EXACT SIMULATION OF COUPLED WRIGHT–FISHER DIFFUSIONS

CELIA GARCÍA-PAREJA ⓘ,* *** AND
HENRIK HULT,* *KTH Royal Institute of Technology*
TIMO KOSKI,* ** *KTH Royal Institute of Technology and University of Helsinki*

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 multivariate Wright–Fisher diffusions that have drifts depending on each other through a coupling term and that find applications in the study of networks of interacting genes. The proposed rejection algorithm uses independent neutral Wright–Fisher diffusions as candidate proposals, which are only needed at a finite number of points. Once a candidate is accepted, the remainder of the path can be recovered by sampling from neutral multivariate Wright–Fisher bridges, for which an exact sampling strategy is also provided. Finally, the algorithm's complexity is derived and its performance demonstrated in a simulation study.

Keywords: Exact simulation; rejection algorithm; multivariate diffusions; population genetics; coupled Wright–Fisher model; epistasis

2010 Mathematics Subject Classification: Primary 60J60; 65C30
Secondary 60H35; 65C05

1. Introduction

Sampling paths of a diffusion process remains a challenging problem in applied probability. The major bottleneck is that its finite-dimensional distributions are seldom available in closed form, and one must often resort 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-discretization grid, which, in turn, increases computational costs. 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 (random) time points in order to take a decision. The candidate needs to be such that it can be simulated without approximation, and it needs to allow for the construction of the acceptance–rejection probability by means of a

Girsanov transformation of measures; see [34]. 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 instant by simulating from suitable diffusion bridges and with no further reference to the unknown target distribution.

In their seminal paper, Beskos and Roberts [5] present an exact rejection algorithm for simulation of paths of a 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 are relaxed to one-sided bounds in a second publication [3], and a further extension based on a layered Brownian bridge construction provides an exact simulation method for boundary crossing and hitting times [4]. Other extensions include algorithms that allow for simulation of killed diffusions and have applications to double-barrier option pricing problems [9], and a localized exact algorithm that relaxes any boundedness assumptions by considering smaller pieces of the target path and can be used to simulate diffusions with boundaries [10]. Another interesting approach is that presented in [40], 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 [37]) to provide unit volatility coefficients, which hinders generalizability to the multivariate case. To overcome these issues, alternative techniques have been proposed, such as the one in [6] which introduces an exact algorithm for simulation of multivariate diffusions based on tolerance-enforced simulation and rough path analysis. This algorithm overcomes the more restrictive assumptions required in [5] and [3] but has, admittedly, 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 thereof). For diffusions with finite boundaries, for example, Brownian candidates may either differ too much from the target, thus providing low acceptance probabilities, or be unsuitable to construct the acceptance–rejection probability itself. Rejection algorithms with candidates other than Brownian motion include [31], which uses Bessel proposals to simulate target diffusions with a finite entrance boundary.

In this context, the recent work in [32] 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. Diffusions of this class, as well as their multivariate counterparts, are extensively used in population genetics, where proliferation of exact simulation algorithms can foster the use of suitable inferential techniques such as approximate Bayesian computation; see [46], [2], and [21].

Along these lines, the main contribution of this paper is to present an exact rejection algorithm for coupled Wright–Fisher diffusions, with candidates built from samples of independent (multivariate) neutral Wright–Fisher diffusions that can be recovered using the techniques presented in [32]. The coupled Wright–Fisher diffusion [1] 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). It incorporates parent-dependent mutation at each locus, interlocus selection in the form of pairwise interactions, and free recombination. The model is based on quasi-linkage equilibrium where the fitness coefficients are inspired by a Potts model (see [1, 38]) and generalize the classical additive fitness under weak selection (see e.g. [7, Ch. II]) to the multi-locus case. The assumptions in the coupled Wright–Fisher model

are suitable, for example, for studying networks of loci in recombining populations of bacteria (e.g. Streptococcus pneumoniae) that evolve under shared selective pressure when the linkage disequilibrium is low across the genome; see [14, 43]. In contrast, the model is unsuitable in populations of bacteria where the amount of homologous recombination is low, which makes it difficult to separate couplings arising from recombination from those arising from selection (e.g. Streptococcus pyogenes).

The coupled Wright–Fisher diffusion corresponds to a haploid version of the two-locus model in the case of weak selection, loose linkage in [17]; see also [20] for a different multi-locus extension. Diffusion models in the presence of epistasis, i.e., allelic interactions between loci, have been used in the analysis of sets of human immune system genes in the presence of balancing selection; see [8]. Conversely, diffusion approximations have been deemed poor in some scenarios, e.g. for low mutation rates where the stationary density is ill-defined at the boundary; see [30].

To complete the proposed exact rejection algorithm, a further contribution of this paper deals with simulation of multivariate Wright–Fisher bridges, for which an exact simulation technique is provided. These bridges allow sampling further points of the path once a skeleton of the coupled Wright–Fisher diffusion has been accepted. Our sampling approach can therefore be viewed as a generalization of that presented in [32] for the one-dimensional Wright–Fisher diffusions to the multivariate case.

The rest of the paper is structured as follows. In Section 2 we briefly present the main properties and structure of the family of coupled Wright–Fisher diffusions, together with a formal overview of 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. This leads us 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 multivariate 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. It also 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, [36] and [24]. 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 a haploid population of constant size. Extensions of the model include consideration of 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 of the family of Wright–Fisher models can be found, for example, in [12] or [18].

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 [44], [43], [45], and [39]. Within this framework, the coupled Wright–Fisher model [1] tracks the evolution of frequencies of allele types located at different loci, and, as well as locus-wise mutation and selection, describes possible selective pairwise allele interactions between loci in a haploid population. Its diffusion approximation can be derived as the weak limit of a sequence of discrete coupled Wright–Fisher models characterized by the assumption that the evolution of the population at one locus is conditionally independent of the other loci given the state of the previous generation. The coupled Wright–Fisher diffusion can be expressed as a system of stochastic differential equations of the form

$$dX_t = [\alpha(X_t) + G(X_t)]dt + D^{\frac{1}{2}}(X_t)dB_t, \ X_0 = x_0, \ t \in [0, T], \tag{2.1}$$

where X_t is a vector of frequencies of allele types, α 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 $i \in \{1, \ldots, L\}$, referring to a specific locus, and $j \in \{1, \ldots, d_i - 1\}$, referring to an allele type, so that $x = \{x^i\}_{i=1}^{L}$ where each $x^i = \{x^{ij}\}_{j=1}^{d_i-1}$. 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

$$\alpha^{ij}(x^{ij}) = \frac{1}{2}(\theta_j^i - |\theta|x^{ij}), \tag{2.2}$$

where $|\theta| = \sum_{k=1}^{d_i} \theta_k^i$ and $\theta_k^i > 0$ 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 these here.

The coupling term $G(x) \in \mathbb{R}^n$ has general form (Svirezhev–Shahshahani gradient)

$$G(x) = D(x)\nabla_x(\overline{V} \circ f)(x),$$

where the square of the diffusion matrix $D(x) = \text{diag}(D^i(x^i)) \in \mathbb{R}^{n \times n}$ is an L-block diagonal matrix with entries

$$D_{jk}^i = \begin{cases} x^{ij}(1 - x^{ij}), & j = k, \\ -x^{ij}x^{ik}, & j \neq k, \end{cases} \quad j, k \in \{1, \ldots, d_i - 1\}, i \in \{1, \ldots, L\};$$

∇_x is the gradient operator with respect to each component of x; f transforms x into the augmented $(n + L)$-dimensional vector $\overline{x}$ that reflects the dependency between allele frequencies at locus i, i.e.,

$$\overline{x}^{ik} := f^{ik}(x) = \begin{cases} x^{ij}, & j = k \in \{1, \ldots, d_i - 1\}, \\ 1 - \sum_{j=1}^{d_i-1} x^{ij}, & k = d_i; \end{cases}$$

and $\overline{V}(\overline{x}) \in \mathbb{R}$,

$$\overline{V}(\overline{x}) = (\overline{x})^T s + \frac{1}{2}(\overline{x})^T H \overline{x},$$

where $s \in \mathbb{R}^{n+L}$ is the vector of within-locus selection parameters and $H \in \mathbb{R}^{(n+L) \times (n+L)}$ is a symmetric between-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 by 0^{ii}), and off-diagonal blocks of the form $H^{il} = (H^{li})^T \in \mathbb{R}^{d_i \times d_l}$, $i \neq l$, $i, l \in \{1, \ldots, L\}$,

$$
H = \begin{pmatrix}
0^{11} & H^{12} & \cdots & \cdots & \cdots & \cdots & H^{1L} \\
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\
H^{i1} & \cdots & H^{i(i-1)} & 0^{ii} & H^{i(i+1)} & \cdots & H^{iL} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
H^{L1} & \cdots & \cdots & \cdots & \cdots & \cdots & 0^{LL}
\end{pmatrix},
$$

