EXACT SIMULATION OF COUPLED WRIGHT-FISHER DIFFUSIONS

CElia GARCÍA-PAREJA, Henrik Hult, and Timo Koski

Abstract. In this paper an exact rejection algorithm for simulating paths of the coupled Wright-Fisher diffusion is introduced. The coupled Wright-Fisher diffusion is a family of multidimensional Wright-Fisher diffusions that have drifts depending on each other through a coupling term and that find applications in the study of interacting genes’ networks as those encountered in studies of antibiotic resistance. Our algorithm uses independent neutral Wright-Fisher diffusions as candidate proposals, which can be sampled exactly by means of existing algorithms and are only needed at a finite number of points. Once a candidate is accepted, the remaining of the path can be recovered by sampling from a neutral multivariate Wright-Fisher bridge, for which we also provide an exact sampling strategy. The technique relies on a modification of the alternating series method and extends existing algorithms that are currently available for the one-dimensional case. Finally, the algorithm’s complexity is derived and its performance demonstrated in a simulation study.

1. Introduction

Sampling paths of a diffusion process remains a challenging problem in applied probability. The major bottleneck is that their finite dimensional distributions are seldom available in closed form, and one often needs resorting to time-discretized numerical approximations. These approximations, however, induce bias and approximation errors that are difficult to quantify. Moreover, reducing such errors requires refining the time grid, which, in turn, increases computational costs dramatically. In this context, exact simulation algorithms, which aim to recover samples from the true finite dimensional distributions of a diffusion, have become increasingly popular.

The standard approach to exact simulation of diffusions is based on the family of exact rejection algorithms, which rely on an acceptance-rejection scheme that requires samples from a candidate diffusion only at a finite collection of time points in order to take a decision. The candidate needs to be such that is possible to simulate without approximation, and that allows for the construction of the acceptance-rejection probability by means of the Girsanov’s transformation of measures, see [23]. Once a candidate is accepted, the algorithm returns a skeleton of the target path, and the remaining segments can be sampled at any other time instance by

MSC2010 subject classifications. Primary 60J60, 65C30; secondary 60H35, 65C05.

Key words and phrases. exact simulation, rejection algorithm, multivariate diffusions, population genetics, coupled Wright-Fisher model.
simulating from suitable diffusion bridges and with no further reference to the unknown target distribution. In their seminal paper, Beskos and Roberts [4] present an exact rejection algorithm for simulation of paths of a certain class of one-dimensional diffusions, which requires imposing some boundedness assumptions on the drift of the target diffusion and its derivative. Acknowledged as too restrictive, these assumptions were later relaxed in a second publication to one-sided bounds, see [2]. A further extension is based on a layered Brownian bridge construction and provides an exact simulation method for boundary crossing and hitting times, see [3]. Other extensions include algorithms that allow for simulation of killed diffusions and have applications to double barrier option pricing problems [6], or a localized exact algorithm that relaxes any boundedness assumptions by considering smaller pieces of the target path, and that can be used to simulate diffusions with boundaries [7]. An interesting approach is also that presented in [28], where the authors provide an exact rejection algorithm for jump diffusions.

Downsides of the exact rejection algorithms presented above include that they impose somewhat strong assumptions on the drift and require the use of the Lamperti transformation, see [26], to obtain unit volatility coefficients, which hinders generalizability to the multivariate case. To overcome these issues, alternative techniques have been proposed, such as the one in [5] that presented an exact algorithm for simulation of multivariate diffusions based on tolerance enforced simulation and rough paths analysis. Their algorithm overcomes the more restrictive assumptions required in [4] and [2] but has, however, infinite expected running time.

Another restrictive feature of exact rejection algorithms is that they rely heavily on the availability of suitable candidates (in all cases mentioned above, Brownian motion or slight modifications therein). For diffusions with finite boundaries, for example, Brownian candidates can either differ too much from the target, thus providing low acceptance probabilities, or be unsuitable to construct the acceptance-rejection probability itself. Rejection algorithms that use candidates other than Brownian motion include that presented in [20], which uses Bessel proposals to simulate a certain class of diffusions with a finite entrance boundary.

In this context, the recent work in [21] extends the class of diffusions for which exact rejection simulation is possible. The authors propose a simulation technique to recover samples from neutral Wright-Fisher diffusions that, in turn, are used as candidates in an exact rejection algorithm for simulating a wider target family of one-dimensional Wright-Fisher diffusions. This class of diffusions, as well as its multivariate counterparts, are extensively used in population genetics.

Along these lines, a main contribution of this paper is to present an exact rejection algorithm for coupled Wright-Fisher diffusions. The coupled Wright-Fisher diffusion is a family of multivariate Wright-Fisher diffusions that models how different allele types (genetic traits) co-evolve across different loci (different locations along the genome), over generations [1]. This type of diffusion model has applications to the study of interactions in network structures such as those encountered in the study of antibiotic resistance, see, for example, [29], among others. Candidates for our algorithm are built from samples of independent neutral Wright-Fisher diffusions, which can be recovered using the techniques presented in [21]. Our second
contribution deals with simulation of multidimensional Wright-Fisher bridges, for which we present an exact simulation technique. These bridges complete the proposed algorithm in that allow sampling further points of the path once a skeleton of the coupled Wright-Fisher diffusion has been accepted. Our proposed sampling approach can therefore be viewed as a generalization of that presented in [21] for the one-dimensional Wright-Fisher diffusions to the multivariate case.

The rest of the paper is structured as follows. In Section 2 main properties and structure of the family of coupled Wright-Fisher diffusions are briefly presented jointly with a formal overview on exact rejection algorithms. In Section 3 we recall and present some revised algorithms for exact simulation of one and multidimensional neutral Wright-Fisher diffusions, i.e., those needed for sampling our candidate processes, leading up to the proposed exact rejection algorithm for coupled Wright-Fisher diffusions (Section 4). Section 5 includes performance results illustrated through several simulation scenarios and in Section 6 the technique for simulating exactly from a multidimensional Wright-Fisher Bridge is provided, which completes the sampling scheme. Finally, Section 7 contains mathematical proofs.

2. Background

This section provides the necessary insights on the structure and main properties of the coupled Wright-Fisher family of diffusions and fixes some notation, as well as provides a brief overview of exact rejection algorithms for diffusions, which constitute the basis of our work.

2.1. Coupled Wright-Fisher diffusions. The family of Wright-Fisher models, and more specifically their diffusion approximations, have been widely used in population genetics, see, for instance, [25] and [14]. In its simplest form, the Wright-Fisher model describes the evolution of the frequency of two allele types in a single locus that have the same fitness, and whose configuration at each new generation of individuals is chosen uniformly and with replacement from that of the current generation in an haploid population of constant size. Extensions of the model include considering more than two allele types that might be located at different loci, and can incorporate other evolutionary forces such as mutation, selection and recombination. A comprehensive overview on the family of Wright-Fisher models can be found, for example, in [8] or [12].

With the proliferation of genome-wide association studies, questions arise about how genetic variants associated to numerous diseases co-evolve or interact over time. Moreover, the increasing availability of allele frequency time series data is fostering the study of evolutionary forces such as mutation or selection, see [30], [29], [31], [27]. Within this framework, the recently proposed coupled Wright-Fisher model [1] tracks the evolution of frequencies of allele types located at different loci, and, besides locus-wise mutations, describes possible selective pairwise interactions between allele frequencies across loci in an haploid population of constant size $N$.

Taking the weak convergence limit as $N$ goes to infinity, and after a proper re-scaling of time, one obtains the coupled Wright-Fisher diffusion approximation that
can be expressed as a system of stochastic differential equations of the form
\[ dX_t = [\alpha(X_t) + G(X_t)]dt + D^{\frac{1}{2}}(X_t)dB_t, \quad X_0 = x_0, \quad t \in [0, T], \tag{2.1} \]

where \( X_t \) is a vector of frequencies of allele types, \( \alpha \) governs their mutations and \( G \) contains the single and pairwise selective locus interactions.

Let \( L \) denote the total number of loci and \( d_i \geq 2 \) the number of different allele types in each of them. For \( n := \sum_{i=1}^{L}[d_i - 1] \), let us index each element of an \( n \)-dimensional vector \( x \) by its referring to a specific locus \( i \in \{1, \ldots, L\} \) and allele type \( j \in \{1, \ldots, d_i - 1\} \), so that \( x = \{x^i\}_{i=1}^{L} \) where each \( x^i = \{x^{ij}\}_{j=1}^{d_i} \). If \( x \in \mathbb{R}^n \) refers to the vector of allele type frequencies \( X_t \), the elements of the drift \( \alpha(x) \in \mathbb{R}^n \) take the form
\[ \alpha^{ij}(x^{ij}) = \frac{1}{2} (\theta_i^j - |\theta|x^{ij}), \tag{2.2} \]

where \( |\theta| = \sum_{k=1}^{d_i} \theta_k^j \) and \( \theta_k^j > 0 \) denote the parent-independent mutation rates to allele type \( k \in \{1, \ldots, d_i\} \) at locus \( i \), so that mutations occur at each locus separately. Wright-Fisher diffusions with drift \( \alpha(x) \) correspond to the reversible neutral mutations allele model. The coupled Wright-Fisher model also admits parent-dependent mutations, but we will not consider them here.

The coupling term \( G(x) \in \mathbb{R}^n \) has general form
\[ G(x) = D(x)\nabla_x(\nabla \circ f)(x), \tag{2.3} \]

where the square of the diffusion matrix \( D(x) = \text{diag}(D^j(x^i)) \in \mathbb{R}^{n \times n} \) is an \( L \)-blocks diagonal matrix with entries
\[ D^j_{ik} = \begin{cases} x^{ij}(1 - x^{ij}), & j = k, \\ -x^{ij}x^{ik}, & j \neq k, \end{cases} \quad j, k \in \{1, \ldots, d_i - 1\}, i \in \{1, \ldots, L\}, \]

\( \nabla_x \) is the gradient operator w.r.t. each component of \( x \), \( f \) transforms \( x \) in the augmented \( (n + L) \)-dimensional vector \( \varpi \) that reflects the dependency between allele frequencies at locus \( i \), i.e.,
\[ \varpi^k := f^k(x) = \begin{cases} x^{ij}, & j = k \in \{1, \ldots, d_i - 1\}, \\ 1 - \sum_{j=1}^{d_i-1} x^{ij}, & k = d_i, \end{cases} \]

and \( \nabla(\varpi) \in \mathbb{R} \),
\[ \nabla(\varpi) = (\varpi)^T s + \frac{1}{2}(\varpi)^TH\varpi, \tag{2.4} \]

where \( s \in \mathbb{R}^{n+L} \) is a within locus selection parameters vector and \( H \in \mathbb{R}^{(n+L)\times(n+L)} \) is a symmetric across-loci pairwise interactions matrix. The matrix \( H \) is in fact built by \( L \) blocks of zeros of size \( d_i \times d_i \) in the main diagonal (denoted \( 0^{ii} \)), and off-diagonal blocks of the form \( H^{il} = (H^{ii})^T \in \mathbb{R}^{d_i \times d_i}, i \neq l, i, l \in \{1, \ldots, L\}, \)
\[
H = \begin{pmatrix}
0^{11} & H^{12} & \cdots & \cdots & \cdots & \cdots & H^{1L} \\
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\
H^{i1} & \cdots & H^{i(i-1)} & 0^{ii} & H^{i(i+1)} & \cdots & H^{iL} \\
\vdots & \ddots & \vdots & \ddots & \ddots & \vdots & \vdots \\
H^{L1} & \cdots & \cdots & \cdots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \vdots & \ddots & \ddots & \ddots & 0^{LL}
\end{pmatrix},
\]
so that interactions of each locus with itself are not permitted. Note that if one removes the coupling term \( G \), (2.1) simply becomes the usual multidimensional neutral mutations Wright-Fisher diffusion, where \( X_t \) describes the evolution of allele frequencies that evolve independently at each locus. The explicit form of \( V(x) := \nabla_x (\nabla \circ f)(x) \) in terms of \( s \) and \( H \) is specified in the following proposition, whose proof can be found in Section 7.

