O((\widetilde{\phi}^+ - \widetilde{\phi}^-)KT e^{(\widetilde{\phi}^+ - \widetilde{\phi}^-)T + \widetilde{A}^+})$, which is linear in K as in Beskos, Papaspiliopoulos and Roberts (2006).

As this is a statement about the complexity of the algorithm as the path length increases, it continues to hold even if we account for the increasing cost associated with each Poisson point as $T \to 0$, as quantified by Proposition 5. In practice, one might wish to optimize the choice of K and T for a given path length KT. In our application, we must be prepared to introduce an additional constraint in order to fix T some distance away from 0 (and, as we recommend above, the choice $T \geq 0.05$ seems adequate).

6. Simulating a nonneutral Wright–Fisher bridge. For completeness, here we provide an algorithm for simulating a nonneutral Wright–Fisher *bridge* (Algorithm 8). This follows immediately from the previous sections; the only modification is to note that the appropriate candidate process is the corresponding neutral bridge, which can be simulated via Algorithm 4. The rest follows exactly as in the Brownian case; see Beskos, Papaspiliopoulos and Roberts (2006), Section 6.2, for details.

Algorithm 8: Exact algorithm for simulating the path of a diffusion process with law $\mathbb{WF}_{\gamma,x}$ conditioned on $X_T = y$

---

1 repeat

2   Simulate $\boldsymbol{\Phi}$, a Poisson point process on $[0, T] \times [0, \widetilde{\phi}^+ - \widetilde{\phi}^-]$

3   Given $\boldsymbol{\Phi} = \{(t_j, \psi_j) : j = 0, 1, \ldots, J\}$, simulate $X \sim (\mathbb{WF}_{\alpha,x} \mid X_T = y)$
     at times $(t_j)_{j=0,1,\ldots,J}$

4   **if** $\widetilde{\phi}(X_{t_j}) - \widetilde{\phi}^- \leq \psi_j$ *for all* $j = 0, 1, \ldots, J$ **then**

5   **return** $\{(t_j, X_{t_j}) : j = 0, 1 \ldots, J\} \cup \{(T, y)\}$

6   **end**

7 until *false*;

---

7. Discussion. In this paper, we have shown how to simulate exactly from the scalar Wright–Fisher diffusion, as well as a number of important and closely related processes: these include the ancestral process of an ∞-leaf Kingman coalescent tree, the Fleming–Viot process with parent-independent mutation, the non-neutral Wright–Fisher diffusion and neutral and nonneutral Wright–Fisher bridges. Some interesting open problems remain, including mutation operators which do not lead to reversible diffusions, and the problem of sampling from $(d-1)$-dimensional $(2 < d \leq \infty)$ Wright–Fisher bridges.

It is also important to notice that, in order to employ the machinery proposed in this paper, the mutation parameters θ_1, θ_2 in the α-component of a general drift γ need to be both positive: models of the form (25) with no mutation ($\alpha = 0$) or with one-directional mutation (with only one mutation parameter positive and the other null) have at least one absorbing boundary and, therefore, there cannot be absolute continuity with respect to a stationary Wright–Fisher diffusion with a transition density expansion of the form of (12). For such cases, series expansions are in fact available with a structure similar to (4) [see, e.g., Ethier and Griffiths (1993)] and we believe it should be relatively simple to adapt our method to encompass selection models with absorbing boundaries, a goal we do not pursue here. For choices of drifts γ more general than (25), arising possibly in models beyond population genetics, it is harder to specify conditions for (WF1) verifiable in a simple way by inspection of γ. The determination of which drift functions γ guarantee (assuming identical diffusion coefficients) absolute continuity with respect to a Wright–Fisher process, seems to be, to our knowledge, an open problem. We expect that most of what is affected by the drift pertains to the behaviour of the process at the boundaries. The problem, however, might be delicate and not just limited to making sure that γ prevents the boundaries from being absorbing: for example, if for the diffusion with drift γ the boundaries are entrance or reflecting, the rate of escape from the boundaries might still make its sample paths qualitatively different from those of any Wright–Fisher diffusion, even within the same

boundary regime, respectively entrance or reflecting. To support this conjecture, we remark that the very same circumstances induce mutual singularity in squared Bessel processes (whose diffusion coefficient is $\sqrt{x}$ hence quite similar near zero to the Wright–Fisher diffusion): it is well known that any two squared Bessel processes starting at 0 are mutually singular whenever their drifts differ, even if they are both within the same boundary regime [see Pitman and Yor (1981), Lemma (2.1) and references therein].