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 multivariate neutral mutations Wright–Fisher diffusion, where X_t describes the evolution of allele frequencies that evolve at each locus without interacting. The explicit form of $V(x) := \nabla_x(\overline{V} \circ f)(x)$ 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 $V(x) := \nabla_x(\overline{V} \circ f)(x)$. Then $\forall i \in \{1, \ldots, L\}, j \in \{1, \ldots, d_i - 1\}$,*

$$
V^{ij}(x) = K_s^{ij} + \sum_{\substack{l=1 \\ l \neq i}}^{L} \left(K_l^{ij} + \sum_{k=1}^{d_l - 1} K_{lk}^{ij} x^{lk} \right),
$$

with

$$
K_s^{ij} := s^{ij} - s^{id_i}, \quad K_l^{ij} := h_{jd_l}^{il} - h_{d_i d_l}^{il}, \text{ and } K_{lk}^{ij} := h_{jk}^{il} - h_{jd_l}^{il} - h_{d_i k}^{il} + h_{d_i d_l}^{il},
$$

where h_{jk}^{il} denotes the entry in row j and column k of the block H^{il} of H.

Following Kimura's formulation [35], 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

$$
P(x) = \frac{1}{Z}\pi(x)e^{2(\overline{V} \circ f)(x)}, \tag{2.3}
$$

where

$$
\pi(x) = \prod_{i=1}^{L} \pi^i(x^i) = \prod_{i=1}^{L} \left((1 - \sum_{j=1}^{d_i - 1} x^{ij})^{\theta_{d_i}^i - 1} \prod_{j=1}^{d_i - 1} (x^{ij})^{\theta_j^i - 1} \right)
$$

and Z is the normalizing constant

$$
Z = \int_{x \in \mathcal{X}} \pi(x)e^{2(\overline{V} \circ f)(x)}dx.
$$

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

The representation (2.3) resembles the stationary density of the haploid version of the model studied by Fearnhead (see [20, Theorem 2]), which in the two-locus case agrees with the coupled Wright–Fisher diffusion.

2.2. Overview of exact rejection algorithms for simulation of diffusions

This subsection provides an overview of the exact rejection algorithm presented in [5], and presents the same sampling scheme followed for simulating from the coupled Wright–Fisher diffusion, as detailed further in Algorithm 4. Let

$$dX_t = \mu(X_t)dt + dB_t, \ X_0 = x_0, \ t \in [0, T], \tag{2.4}$$

where $\mu(\cdot)$ is such that (2.4) admits a weakly unique 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 a Girsanov transformation of measures one can write the Radon–Nikodym derivative of $\mathbb{Q}_{x_0}$ with respect to $\mathbb{P}_{x_0}$ as

$$\frac{d\mathbb{Q}_{x_0}}{d\mathbb{P}_{x_0}} = \exp\left\{ \int_0^T \mu(B_t)dX_t - \frac{1}{2} \int_0^T \mu^2(B_t)dt \right\}. \tag{2.5}$$

Assuming that $\mu(\cdot)$ is differentiable everywhere and using Itô's lemma, (2.5) can be rewritten as

$$\frac{d\mathbb{Q}_{x_0}}{d\mathbb{P}_{x_0}} = \exp\left\{ \tilde{A}(B_T) - \tilde{A}(x_0) \right\} \exp\left\{ -\frac{1}{2} \int_0^T (\mu^2(B_t) + \mu'(B_t))dt \right\},$$

where $\tilde{A}(x) := \int_0^x \mu(u)du$. Imposing the further conditions that $\tilde{A}(x)$ be bounded above by a constant K^A, and that $(\mu^2 + \mu')/2$ be bounded between constants K^- and K^+, we find that

$$\frac{d\mathbb{Q}_{x_0}}{d\mathbb{P}_{x_0}} \propto \exp\left\{ \tilde{A}(B_T) - K^A \right\} \exp\left\{ -\int_0^T (\tilde{\phi}(B_t) - K^-)dt \right\}, \tag{2.6}$$

with $\tilde{A}(x) \le K^A$ and $K^- \le \tilde{\phi}(x) := \frac{1}{2}[\mu^2(x) + \mu'(x)] \le K^+$, is a suitable acceptance–rejection probability.

Note that an exact evaluation of the integral in (2.6) is not possible without any approximation error, because this would require the storage of 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 [5] and references therein is that an event occurring with probability (2.6) can be evaluated by sampling B only at a finite number of time points. This follows because the last term in (2.6) can be interpreted as the probability of the event $\omega_{\tilde{\phi}}$ that no points from a 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 $t \mapsto \tilde{\phi}(B_t)$. A formal statement and proof of this observation can be found in Theorem 1 of [3].

Therefore, the exact sampling procedure (detailed in Algorithm 1) starts by drawing a sample from Φ that will determine the (random) time points at which the candidate B will be drawn, and then provides a skeleton of $(X_t)_{t \in [0,T]}$ 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 previously presented versions of the algorithm, B_T is considered to follow a specific probability distribution, thus slightly modifying the candidate to be a certain biased Brownian motion, which is shown to improve the algorithm's efficiency. Such an option is not needed for our purposes and is therefore not fully described here. In brief, this modification permits the boundedness condition on $\tilde{A}(x)$ to be relaxed, but the equivalent function in our proposed exact algorithm is already bounded.

Algorithm 1 Exact algorithm for simulating skeletons of paths $(X_t)_{t\in[0,T]}$ of a diffusion process with law $\mathbb{Q}_{x_0}$

---

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

---

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 a weakly unique solution of

$$dX_t = \alpha(X_t)dt + D^{\frac{1}{2}} dB_t, \quad X_0 = x_0, \ t \in [0, T], \tag{3.1}$$

with $\alpha(X_t)$ a $(d_i - 1)$-dimensional vector with $\alpha^{ij}(x) = \frac{1}{2}(\theta_j^i - |\theta|x^{ij})$.

3.1. Transition density function expansions

Exact simulation of each neutral Wright–Fisher diffusion is possible by exploiting an available eigenfunction expansion of its transition density function that allows a probabilistic representation; see, for example, [29].

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 $\theta_j + l_j$, $j \in \{1, \ldots, d\}$. Then the probabilistic representation of the transition density function of $(X_t)_{t\in[0,T]}$ in (3.1) is given by

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

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

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

and

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

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

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 $M \sim A_\infty^\theta(t)$.
2 Given $A_\infty^\theta(t) = m$, 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** $Y = (y_1, \ldots, y_{d-1})$.

In the case of one-dimensional $(d = 2)$ Wright–Fisher diffusions, the multivariate components of the mixture in (3.2) reduce to their one-dimensional counterparts, i.e., a binomial and a beta random variable, respectively [29].

A sampling strategy for $g(x, \cdot \, ; t)$ is summarized in Algorithm 2; see [28] or [32] for an analogous version in the one-dimensional case (the latter also includes 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 $q_m^\theta(t)$ poses some difficulty because it is only known in infinite series form. Previous approaches for simulating from approximated versions of $q_m^\theta(t)$ can be found in [28], [25], [26], and [33]. In the next section, we review the exact simulation procedure presented in [32], which is the one used here.

3.2. Exact simulation of the ancestral process A_∞^θ

We describe here the sampling procedure for recovering exact samples of $q_m^\theta(t)$, the transition functions of the aforementioned death process A_∞^θ. In more detail, 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 with its only possible transition $m \to m - 1$ 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 $q_m^\theta(t) = \lim_{n \to \infty} \mathbb{P}(A_n^\theta(t) = m)$.

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 [16], 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 [15] or [29] for a complete derivation and details. The corresponding dual (coalescent) process for coupled Wright–Fisher diffusions is derived in [19].