**Proposition 2.1.** Let \( V(x) \in \mathbb{R}^n \) be the \( n \)-dimensional vector such that \( V(x) := \nabla_x (\nabla \circ f)(x) \). Then \( \forall \{1, \ldots, L\}, j \in \{1, \ldots, d_i - 1\} \)

\[
V^{ij}(x) = K^{ij}_s + \sum_{l=1}^L \sum_{l \neq j} K^{ij}_{lk} x^{lk},
\]

with

\[
K^{ij}_s := s^{ij} - s^{id}, \quad K^{ij}_l := h^{il}_{jd} - h^{il}_{d,dl} \quad \text{and} \quad K^{ij}_{lk} := h^{il}_{jk} - h^{il}_{jd} - h^{il}_{dl} + h^{il}_{dl},
\]

where \( h^{il}_{jk} \) denotes the \( j \)-th-row-\( k \)-th-column entry of the block \( H^{il} \) of \( H \).

Following Kimura’s formulation [24], the explicit stationary density of (2.1) can also be obtained by solving the corresponding Fokker-Planck equation, see [1], whose solution takes the form

\[
P(x) = \frac{1}{Z} \pi(x) e^{2(\nabla \circ f)(x)}, \quad (2.5)
\]

where

\[
\pi(x) = \prod_{i=1}^L \pi^i(x^i) = \prod_{i=1}^L \left( 1 - \sum_{j=1}^{d_i-1} x^{ij} \right)^{\theta_i^{ij}-1} \prod_{j=1}^{d_i-1} (x^{ij})^{\theta_i^{ij}-1}
\]

and \( Z \) is the normalizing constant

\[
Z = \int_{x \in \mathcal{X}} \pi(x) e^{2(\nabla \circ f)(x)} dx.
\]

Here \( \mathcal{X} = \{x \in \mathbb{R}^n \mid x^{ij} \geq 0, \sum_{j=1}^{d_i-1} x^{ij} \leq 1\} \).

2.2. Overview of exact rejection algorithms for simulation of diffusions.

This subsection provides an overview on the exact rejection algorithm presented in [4], and presents the same sampling scheme followed for simulating from the coupled Wright-Fisher diffusion, detailed further in Algorithm 4. Let

\[
dX_t = \mu(X_t)dt + dB_t, \quad X_0 = x_0, \quad t \in [0, T], \quad (2.6)
\]

where \( \mu(\cdot) \) is such that (2.6) admits a unique weak solution \((X_t)_{t \in [0,T]}\). Let \( \mathbb{Q}_{x_0} \) be the law of the process \((X_t)_{t \in [0,T]}\) and let \( \mathbb{P}_{x_0} \) denote the law of a Brownian motion \((B_t)_{t \in [0,T]}\) starting at \( B_0 = x_0 \). By means of the Girsanov’s transformation of measures we can write the Radon-Nykodým derivative of \( \mathbb{Q}_{x_0} \) w.r.t. \( \mathbb{P}_{x_0} \)

\[
\frac{d\mathbb{Q}_{x_0}}{d\mathbb{P}_{x_0}} = \exp \left\{ \int_0^T \mu(B_t)dX_t - \frac{1}{2} \int_0^T \mu^2(B_t)dt \right\}. \quad (2.7)
\]
Assuming that \( \mu(\cdot) \) is differentiable everywhere, we can rewrite (2.7) using Itô’s lemma as

\[
\frac{dQ_{x_0}}{dP_{x_0}} = \exp \left\{ \tilde{A}(B_T) - \tilde{A}(x_0) \right\} \exp \left\{ -\frac{1}{2} \int_0^T (\mu^2(B_t) + \mu'(B_t)) dt \right\},
\]

where \( \tilde{A}(x) := \int_0^x \mu(u) du \). Imposing the further conditions of \( \tilde{A}(x) \) to be bounded above by a constant \( K^A \), and \((\mu^2 + \mu')/2 \) to be bounded between constants \( K^- \) and \( K^+ \), the acceptance-rejection probability is

\[
\frac{dQ_{x_0}}{dP_{x_0}} \propto \exp \left\{ \tilde{A}(B_T) - K^A \right\} \exp \left\{ -\int_0^T (\tilde{\phi}(B_t) - K^-) dt \right\},
\]

with \( \tilde{A}(x) \leq K^A \) and \( K^- \leq \tilde{\phi}(x) := \frac{1}{2} |\mu^2(x) + \mu'(x)| \leq K^+ \).

Note that, a priori, an exact evaluation of the integral in (2.9) is not possible without any approximation error because it would require to store infinitely many points of the sample candidate path \( B = (B_t)_{t \in [0,T]} \). However, the key point of the exact algorithms proposed in [4] and publications therein, lies on the fact that an event occurring with probability (2.9) can be evaluated sampling \( B \) only at a finite number of time points. This follows because the last term in (2.9) can be interpreted as the probability of the event that no points from an homogeneous spatial Poisson process \( \Phi = \{(t_j, \psi_j) : j = 1, \ldots, J \} \) with unit intensity on \([0,T] \times [0,K^+ - K^-]\) lie below the graph of \( t \mapsto \tilde{\phi}(B_t) \), which we refer to as \( \omega^{\tilde{\phi}} \). A formal statement and proof of this observation can be found in Theorem 1 of [2].

Therefore, the exact sampling procedure (detailed in Algorithm 1) starts by drawing a sample from \( \Phi \) that will determine the time points at which the candidate \( B \) will be drawn, and then provides a skeleton of \( X_t \) at such time points upon acceptance of the sampled candidate (that is, if all the evaluated \( \tilde{\phi}(B_t) \) lie above the sampled Poisson points). Note that the last point on the candidate path, \( B_T \), serves to evaluate an event that occurs with probability \( \exp\{\tilde{A}(B_T) - K^A\} \) and that is independent of \( \omega^{\tilde{\phi}} \). In an original version of the algorithm, \( B_T \) is sampled from a slightly modified distribution; a biased Brownian motion that serves as a valid candidate and improves the algorithm’s efficiency. Such option is not needed for our purposes and is therefore not fully described here. In brief, this modification permits relaxing the boundedness condition on \( \tilde{A}(x) \), but the equivalent function in our proposed exact algorithm is already bounded.

3. Simulation of the candidate processes

This section is devoted to describing existing simulation strategies for the candidate processes in the exact rejection algorithm that will be presented later in Section 4. Suitable candidate processes in our setting will be \( L \) independent \((d_i - 1)\)-dimensional neutral Wright-Fisher diffusions \((X_t)_{t \in [0,T]}\), each one unique weak solution of

\[
\frac{dX_t}{dt} = \alpha(X_t) dt + D^{1/2} dB_t, \quad X_0 = x_0, \quad t \in [0,T],
\]

with \( \alpha(X_t) \) a \((d_i - 1)\)-dimensional vector with \( \alpha^{ij}(x) = \frac{1}{2} (\theta^j_i - |\theta^{x^{ij}}|). \)
Algorithm 1 Exact algorithm for simulating skeletons of paths \((X_t)_{t \in [0,T]}\) of a diffusion process with law \(Q_{x_0}\):

1. Simulate \(\Phi\), a Poisson process on \([0,T] \times [0,K^+ - K^-]\).
2. Simulate \(U \sim \text{Uniform}(0,1)\).
3. Given \(\Phi = \{(t_j, \psi_j) : j = 1, \ldots, J\}\), simulate \(B \sim \mathbb{P}_{x_0}\) at times \(t_1, \ldots, t_J\) and at time \(T\).
4. If \(\tilde{\phi}(B_{t_j}) - K^- \leq \psi_j, \forall j\) and \(U \leq \exp\{\tilde{A}(B_T) - K^-A\}\) then return \(\{(t_j, B_{t_j}), \forall j\} \cup \{(T, B_T)\}\).
5. Else go back to Step 1.

3.1. Transition density function expansions. Exact simulation of each neutral Wright-Fisher diffusion is possible by exploiting available transition density’s eigenfunction expansions that allow a probabilistic representation, see, for example, [19].

For a fixed locus \(i \in \{1, \ldots, L\}\), let \(x = (x_1, \ldots, x_{d-1})\) be a vector of initial frequencies and \(\theta + l\) a \(d\)-dimensional vector with entries \(\theta_k + l_k, k \in \{1, \ldots, d\}\). Then, the probabilistic representation of the transition density function of \((X_t)_{t \in [0,T]}\) in (3.1) is given by

\[
g(x, y; t) = \sum_{m=0}^{\infty} q_m^\theta(t) \sum_{|l|=m} \mathcal{M}_{m,x}(l) \mathcal{D}_{\theta+l}(y), \tag{3.2}
\]

where \(q_m^\theta(t)\) are transition functions of a pure death process \(A_{\infty}^\theta(t)\) with an entrance boundary at \(\infty\), \(\mathcal{M}_{m,x}(\cdot)\) is the probability mass function (PMF) of a multinomial random variable, and \(\mathcal{D}_{\theta+l}(\cdot)\) the probability density function (PDF) of a Dirichlet random variable, that is,

\[
\mathcal{M}_{m,x}(l) = \frac{m!}{\prod_{j=1}^{d} l_j!} (1 - \sum_{j=1}^{d-1} x_j)^{l_d} \prod_{j=1}^{d-1} x_j^{l_j},
\]

and

\[
\mathcal{D}_{\theta+l}(y) = \frac{\Gamma([\theta + l])}{\prod_{j=1}^{d} \Gamma(\theta_j + l_j)} (1 - \sum_{j=1}^{d-1} y_j)^{\theta_d + l_d} \prod_{j=1}^{d-1} y_j^{\theta_j + l_j - 1},
\]

with \(l_d = m - \sum_{j=1}^{d-1} l_j\). A more detailed description of the process \(A_{\infty}^\theta(t)\) as well as an exact sampling technique are provided in detail later on.

In case of one-dimensional \((d = 2)\) Wright-Fisher diffusions, the multivariate components on the mixture in (3.2) reduce to their one-dimensional counterparts, i.e., a binomial and a beta random variables respectively [19].