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

$$(26) \qquad \frac{b_{k+1}^{(t,\theta)}(m)}{b_k^{(t,\theta)}(m)} = \frac{\theta + m + k - 1}{k - m + 1} \cdot \frac{\theta + 2k + 1}{\theta + 2k - 1} e^{-\frac{(2k+\theta)t}{2}} =: f_m^\theta(k) e^{-\frac{(2k+\theta)t}{2}},$$

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 (26) is subsequently decreasing in k monotonically to 0. Part (i) follows for the finite k $(= m + i)$ for which the right-hand side of (26) 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 (26) can be made independent of m by noting that

$$(27) \qquad\qquad\qquad f_m^\theta(k) < f_k^\theta(k) = \theta + 2k + 1.$$

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 $D_0^{(t,\theta)}$ is one way to express this condition, since the maximum of $(\theta + 2k + 1)e^{-(2k+\theta)t/2}$ which separates the two solutions is at $k = (\frac{1}{t} - \frac{\theta+1}{2})$. 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 (4) into (11), 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 $\Gamma(x + 1) = x\Gamma(x)$,

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

$$= \left[\frac{m+1}{m+1-l}(1-x)\frac{\theta+m}{\theta_2+m-l}(1-z) \right]\mathbb{P}(L_m = l)\mathcal{D}_{\theta_1+l,\theta_2+m-l}(z)$$

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

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

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 $l = 0, \ldots, \lfloor mz \rfloor$,

(28)
$$\sum_{l=0}^{\lfloor mz \rfloor} \mathbb{P}(L_{m+1} = l)\mathcal{D}_{\theta_1+l,\theta_2+m+1-l}(z)$$

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

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

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

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

Finally, sum (28) and (29) to yield (18), noting that the overlap on the right-hand side at $l = \lfloor mz \rfloor$ necessitates the given definition of $K^{(x,z)}$ (instead of the simpler bound $\frac{1-x}{1-z} \vee \frac{x}{z}$). □

PROOF OF PROPOSITION 3.   First, since $m \geq E^{(t,\theta)}$ 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

$$(30) \qquad c_{m+j+1,m-j}^{(x,z,t,\boldsymbol{\theta})} < c_{m+j,m-j}^{(x,z,t,\boldsymbol{\theta})}.$$

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

$$(31) \qquad \sum_{j=0}^{m} c_{m+1+j,m-j}^{(x,z,t,\boldsymbol{\theta})} < \sum_{j=0}^{m} c_{m+j,m-j}^{(x,z,t,\boldsymbol{\theta})},$$

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 (30) and then sum over $j = 1, \ldots, m$:

$$(32) \qquad \sum_{j=1}^{m} c_{m+2+j,m-j}^{(x,z,t,\boldsymbol{\theta})} < \sum_{j=1}^{m} c_{m+1+j,m-j}^{(x,z,t,\boldsymbol{\theta})}.$$

It then suffices to show

$$(33) \qquad c_{m+1,m+1}^{(x,z,t,\boldsymbol{\theta})} + c_{m+2,m}^{(x,z,t,\boldsymbol{\theta})} < c_{m+1,m}^{(x,z,t,\boldsymbol{\theta})},$$

for if we sum (32) and (33) we obtain $d_{2m+2} < d_{2m+1}$ as required (Figure 2). To show (33), 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 (26) and the inequality following from (27). Hence, choosing $k = m + 1$ and noting that $m \geq C_\varepsilon^{(t,\theta)}$,

$$(34) \qquad c_{m+2,m}^{(x,z,t,\boldsymbol{\theta})} \leq (\theta + 2k + 1)e^{-\frac{(2k+\theta)t}{2}} c_{m+1,m}^{(x,z,t,\boldsymbol{\theta})} < (1 - \varepsilon)c_{m+1,m}^{(x,z,t,\boldsymbol{\theta})}.$$