An expression for $q_m^\theta(t)$ starting from the entrance boundary at infinity is as follows (see [24]):

$$q_m^\theta(t) = \sum_{i=0}^{\infty} (-1)^i b_{m+i}^{(t,\theta)}(m), \tag{3.3}$$

where

$$b_{m+i}^{(t,\theta)}(m) = \frac{(|\theta| + 2(m+i) - 1)}{m! \, i!} \frac{\Gamma(|\theta| + 2m + i - 1)}{\Gamma(|\theta| + m)} e^{(m+i)(m+i+|\theta|-1)t/2}. \tag{3.4}$$

As shown in [32], samples from $q_m^\theta(t)$ can be recovered exactly by means of a variant of the alternating series method, described in [13, Chapter 4]. In brief, the alternating series method would require the sequence of coefficients $b_{m+i}^{(t,\theta)}(m)$ to be decreasing in i for each m, a condition

that is not always met here. Nonetheless, one can exploit the fact 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 $b_{m+i}^{(t,\theta)}(m)$ decreases monotonically as i tends to ∞. More explicitly, there exists

$$C_m^{(t,\theta)} = \inf \left\{ i \geq 0 \colon (b_{m+i+1}^{(t,\theta)}(m)/b_{m+i}^{(t,\theta)}(m)) < 1 \right\} < \infty. \tag{3.5}$$

Then, once $C_m^{(t,\theta)}$ 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 $C_m^{(t,\theta)}$.

Proposition 3.1. ([32, Proposition 1]) *Let $b_{m+i}^{(t,\theta)}(m)$ be the coefficients defined in (3.4) and let $C_m^{(t,\theta)}$ be as in (3.5). Then*

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

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

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

$$D_\epsilon^{(t,\theta)} = \inf \left\{ u \geq \left(\frac{1}{t} - \frac{|\theta|+1}{2} \right) \vee 0 \colon (|\theta| + 2u - 1)e^{u(u+|\theta|-1)t/2} < 1 - \epsilon \right\}, \tag{3.6}$$

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

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

Following Proposition 3.1, one can then recover exact samples from $q_m^\theta(t)$ because the terms $b_{m+i}^{(t,\theta)}(m)$ become monotonically smaller with increasing i, and for each m there exists k_m, an element of $\tilde{k} \in \mathbb{R}^{M+1}$ (i.e., $\tilde{k} = \{k_m\}_{m=0}^M$), such that

$$S_{\tilde{k}}^-(M) := \sum_{m=0}^M \sum_{i=0}^{2k_m+1} (-1)^i b_{m+i}^{(t,\theta)}(m) \leq \sum_{m=0}^M q_m^\theta(t) \leq \sum_{m=0}^M \sum_{i=0}^{2k_m} (-1)^i b_{m+i}^{(t,\theta)}(m) =: S_{\tilde{k}}^+(M).$$

Because

$$\limsup_{\tilde{k} \to (\infty,\dots,\infty)} S_{\tilde{k}}^-(M) = \mathbb{P}(A_\infty^\theta(t) \leq M) \quad \text{and} \quad \liminf_{\tilde{k} \to (\infty,\dots,\infty)} S_{\tilde{k}}^+(M) = \mathbb{P}(A_\infty^\theta(t) \leq M),$$

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

$$k_m^0 = \inf \left\{ k_m \in \mathbb{N} \colon S_{\tilde{k}}^-(M) > U \text{ or } S_{\tilde{k}}^+(M) < U \right\}$$

for each $m \in \{0, \dots, M\}$.

Now, if $\tilde{k}^0$ is such that $S_{\tilde{k}^0}^-(M) > U$, standard inverse sampling provides

$$\inf \left\{ M \in \mathbb{N} \colon \sum_{m=0}^M q_m^\theta(t) \geq S_{\tilde{k}^0}^-(M) > U \right\}, \tag{3.7}$$

Algorithm 3 Exact simulation of samples from $q_m^\theta(t)$, transition functions of the ancestral process A_∞^θ

---

1 Set $M \leftarrow \hat{q}_{\mathrm{mod}}, \tilde{k} \leftarrow (0, \ldots, 0), j \leftarrow 1$
2 Simulate $U \sim \mathrm{Uniform}(0, 1)$
3 **repeat**
4   **for all** $m \in \{0, \ldots, M\}$ **do**
5   Set $k_m \leftarrow \lceil C_m^{(t,\theta)}/2 \rceil$
6   **end for**
7   **while** $S_{\tilde{k}}^-(M) < U < S_{\tilde{k}}^+(M)$ **do**
8   Set $\tilde{k} \leftarrow \tilde{k} + (1, \ldots, 1)$
9   **end while**
10   **if** $S_{\tilde{k}}^-(M) > U$, check if inf M satisfied **then**
11   **return** M
12   **else if** $S_{\tilde{k}}^+(M) < U$ **then**
13   Set $M \leftarrow \hat{q}_{\mathrm{mod}} + (-1)^j \lceil \frac{j}{2} \rceil$
14   **if** j odd **then**
15     $\tilde{k} \leftarrow (k_0, \ldots, k_M)$
16   **else if** j even **then**
17       $\tilde{k} \leftarrow (\tilde{k}, 0, \ldots, 0)$
18   **end if**
19   **end if**
20   Set $j \leftarrow j + 1$
21 **until** false

---

with M exactly distributed following $q_m^\theta(t)$. The sampling strategy will consist in exploring the summands in $S_{\tilde{k}}^-(M)$ and $S_{\tilde{k}}^+(M)$ through their respective indexes m and k_m, until, for a given realization of U, the condition (3.7) is satisfied.

A complete simulation procedure is presented in Algorithm 3, where several improvements mentioned in [32] 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 $\hat{q}_{\mathrm{mod}}$ of $q_m^\theta(t)$, a modification that decreases computation times substantially. Such initialization originates from an asymptotic approximation of the transition functions $q_m^\theta(t)$, see [25], which states that as $t \to 0$, $A_\infty^\theta(t)$ converges to a normal distribution, that is,

$$\frac{A_\infty^\theta(t) - \mu^{(t,\theta)}}{\sigma^{(t,\theta)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \text{ as } t \to 0, \tag{3.8}$$

where $\mu^{(t,\theta)} = 2\eta/t$ and

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

with $\eta = \beta/e^\beta - 1$ for $\beta \neq 0$ or $\eta = 1$ otherwise, $\beta = \frac{1}{2}(|\theta| - 1)t$, and where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution; see [32, Theorem 1].

Note that this initialization is possible because the exploration of the summands in $S^-_{\tilde{k}}(M)$ and $S^+_{\tilde{k}}(M)$ does not need to occur in any specific order. It is also worth mentioning that, when M is initialized at 0, the vector $\tilde{k}$ is updated increasingly, i.e. a new element of the vector is added at every new iteration where M is increased by one unit. In Algorithm 3, however, M is updated telescopically; that is, at each new iteration j, M moves farther from $\hat{q}_{\mathrm{mod}}$ by one unit alternatingly above or below. This in turn entails updating the corresponding $M + 1$ elements of $\tilde{k}$ accordingly; i.e., the number of elements might either increase or decrease at each iteration. In addition, because M is updated telescopically, when an M such that $S^-_{\tilde{k}}(M) > U$ is found, one needs to ensure whether this is the infimum M for which this condition is satisfied. For precise results on the complexity of Algorithm 3 and simulation performance we refer the reader to [32].

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 diffusions, each one a weakly unique solution of (3.1) and sampled exactly 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 $\mathbb{CWF}_{\alpha,G,x_0}$ be the law of X, solution of (2.1), and let $\mathbb{WFL}_{\alpha,x_0}$ be the joint law of L independent $(d_i - 1)$-dimensional neutral Wright–Fisher diffusions weakly unique solutions of (3.1), $i \in \{1, \ldots, L\}$. Then the Radon–Nikodym derivative of $\mathbb{CWF}_{\alpha,G,x_0}$ with respect to $\mathbb{WFL}_{\alpha,x_0}$ is of the form*

$$\frac{d\mathbb{CWF}_{\alpha,G,x_0}}{d\mathbb{WFL}_{\alpha,x_0}} = \exp\{A(X_0, X_T)\} \exp\left\{-\int_0^T \phi(X_t)dt\right\},$$

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

$$A(X_0, X_T) := \int_0^T V(X_t) \cdot dX_t = \sum_{i=1}^L \sum_{j=1}^{d_i-1} \left(K_s^{ij}\left(X_T^{ij} - X_0^{ij}\right) + \sum_{l=1}^L K_l^{ij}\left(X_T^{ij} - X_0^{ij}\right) \right.$$
$$\left. + \sum_{l=i+1}^L \sum_{k=1}^{d_l-1} K_{lk}^{ij}\left(X_T^{ij}X_T^{lk} - X_0^{ij}X_0^{lk}\right) \right)$$

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 3 (or the corresponding modification for one-dimensional diffusions, if $d_i = 2$ for some i), Algorithm 4 returns an exactly simulated skeleton of $(X_t)_{t \in [0,T]}$ solution of (2.1).