A sampling strategy for \(g(x, \cdot; t)\) is summarized in Algorithm 2, see [18] or [21] for an analogous version in the one-dimensional case, the latter also including a modification for the infinite-dimensional case, that is, for Fleming-Viot diffusions. Once expressed in probabilistic terms and given the simplicity of Algorithm 2, recovering samples from \(g(x, \cdot; t)\) seems straightforward. However, sampling exactly from \(q_m^\theta(t)\)
Algorithm 2 Exact simulation of samples from \( g(x, \cdot ; t) \), transition density of the \((d - 1)\)-dimensional neutral Wright-Fisher diffusion with recursive mutation

1. Simulate \( M \sim A^\theta_\infty(t) \).
2. Given \( A^\theta_\infty(t) = m \), simulate \( L \sim \text{Multinomial}(m, x) \).
3. Given \( L = (l_1, \ldots, l_{d-1}) \), simulate \( Y \sim \text{Dirichlet}(\theta + l) \).
4. return \( Y = (y_1, \ldots, y_{d-1}) \).

poses some difficulty because it is only known in infinite series form. Previous approaches for simulating from approximated versions of \( q^\theta_m(t) \) can be found in [18], [15], [16] or [22]. In the next section, we review the exact simulation procedure presented in [21], which is the one used here.

3.2. Exact simulation of the ancestral process \( A^\theta_\infty \). We describe here the sampling procedure for recovering exact samples of \( q^\theta_m(t) \), transition functions of the aforementioned death process \( A^\theta_\infty \). In more detail, let \( \{A^\theta_n(t) : t \geq 0\} \) be a pure death process on \( \mathbb{N} \) such that \( A^\theta_n(0) = n \) almost surely and with its only possible transition \( m \rightarrow m - 1 \) occurring at rate \( m(m + |\theta| - 1)/2 \) for each \( m = 1, \ldots, n \), that is, it represents the number of non-mutant lineages that coalesce backwards in time in the coalescent process with mutation. Then, let \( q^\theta_m(t) = \lim_{n \to \infty} \Pr(A^\theta_n(t) = m) \).

For a more thorough interpretation of the transition density \( g(x, \cdot ; t) \) and its one-dimensional counterpart, it is worth noting that the expansion in (3.2) is derived via a duality principle for Markov processes [11], that is, from the moment dual process of the Wright-Fisher diffusion, which is also a pure death process representing lineages backwards in time, see for example [10] or [19] for a complete derivation and details.

An expression for \( q^\theta_m(t) \) starting from the entrance boundary at infinity is, see [14],

\[
q^\theta_m(t) = \sum_{i=0}^{\infty} (-1)^i b^{(t,\theta)}_{m+i}(m),
\]

where

\[
b^{(t,\theta)}_{m+i}(m) = \frac{(|\theta| + 2m + i - 1)!}{m!} \frac{\Gamma(|\theta| + 2m + i - 1) \Gamma(m + i + |\theta| - 1)t/2}{\Gamma(|\theta| + m)} c^{(m+i)(m+i+|\theta|-1)t/2}.
\]

As shown in [21], samples from \( q^\theta_m(t) \) can be recovered exactly by means of a variant of the alternating series method, described in [9], Chapter 4. In brief, the alternating series method would require the sequence of coefficients \( b^{(t,\theta)}_{m+i}(m) \) to be decreasing in \( i \) for each \( m \), condition that is not always met here. Nonetheless, one can exploit that there exists a finite \( C^m_{(t,\theta)} \) such that for all \( m \) and \( i \geq C^m_{(t,\theta)} \) the sequence of coefficients \( b^{(t,\theta)}_{m+i+1}(m) \) decreases monotonically as \( i \) tends to \( \infty \). More explicitly, there exists

\[
C^m_{(t,\theta)} = \inf \left\{ i \geq 0 : \frac{b^{(t,\theta)}_{m+i+1}(m)}{b^{(t,\theta)}_{m+i}(m)} < 1 \right\} < \infty.
\]
Then, once \( C^{(t, \theta)}_m \) is available, the remaining of the sequence of coefficients is ensured to be decreasing and the alternating series method can be applied. The following proposition summarizes the main properties of the bound \( C^{(t, \theta)}_m \).

**Proposition 3.1.** [Proposition 1 in [21]] Let \( b_{m+i}^{(t, \theta)}(m) \) be the coefficients defined in (3.4) and \( C^{(t, \theta)}_m \) be as in (3.5). Then

i) \( C^{(t, \theta)}_m < \infty \) for all \( m \).
ii) \( b_{m+i}^{(t, \theta)}(m) \downarrow 0 \) as \( i \rightarrow \infty \) for all \( i \geq C^{(t, \theta)}_m \).
iii) \( C^{(t, \theta)}_m = 0 \) for all \( m > D^{(t, \theta)}_\epsilon \), where

\[
D^{(t, \theta)}_\epsilon = \inf \left\{ u \geq \frac{1}{t} - \frac{|\theta| + 1}{2} \right\} \forall 0 : (|\theta| + 2u - 1)e^{u(|\theta| - 1)t/2} < 1 - \epsilon \right\}, \quad (3.6)
\]

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

Property iii) in Proposition 3.1 will be of interest later when proposing the exact sampling algorithm for \( (d - 1) \)-dimensional Wright-Fisher bridges (Section 6), where an explicit bound on \( m \) is needed.

One can then recover exact samples from \( q^\theta_m(t) \) because the terms \( b_{m+i}^{(t, \theta)}(m) \) become monotonically smaller with increasing \( i \), and for each \( m \) there exist \( k_m \), elements of \( k \in \mathbb{R}^{M+1} \) (i.e., \( k = \{k_m\}_{m=0}^{M} \)) such that

\[
S^-_k(M) := \sum_{m=0}^{M} \sum_{i=0}^{2k_0+1} (-1)^i b_{m+i}^{(t, \theta)}(m) \leq \sum_{m=0}^{M} q^\theta_m(t) \leq \sum_{m=0}^{M} \sum_{i=0}^{2k_0} (-1)^i b_{m+i}^{(t, \theta)}(m) =: S^+_k(M).
\]

Because

\[
\limsup_{k \to (\infty, \ldots, \infty)} S^-_k(M) = \Pr(A^\theta_\infty(t) \leq M) \quad \text{and} \quad \liminf_{k \to (\infty, \ldots, \infty)} S^+_k(M) = \Pr(A^\theta_\infty(t) \leq M),
\]

and both \( S^-_k(M) \) and \( S^+_k(M) \) can be computed from finitely many terms, given \( U \sim \text{Uniform}(0, 1) \) we can find \( k_0 \in \mathbb{R}^{M+1} \) with elements \( k_m \) such that

\[
k_m = \inf \left\{ k_m \in \mathbb{N} : S^-_k(M) > U \text{ or } S^+_k(M) < U \right\},
\]

for each \( m \in \{0, \ldots, M\} \).

Now, if \( k_0 \) is such that \( S^-_{k_0}(M) > U \), standard inverse sampling provides

\[
\inf \left\{ M \in \mathbb{N} : \sum_{m=0}^{M} q^\theta_m(t) \geq S^-_{k_0}(M) > U \right\}, \quad (3.7)
\]

with \( M \) exactly distributed following \( q^\theta_m(t) \). The sampling strategy will consist in exploring the summands in \( S^+_k(M) \) and \( S^-_k(M) \) through their respective indexes \( m \) and \( k_m \), until, for a given realization of \( U \), condition (3.7) is satisfied.

A complete simulation procedure is presented in Algorithm 3, where several improvements mentioned in [21] have been incorporated. Most notably, the variable \( M \) is initialized at the nearest integer around the mean of a certain distribution (not necessarily at 0) that serves as an estimate of the mode \( q^\theta_m(t) \), a modification that decreases computation times substantially. Such initialization originates from
an asymptotic approximation of the transition functions $q_\theta^m(t)$, that first appeared in [15], which states that as $t \to 0$, $A_\infty^\theta(t)$ converges to a normal distribution, that is,

$$
\frac{A_\infty^\theta(t) - \mu^{(t,\theta)}}{\sigma^{(t,\theta)}} \xrightarrow{D} \mathcal{N}(0, 1) \text{ as } t \to 0,
$$

(3.8)

where $\mu^{(t,\theta)} = \frac{2\eta}{t}$, $(\sigma^{(t,\theta)})^2 = \begin{cases} \frac{2\eta}{t^2} (\eta + \beta)^2 \left(1 + \frac{\eta}{\eta + \beta} - 2\eta\right), & \beta \neq 0 \\ \frac{2}{3t}, & \beta = 0 \end{cases}$,

with $\eta = \beta/e^{\beta} - 1$ for $\beta \neq 0$ or $\eta = 1$ otherwise, and $\beta = \frac{1}{2} (|\theta| - 1)t$, and $\xrightarrow{D}$ denotes convergence in distribution, see Theorem 1 in [21].

**Algorithm 3**

Exact simulation of samples from $q_\theta^m(t)$, transition functions of the ancestral process $A_\infty^\theta$

1. Set $M \leftarrow \hat{q}_{\text{mod}}, \bar{k} \leftarrow (0, \ldots, 0), j \leftarrow 1$
2. Simulate $U \sim \text{Uniform}(0, 1)$
3. repeat
   for all $m \in \{0, \ldots, M\}$ do
   5. Set $k_m \leftarrow \left[\frac{C_m^{(t,\theta)}}{2}\right]$
   end for
   7. while $S^-_k(M) < U < S^+_k(M)$ do
   8. Set $\bar{k} \leftarrow \bar{k} + (1, \ldots, 1)$
   end while
   10. if $S^-_k(M) > U$ then
       return $M$
   12. else if $S^+_k(M) < U$ then
       Set $M \leftarrow \hat{q}_{\text{mod}} + (-1)^j j$
       if $j$ odd then
           $\bar{k} \leftarrow (k_0, \ldots, k_M)$
       else if $j$ even then
           $\bar{k} \leftarrow (\bar{k}, 0, \ldots, 0)$
       end if
   18. end if
   19. end if
20. Set $j \leftarrow j + 1$
21. until false

Note that this initialization is possible because exploring the summands in $S^-_k(M)$ and $S^+_k(M)$ does not require to follow any specific order. It is also worth mentioning that, when $M$ is initialized at 0, the vector $\bar{k}$ is updated increasingly, i.e., a new element of the vector is added at every new iteration where $M$ is increased one unit. In Algorithm 3 however, $M$ is updated telescopicly, that is, at each new iteration $j$, $M$ either increases or decreases by one unit. This in turn, entails updating the corresponding $M + 1$ elements of $k$ accordingly, i.e., the number of elements might either increase or decrease at each iteration. For precise results on the complexity of Algorithm 3 and simulation performance we refer the reader to [21].
At this stage, Algorithm 3 can be used in step 1 of Algorithm 2 and an exact sampling procedure for \((d - 1)\)-dimensional neutral Wright-Fisher diffusions is completed.