Second, note that

$$(35) \qquad \begin{aligned} \frac{c_{m+1,m+1}^{(x,z,t,\boldsymbol{\theta})}}{c_{m+1,m}^{(x,z,t,\boldsymbol{\theta})}} &= \frac{1}{m+1}\frac{\theta+2m}{\theta+m}\frac{\mathbb{E}[\mathcal{D}_{\theta_1+L_{m+1},\theta_2+m+1-L_{m+1}}(z)]}{\mathbb{E}[\mathcal{D}_{\theta_1+L_m,\theta_2+m-L_m}(z)]} \\ &\leq \frac{1}{m+1}\cdot 2\frac{\mathbb{E}[\mathcal{D}_{\theta_1+L_{m+1},\theta_2+m+1-L_{m+1}}(z)]}{\mathbb{E}[\mathcal{D}_{\theta_1+L_m,\theta_2+m-L_m}(z)]} \\ &< \varepsilon, \end{aligned}$$

using Lemma 1 and $m + 1 \geq 2K^{(x,z)}/\varepsilon$ for the last inequality. Rearrange (35) and sum with (34) to get (33). $\square$
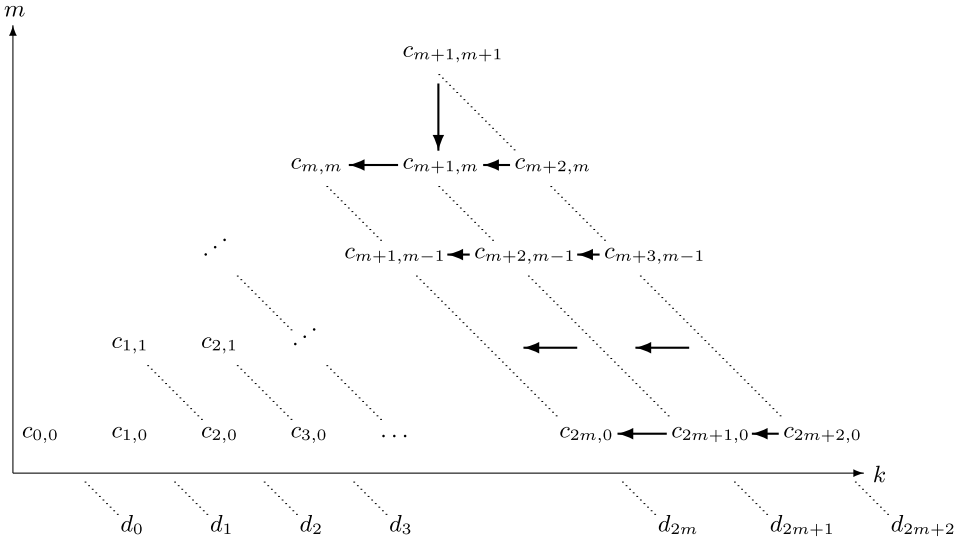
FIG. 2. *Computation of $c_{k,m}^{(x,z,t,\boldsymbol{\theta})}$. The sum of each antidiagonals (dashed) defines the sequence $(d_i)_{i=0,1,\ldots}$. To show that $d_{i+1} < d_i$ terms are paired off as shown by the arrows; that is, the coefficient at the head of a set of arrows is greater in magnitude than the sum of the terms at its tails.*

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, Papaspiliopoulos and Roberts (2008) to a function $g(u_1, u_2, u_3) \propto u_1 u_2 / u_3$. $\square$

PROOF OF PROPOSITION 5. (i) Using (26) and (27),

$$(36) \qquad \frac{b_{K+1}^{(t,\theta)}(m)}{b_K^{(t,\theta)}(m)} = f_m^\theta(K) e^{-\frac{(2K+\theta)t}{2}} < (\theta + 2K + 1) e^{-\frac{(2K+\theta)t}{2}}.$$

The right-hand side of (36) is less than 1 provided

$$K > \frac{\log(\theta + 2m + 1)}{t} - \frac{\theta}{2},$$

which implies that $C_m^{(t,\theta)} < \frac{\log(\theta + 2m + 1)}{t} - \frac{\theta}{2} = O(t^{-1})$ as $t \to 0$.