Algorithm 4 Exact rejection algorithm for simulating skeletons of the paths $(X_t)_{t \in [0,T]}$ of a diffusion process with law $\mathbb{CWF}_{\alpha,G,x_0}$

---

1 Simulate Φ, 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 $X \sim \mathbb{WFL}_{\alpha,x_0}$ at times $\{t_1, \ldots, t_J, T\}$.
4 **if** $\phi(X_{t_j}) - C^- \leq \psi_j, \ \forall j$ and $U \leq \exp\{A(X_0, X_T) - A^+\}$ **then**
5   **return** $\{(t_j, X_{t_j}), \ \forall j\} \cup \{(T, X_T)\}$
6 **else**
7   Go back to Step 1.
8 **end if**

The algorithm's computational complexity can also be established; this 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; let 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 is accepted. Then $\mathrm{E}[LM(t)] < \infty$ and $\mathrm{E}[N(T)] < \infty$, and more specifically, there exists $\kappa > 0$ such that*

$$\mathrm{E}[LM(t)] = o(t^{-(1+\kappa)}) \quad \text{as } t \to 0, \text{ and}$$

$$\mathrm{E}[N(T)] \leq T(C^+ - C^-)e^{T(C^+ - C^-) + A^+ - A^-}.$$

In summary, the complexity of Algorithm 4 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 until acceptance increases exponentially.

The latter is easily solvable, simply by considering shorter intervals $[t_{k-1}, t_k]$ such that

$$\bigcup_{k=1}^{K} [t_{k-1}, t_k] = [0, T], \quad \text{with } t_0 = 0, \ t_K = T, \text{ and } t_{k-1} < t_k, \ \forall k \in \{1, \ldots, K\},$$

and then concatenating the accepted skeletons in each of them. To solve the problem when $t \to 0$, we follow the recommendation in [32], and whenever $t < 0.05$, resort to the approximation in (3.8).

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 parameters, but different selective pairwise interactions.

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

			$h = 0.1$				
T	(x_0^{11}, x_0^{21})	Att.	Poisson points	Coeffs	Approx.	Acc. prob.	Time (s)
0.1	(0.5, 0.5)	1.07	0.001	300.68	0.001	0.93	0.146
0.1	(0.02, 0.8)	1.11	0.001	311.04	0.001	0.90	0.149
1.0	(0.5, 0.5)	1.08	0.004	7.79	0.001	0.92	0.001
1.0	(0.02, 0.8)	1.08	0.004	7.78	0.001	0.92	0.001
5.0	(0.5, 0.5)	1.09	0.029	3.78	0.002	0.92	0.001
5.0	(0.02, 0.8)	1.12	0.031	4.23	0.004	0.89	0.001

5. Numerical experiments

In the following, several implementations of Algorithm 4 are shown along with their simulation results. The examples below represent plausible network structures present in biological applications, and whose interaction parameters are of interest. Examples of these include frequency-dependent selection in population dynamics with applications to vaccine interventions [11], genome-wide discovery of interdependent loci affecting antibiotic resistance [42], co-evolution of interacting human gamete-recognition genes [41], and analysis of sequence data [23].

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 mutation, one-type allele interaction between loci, and no within-locus selection reduces (2.1) to

$$
\begin{cases}
dX_t^{11} = \alpha^{11}(X_t^{11})dt + X_t^{11}(1 - X_t^{11})hX_t^{21}dt + \sqrt{X_t^{11}(1 - X_t^{11})}dB_t^1, \\
dX_t^{21} = \alpha^{21}(X_t^{21})dt + X_t^{21}(1 - X_t^{21})hX_t^{11}dt + \sqrt{X_t^{21}(1 - X_t^{21})}dB_t^2,
\end{cases}
\tag{5.1}
$$

where the $\alpha^{i1}(\cdot)$ are as in (2.2), $B = (B_t^1, B_t^2)$ is a vector of independent Brownian motions, $H^{12} = H^{21} = \begin{pmatrix} h & 0 \\ 0 & 0 \end{pmatrix}$, and H is 0 elsewhere.

In this case, $A(X_0, X_T) := h(X_T^{11}X_T^{21} - X_0^{11}X_0^{21})$ and

$$
\phi(X_t) = \frac{1}{2}\left(h^2\left[(X_t^{21})^2 X_t^{11}(1 - X_t^{11}) + (X_t^{11})^2 X_t^{21}(1 - X_t^{21}) \right] \right.
$$
$$
\left. + 2h\left[X_t^{21}\alpha^{11}(X_t^{11}) + X_t^{11}\alpha^{21}(X_t^{21}) \right] \right),
$$

and the bound constants were set to $A^+ = h(1 - x_0^{11}x_0^{21})$, $C^- = -\frac{h}{2}(|\theta^1| + |\theta^2|)$, and $C^+ = \frac{h}{2}\left(\frac{h}{2} + \theta_1^1 + \theta_1^2\right)$, where $|\theta^i| = \theta_1^i + \theta_2^i$.

The results of several simulation scenarios for sampled skeletons of the diffusion solution of (5.1) are shown in Table 1 and Table 2, which report the total length T of the considered interval $[0, T]$, the initialization of the path (x_0^{11}, x_0^{21}), the average number of attempts (drawn skeletons) until acceptance, the average number of Poisson points needed until acceptance, the average number of coefficients computed in Algorithm 3, the acceptance probability, the

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

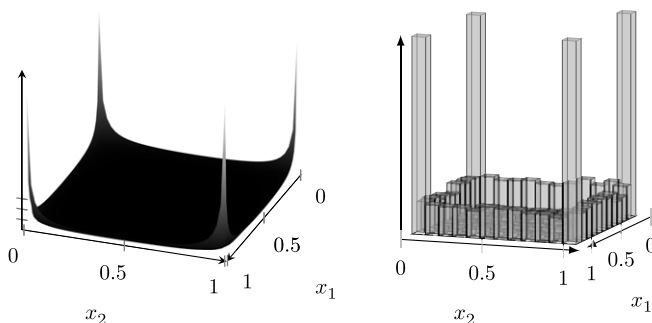|			$h = 1$				
T	(x_0^{11}, x_0^{21})	Att.	Poisson points	Coeffs	Approx.	Acc. prob.	Time (s)
0.1	(0.5, 0.5)	2.11	0.059	610.96	0.005	0.47	0.318
0.1	(0.02, 0.8)	2.69	0.076	778.90	0.005	0.37	0.393
1.0	(0.5, 0.5)	2.23	0.615	90.19	0.005	0.45	0.058
1.0	(0.02, 0.8)	2.77	0.774	111.10	0.001	0.36	0.063

FIGURE 1. Stationary distribution (left) and histogram of 10,000 samples (right) from (X_t^{11}, X_t^{21}) satisfying (5.1) at $T = 5$ with $(x_0^{11}, x_0^{21}) = (0.5, 0.5)$, $h = 0.1$, and $\theta_j^i = 0.01$.

number of approximations needed due to small ts in between drawn points of the skeleton, and the average time in seconds per accepted path.

As shown in Table 1 and Table 2, the average number of coefficients needed is larger for shorter intervals, where the sampled Poisson points are more likely to be close to each other ($t \to 0$), than for longer intervals, as 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. In the case $h = 1$ and $T = 5$ the total running times became prohibitive.