4. Exact rejection algorithm for simulating coupled Wright-Fisher diffusions

Let \(X_t\) be the \(n\)-dimensional vector of allele frequencies satisfying (2.1). Following the same scheme as Algorithm 1 in Section 2, Algorithm 4 simulates exact skeletons of paths of coupled Wright-Fisher diffusions with \(L\) loci and \(d_i\) allele types each, \(i \in \{1, \ldots, L\}\). Candidate processes in this case are \(L\) independent \((d_i - 1)\)-dimensional neutral Wright-Fisher diffusion, each one unique weak solution of (3.1) and sampled following Algorithm 2.

The exact rejection algorithm proposed in this paper relies on the existence and characterization of the following acceptance-rejection probability, which is detailed in Theorem 4.1 and whose proof is deferred to Section 7.

Theorem 4.1. Let \(CWF_{\alpha,G,x_0}\) be the law of \(X\), the solution of (2.1) and \(WF_{L\alpha,x_0}\) be the joint law of \(L\) independent \((d_i - 1)\)-dimensional neutral Wright-Fisher diffusions, \(i \in \{1, \ldots, L\}\). Then, the Radon-Nykodým derivative of \(CWF_{\alpha,G,x_0}\) w.r.t. \(WF_{L\alpha,x_0}\) is of the form

\[
\frac{dCWF_{\alpha,G,x_0}}{dWF_{L\alpha,x_0}} = \exp \left\{ - \int_0^T \phi(X_t) \, dt \right\}
\]

and there exist constants \(A^-\), \(A^+\), \(C^-\) and \(C^+\) such that \(A(X_0, X_T)\) is bounded on \([0,1]^n \times [0,1]^n\) by \(A^- \leq A(X_0, X_T) \leq A^+\) and \(\phi(X_t)\) is bounded on \([0,1]^n\) by \(C^- \leq \phi(X_t) \leq C^+\), with

\[
A(X_0, X_T) := \int_0^T V(X_t) \cdot dX_t = \sum_{i=1}^L \sum_{j=1}^{d_i-1} \left( K^{ij}_s (X_T^{ij} - X_0^{ij}) + \sum_{l=1}^L K^{ij}_l (X_T^{ij} - X_0^{ij}) \right)
\]

\[
+ \sum_{l=i+1}^L \sum_{k=1}^{d_l-1} K^{ij}_{lk} (X_T^{ij} X_T^{lk} - X_0^{ij} X_0^{lk})
\]

and

\[
\phi(X_t) := \frac{1}{2} \left[ (V(X_t))^T D(X_t) V(X_t) + 2(V(X_t))^T \alpha(X_t) \right].
\]

Using Algorithm 2 in Step 4 (or in case \(d_i = 2\) for some \(i\), the corresponding modification for one-dimensional diffusions), Algorithm 4 returns an exactly simulated skeleton of the solution of (2.1).

The algorithm’s computational complexity can also be established, and is made precise in Proposition 4.1, whose proof can be found in Section 7.

Proposition 4.1. Let \(L\) be the number of loci, \(M(t)\) denote the total number of coefficients that must be computed in the implementation of Algorithm 3, where \(t \in (0,T)\) is the time distance between two sampled skeleton points, and let \(N(T)\) denote the number of Poisson points required until the first skeleton in Algorithm 4
Algorithm 4 Exact rejection algorithm for simulating skeletons of the paths \((X_t)_{t \in [0,T]}\) of a diffusion process with law \(CWF_{\alpha,G,x_0}\)

1. Simulate \(\Phi\), a Poisson process on \([0,T] \times [0,C^+ - C^-]\)
2. Simulate \(U \sim \text{Uniform}(0,1)\)
3. Given \(\Phi = \{(t_j, \psi_j) : j = 1, \ldots, J\}\), simulate \(X \sim WF_{L_{\alpha,x_0}}\) at times \(\{t_1, \ldots, t_J, T\}\).
4. if \(\phi(X_{t_j}) - C^- \leq \psi_j, \forall j\) and \(U \leq \exp\{A(X_T) - A^+\}\) then
5. \(\text{return } \{(t_j, X_{t_j}), \forall j\} \cup \{(T, X_T)\}\)
6. else
7. Go back to Step 1.
8. end if

is accepted. Then, \(E[LM(t)] < \infty\) and \(E[N(T)] < \infty\), and more specifically, there exists \(\kappa > 0\) such that
\[ E[LM(t)] = o(t^{-(1+\kappa)}) \text{ as } t \to 0, \quad \text{and } E[N(T)] \leq T(C^+ - C^-)e^{T(C^+ - C^-) + A^+ - A^-}. \]

In summary, Algorithm’s complexity increases either as \(t \to 0\), when the average number of coefficients to be computed in Algorithm 3 increases as \(1/t\), or with increasing \(T\), when the average number of Poisson points needed before acceptance increases exponentially.

The latter is easily solvable, simply by considering shorter intervals \([t_{k-1}, t_k]\) such that
\[ \bigcup_{k=1}^K [t_{k-1}, t_k] = [0, T], \text{ with } t_0 = 0, t_K = T, \text{ and } t_{k-1} < t_k, \forall k \in \{1, \ldots, K\}, \]
and then concatenating the accepted skeletons in each of them. To solve the problem when \(t \to 0\), we followed the recommendation in [21], and whenever \(t < 0.05\), resort to the approximation in [33].

While asymptotically, the algorithm’s growth rate does not depend on the number of loci, it is worth mentioning that with increasing \(L\) the acceptance probability decreases, as there are a larger number of skeletons that need to be accepted simultaneously, which naturally affects the algorithm’s feasibility. This will be clearer in the next section, where simulation results for examples with \(L = 2\) and \(L = 4\) are provided. As expected, the acceptance probability also decreases whenever the target diffusion differs more from the neutral Wright-Fisher candidate. This is exemplified in the next section, where results are shown for two coupled Wright-Fisher models with the same number of loci and mutation, but different selective pairwise interactions parameters.

5. Numerical experiments

In the following, several implementations of Algorithm 4 are shown along with their simulation results. Consider the case of two loci with two allele types each, i.e., \(L = 2\) and \(d_1 = d_2 = 2\). A particular example with one type allele interaction
and the bound constants were set to \(A = \frac{h}{2} (\theta_1 + \theta_2)\), \(C^- = -\frac{h}{2} (|\theta_1| + |\theta_2|)\) and \(C^+ = \frac{h}{2} \left(\frac{1}{2} + \theta_1 + \theta_2\right)\), where \(|\theta| = \theta_1 + \theta_2\).

Results of several simulation scenarios for sampled skeletons of paths solution of (5.1) are shown in Table 1 and Table 2, where the total length \(T\) of the considered interval \([0, T]\), the initialization of the path \((x_0^1, x_0^2)\), the average number of attempts (drawn skeletons) before acceptance, average number of Poisson points needed before acceptance, average number of coefficients computed in Algorithm 3, percentage of approximations needed due to small \(ts\) in between drawn points of the skeleton, acceptance probability and average time in seconds per accepted path are reported.

As shown in Table 1 and Table 2, the average number of coefficients needed in shorter intervals where the sampled Poisson points are more likely to be close to each other \((t \rightarrow 0)\) are larger than for longer intervals, as was expected from the results presented in Proposition 4.1. Also, the acceptance probabilities when \(h = 1\) drop to around half compared with the simulations when \(h = 0.1\). This is also expected, as the model with larger pairwise interaction parameter differs more from the candidate paths, so acceptance of the candidate becomes harder. This is also reflected in the average number of attempts, needed Poisson points and coefficients, which increase consistently in the case \(h = 1\). Running time also increases with increasing \(T\), so as to in the case for \(h = 1\) and \(T = 5\), total running times became prohibitive.

In order to establish the correctness of Algorithm 4 and with the aim of providing a qualitative comparison, samples from the paths of (5.1) at a large \(T\) were compared with the corresponding stationary density, and this was done for different mutation

### Table 1. Table for 10,000 sampled paths satisfying (5.1) with \(h = 0.1\) and \(\theta_1 = \theta_2 = \theta_1^2 = \theta_2^2 = 0.01\)

| \(T\) \(x_0^1, x_0^2\) | Attempts | Poisson points | Coeffs | Approx | Acc prob | Time (s) |
|---|---|---|---|---|---|---|
| 0.1 \((0.5, 0.5)\) | 1.08 | 0.0040 | 302.46 | 0.0018 | 0.93 | 0.14 |
| 0.1 \((0.02, 0.8)\) | 1.10 | 0.0004 | 307.77 | 0.014 | 0.91 | 0.14 |
| 1.0 \((0.5, 0.5)\) | 1.08 | 0.0028 | 7.95 | 0.0012 | 0.92 | 0.001 |
| 1.0 \((0.02, 0.8)\) | 1.10 | 0.0025 | 8.09 | 0.0022 | 0.91 | 0.001 |
| 5.0 \((0.5, 0.5)\) | 1.09 | 0.0148 | 4.10 | 0.0008 | 0.92 | 0.0007 |
| 5.0 \((0.02, 0.8)\) | 1.11 | 0.0121 | 4.14 | 0.0006 | 0.89 | 0.0009 |
Table 2. Table for 10,000 sampled paths satisfying (5.1) with $h = 1$ and $\theta_1^i = \theta_2^i = \theta_1^2 = \theta_2^2 = 0.01$.

| $T$ | $(x_{10}^0, x_{20}^0)$ | Attempts | Poisson points | Coeffs | Approx | Acc prob | Time (s) |
|-----|----------------------|----------|---------------|--------|--------|----------|----------|
| 0.1 | (0.5, 0.5)           | 2.11     | 0.0195        | 612.64 | 0.0001 | 0.47     | 0.298    |
| 0.1 | (0.02, 0.8)          | 2.7      | 0.025         | 775.52 | 0.0001 | 0.37     | 0.357    |
| 1.0 | (0.5, 0.5)           | 2.18     | 0.194         | 86.93  | 0.0001 | 0.46     | 0.045    |
| 1.0 | (0.02, 0.8)          | 2.72     | 0.25          | 107.24 | 0.0001 | 0.37     | 0.056    |

Figure 1. Stationary distribution (left) and histogram of 10,000 samples (right) from $(X_{1t}^1, X_{2t}^2)$ satisfying (5.1) at $T = 5$ with $(x_{10}^0, x_{20}^0) = (0.5, 0.5), h = 0.1$ and $\theta_j^i = 0.01$.

parameters (see Figure 1 and Figure 2) showing satisfactory results. Note that the stationary density for (5.1) can be explicitly calculated and its normalizing constant reads, see [1],

$$Z = \frac{\Gamma(\theta_1^2)\Gamma(\theta_2^2)}{\Gamma(\theta_1^2 + \theta_2^2)} \sum_{n=0}^{\infty} \frac{(\theta_2^2)^{n}(2h)^n}{\theta_2^{2n}n!} \frac{\Gamma(\theta_1^2)}{\Gamma(|\theta_1^2| + n)} \Gamma(\theta_1^2 + n),$$

(5.2)

where $a^{(n)} = a(a + 1) \ldots (a + n - 1)$. In this case, the sum in (5.2) is truncated at $n = 70$, from which the rest of the sum can be considered negligible.