(ii) Continuing,

$$C_m^{(t,\theta)} < \frac{\log(\theta + 2m + 1)}{t} - \frac{\theta}{2} < \frac{\log(\theta + 2D_0^{(t,\theta)} + 1)}{t} - \frac{\theta}{2},$$

with the right-hand side independent of m and asymptotically $O(t^{-1} \log(t^{-1}))$ as $t \to 0$ by (iii) below.

(iii) By inspection of the definition (8) of $D_0^{(t,\theta)}$, in order to ensure $K > D_0^{(t,\theta)}$ it suffices that

$$K > \frac{1 \vee \log(\theta + 2K + 1)}{t},$$

which holds for sufficiently small t if $K \sim t^{-(1+\kappa)}$, for any fixed $\kappa > 0$. Hence, $D_0^{(t,\theta)} = o(t^{-(1+\kappa)})$ as $t \to 0$.

(iv) Parts (i)–(iii) cover those terms that must be precomputed in Algorithm 2: For the random number of remaining terms, we may assume $K \geq m + C_m^{(t,\theta)}$, so that $b_{K+1}(m) < b_K(m)$. Write $Q_M^\theta(t) := \mathbb{P}(A_\infty^\theta(t) \leq M)$. Our aim is to show that these random remaining terms do not add to the complexity of the calculation beyond (iii); we achieve this by determining the complexity of

$$\widetilde{C}_\delta^{(t,\theta)}(M) := \inf\left\{ K \geq \max_{m \in \{0,\dots,M\}} (m + C_m^{(t,\theta)}) : \sum_{m=0}^{M} \sum_{k=K}^{\infty} (-1)^{k-m} b_k(m) < \delta \right\}$$

as $t \to 0$, for a fixed $\delta > 0$. This is the appropriate quantity to look at, since if $K \geq \widetilde{C}_\delta^{(t,\theta)}(M)$ then $|Q_M^\theta(t) - S_{\mathbf{k}}^\pm(M)| < \delta$, when $2k_i \geq K$; $i = 0, \dots, M$. Furthermore, averaging over the uniform random variable U drawn for inversion sampling, the total number of coefficients $N^{(t,\theta)} \mid U = u$ needed to determine that $Q_{m-1}^\theta(t) < u < Q_m^\theta(t)$ is given by the number of terms needed to bound both our estimates of $Q_{m-1}^\theta(t)$ and $Q_m^\theta(t)$ away from u:

$$(37) \quad \begin{aligned} \mathbb{E}(N^{(t,\theta)}) &= \mathbb{E}\big[\mathbb{E}(N^{(t,\theta)} \mid U)\big] \\ &\leq \sum_{m=0}^{\infty} \int_{Q_{m-1}^\theta(t)}^{Q_m^\theta(t)} \big[\widetilde{C}_{Q_m^\theta(t)-u}^{(t,\theta)}(m) + \widetilde{C}_{u-Q_{m-1}^\theta(t)}^{(t,\theta)}(m-1)\big]\,du. \end{aligned}$$

We will show that if $K \sim t^{-(1+\kappa)}$ as $t \to 0$, for a fixed $\kappa > 0$, then we can attain the stated growth condition on $\mathbb{E}(N^{(t,\theta)})$. Using that the right-hand side of (36) is decreasing in K for $K \geq m + C_m^{(t,\theta)}$, for each $\zeta < 1$ we can find a constant $c_1^{(\theta)}$ such that for $K > c_1^{(\theta)} t^{-1}$, $f_m^\theta(K) e^{-(2K+\theta)t/2} < \zeta$. Hence,

$$(38) \quad \begin{aligned} \sum_{m=0}^{M} \sum_{k=K}^{\infty} (-1)^{k-m} b_k(m) &< \sum_{m=0}^{M} \sum_{k=K}^{\infty} b_k(m) \\ &< \sum_{m=0}^{M} \sum_{k=K}^{\infty} \zeta^{k-K} b_K(m) = \sum_{m=0}^{M} \frac{b_K(m)}{1-\zeta}. \end{aligned}$$