In order to establish the correctness of Algorithm 4 and with the aim of providing a qualitative comparison, sampled paths of the solution of (5.1) at large T are compared with the corresponding stationary density, and this is done for different mutation parameters (see Figure 1 and Figure 2) with satisfactory results. Note that the stationary density for (5.1) can be explicitly computed, and its normalizing constant reads as follows, see [1]:

$$Z = \frac{\Gamma(\theta_1^2)\Gamma(\theta_2^2)}{\Gamma(\theta_1^2 + \theta_2^2)} \sum_{n=0}^{\infty} \frac{(\theta_1^2)^{(n)}(2h)^n}{|\theta^2|^{(n)}n!} \frac{\Gamma(\theta_2^1)}{\Gamma(|\theta^1| + n)} \Gamma(\theta_1^1 + n),$$

where $a^{(n)} = a(a+1)\dots(a+n-1)$. For the qualitative comparisons in Figure 1 and Figure 2, the infinite sum is truncated at a sufficiently high n so that the remainder is negligible. For the parameters in Table 1 and 2 the truncation level is set to $n = 70$.
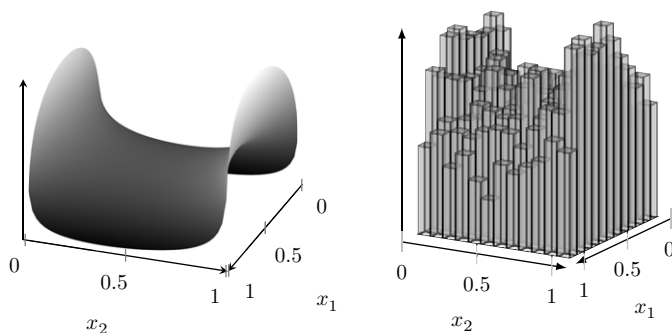
FIGURE 2. Stationary distribution (left) and histogram of 10,000 samples (right) from (X_t^{11}, X_t^{21}) satisfying (5.1) at $T = 5$ with $(x_0^{11}, x_0^{21}) = (0.5, 0.5)$, $h = 0.1$, and $\theta_j^1 = 1.2$, $\theta_j^2 = 0.8$.

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

$$
\begin{cases}
dX_t^{11} = \alpha^{11}(X_t^{11})dt + X_t^{11}(1 - X_t^{11})(h_1 X_t^{21} + h_2 X_t^{31} h_3 X_t^{41})dt + \\
\qquad\quad + \sqrt{X_t^{11}(1 - X_t^{11})}dB_t^1, \\
dX_t^{21} = \alpha^{21}(X_t^{21})dt + X_t^{21}(1 - X_t^{21})h_1 X_t^{11}dt + \sqrt{X_t^{21}(1 - X_t^{21})}dB_t^2, \\
dX_t^{31} = \alpha^{31}(X_t^{31})dt + X_t^{31}(1 - X_t^{31})h_2 X_t^{11}dt + \sqrt{X_t^{31}(1 - X_t^{31})}dB_t^3, \\
dX_t^{41} = \alpha^{41}(X_t^{41})dt + X_t^{41}(1 - X_t^{41})h_3 X_t^{11}dt + \sqrt{X_t^{41}(1 - X_t^{41})}dB_t^4,
\end{cases}
\tag{5.2}
$$

where the $\alpha^{i1}(\cdot)$ are as in (2.2), $B = (B_t^1, B_t^2, B_t^3, B_t^4)$ is a vector of independent Brownian motions,

$$
H^{12} = H^{21} = \begin{pmatrix} h_1 & 0 \\ 0 & 0 \end{pmatrix}, \; H^{13} = H^{31} = \begin{pmatrix} h_2 & 0 \\ 0 & 0 \end{pmatrix}, \; \text{and } H^{14} = H^{41} = \begin{pmatrix} h_3 & 0 \\ 0 & 0 \end{pmatrix},
$$

and H is 0 elsewhere. In this case,

$$
A(X_0, X_T) = h_1(X_T^{11} X_T^{21} - X_0^{11} X_0^{21}) + h_2[X_T^{11} X_T^{31} - X_0^{11} X_0^{31}] + h_3[X_T^{11} X_T^{41} - X_0^{11} X_0^{41})
$$

and

$$
\begin{aligned}
\phi(X_t) = \frac{1}{2}(&[(h_1)^2 X_t^{21}(1 - X_t^{21}) + (h_2)^2 X_t^{31}(1 - X_t^{31}) + (h_3)^2 X_t^{41}(1 - X_t^{41})](X_t^{11})^2 \\
&+ (h_1 X_t^{21} + h_2 X_t^{31} + h_3 X_t^{41})^2 X_t^{11}(1 - X_t^{11}) + 2[(h_1 X_t^{21} + h_2 X_t^{31} + h_3 X_t^{41})\alpha^{11}(X_t^{11}) \\
&+ (h_1 \alpha^{21}(X_t^{21}) + h_2 \alpha^{31}(X_t^{31}) + h_3 \alpha^{41}(X_t^{41}))X_t^{11}]),
\end{aligned}
$$

TABLE 3. Table for 10,000 sampled paths with $h_1 = 0.1$, $h_2 = 0.15$, $h_3 = 0.2$, and $\theta_j^i = 0.2$.

T	x_0^{i1}	Att.	Poisson points	Coeffs	Approx.	Acc. prob.	Time (s)
			$h_1 = 0.1$, $h_2 = 0.15$, $h_3 = 0.2$				
0.1	0.5	1.45	0.054	412.65	0.001	0.69	0.402
1.0	0.5	2.08	0.811	112.60	0.005	0.48	0.129
5.0	0.5	9.96	19.648	881.37	0.004	0.10	1.007

and the bound constants were set to

$$A^+ = h_1(1 - x_0^{11}x_0^{21}) + h_2(1 - x_0^{11}x_0^{31}) + h_3(1 - x_0^{11}x_0^{41}),$$

$$C^- = -\frac{1}{2}(|\theta^1|(h_1 + h_2 + h_3) + |\theta^2|h_1 + |\theta^3|h_2 + |\theta^4|h_3), \text{ and}$$

$$C^+ = \frac{1}{2}\left(\theta_1^1(h_1 + h_2 + h_3) + \frac{(h_1 + h_2 + h_3)^2 + (h_1)^2 + (h_2)^2 + (h_3)^2}{4} + h_1\theta_2^1 + h_2\theta_3^1 + h_3\theta_4^1\right).$$

Similarly to the previous example with $L = 2$ and interaction parameter $h = 1$, the model in (5.2) differs more from the candidate process than, 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 in the low average acceptance probabilities, or, equivalently, in the average number of attempts or simulated Poisson points needed until acceptance; see Table 3. Moreover, simulating from the model in (5.2) implicitly requires the simultaneous acceptance of four candidate paths, which makes acceptance more difficult. Nonetheless, it is still feasible to use Algorithm 4 in these scenarios, as other approximate simulation strategies would be affected by similar problems.

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 remainder 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, [3].

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

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

where $g(\cdot, \cdot; \cdot)$ is as in (3.2), see [22]: The precise eigenfunction expansion for $g_{z,t}(x, \cdot; s)$ is provided in the following proposition, whose proof can be found in Section 7.

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

$$g_{z,t}(x, y; s) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{\substack{l \\ |l|=m}} \sum_{\substack{r \\ |r|=n}} p_{m,n,l,r}^{(x,z,s,t,\theta)} \mathcal{D}_{\theta+l+r}(y),$$

Algorithm 5 Exact simulation of samples from $g_{z,t}(x, \cdot;s)$, transition density of the $(d-1)$-dimensional neutral Wright–Fisher bridge

---

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

---

with

$$p^{(x,z,s,t,\theta)}_{m,n,l,r} = \frac{q^\theta_m(s)q^\theta_n(t-s)}{g(x, z;t)} \mathcal{M}_{m,x}(l)\mathcal{D}_{\theta+r}(z)\mathcal{DM}_{\theta+l;n}(r),$$

where $\mathcal{DM}_{\theta+l;n}(\cdot)$ denotes the PMF of a Dirichlet–multinomial random variable, with

$$\mathcal{DM}_{\theta+l;n}(r) = \frac{n!\Gamma(|\theta+l|)}{\Gamma(|\theta+l+r|)} \prod_{j=1}^d \frac{\Gamma(\theta_j + l_j + r_j)}{l_j!\Gamma(\theta_j + l_j)}.$$

Following the result in Proposition 6.1, a sampling scheme for $g_{z,t}(x, y;s)$ is provided in Algorithm 5.

Similarly to Step 1 in Algorithm 2, sampling exactly from the discrete random variable with PMF $p^{(x,z,s,t,\theta)}_{m,n,l,r}$ is not straightforward. However, given the results obtained so far, it only remains to find how to evaluate $g(x, z;t)$ without approximation error, and the sampling strategy will be complete. As pointed out in [32] for the one-dimensional case, note that the problem of evaluating $g(x, z;t)$ at x and z is different from that of sampling from it.