Consider now the case of four loci with two allele types each, i.e., $L = 4$ and $d_l = 2$ for $l \in \{1, 2, 3, 4\}$. A particular example with one type allele interaction between loci with no selection reduces (2.1) to

$$
\begin{align*}
    dX_1^1 &= \alpha(X_1^1)dt + X_1^1(1 - X_1^1)(h_1X_2^2 + h_2X_3^3 + h_3X_4^4)dt + \sqrt{X_1^1(1 - X_1^1)}dB_1^1 \\
    dX_2^2 &= \alpha(X_2^2)dt + X_2^2(1 - X_2^2)h_1X_1^1dt + \sqrt{X_2^2(1 - X_2^2)}dB_2^2 \\
    dX_3^3 &= \alpha(X_3^3)dt + X_3^3(1 - X_3^3)h_2X_1^1dt + \sqrt{X_3^3(1 - X_3^3)}dB_3^3 \\
    dX_4^4 &= \alpha(X_4^4)dt + X_4^4(1 - X_4^4)h_3X_1^1dt + \sqrt{X_4^4(1 - X_4^4)}dB_4^4 \\
\end{align*}
$$

(5.3)
The bound constants were set to $\mathbf{A}$ to $\mathbf{C}$ similarly to the previous example with $\alpha$ where model (5.3) differs more from the candidate process, as say, a model with only one $\mathbf{H}$ to $\mathbf{X}$ means that all accepted paths used only the initial and last points, corresponding to $X_0$ and $X_T$. Thus, it is still feasible to use Algorithm 4 in these scenarios, as other approximate simulation strategies would be affected by the same problems. It is also of importance to note that the average number of Poisson points needed is 0, which means that all accepted paths used only the initial and last points, corresponding to $X_0$ and $X_T$.

where $\alpha(\cdot)$ is as in (2.1), $H^{12} = H^{21} = \begin{pmatrix} h_1 & 0 \\ 0 & 0 \end{pmatrix}$, $H^{13} = H^{31} = \begin{pmatrix} h_2 & 0 \\ 0 & 0 \end{pmatrix}$ and $H^{14} = H^{41} = \begin{pmatrix} h_3 & 0 \\ 0 & 0 \end{pmatrix}$, and $H$ is 0 everywhere else.

In this case, $A(X_0, X_T) := h_1[X_1^2X_2^2 - X_0^2X_0^2] + h_2[X_1^3X_2^3 - X_0^3X_0^3] + h_3[X_1^4X_2^4 - X_0^4X_0^4]$ and

$$
\phi(X_t) = \frac{1}{2} \left[ (h_1X_t^2 + h_2X_t^3 + h_3X_t^4)^2X_t(1 - X_t^4) + (X_t^4)^2[(h_1)X_t^2(1 - X_t^2) + (h_2)^2X_t^3(1 - X_t^2) + (h_3)^2X_t^4(1 - X_t^4)] + 2[(h_1X_t^2 + h_2X_t^3 + h_3X_t^4)\alpha(X_t^4) + X_t^4(h_1\alpha(X_t^2) + h_2\alpha(X_t^3) + h_3\alpha(X_t^4))] \right].
$$

The bound constants were set to $A^+ = h_1(1 - x_0^2x_0^2) + h_2(1 - x_0^3x_0^3) + h_3(1 - x_0^4x_0^4)$, $C^- = -\frac{1}{2}[(h_1 + h_2 + h_3)|\theta|^4 + h_1|\theta|^2 + h_2|\theta|^3 + h_3|\theta|^4]$ and $C^+ = \frac{1}{2} \left[ \frac{(h_1 + h_2 + h_3)^2 + (h_1)^2 + (h_2)^2 + (h_3)^2}{4} + (h_1 + h_2 + h_3)\theta_1 + h_1\theta_2 + h_2\theta_1 + h_3\theta_4 \right].$

Similarly to the previous example with $L = 2$ and interaction parameter $h = 1$, model (5.3) differs more from the candidate process, as say, a model with only one pairwise interaction parameter (that is, a model with, for example, $h_2 = h_3 = 0$). This is clearly reflected by the average low acceptance probabilities (see Table 3). Nonetheless, it is still feasible to use Algorithm 4 in these scenarios, as other approximate simulation strategies would be affected by the same problems. It is also of importance to note that the average number of Poisson points needed is 0, which means that all accepted paths used only the initial and last points, corresponding to $X_0$ and $X_T$. 

**Figure 2.** Stationary distribution (left) and histogram of 10,000 samples (right) from $(X_1^T, X_2^T)$ satisfying (5.1) at $T = 5$ with $(x_0^1, x_0^2) = (0.5, 0.5)$, $h = 0.1$ and $\theta_1^j = 1.2$, $\theta_2^j = 0.8$.
Table 3. Table for 10,000 sampled paths with $h_1 = 0.1, h_2 = 0.15, h_3 = 0.2$ and $\theta_j = 0.2$

| $T$ | $x_{10}^i$ | Attempts | Poisson points | Coeffs | Approx | Acc prob | Time (s) |
|-----|------------|----------|---------------|---------|--------|----------|----------|
| 0.1 | 0.5        | 1.46     | 0             | 415.88  | 0.0001 | 0.68     | 0.36     |
| 1.0 | 0.5        | 2.08     | 0             | 110.67  | 0.0001 | 0.48     | 0.123    |
| 5.0 | 0.5        | 9.94     | 0             | 871.71  | 0.0001 | 0.10     | 1.45     |

6. Simulation of multidimensional neutral Wright-Fisher bridges

To complete our simulation scheme, this section presents an exact simulation technique for sampling from neutral $(d-1)$-dimensional Wright-Fisher bridges. As mentioned before, once Algorithm 4 recovers a skeleton of the desired coupled Wright-Fisher diffusion, the remaining of the path can be filled by sampling from the corresponding neutral Wright-Fisher bridges, with no further reference to the target distribution needed, see, for example, [2].

Consider a $(d-1)$-dimensional Wright-Fisher bridge, between $x$ at time 0 and $z$ at time $t$. Its transition density is given by, see [13],

$$g_{z,t}(x,y;s) = g(x,y;s) g(y,z;t-s) g(x,z;t), \quad 0 < s < t, \quad (6.1)$$

where $g(\cdot,\cdot,\cdot)$ is as in (3.2). The precise eigenfunction expansion for $g_{z,t}(x,\cdot;s)$ is provided in the following proposition, whose prove can be found in Section 7.

**Proposition 6.1.** Let $g_{z,t}(x,\cdot;s)$ be the transition density function of a $(d-1)$-dimensional Wright-Fisher bridge. Then, its eigenfunction expansion reads

$$g_{z,t}(x,y;s) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{l=0}^{l_m} \sum_{r=0}^{r_m} \sum_{l_1, n_1}^{l_m, n_m} p_{m,n,l,r}^{(x,z,s,t,\theta)} \mathcal{D}_{\theta+1} \mathcal{M}_{\theta+1}(r),$$

with

$$p_{m,n,l,r}^{(x,z,s,t,\theta)} = \frac{q_{m}^{d}(s) q_{m}^{d}(t-s) \mathcal{M}_{m,x}(l) \mathcal{D}_{\theta+r}(z) \mathcal{D}_{\theta+r}(r)}{g(x,z;t)}$$

where $\mathcal{D}_{\theta+r}(\cdot)$ denotes the PMF of a Dirichlet-Multinomial random variable, with

$$\mathcal{D}_{\theta+r}(r) = \frac{n! \Gamma((\theta+l+r)_{l=1}^{d})}{\Gamma(\theta+l+r_{l=1}^{d})} \prod_{j=1}^{d} \frac{\Gamma(\theta_{j}+l_{j}+r_{j})}{l_{j}! \Gamma(\theta_{j}+l_{j})}.$$
Algorithm 5

1. Simulate $(M, N, L, R) \sim \{p_{m,n,l,r}^{(x,z,s,t,\theta)} : (m, n, l, r) \in \mathbb{N}^4\}$.
2. Given $(M, N, L, R) = (m, n, l, r)$, simulate $Y \sim \text{Dirichlet}(\theta + l + r)$.
3. return $Y = (y_1, \ldots, y_{d-1})$.

one-dimensional case, note that evaluating $g(x, z; t)$ at $x$ and $z$ is a different problem than sampling from it.

By (3.2) and (3.3), one obtains

$$g(x, z; t) = \sum_{m=0}^{\infty} \sum_{i=0}^{\infty} (-1)^i c_{m+i, m}^{(x, z, t, \theta)}(m) = \sum_{m=0}^{\infty} \sum_{i=0}^{\infty} (-1)^i b_{m+i}^{(t, \theta)}(m) E[D_{\theta + L_m}(z)],$$

where $L_m \sim \text{Multinomial}(m, x)$.

As in Section 3, the aim is to find monotonically converging bounds on $p_{m,n,l,r}^{(x,z,s,t,\theta)}$ so that the alternating series method can be applied.

Let

$$d_{2m} = \sum_{i=0}^{m} c_{m+i, m-i}^{(x, z, t, \theta)}(m) \quad \text{and} \quad d_{2m+1} = \sum_{i=0}^{m} c_{m+i+1, m-i}^{(x, z, t, \theta)}(m), \quad \text{for } m = 0, 1, \ldots,$$

which rearranging the terms in (6.2), gives the alternating series

$$g(x, z; t) = \sum_{m=0}^{\infty} (d_{2m} - d_{2m+1}) = d_0 - d_1 + d_2 - \ldots$$

Indeed, it can be proved that the terms $(d_i)_{i \geq 0}$ are monotonically decreasing from a certain threshold that is characterized in the following results, a required condition to apply the alternating series method.

First, note that the strategy presented in [21] for the one-dimensional case applies here almost \textit{mutatis mutandis}, with the exception of the terms involving $E[D_{\theta + L_m}(z)]$ which by an analogous (generalized) argument can be shown to decrease in $m$, as shown in the following lemma, proved in Section 7.

\textbf{Lemma 6.1.} Let $L_m \sim \text{Multinomial}(m, x)$. Then for all $m \in \mathbb{N}$

$$E[D_{\theta + L_{m+1}}(z)] \leq \tilde{K}^{(\theta, x, z)} E[D_{\theta + L_m}(z)],$$

where

$$\tilde{K}^{(\theta, x, z)} = \left( \frac{\theta}{\theta_d} \lor 2(1 + |\theta|) \frac{1 - \sum_{j=1}^{d-1} x_j}{1 - \sum_{j=1}^{d-1} z_j} \right) + \sum_{j=1}^{d-1} \left( \frac{|\theta|}{\theta_j} \lor 2(1 + |\theta|) \frac{x_j}{z_j} \right).$$

Now, similarly to (6.3), the following bound is defined

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

and used in the next Proposition 6.2 that fully characterizes the bound on $m$. The proof is again deferred to Section 7.
Proposition 6.2. Let the sequence \((d_i)_{i \geq 0}\) be as defined in (6.3), and consider the bounds \(E^{(t, \theta)}\), and \(D^{(t, \theta)}_\epsilon\) as in (6.4), and (3.6), respectively. Then, for \(\epsilon \in (0, 1)\)
\[
d_{2m+2} < d_{2m+1} < d_{2m}
\]
for all \(m \geq E^{(t, \theta)} \lor D^{(t, \theta)}_\epsilon \lor 2\tilde{K}^{(\theta, x, z)}/\epsilon\).