Routine calculations show that, for $m \geq 1$,

$$\begin{aligned} a_{k,m+1}^\theta &= a_{km}^\theta \frac{(k-m)(\theta+m+k-1)}{(m+1)(\theta+m)} \\ &\leq a_{km}^\theta \frac{(k-1)(\theta+k)}{2(\theta+1)} \leq a_{k1}^\theta \left[\frac{(k-1)(\theta+k)}{2(\theta+1)}\right]^{k-1}, \end{aligned}$$

and thus, applying Stirling's formula to $a_{k1}^\theta \sim k^{c_2^{(\theta)}}$,

$$(39) \quad b_K(m) = a_{Km}^\theta e^{-K(K+\theta-1)t/2} \leq e^{c_3^{(\theta)} K \log K - K(K+\theta-1)t/2} \leq c_4^{(\theta)} e^{-K^2 t/2},$$

for some constants $c_2^{(\theta)}, c_3^{(\theta)}, c_4^{(\theta)}$ (which again exist by our assumption about the asymptotic growth of K). In the special case $m = 0$, Stirling's formula also yields $a_{k0}^\theta \sim k^{c_2^{(\theta)}}$ and so the inequalities in (39) continue to hold. Combining (38) with (39), we find

$$\sum_{m=0}^{M} \sum_{k=K}^{\infty} (-1)^{k-m} b_k(m) < c_5^{(\theta)}(M+1)e^{-K^2 t/2},$$

for some $c_5^{(\theta)}$, which is less than δ if

$$(40) \qquad K > \sqrt{\frac{2}{t} \log\left(\frac{c_5^{(\theta)}(M+1)}{\delta}\right)}.$$

(This is not a tight bound but suffices in the calculations below.) In summary, if K satisfies both $K > c_1^{(\theta)} t^{-1}$ and (40) then $K > \widetilde{C}_\delta^{(t,\theta)}(M)$. Integrating over δ:

$$
\begin{aligned}
(41) \qquad & \int_{Q_{m-1}^\theta(t)}^{Q_m^\theta(t)} \widetilde{C}_{Q_m^\theta(t)-u}^{(t,\theta)}(m) \, du \\
& = \int_0^{q_m^\theta(t)} \widetilde{C}_\delta^{(t,\theta)}(m) \, d\delta \\
& < \int_0^{q_m^\theta(t)} \left[\frac{c_1^{(\theta)}}{t^{1+\kappa}} + \sqrt{\frac{2}{t} \log\left(\frac{c_5^{(\theta)}(m+1)}{\delta}\right)} \right] d\delta \\
& = q_m^\theta(t) \left[\frac{c_1^{(\theta)}}{t^{1+\kappa}} + \sqrt{\frac{2}{t}} \left[\log(c_5^\theta(m+1)) + 1 - \log q_m^\theta(t) \right] \right).
\end{aligned}
$$

It remains to show that the resulting expression (41) can be summed over m, which is possible by Theorem 1; in particular,

$$q_m^\theta(t) = \frac{1}{\sqrt{2\pi(\sigma^{(t,\theta)})^2}} \exp\left(\frac{(m - \mu^{(t,\theta)})^2}{2(\sigma^{(t,\theta)})^2}\right) + o(1),$$

$$\mu^{(t,\theta)} = \frac{2}{t} + O(1), \qquad (\sigma^{(t,\theta)})^2 = \frac{2}{3t} + O(1).$$