By (3.2) and (3.3), one obtains

$$g(x, z;t) = \sum_{m=0}^{\infty} \sum_{i=0}^{\infty} (-1)^i c^{(x,z,t,\theta)}_{m+i,m}(m) = \sum_{m=0}^{\infty} \sum_{i=0}^{\infty} (-1)^i b^{(t,\theta)}_{m+i}(m)\mathbb{E}[\mathcal{D}_{\theta+L_m}(z)], \quad (6.2)$$

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

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

Let

$$d_{2m} = \sum_{i=0}^{m} c^{(x,z,t,\theta)}_{m+i,m-i}(m), \qquad d_{2m+1} = \sum_{i=0}^{m} c^{(x,z,t,\theta)}_{m+1+i,m-i}(m),$$

for $m = 0, 1, \ldots$, which, rearranging the terms in (6.2), gives the alternating series

$$g(x, z;t) = \sum_{m=0}^{\infty} (d_{2m} - d_{2m+1}) = d_0 - d_1 + d_2 - \ldots. \quad (6.3)$$

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, which is a condition required to apply the alternating series method.

First, note that the strategy presented in [32] for the one-dimensional case applies here almost without change, with the exception of the terms involving $\mathbb{E}[\mathcal{D}_{\theta+L_m}(z)]$ which by an

analogous (generalized) argument can be shown to decrease in m, as shown in the following lemma, proved in Section 7.

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

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

where

$$\tilde{K}^{(\theta,x,z)} = \left(\frac{|\theta|}{\theta_d} \left(1 - \sum_{j=1}^{d-1} z_j\right) \vee \frac{2(1+|\theta|)}{1 - \sum_{j=1}^{d-1} z_j} \right) \left(1 - \sum_{j=1}^{d-1} x_j\right)$$
$$+ \sum_{j=1}^{d-1} \left(\frac{|\theta|}{\theta_j} z_j \vee \frac{2(1+|\theta|)}{z_j} \right) x_j. \tag{6.4}$$

Now, similarly to (3.5), the following bound is defined:

$$E^{(t,\theta)} = \inf \left\{ m \geq 0 \colon 2j \geq C_{m-j}^{(t,\theta)} \text{ for all } j = 0, \ldots, m \right\}. \tag{6.5}$$

This is used in Proposition 6.2 below, which 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* $E^{(t,\theta)}$, $D_\epsilon^{(t,\theta)}$, *and* $\tilde{K}^{(\theta,x,z)}$ *as in (6.5), (3.6), and (6.4) respectively. Then, for* $\epsilon \in (0, 1)$,

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

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

Once the bound in Proposition 6.2 is established, exact simulation of $(d-1)$-dimensional Wright–Fisher bridges ($d > 2$) is possible by setting

$$\tilde{F}_{m,n,l,r}^{(s,t,\theta)} := C_m^{(s,\theta)} \vee C_n^{(t-s,\theta)} \vee E^{(t,\theta)} \vee D_\epsilon^{(t,\theta)} \vee 2\tilde{K}^{(\theta,x,z)}/\epsilon,$$

which for $2u \geq \tilde{F}_{m,n,l,r}^{(s,t,\theta)}$ provides the monotonically converging bounds

$$e_{m,n,l,r}(2u+1) < e_{m,n,l,r}(2u+3) < p_{m,n,l,r}^{(x,z,s,t,\theta)} < e_{m,n,l,r}(2u+2) < e_{m,n,l,r}(2u),$$

where

$$e_{m,n,l,r}(u) = \frac{\sum_{i=0}^u (-1)^i b_{m+i}^{(s,\theta)}(m) \sum_{i=0}^u (-1)^i b_{n+i}^{(t-s,\theta)}}{\sum_{i=0}^{u+1} (-1)^i d_i} \mathcal{M}_{m,x}(l) \mathcal{D}_{\theta+r}(z) \mathcal{DM}_{\theta+l;n}(r);$$

see [32, Proposition 4].

To recover exact samples from $p_{m,n,l,r}^{(x,z,s,t,\theta)}$, consider the pairing bijective function $\Sigma \colon \mathbb{N} \to \mathbb{N} \times \mathbb{N} \times \mathbb{N}^d \times \mathbb{N}^d$ such that $\Sigma(j) = (m, n, l, r)$. Now, for each j there exists v_j, an element of $\tilde{v} \in \mathbb{R}^{J+1}$ (i.e., $\tilde{v} = \{v_j\}_{j=0}^J$), such that

$$R_{\tilde{v}}^-(J) := \sum_{j=0}^J e_{\Sigma(j)}(2v_j + 1) \leq \sum_{j=0}^J p_{\Sigma(j)}^{(x,z,s,t,\theta)} \leq \sum_{j=0}^J e_{\Sigma(j)}(2v_j) := R_{\tilde{v}}^+(J),$$

Algorithm 6 Exact simulation of samples from the discrete random variable with PMF $\{p_{m,n,l,r}^{(x,z,s,t,\theta)}:(m, n, l, r) \in \mathbb{N} \times \mathbb{N} \times \mathbb{N}^d \times \mathbb{N}^d\}$

---

1 Set $j \leftarrow 0, v_0 \leftarrow 0, \tilde{v} \leftarrow (v_0)$
2 Simulate $U \sim \text{Uniform}(0, 1)$
3 **repeat**
4   Set $v_j \leftarrow \lceil \tilde{F}_{\Sigma(j)}^{(s,t,\theta)}/2 \rceil$
5   **while** $R_{\tilde{v}}^-(j) < U < R_{\tilde{v}}^+(j)$ **do**
6   Set $\tilde{v} \leftarrow \tilde{v} + (1, \dots, 1)$
7   **end while**
8   **if** $R_{\tilde{v}}^-(j) > U$ **then**
9   **return** $\Sigma(j)$
10   **else if** $R_{\tilde{v}}^+(j) < U$ **then**
11   Set $\tilde{v} \leftarrow (v_0, \dots, v_j, 0)$
12   Set $j \leftarrow j + 1$
13   **end if**
14 **until** false

---

providing a setting analogous to the one presented in Section 3. The proposed exact sampling scheme can be found in [32, Algorithm 5], which we reproduce here as Algorithm 6 for completeness.

Other approaches to exact simulation of one-dimensional Wright–Fisher bridges include that recently proposed in [27], which restricts to the case where either θ_1 or θ_2 is 0, and one or both of x and z are 0, which is not applicable here.

7. Proofs

Proof of Proposition 2.1. Let $V:\mathbb{R}^n \to \mathbb{R}^n$ be such that $V := \nabla_x(\overline{V} \circ f)(x)$, where we recall that

$$f^{ik}(x) = \overline{x}^{ik} = \begin{cases} x^{ij}, & j, k = \{1, \dots, d_i - 1\}, \\ 1 - \sum_{j=1}^{d_i-1} x^{ij}, k = d_i. \end{cases}$$

Then for all $i \in \{1, \dots, L\}, j \in \{1, \dots, d_i - 1\}$, we have

$$V^{ij}(x) = \frac{\partial(\overline{V} \circ f)(x)}{\partial x^{ij}} = \frac{\partial}{\partial \overline{x}^{ij}} \overline{V}(\overline{x}) - \frac{\partial}{\partial \overline{x}^{id_i}} \overline{V}(\overline{x})$$

$$= s^{ij} + \sum_{l=1}^{L} \sum_{k=1}^{d_l} h_{jk}^{il} \overline{x}^{lk} - s^{id_i} - \sum_{l=1}^{L} \sum_{k=1}^{d_l} h_{d_ik}^{il} \overline{x}^{lk},$$

$$= s^{ij} - s^{id_i} + \sum_{l=1}^{L} \left(h_{jd_l}^{il} - h_{d_id_l}^{il} + \sum_{k=1}^{d_l-1} \left(h_{jk}^{il} - h_{jd_l}^{il} - h_{d_ik}^{il} + h_{d_id_l}^{il} \right) x^{lk} \right)$$

$$= K_s^{ij} + \sum_{l=1}^{L} \left(K_l^{ij} + \sum_{k=1}^{d_l-1} K_{lk}^{ij} x^{lk} \right) = K_s^{ij} + \sum_{\substack{l=1 \\ l \neq i}}^{L} \left(K_l^{ij} + \sum_{k=1}^{d_l-1} K_{lk}^{ij} x^{lk} \right),$$

where the last equality holds because whenever $i = l$, all entries of the blocks H^{ii} are 0. □