Once the bound in Proposition 6.2 is established, exact simulation of \((d-1)\)-dimensional Wright-Fisher bridges \((d > 2)\) is possible by setting
\[
\tilde{F}^{(s, t, \theta)}_{m, n, l, r} := C^{(s, \theta)}_m \lor C^{(t-s, \theta)}_n \lor E^{(t, \theta)} \lor D^{(t, \theta)}_\epsilon \lor 2|\theta|/\epsilon,
\]
which for \(2u \geq \tilde{F}^{(s, t, \theta)}_{m, n, l, r}\) provides the monotonically converging bounds
\[
e_{m, n, l, r}(2u + 1) < e_{m, n, l, r}(2u + 3) < p^{(x, z, s, t, \theta)}_{m, n, l, r} < e_{m, n, l, r}(2u + 2) < e_{m, n, l, r}(2u),
\]
where
\[
e_{m, n, l, r}(u) = \sum_{i=0}^{u} (-1)^i b^{(s, \theta)}_{m+i}(m) \sum_{i=0}^{u} (-1)^i b^{(t-s, \theta)}_{n+i} M_{m, x}(l) D_{\theta+r}(z) D M_{\theta+l:n}(r),
\]
see Proposition 4 in [21].

To recover exact samples from \(p^{(x, z, s, t, \theta)}_{m, n, l, r}\), consider the pairing bijective function \(\Sigma : \mathbb{N} \rightarrow \mathbb{N}^4\) such that \(\Sigma(j) = (m, n, l, r)\). Now, for each \(j\) there exist \(v_j\), elements of \(\tilde{v} \in \mathbb{R}^{J+1}\) (i.e., \(\tilde{v} = \{v_j\}_{j=0}^{J}\)) such that
\[
R^{-}_{\tilde{v}}(J) := \sum_{j=0}^{J} e^{\Sigma(j)}(2v_j + 1) \leq \sum_{j=0}^{J} p^{(x, z, s, t, \theta)}_{\Sigma(j)} \leq \sum_{j=0}^{J} e^{\Sigma(j)}(2v_j) := R^{+}_{\tilde{v}}(J),
\]
providing an analogous setting as the one presented in Section 3. The proposed exact sampling scheme can be found in the following Algorithm 6, see Algorithm 5 in [21], that we reproduce here for completeness.

Other approaches to exact simulation of one-dimensional Wright-Fisher bridges include that recently proposed in [17] that restricts to the case where either \(\theta_1\) or \(\theta_2\) are 0, and one or both of \(x\) and \(z\) are 0, which is not applicable here.

7. Proofs

Proof of Proposition 2.1. Let \(V : \mathbb{R}^n \rightarrow \mathbb{R}^n\) be such that \(V := \nabla_x (\nabla \circ f)(x)\) where recall that
\[
f^{ik}(x) = \pi^{ik} = \begin{cases} x^{ij}, & j, k = \{1, \ldots, d_i - 1\}, \\ 1 - \sum_{j=1}^{d_i-1} x^{ij}, & k = d_i. \end{cases}
\]
Algorithm 6 Exact simulation of samples from the discrete random variable with PMF \( p_{m,n,d,r}^{(x,z,s,t,θ)} : (m,n,l,r) \in \mathbb{N}^4 \)

1. Set \( j \leftarrow 0, v_0 \leftarrow 0, \bar{v} \leftarrow (v_0) \)
2. Simulate \( U \sim \text{Uniform}(0,1) \)
3. repeat
   4. Set \( v_j \leftarrow [F_{Σ(j)}(v) / 2] \)
   5. while \( R^-_j(j) < U < R^+_j(j) \) do
      6. Set \( \bar{v} \leftarrow \bar{v} + (1, \ldots, 1) \)
   7. end while
   8. if \( R^-_j(j) > U \) then
      9. return \( Σ(j) \)
   10. else if \( R^+_j(j) < U \) then
      11. Set \( \bar{v} \leftarrow (v_0, \ldots, v_j, 0) \)
      12. Set \( j \leftarrow (j + 1) \)
   13. end if
5. until false

Then \( ∀ i \in \{1, \ldots, L\}, j \in \{1, \ldots, d_i - 1\}, \)

\[
V^{ij}(x) = \frac{∂(V \circ f)(x)}{∂x^j} = \frac{∂}{∂x^j} V(\varphi) - \frac{∂}{∂x^j} f(\varphi)
\]

\[
= s^{ij} + \sum_{l=1}^{L} \sum_{k=1}^{d_l} h_{lk}^{ij} x_k - s^{ij} - \sum_{l=1}^{L} \sum_{k=1}^{d_l} h_{id_l}^{ij} x_k
\]

\[
= s^{ij} - s^{id_i} + \sum_{l=1}^{L} \left( h_{jd_l}^{il} - h_{id_l}^{il} + \sum_{k=1}^{d_l-1} \left( h_{lk}^{il} - h_{j(k+1)}^{il} + h_{j(k-1)d_i}^{il} + h_{j(k)d_i}^{il} \right) \right)
\]

\[
= K_{i,i}^{ij} + \sum_{l=1}^{L} \left( K_{i,i}^{ij} + \sum_{k=1}^{d_l-1} K_{k,i}^{ij} x_k \right) = K_{i,i}^{ij} + \sum_{l=1}^{L} \left( K_{i,i}^{ij} + \sum_{k=1}^{d_l-1} K_{k,i}^{ij} x_k \right)
\]

where the last equality holds because whenever \( i = l \), all entries of the blocks \( H^{li} \) are 0.

Proof of Theorem 4.1 By definition of \( A(\cdot), φ(\cdot), \) and \( X_t \),

\[
\exp \{ A(X_0, X_T) \} \exp \left\{ - \int_0^T φ(X_t) dt \right\} = \exp \left\{ \int_0^T V(X_t) \cdot dX_t - \int_0^T \frac{1}{2} [ (V(X_t))^T D(X_t) V(X_t) + 2 V(X_t))^T α_X(X_t) ] dt \right\}.
\]

\[
= \exp \left\{ \int_0^T D_{\tau} V(X_t) \cdot dB_t - \int_0^T \frac{1}{2} [ (V(X_t))^T D(X_t) V(X_t) ] dt \right\}.
\]

Because \( V(\cdot) \) and \( D(\cdot) \) are continuous on \([0,1]^n\), there exist constants \( C^- \) and \( C^+ \) such that \( C^- ≤ \frac{1}{2} V(X_t)^T D(X_t) V(X_t) ≤ C^+ \), a.s.
Consequently, \[ \int_0^T \frac{1}{2} (V(X_t))^T D(X_t) V(X_t) dt < \infty, \quad \text{a.s.,} \]
so Novikov’s condition is fulfilled and (7.1) can be identified as a Girsanov transformation \([23]\) with Girsanov kernel \(D^\alpha(X_t)V(X_t)\).

Let \(Q\) be the probability measure with 
\[
\frac{dQ}{d\mathcal{WF}\alpha,x_0} = \exp \{ A(X_0, X_T) \} \exp \left\{ -\int_0^T \phi(X_t) dt \right\}.
\]
It follows that the law of \(X\) under \(Q\) coincides with \(\mathcal{WF}\alpha,G,x_0\). Indeed, by Girsanov’s theorem
\[
\tilde{B}_t = B_t - \int_0^t D^\alpha(X_s)V(X_s)ds,
\]
is a \(Q\)-Brownian motion and
\[
dX_t = [\alpha(X_t) + D(X_t)V(X_t)]dt + D^\alpha(X_t)d\tilde{B}_t = [\alpha(X_t) + G(X_t)]dt + D^\alpha(X_t)d\tilde{B}_t.
\]
Now by Proposition 2.1 and for \(K_s, K_l, K_{lk} \in \mathbb{R}^n\)
\[
\int_0^T V(X_t) \cdot dX_t = \int_0^T \left( K_s + \sum_{l=1}^L \left( K_l + \sum_{k=1}^{d_l-1} K_{lk} X_t^{lk} \right) \right) \cdot dX_t
\]
\[
= \int_0^T K_s \cdot dX_t + \int_0^T \sum_{l=1}^L K_l \cdot dX_t + \int_0^T \sum_{k=1}^{d_l-1} K_{lk} X_t^{lk} \cdot dX_t
\]
\[
= \sum_{i=1}^L \sum_{j=1}^{d_i} \left( \int_0^T K_i^{ij} dX_t^i + \int_0^T \sum_{l=1}^L K_i^{ij} dX_t^{ij} + \int_0^T \sum_{l=i+1}^L \sum_{k=1}^{d_l} K_{lk}^{ij} d(X_t^{lk} X_t^{ij}) \right)
\]
where the farmost right term comes from pairing terms of the form
\[
K_{lk}^{ij} X_t^{lk} dX_t^i + K_{ij}^{lk} X_t^{ij} dX_t^{lk} = K_{lk}^{ij} d(X_t^{lk} X_t^{ij}),
\]
and recalling that \(K_{lk}^{ij} = K_{lk}^{ij}\), and \(i \neq l\) prevents squared terms. Hence,
\[
\int_0^T V(X_t) \cdot dX_t = \sum_{i=1}^L \sum_{j=1}^{d_i} \left( K_{ij}^{ij} (X_t^i - X_0^i) + \sum_{l=1}^L K_{ij}^{ij} (X_t^{ij} - X_0^{ij}) \right)
\]
\[
+ \sum_{l=i+1}^L \sum_{k=1}^{d_l} K_{lk}^{ij} (X_t^{lk} X_t^{ij} - X_0^{lk} X_0^{ij}) \right).
\]
The fact that
\[
A(X_0, X_T) := \sum_{i=1}^L \sum_{j=1}^{d_i} \left( K_{ij}^{ij} (X_t^i - X_0^i) + \sum_{l=1}^L K_{ij}^{ij} (X_t^{ij} - X_0^{ij}) \right)
\]
Proof of Proposition 6.1. Let $\mathcal{D}$ be the $1$-dimensional Wright-Fisher bridge, between $x$ and (3.2) $x_0$. Then applying the law of total expectation concludes the proof.

Proof of Proposition 4.1. Let $M(t)$ be the total number of coefficients that must be computed in Algorithm 3 with $t \in (0, T)$ is the time distance between two sampled skeleton points. By Proposition 5 (iv) in [21], there exists a $\kappa > 0$ such that $E[M(t)] = o(t^{-(1+\kappa)})$ as $t \to 0$, and further random coefficients needed in Algorithm 2 do not add to the algorithms’ complexity. Similarly, although our rejection scheme uses Algorithm 3 $L$ times,

$$E[LM(t)] = LE[M(t)] = o(t^{-(1+\kappa)})$$

so the algorithm’s complexity is proportional to $L$, but its growth rate as $t \to 0$ remains the same as with $L = 1$.