Hence, by Jensen's inequality,

$$\sum_{m=0}^{\infty} q_m^\theta(t) \log m = \mathbb{E}\left[\log A_\infty^\theta(t)\right] \le \log \mathbb{E}\left[A_\infty^\theta(t)\right] = O\left(\log t^{-1}\right),$$

$$
\begin{aligned}
\sum_{m=0}^{\infty} q_m^\theta(t)\left[-\log q_m^\theta(t)\right] & = \mathbb{E}\left[\log(\sqrt{2\pi}\sigma^{(t,\theta)}) + \frac{1}{2}\left(\frac{A_\infty^\theta(t) - \mu^{(t,\theta)}}{\sigma^{(t,\theta)}}\right)^2\right] \\
& = O\left(\log t^{-1}\right) + \frac{1}{2}\mathbb{E}[X^2] + O(1) = O\left(\log t^{-1}\right),
\end{aligned}
$$

where $X \sim N(0, 1)$. Combining these results with (41), we obtain

$$\sum_{m=0}^{\infty} \int_{Q_{m-1}^{\theta}(t)}^{Q_m^{\theta}(t)} \widetilde{C}_{Q_m^{\theta}(t)-u}^{(t,\theta)}(m) \, du = O\big(t^{-(1+\kappa)}\big) + O\big(t^{-1/2} \log t^{-1}\big)$$

$$= O\big(t^{-(1+\kappa)}\big),$$

showing that the first term on the right of (37) is $O(t^{-(1+\kappa)})$. The second term is also $O(t^{-(1+\kappa)})$ by a similar argument. Since κ was arbitrary, $\mathbb{E}[N^{(t,\theta)}] = o(t^{-(1+\kappa)})$. $\quad\square$

PROOF OF PROPOSITION 6.   For a diffusion with drift γ, since η is continuous on $(0, 1)$, then

$$\int_0^T \frac{\gamma^2(X_t) - \alpha^2(X_t)}{X_t(1 - X_t)} \, dt = \int_0^T \big[\eta^2(X_t)X_t(1 - X_t) + 2\eta(X_t)\alpha(X_t)\big] dt < \infty,$$

then Novikov's condition is satisfied and a Girsanov transform exists with respect to the neutral WF diffusion with drift α, that is, (WF1) holds. In particular, (23) reads

$$(42) \qquad \exp\bigg\{\int_0^T \eta(X_t) \, dX_t - \frac{1}{2} \int_0^T \big[\eta^2(X_t)X_t(1 - X_t) + 2\eta(X_t)\alpha(X_t)\big] dt\bigg\}.$$

The function η is also continuously differentiable in $(0, 1)$, so $\gamma'(x) - \alpha'(x) = \eta'(x)x(1 - x) + \eta(x)(1 - 2x)$ is continuous and

$$(43) \qquad \widetilde{\phi}(x) = \frac{1}{2}\big[x(1 - x)(\eta^2(x) + \eta'(x)) + 2\eta(x)\alpha(x)\big]$$

which, being itself continuous in $[0, 1]$, is then bounded in $(0, 1)$, and (WF3) follows.

(WF2) and (WF4) are obvious. Thus, (42) has a version of the form (24) with $\widetilde{\phi}$ as in (43) and $\widetilde{A}(x) = \int_0^x \eta(z) \, dz$. $\quad\square$

REFERENCES

BARBOUR, A. D., ETHIER, S. N. and GRIFFITHS, R. C. (2000). A transition function expansion for a diffusion model with selection. *Ann. Appl. Probab.* **10** 123–162. MR1765206

BESKOS, A., PAPASPILIOPOULOS, O. and ROBERTS, G. O. (2006). Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli* **12** 1077–1098. MR2274855

BESKOS, A., PAPASPILIOPOULOS, O. and ROBERTS, G. O. (2008). A factorisation of diffusion measure and finite sample path constructions. *Methodol. Comput. Appl. Probab.* **10** 85–104. MR2394037

BESKOS, A. and ROBERTS, G. O. (2005). Exact simulation of diffusions. *Ann. Appl. Probab.* **15** 2422–2444. MR2187299

BLADT, M., FINCH, S. and SØRENSEN, M. (2016). Simulation of multivariate diffusion bridges. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 343–369. MR3454200

BLADT, M. and SØRENSEN, M. (2014). Simple simulation of diffusion bridges with application to likelihood inference for diffusions. *Bernoulli* **20** 645–675. MR3178513

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 Probab. Stat.* **13** 197–217. MR2518546

CHIB, S., PITT, M. and SHEPHARD, N. (2010). Likelihood based inference for diffusion driven state space models. Working Paper.