Proof of Theorem 4.1. By definition of $A(\cdot)$, $\phi(\cdot)$, and X_t,

$$\exp\{A(X_0, X_T)\} \exp\left\{-\int_0^T \phi(X_t)dt\right\} =$$

$$= \exp\left\{\int_0^T V(X_t) \cdot dX_t - \int_0^T \frac{1}{2}\left[(V(X_t))^T D(X_t)V(X_t) + 2(V(X_t))^T \alpha(X_t)\right]dt\right\}$$

$$= \exp\left\{\int_0^T D^{\frac{1}{2}}(X_t)V(X_t) \cdot dB_t - \int_0^T \frac{1}{2}\left[(V(X_t))^T D(X_t)V(X_t)\right]dt\right\}. \tag{7.1}$$

Because $V(\cdot)$ and $D(\cdot)$ are continuous on $[0, 1]^n$, there exist constants C^- and C^+ such that

$$C^- \le \frac{1}{2}V(X_t)^T D(X_t)V(X_t) \le C^+, \quad \text{almost surely.}$$

Consequently,

$$\int_0^T \frac{1}{2}(V(X_t))^T D(X_t)V(X_t)dt < \infty, \quad \text{almost surely,}$$

so Novikov's condition is fulfilled and (7.1) can be identified as a Girsanov transformation [34] with Girsanov kernel $(V(X_t))^T D^{\frac{1}{2}}(X_t)$.

Let $\mathbb{Q}$ be the probability measure with

$$\frac{d\mathbb{Q}}{d\mathbb{WFL}_{\alpha,x_0}} = \exp\{A(X_0, X_T)\} \exp\left\{-\int_0^T \phi(X_t)dt\right\}.$$

It follows that the law of X under $\mathbb{Q}$ coincides with $\mathbb{CWF}_{\alpha,G,x_0}$. Indeed, by Girsanov's theorem,

$$\tilde{B}_t = B_t - \int_0^t D^{\frac{1}{2}}(X_s)V(X_s)ds$$

is a $\mathbb{Q}$-Brownian motion and

$$dX_t = [\alpha(X_t) + D(X_t)V(X_t)]dt + D^{\frac{1}{2}}(X_t)d\tilde{B}_t = [\alpha(X_t) + G(X_t)]dt + D^{\frac{1}{2}}(X_t)d\tilde{B}_t.$$

Now, by Proposition 2.1, and for $K_s, K_l, K_{lk} \in \mathbb{R}^n$,

$$\int_0^T V(X_t) \cdot dX_t = \int_0^T \left(K_s + \sum_{l=1}^L \left(K_l + \sum_{k=1}^{d_l-1} K_{lk}X_t^{lk}\right)\right) \cdot dX_t$$

$$= \int_0^T K_s \cdot dX_t + \int_0^T \sum_{l=1}^L K_l \cdot dX_t + \int_0^T \sum_{l=1}^L \sum_{k=1}^{d_l-1} K_{lk}X_t^{lk} \cdot dX_t$$

$$= \sum_{i=1}^L \sum_{j=1}^{d_i-1} \left(\int_0^T K_s^{ij}dX_t^{ij} + \int_0^T \sum_{l=1}^L K_l^{ij}dX_t^{ij} + \int_0^T \sum_{l=i+1}^L \sum_{k=1}^{d_l-1} K_{lk}^{ij} d\left(X_t^{lk}X_t^{ij}\right)\right),$$

where the rightmost term comes from pairing terms of the form

$$K_{lk}^{ij}X_t^{lk}dX_t^{ij} + K_{ij}^{lk}X_t^{ij}dX_t^{lk} = K_{lk}^{ij}d\left(X_t^{lk}X_t^{ij}\right),$$

and recalling that $K_{lk}^{ij} = K_{ij}^{lk}$, and $i \neq l$ prevents squared terms. Hence,

$$\int_0^T V(X_t) \cdot dX_t = \sum_{i=1}^L \sum_{j=1}^{d_i-1} \left(K_s^{ij}\left(X_T^{ij} - X_0^{ij}\right) + \sum_{l=1}^L K_l^{ij}\left(X_T^{ij} - X_0^{ij}\right) \right.$$
$$\left. + \sum_{l=i+1}^L \sum_{k=1}^{d_l-1} K_{lk}^{ij}\left(X_T^{lk}X_T^{ij} - X_0^{lk}X_0^{ij}\right) \right).$$

The fact that

$$A(X_0, X_T) := \sum_{i=1}^L \sum_{j=1}^{d_i-1} \left(K_s^{ij}\left(X_T^{ij} - X_0^{ij}\right) + \sum_{l=1}^L K_l^{ij}\left(X_T^{ij} - X_0^{ij}\right) + \sum_{l=i+1}^L \sum_{k=1}^{d_l-1} K_{lk}^{ij}\left(X_T^{ij}X_T^{lk} - X_0^{ij}X_0^{lk}\right) \right)$$

is bounded follows immediately, concluding the proof. □

Proof of Proposition 4.1. Let $M(t)$ be the total number of coefficients that must be computed in Algorithm 3, with $t \in (0, T)$ the time distance between two sampled skeleton points. By [32, Proposition 5(iv)], there exists a $\kappa > 0$ such that $\mathbb{E}[M(t)] = o(t^{-(1+\kappa)})$ 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,

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

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

Now let $\epsilon := d\mathbb{CWF}_{\alpha,G,x_0}/d\mathbb{WFL}_{\alpha,x_0}$ be the acceptance–rejection probability in Algorithm 4. Because $A^- \leq A(X_T) \leq A^+$ and $C^- \leq \phi(X_t) \leq C^+$,

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

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

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

where the first equality follows from considering the expectation of the sum of all drawn Poisson points $\sum_{i=1}^I D_i$ over I iterations of the algorithm until the first path is accepted. First conditioning 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),

$$
\begin{aligned}
g_{z,t}(x, y; s) &= \frac{g(x, y; s) g(y, z; t-s)}{g(x, z; t)} \\
&= \frac{1}{g(x, z; t)} \sum_{m=0}^{\infty} q_m^{\theta}(s) \sum_{\substack{l \\ |l|=m}} \mathcal{M}_{m,x}(l) \mathcal{D}_{\theta+l}(y) \sum_{n=0}^{\infty} q_n^{\theta}(t-s) \sum_{\substack{r \\ |r|=n}} \mathcal{M}_{n,y}(r) \mathcal{D}_{\theta+r}(z) \\
&= \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{\substack{l \\ |l|=m}} \sum_{\substack{r \\ |r|=n}} \frac{q_m^{\theta}(s) q_n^{\theta}(t-s)}{g(x, z; t)} \mathcal{M}_{m,x}(l) \mathcal{D}_{\theta+r}(z) \mathcal{M}_{n,y}(r) \mathcal{D}_{\theta+l}(y).
\end{aligned}
$$

By definition,

$$
\begin{aligned}
\mathcal{M}_{n,y}(r) \mathcal{D}_{\theta+l}(y) &= \frac{n!}{\prod_{j=1}^{d} l_j!} \left(1 - \sum_{j=1}^{d-1} y_j\right)^{l_d} \prod_{j=1}^{d-1} y_j^{r_j} \\
&\quad \times \frac{\Gamma(|\theta + l|)}{\prod_{j=1}^{d} \Gamma(\theta_j + l_j)} \left(1 - \sum_{j=1}^{d-1} y_j\right)^{\theta_d + l_d - 1} \prod_{j=1}^{d-1} y_j^{\theta_j + l_j - 1}.
\end{aligned}
$$

Multiplying and dividing by $\dfrac{\Gamma(|\theta + l + r|)}{\prod_{j=1}^{d} \Gamma(\theta_j + l_j + r_j)}$ and rearranging shows that

$$
\begin{aligned}
\frac{n! \Gamma(|\theta + l|)}{\Gamma(|\theta + l + r|)} \prod_{j=1}^{d} \frac{\Gamma(\theta_j + l_j + r_j)}{\Gamma(\theta_j + l_j) r_j!} \frac{\Gamma(|\theta + l + r|)}{\prod_{j=1}^{d} \Gamma(\theta_j + l_j + r_j)} \left(1 - \sum_{j=1}^{d-1} y_j\right)^{\theta_d + l_d + r_d - 1} & \\
\times \prod_{j=1}^{d-1} y_j^{\theta_j + l_j + r_j - 1} = \mathcal{D}\mathcal{M}_{\theta+l;n}(r) \mathcal{D}_{\theta+l+r}(y). &
\end{aligned}
$$