Let now $\epsilon := dCWF_{\alpha,G,x_0}/d\mathcal{W}F_{\alpha,x_0}$ be the acceptance-rejection probability in Algorithm 4. Because $A^- \leq A(X_T) \leq A^+$ and $C^- \leq \phi(X_t) \leq C^+$,

$$\epsilon \propto \exp \{A(X_T) - A^+\} \exp \{-\int_0^T (\phi(X_t) - C^-) dt\} \geq \exp \{-T(C^+ - C^-) + A^+ - A^-\}.$$ 

Let $D$ refer to the number of Poisson points needed to decide upon acceptance or rejection of a proposed path. Then, following Proposition 3 in [2]

$$E[N(T)] = E[D]/\epsilon = T(C^+ - C^-)/\epsilon \leq T(C^+ - C^-)e^{T(C^+ - C^-) + A^+ - A^-},$$

where the first equality follows from considering the expectation of the sum of all drawn Poisson points $\sum_{i=1}^T D_i$ over $I$ iterations of the algorithm until the first path is accepted. Conditioning first on $I$ and applying the law of iterated expectations, and then applying the law of total expectation concludes the proof.

Proof of Proposition 6.1. Let $g_{z,t}(x, y; s)$ be the transition density function of a $(d - 1)$-dimensional Wright-Fisher bridge, between $x$ at time 0 and $z$ at time $t$. By (6.1) and (3.2)

$$g_{z,t}(x, y; s) = \frac{g(x, y; s)g(y, z; t - s)}{g(x, z; t)} = \frac{1}{g(x, z; t)} \sum_{m=0}^{\infty} q_{m+1}^\theta(s) \sum_{|l|=m} \mathcal{M}_{m,x} (l) D_{\theta+l(y)} \sum_{n=0}^{\infty} q_n^\theta(t - s) \sum_{|r|=n} \mathcal{M}_{n,y} (r) D_{\theta+r(z)}$$

$$= \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{|l|=m} \sum_{|r|=n} q_{m+1}^\theta(s) q_n^\theta(t - s) g(x, z; t) \mathcal{M}_{m,x} (l) D_{\theta+l(y)} \mathcal{M}_{n,y} (r) D_{\theta+r(y)}.$$ 

By definition
\[ \mathcal{M}_{n,y}(r)D_{\theta + l}(y) = \frac{n!}{\prod_{j=1}^{d} l_j!} (1 - \sum_{j=1}^{d-1} y_j)^{l_j} \prod_{j=1}^{d-1} y_j^{r_j} \]
\[ \times \frac{\Gamma((\theta + l))}{\prod_{j=1}^{d} \Gamma(\theta_j + l_j)} (1 - \sum_{j=1}^{d-1} y_j)^{\theta_d + l_d - 1} \prod_{j=1}^{d-1} y_j^{\theta_j + l_j - 1}. \]

Multiplying and dividing by \[ \frac{\Gamma((\theta + l + r))}{\prod_{j=1}^{d} \Gamma(\theta_j + l_j + r_j)} \]
and rearranging, shows that
\[ \frac{n!\Gamma((\theta + l))}{\Gamma((\theta + l + r))} \prod_{j=1}^{d} \Gamma(\theta_j + l_j + r_j)! \frac{\Gamma((\theta + l + r))}{\prod_{j=1}^{d} \Gamma(\theta_j + l_j + r_j)} (1 - \sum_{j=1}^{d-1} y_j)^{\theta_d + l_d + r_d - 1} \prod_{j=1}^{d-1} y_j^{\theta_j + l_j + r_j - 1} = D_M \mathcal{M}_{\theta + l,m}(r) D_{\theta + l+r}(y), \]

Now, identifying the coefficients of \( p_{m,n,l,d}^{(x,z,s,t)} \), the proof is complete. \( \square \)

**Proof of Lemma 6.1** For the sake of simplicity, we will first consider the case \( d = 3 \) and show that by analogous arguments our results can be extended to the general \((d - 1)\)-dimensional case. First, note the indexes of the sum on the right hand side of
\[ E[D_{\theta + L_m}(z)] = \sum_{|l| = m} \text{Pr}(L_m = l) D_{\theta + l}(z), \]
can be seen as placed on the \((d - 1)\)-face of the \(d\)-simplex \( S_d = \{ l \in \mathbb{R}^d \mid l_j \geq 0, \sum_{j=1}^{d} l_j = m \} \). For example, if \( d = 3 \), we only need to consider indexes \( l_1 \) and \( l_2 \) (see Figure 3). Thus,
\[ E[D_{\theta + L_m}(z)] = \sum_{l_1=0}^{m} \sum_{l_2=0}^{m - l_1} \text{Pr}(L_m = l) D_{\theta + l}(z). \quad (7.2) \]

Let us define the quantity
\[ Q_m := \text{Pr}(L_m = l) D_{\theta + l}(z). \]

Then, for \( l_1 \leq |mz_1|, l_2 \leq |mz_2|, \)
\[ Q_{m+1} = \frac{m + 1}{m + 1 - l_1 - l_2} \frac{|\theta| + m}{\theta_3 + m - l_1 - l_2} (1 - x_1 - x_2)(1 - z_1 - z_2) Q_m \quad (7.3) \]
\[ \leq \frac{m + 1}{1 + m(1 - z_1 - z_2) \theta_3 + m(1 - z_1 - z_2)} (1 - x_1 - x_2)(1 - z_1 - z_2) Q_m \]
\[ \leq \frac{m^2 + m(1 + |\theta|) + |\theta|}{m^2(1 - z_1 - z_2)^2} (1 - x_1 - x_2)(1 - z_1 - z_2) Q_m \]
\[ \leq \left( \frac{|\theta|}{\theta_3} \vee 2(1 + |\theta|) \frac{1}{1 - z_1 - z_2} \right) Q_m. \]
Figure 3. Graphical representation of the 2-dimensional case ($d = 3$). Each point in the figure represents an element of the sum in (7.2). For every and up to each $m$, all points depicted in the plane $l_1 - l_2$ (left) correspond to points depicted in the $m$th $(l_1 - l_2 - l_3)$-face (right), which in turn represent each and every one of the summands in (7.2).

where for the last inequality, note that the function $g(m) := \frac{m^2 + m(1 + |\theta|) + |\theta|}{m^2(1 - z_1 - z_2)}$ is decreasing in $m$, and thus for $m \geq 1$, it attains its maximum value at $g(1)$. Then, if $f(m) := \frac{m + 1}{1 + m(1 - z_1 - z_2)} + m(1 + |\theta|)$, one obtains $f(m) \leq f(0) \lor g(m) \leq f(0) \lor g(1)$ yielding the desired result.

Similarly, for $l_1 \leq \lfloor mz_1 \rfloor$, $\lfloor mz_2 \rfloor \leq l_2$,

$$Q_{m+1} = \frac{m + 1 + |\theta| + m}{l_1 + 1 + \theta_1 + l_1} \frac{1}{m^2 z_2^2} Q_m \leq \frac{m + 1 + |\theta| + m}{m z_2 + 1 + \theta_2 + m z_2} x_2 z_2 Q_m \leq \left(\frac{|\theta|}{\theta_2} + 2(1 + |\theta|) \frac{x_2}{z_2}\right) Q_m,$$

and for $\lfloor mz_1 \rfloor \leq l_1$ and all $l_2$,

$$Q_{m+1} = \frac{m + 1 + |\theta| + m}{l_1 + 1 + \theta_1 + l_1} x_1 z_1 Q_m \leq \frac{m + 1 + |\theta| + m}{m z_2 + 1 + \theta_1 + m z_1} x_1 z_1 Q_m \leq \left(\frac{|\theta|}{\theta_1} + 2(1 + |\theta|) \frac{x_1}{z_1}\right) Q_m.$$

Combining the inequalities in (7.3), (7.4) and (7.5),

$$E[D_{\theta + L_{m+1}}(z)] = \sum_{l_1=0}^{m+1} \sum_{l_2=0}^{m+1-l_1} Q_{m+1} \leq \left(\frac{|\theta|}{\theta_3} \lor 2(1 + |\theta|) \frac{1 - x_1 - x_2}{1 - z_1 - z_2}\right) \sum_{l_1=0}^{mz_1} \sum_{l_2=0}^{mz_2} Q_m.$$
Note that this partition covers all (non exclusive) combinations and includes all the elements of the sum. Now, comparing \( E[D] \leq \) where 
\[
\sum_{l_1=0}^{m} \sum_{l_2=|mz_2|}^{m-l_1} Q_m + \sum_{l_1=|mz_1|}^{m} \sum_{l_2=0}^{m-l_1} Q_m
\]
\[
\leq \left[ \left( \frac{|\theta|}{\theta_3} \lor \frac{1}{2(1 + |\theta|)} \right) \frac{1 - z}{1 - z_1 - z_2} \right] + \sum_{j=1}^{2} \left( \frac{|\theta|}{\theta_j} \lor \frac{2(1 + |\theta|) x_j}{z_j} \right) \right] E[D_{\theta+l}(z)],
\]
where in the first inequality the terms of the sums starting in \( l_j = |mz_j| \) are shifted by one index, and the last inequality holds after taking common factors and noting that the terms for \( l_j = |mz_j| \) are bounded by both 
\[
\left( \frac{|\theta|}{\theta_3} \lor \frac{1}{2(1 + |\theta|)} \right) \frac{1 - x - x_2}{1 - z_1 - z_2}
\]
and 
\[
\left( \frac{|\theta|}{\theta_j} \lor \frac{2(1 + |\theta|) x_j}{z_j} \right).
\]

The proof for the general \((d - 1)\)-dimensional case follows analogously. Consider

\[
E[D_{\theta} + \sum_{l_m=1}^{m} \ldots \sum_{l_{d-1}=0}^{m-|l|d-2} \Pr(L_m = l)D_{\theta+l}(z),
\]
where \( |l|d-2 = \sum_{j=1}^{d-2} l_j \).

Following the strategy used above, the sums can be partitioned in terms such that either

a) \( l_j \leq |mz_j|, \forall j \),

b) \( |mz_j| \leq l_j, j \neq 1 \) and \( l_i \leq |mz_i|, \forall i \neq j \) or

c) \( |mz_1| \leq l_1 \) and \( l_i, \forall i \neq 1 \) free.