COOP, G. and GRIFFITHS, R. C. (2004). Ancestral inference on gene trees under selection. *Theor. Popul. Biol.* **66** 219–232.

DANGERFIELD, C. E., KAY, D. and BURRAGE, K. (2010). Stochastic models and simulation of ion channel dynamics. *Proc. Comput. Sci.* **1** 1581–1590.

DANGERFIELD, C. E., KAY, D., MACNAMARA, S. and BURRAGE, K. (2012). A boundary preserving numerical algorithm for the Wright–Fisher model with mutation. *BIT* **52** 283–304. MR2931351

DAWSON, D. A. (1978). Geostochastic calculus. *Canad. J. Statist.* **6** 143–168. MR0532855

DELBAEN, F. and SHIRAKAWA, H. (2002). An interest rate model with upper and lower bounds. *Asia-Pac. Financ. Mark.* **9** 191–209.

DEVROYE, L. (1986). *Nonuniform Random Variate Generation*. Springer, New York. MR0836973

ETHIER, S. N. and GRIFFITHS, R. C. (1993). The transition function of a Fleming-Viot process. *Ann. Probab.* **21** 1571–1590. MR1235429

ETHIER, S. N. and KURTZ, T. G. (1993). Fleming–Viot processes in population genetics. *SIAM J. Control Optim.* **31** 345–386. MR1205982

FAVARO, S., RUGGIERO, M. and WALKER, S. G. (2009). On a Gibbs sampler based random process in Bayesian nonparametrics. *Electron. J. Stat.* **3** 1556–1566. MR2578838

FITZSIMMONS, P., PITMAN, J. and YOR, M. (1993). Markovian bridges: Construction, Palm interpretation, and splicing. In *Seminar on Stochastic Processes*, 1992 (*Seattle, WA*, 1992). *Progress in Probability* **33** 101–134. Birkhäuser, Boston, MA. MR1278079

GOLIGHTLY, A. and WILKINSON, D. J. (2006). Bayesian sequential inference for nonlinear multivariate diffusions. *Stat. Comput.* **16** 323–338. MR2297534

GOLIGHTLY, A. and WILKINSON, D. J. (2008). Bayesian inference for nonlinear multivariate diffusion models observed with error. *Comput. Statist. Data Anal.* **52** 1674–1693. MR2422763

GOURIEROUX, C. and JASIAK, J. (2006). Multivariate Jacobi process with application to smooth transitions. *J. Econometrics* **131** 475–505. MR2276008

GRIFFITHS, R. C. (1979). A transition density expansion for a multi-allele diffusion model. *Adv. in Appl. Probab.* **11** 310–325. MR0526415

GRIFFITHS, R. C. (1980). Lines of descent in the diffusion approximation of neutral Wright–Fisher models. *Theor. Popul. Biol.* **17** 37–50. MR0568666

GRIFFITHS, R. C. (1984). Asymptotic line-of-descent distributions. *J. Math. Biol.* **21** 67–75. MR0770713

GRIFFITHS, R. C. (2006). Coalescent lineage distributions. *Adv. in Appl. Probab.* **38** 405–429. MR2264950

GRIFFITHS, R. C. and LI, W. H. (1983). Simulating allele frequencies in a population and the genetic differentiation of populations under mutation pressure. *Theor. Popul. Biol.* **23** 19–33.

GRIFFITHS, R. C. and SPANÒ, D. (2010). Diffusion processes and coalescent trees. In *Probability and Mathematical Genetics*, *Papers in Honour of Sir John Kingman*. (N. H. Bingham and

C. M. Goldie, eds.). *London Mathematical Society Lecture Note Series* **378** 358–379. Cambridge Univ. Press, Cambridge. MR2744247

GRIFFITHS, R. C. and SPANÒ, D. (2013). Orthogonal polynomial kernels and canonical correlations for Dirichlet measures. *Bernoulli* **19** 548–598. MR3037164

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 Genet.* **5** e1000695.

JENKINS, P. A. (2013). Exact simulation of the sample paths of a diffusion with a finite entrance boundary. Preprint. Available at arXiv:1311.5777.

JEWETT, E. M. and ROSENBERG, N. A. (2014). Theory and applications of a deterministic approximation to the coalescent model. *Theor. Popul. Biol.* **93** 14–29.

KARLIN, S. and TAYLOR, H. M. (1981). *A Second Course in Stochastic Processes*. Academic Press, New York. MR0611513

KLOEDEN, P. E. and PLATEN, E. (1999). *Numerical Solution of Stochastic Differential Equations*, 3rd printing. Springer, Berlin.

LINETSKY, V. (2005). On the transition densities for reflected diffusions. *Adv. in Appl. Probab.* **37** 435–460. MR2144561

LUKIĆ, S., HEY, J. and CHEN, K. (2011). Non-equilibrium allele frequency spectra via spectral methods. *Theor. Popul. Biol.* **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. H. and RUGGIERO, M. (2016). Dynamic density estimation with diffusive Dirichlet mixtures. *Bernoulli* **22** 901–926. MR3449803

NEUENKIRCH, A. and SZPRUCH, L. (2014). First order strong approximations of scalar SDEs defined in a domain. *Numer. Math.* **128** 103–136. MR3248050

ØKSENDAL, B. (2003). *Stochastic Differential Equations. An Introduction with Applications*, 6th ed. Springer, Berlin. MR2001996

PAPASPILIOPOULOS, O. and ROBERTS, G. O. (2008). Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika* **95** 169–186. MR2409721

PAPASPILIOPOULOS, O. and RUGGIERO, M. (2014). Optimal filtering and the dual process. *Bernoulli* **20** 1999–2019. MR3263096

PAPASPILIOPOULOS, O., RUGGIERO, M. and SPANÒ, D. (2014). Filtering hidden Markov measures. Preprint. Available at arXiv:1411.4944.

PARDOUX, E. (2009). Probabilistic models of population genetics. Lecture Notes, available from http://www.latp.univ-mrs.fr/~pardoux/enseignement/cours_genpop.pdf.

PITMAN, J. and YOR, M. (1981). Bessel processes and infinitely divisible laws. In *Stochastic Integrals* (*Proc. Sympos.*, *Univ. Durham*, *Durham*, 1980) (D. Williams, ed.). *Lecture Notes in Math.* **851** 285–370. Springer, Berlin. MR0620995

POLLOCK, M., JOHANSEN, A. M. and ROBERTS, G. O. (2016). On the exact and ε-strong simulation of (jump) diffusions. *Bernoulli* **22** 794–856. MR3449801

SCHRAIBER, J., GRIFFITHS, R. C. and EVANS, S. N. (2013). Analysis and rejection sampling of Wright-Fisher diffusion bridges. *Theor. Popul. Biol.* **89** 64–74.

SCHURZ, H. (1996). Numerical regularization for SDEs: Construction of nonnegative solutions. *Dynam. Systems Appl.* **5** 323–351. MR1410392

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. *Theor. Popul. Biol.* **83** 1–14.

TAVARÉ, S. (1984). Line-of-descent and genealogical processes, and their applications in population genetics models. *Theor. Popul. Biol.* **26** 119–164. MR0770050

WALKER, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Comm. Statist. Simulation Comput.* **36** 45–54. MR2370888

WALKER, S. G., HATJISPYROS, S. J. and NICOLERIS, T. (2007). A Fleming–Viot process and Bayesian nonparametrics. *Ann. Appl. Probab.* **17** 67–80. MR2292580

WILLIAMSON, S. H., HERNANDEZ, R., FLEDEL-ALON, A., ZHU, L., NIELSEN, R. and BUSTA-MANTE, C. D. (2005). Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proc. Natl. Acad. Sci. USA* **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.

DEPARTMENT OF STATISTICS
  AND DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF WARWICK
COVENTRY CV4 7AL
UNITED KINGDOM
E-MAIL: p.jenkins@warwick.ac.uk

DEPARTMENT OF STATISTICS
UNIVERSITY OF WARWICK
COVENTRY CV4 7AL
UNITED KINGDOM
E-MAIL: d.spano@warwick.ac.uk