Note that this partition covers all (non exclusive) combinations and includes all the elements of the sum. Now, comparing \( E[D_{\theta} + \sum_{l_m=1}^{m} \ldots \sum_{l_{d-1}=0}^{m-|l|d-2} \Pr(L_m = l)D_{\theta+l}(z), \) the bounding constants for each case are

a) \( \left( \frac{|\theta|}{\theta_d} \lor \frac{1}{2(1 + |\theta|)} \right) \frac{1 - x_1}{1 - x_2} \), b) \( \left( \frac{|\theta|}{\theta_j} \lor \frac{2(1 + |\theta|) x_j}{z_j} \right), \) c) \( \left( \frac{|\theta|}{\theta_1} \lor \frac{2(1 + |\theta|) x_1}{z_1} \right) \),

yielding

\[
E[D_{\theta} + \sum_{l_m=1}^{m} \ldots \sum_{l_{d-1}=0}^{m-|l|d-2} \Pr(L_m = l)D_{\theta+l}(z),
\]
with

\[
\tilde{K}^{(\theta,x,z)} = \left( \frac{|\theta|}{\theta_d} \lor \frac{2(1 + |\theta|)}{z_1} \right) \frac{1 - x}{1 - x_2} + \sum_{j=1}^{d-1} \left( \frac{|\theta|}{\theta_j} \lor \frac{2(1 + |\theta|) x_j}{z_j} \right).
\]

\( \square \)

**Proof of Proposition 6.2.** The proof follows from that of Proposition 3 in [21], which is reproduced here for completeness.
The inequality $d_{2m+1} < d_{2m}$ follows because if $m \geq E^{(t,\theta)}$ then $2j \geq C_{m-j}^{(t,\theta)}$ for all $j = 0, \ldots, m$, which, by Proposition 3.1, implies $b_{m+j+1}^{(t,\theta)}(m-j) < b_{m+j+1}^{(t,\theta)}(m-j)$. Multiplying by $E[D_{\theta+L_m}^{(z)}]$ and then summing over $j = 0, \ldots, m$ gives

$$d_{2m+1} = \sum_{j=0}^{m} (x,z,t,\theta) c_{m+j+1,m-j} < \sum_{j=0}^{m} (x,z,t,\theta) c_{m+j,m-j} = d_{2m}. \quad (7.6)$$

Proving $d_{2m+2} < d_{2m+1}$ requires some extra steps. First, note that

$$d_{2m+2} = d_{2(m+1)} = \sum_{r=0}^{m+1} (x,z,t,\theta) c_{m+1+r,m+1-r} = \sum_{j=-1}^{m} (x,z,t,\theta) c_{m+2+j,m-j},$$

where $j = r - 1$. Noting now that $2j + 1 > 2j \geq C_{m-j}^{(t,\theta)}$ for all $j = 0, \ldots, m$, which implies $b_{m+j+2}^{(t,\theta)}(m-j) < b_{m+j+1}^{(t,\theta)}(m-j)$, and using the same argument as in (7.6) yields

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

where the sum is taken only over $j = 1, \ldots, m$ so that the remaining terms in $d_{2m+2}$ and $d_{2m+1}$ can be compared. Indeed, it only remains to prove that

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

Note that

$$\frac{c_{k+1,m}^{(x,z,t,\theta)}}{c_{k,m}^{(x,z,t,\theta)}} = \frac{b_{k+1}^{(t,\theta)}(m)}{b_k^{(t,\theta)}(m)} = h_m(k) e^{(2k+|\theta|)t/2} \leq (|\theta| + 2k + 1) e^{(2k+|\theta|)t/2}, \quad (7.7)$$

where $h_m(k) := \frac{|\theta| + m + k - 1}{k - m + 1} |\theta| + 2k + 1$, the second equality follows from (3.4), and the last inequality holds because $h_m(k) < h_k(k) = |\theta| + 2k + 1$, see the proof of Proposition 1 in [21].

Because by hypothesis $m \geq D_{\theta}^{(t,\theta)}$, recalling the definition of $D_{\theta}^{(t,\theta)}$ in (3.6) and choosing $k = m + 1$ in (7.7) yields

$$c_{m+2,m}^{(x,z,t,\theta)} < (|\theta| + 2k + 1) e^{(2k+|\theta|)t/2} c_{m+1,m}^{(x,z,t,\theta)} < (1 - \epsilon)c_{m+1,m}^{(x,z,t,\theta)}. \quad (7.8)$$

Finally,

$$\frac{c_{m+1,m+1}^{(x,z,t,\theta)}}{c_{m+1,m}^{(x,z,t,\theta)}} = \frac{b_{m+1}^{(t,\theta)}(m+1)}{b_{m}^{(t,\theta)}(m+1)} \frac{E[D_{\theta+L_m}^{(z)}]}{E[D_{\theta+L_{m+1}}^{(z)}]} = \frac{|\theta| + 2m}{(m+1)(|\theta| + m)} \frac{E[D_{\theta+L_{m+1}}^{(z)}]}{E[D_{\theta+L_m}^{(z)}]}$$

$$= \frac{1}{m+1} \left(1 + \frac{m}{|\theta| + m}\right) \frac{E[D_{\theta+L_{m+1}}^{(z)}]}{E[D_{\theta+L_m}^{(z)}]} < \frac{2}{(m+1)} \frac{E[D_{\theta+L_{m+1}}^{(z)}]}{E[D_{\theta+L_m}^{(z)}]} < \epsilon,$$

where the last inequality follows because $m + 1 > m \geq 2K^{(\theta,x,z)}/\epsilon$ and using Lemma 6.1 yielding

$$c_{m+1,m+1}^{(x,z,t,\theta)} + c_{m+2,m}^{(x,z,t,\theta)} < \epsilon c_{m+1,m}^{(x,z,t,\theta)} + (1 - \epsilon)c_{m+1,m}^{(x,z,t,\theta)} = c_{m+1,m}^{(x,z,t,\theta)},$$

which concludes the proof. \qed
References

[1] Erik Aurell, Magnus Ekeberg, and Timo Koski. Networks of loci by a multilocus Wright-Fisher model with selection and mutation, and a conjecture by Motoo Kimura. arXiv e-print arXiv:1906.00716, 2019.
[2] Alexandros Beskos, Omiros Papaspiliopoulos, and Gareth O. Roberts. Retrospective exact simulation of diffusion sample paths with applications. Bernoulli, 12(6):1077–1098, 2006.
[3] Alexandros Beskos, Omiros Papaspiliopoulos, and Gareth O. Roberts. A factorisation of diffusion measure and finite sample path constructions. Methodology and Computing in Applied Probability, 10(1):85–104, Mar 2008.
[4] Alexandros Beskos and Gareth O. Roberts. Exact simulation of diffusions. Annals of Applied Probability, 15(4):2422–2444, 2005.
[5] Jose Blanchet and Fan Zhang. Exact simulation for multivariate Itô diffusions. arXiv preprint arXiv:1706.05124, 2017.
[6] Bruno Casella and Gareth O. Roberts. Exact Monte Carlo simulation of killed diffusions. Advances in Applied Probability, 40(1):273–291, 2008.
[7] Nan Chen and Zhengyu Huang. Localization and exact simulation of Brownian motion-driven stochastic differential equations. Mathematics of Operations Research, 38(3):591–616, 2013.
[8] James F. Crow and Motoo Kimura. An introduction to population genetics theory. New York, Evanston and London: Harper & Row, Publishers, 1970.
[9] Luc Devroye. Nonuniform random variate generation. Handbooks in operations research and management science, 13:83–121, 2006.
[10] Alison M. Etheridge and Robert C. Griffiths. A coalescent dual process in a Moran model with genic selection. Theoretical Population Biology, 75(4):320 – 330, 2009.
[11] Stewart N. Ethier and Thomas G. Kurtz. Markov processes: characterization and convergence. John Wiley & Sons, 2009.
[12] Warren J. Ewens. Mathematical population genetics. Interdisciplinary applied mathematics. Springer, New York, 2nd edition, 2004.
[13] Pat Fitzsimmons, Jim Pitman, and Marc Yor. Seminar on stochastic processes, progress in probability, chapter Markovian Bridges: Construction, Palm Interpretation, and Splicing, pages 101–134. Number 33. Birkhäuser Boston, Boston, MA, 1993.
[14] Robert C. Griffiths. Lines of descent in the diffusion approximation of neutral wright-fisher models. Theoretical Population Biology, 17(1):37 – 50, 1980.
[15] Robert C. Griffiths. Asymptotic line-of-descent distributions. Journal of Mathematical Biology, 21(1):67–75, 1984.
[16] Robert C. Griffiths. Coalescent lineage distributions. Advances in Applied Probability, 38(2):405–429, 2006.
[17] Robert C. Griffiths, Paul A. Jenkins, and Dario Spanò. Wright–fisher diffusion bridges. Theoretical Population Biology, 122:67 – 77, 2018.
[18] Robert C. Griffiths and Wen-Hsiung Li. Simulating allele frequencies in a population and the genetic differentiation of populations under mutation pressure. Theoretical Population Biology, 23(1):19 – 33, 1983.
[19] Robert C. Griffiths and Dario Spanò. 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.), volume 378 of LMS Lecture Note Series, chapter 15, pages 358–375. Cambridge University Press, 2010.
[20] Paul A Jenkins. Exact simulation of the sample paths of a diffusion with a finite entrance boundary. arXiv preprint arXiv:1311.5777, 2013.
[21] Paul A. Jenkins and Dario Spanò. Exact simulation of the Wright—Fisher diffusion. Annals of Applied Probability, 27(3):1478–1509, 2017.
[22] Ethan M. Jewett and Noah A. Rosenberg. Theory and applications of a deterministic approximation to the coalescent model. Theoretical Population Biology, 93:14 – 29, 2014.
[23] Ioannis Karatzas and Steven E. Shreve. Brownian Motion and Stochastic Calculus, volume 113 of Graduate Texts in Mathematics. Springer New York, NY, 2nd edition, 1998.
[24] Motoo Kimura. Stochastic processes and distribution of gene frequencies under natural selection. *Cold Spring Harbor Symposia on Quantitative Biology*, (20):33–53, 1955.

[25] Motoo Kimura. Diffusion models in population genetics. *Journal of Applied Probability*, 1(2):177–232, 1964.

[26] Peter E. Kloeden and Eckhard Platen. Numerical solution of stochastic differential equations. 1992.

[27] Nuno R. Nené, Ville Mustonen, and Christopher J. R. Illingworth. Evaluating genetic drift in time-series evolutionary analysis. *Journal of Theoretical Biology*, 437:51 – 57, 2018.

[28] Murray Pollock, Adam M. Johansen, and Gareth O. Roberts. On the exact and \( \varepsilon \)-strong simulation of (jump) diffusions. *Bernoulli*, 22(2):794–856, 2016.

[29] Marcin J. Skwark, Nicholas J. Croucher, Santeri Puranen, Claire Chewapreecha, Maiju Pesonen, Ying Ying Xu, Paul Turner, Simon R. Harris, Stephen B. Beres, James M. Musser, et al. Interacting networks of resistance, virulence and core machinery genes identified by genome-wide epistasis analysis. *PLoS genetics*, 13(2):e1006508, 2017.

[30] Matthias Steinrücken, Anand Bhaskar, and Yun S. Song. A novel spectral method for inferring general diploid selection from time series genetic data. *Annals of Applied Statistics*, 8(4):2203–2222, 2014.

[31] Paula Tataru, Maria Simonsen, Thomas Bataillon, and Asger Hobolth. Statistical inference in the Wright–Fisher model using allele frequency data. *Systematic Biology*, 66(1):e30–e46, 2017.

(Celia García-Pareja*) UNIT OF BIOSTATISTICS, IMM, KAROLINSKA INSTITUTET

*E-mail address*, Corresponding author*: celia.garcia.pareja@ki.se

(Henrik Hult and Timo Koski) DEPARTMENT OF MATHEMATICS, KTH ROYAL INSTITUTE OF TECHNOLOGY

*E-mail address*: hult@kth.se, tjtkoski@kth.se