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

Proof of Lemma 6.1. For the sake of simplicity, we will first consider the case $d = 3$ and then show that by analogous arguments the results can be extended to the general $(d-1)$-dimensional case. First, note that the indexes of the sum on the right-hand side of

$$
\mathbb{E}[\mathcal{D}_{\theta+L_m}(z)] = \sum_{\substack{l \\ |l|=m}} \mathbb{P}(L_m = l) \mathcal{D}_{\theta+l}(z)
$$

can be seen as being placed on the $(d-1)$-face of the d-simplex $\mathcal{S}_d = \{l \in \mathbb{R}^d \mid l_j \geq 0, \sum_{j=1}^{d} l_j = m\}$. For example, if $d = 3$, we only need to consider indexes l_1 and l_2 (see Figure 3). Thus,

$$
\mathbb{E}[\mathcal{D}_{\theta+L_m}(z)] = \sum_{l_1=0}^{m} \sum_{l_2=0}^{m-l_1} \mathbb{P}(L_m = l) \mathcal{D}_{\theta+l}(z). \tag{7.2}
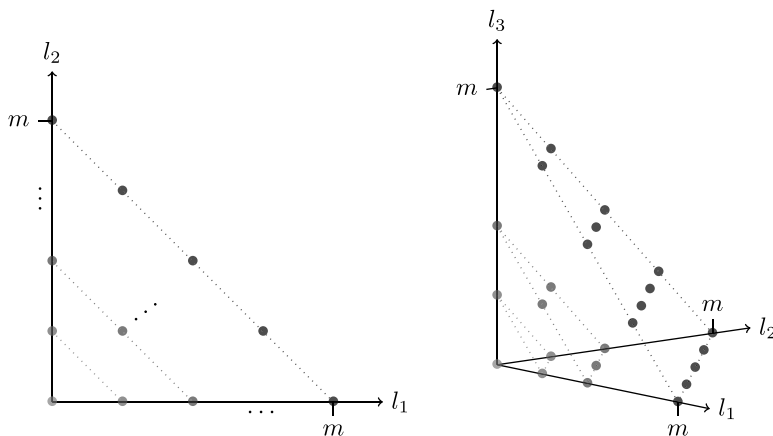$$

FIGURE 3. Graphical representation of the two-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 mth ($l_1 - l_2 - l_3$)-face (right), which in turn represent each and every one of the summands in (7.2).

Let us define the quantities

$$\mathcal{Q}_m := \mathbb{P}(L_m = l)\mathcal{D}_{\theta+l}(z), \qquad \mathcal{Q}_{m+1}^{(j)} := \mathbb{P}(L_{m+1} = \tilde{l})\mathcal{D}_{\theta+\tilde{l}}(z),$$

where $\tilde{l} \in \mathbb{R}^d$ is equal to l except in its jth component, for which $\tilde{l}_j = l_j + 1$. As a special case, when $j = d$, one can write $\tilde{l}_d = m - \sum_{j=1}^{d-1} l_j + 1$.

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

$$\mathcal{Q}_{m+1}^{(d)} = \frac{m+1}{m+1-l_1-l_2} \frac{|\theta|+m}{\theta_3+m-l_1-l_2}(1-x_1-x_2)(1-z_1-z_2)\mathcal{Q}_m \tag{7.3}$$

$$\leq \frac{m+1}{1+m(1-z_1-z_2)} \frac{|\theta|+m}{\theta_3+m(1-z_1-z_2)}(1-x_1-x_2)(1-z_1-z_2)\mathcal{Q}_m$$

$$\leq \left(\frac{|\theta|}{\theta_3} \vee \frac{m^2+m(1+|\theta|)+|\theta|}{m^2(1-z_1-z_2)^2}\right)(1-x_1-x_2)(1-z_1-z_2)\mathcal{Q}_m$$

$$\leq \left(\frac{|\theta|}{\theta_3}(1-z_1-z_2) \vee \frac{2(1+|\theta|)}{1-z_1-z_2}\right)(1-x_1-x_2)\mathcal{Q}_m.$$

Here, in the second inequality, note that the function

$$g(m) := \frac{m^2+m(1+|\theta|)+|\theta|}{m^2(1-z_1-z_2)^2}$$

is decreasing in m, and thus for $m \geq 1$, it attains its maximum value at $g(1)$. Then, if

$$f(m) := \frac{m+1}{1+m(1-z_1-z_2)} \frac{|\theta|+m}{\theta_3+m(1-z_1-z_2)},$$

one obtains $f(m) \leq f(0) \vee g(m) \leq f(0) \vee g(1)$ yielding the desired result.

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

$$\mathcal{Q}_{m+1}^{(2)} = \frac{m+1}{l_2+1}\frac{|\theta|+m}{\theta_2+l_2}x_2z_2\mathcal{Q}_m \leq \frac{m+1}{mz_2+1}\frac{|\theta|+m}{\theta_2+mz_2}x_2z_2\mathcal{Q}_m$$

$$\leq \left(\frac{|\theta|}{\theta_2} \vee \frac{m^2+m(1+|\theta|)+|\theta|}{m^2z_2^2}\right)x_2z_2\mathcal{Q}_m \leq \left(\frac{|\theta|}{\theta_2}z_2 \vee \frac{2(1+|\theta|)}{z_2}\right)x_2\mathcal{Q}_m, \quad (7.4)$$

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

$$\mathcal{Q}_{m+1}^{(1)} = \frac{m+1}{l_1+1}\frac{|\theta|+m}{\theta_1+l_1}x_1z_1\mathcal{Q}_m \leq \frac{m+1}{mz_1+1}\frac{|\theta|+m}{\theta_1+mz_1}x_1z_1\mathcal{Q}_m$$

$$\leq \left(\frac{|\theta|}{\theta_1} \vee \frac{m^2+m(1+|\theta|)+|\theta|}{m^2z_1^2}\right)x_1z_1\mathcal{Q}_m \leq \left(\frac{|\theta|}{\theta_1}z_1 \vee \frac{2(1+|\theta|)}{z_1}\right)x_1\mathcal{Q}_m. \quad (7.5)$$

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

$$\mathbb{E}[\mathcal{D}_{\theta+L_{m+1}}(z)] = \sum_{l_1=0}^{\lfloor mz_1 \rfloor}\sum_{l_2=0}^{\lfloor mz_2 \rfloor}\mathcal{Q}_{m+1}^{(d)} + \sum_{l_1=0}^{\lfloor mz_1 \rfloor}\sum_{l_2=\lfloor mz_2 \rfloor+1}^{m+1-l_1}\mathcal{Q}_{m+1}^{(2)} + \sum_{l_1=\lfloor mz_1 \rfloor+1}^{m+1}\sum_{l_2=0}^{m+1-l_1}\mathcal{Q}_{m+1}^{(1)}$$

$$\leq \left(\frac{|\theta|}{\theta_3}(1-z_1-z_2) \vee \frac{2(1+|\theta|)}{1-z_1-z_2}\right)(1-x_1-x_2)\sum_{l_1=0}^{\lfloor mz_1 \rfloor}\sum_{l_2=0}^{\lfloor mz_2 \rfloor}\mathcal{Q}_m$$

$$+ \left(\frac{|\theta|}{\theta_2}z_2 \vee \frac{2(1+|\theta|)}{z_2}\right)x_2\sum_{l_1=0}^{\lfloor mz_1 \rfloor}\sum_{l_2=\lfloor mz_2 \rfloor}^{m-l_1}\mathcal{Q}_m$$

$$+ \left(\frac{|\theta|}{\theta_1}z_1 \vee \frac{2(1+|\theta|)}{z_1}\right)x_1\sum_{l_1=\lfloor mz_1 \rfloor}^{m}\sum_{l_2=0}^{m-l_1}\mathcal{Q}_m$$

$$\leq \left[\left(\frac{|\theta|}{\theta_3}(1-z_1-z_2) \vee \frac{2(1+|\theta|)}{1-z_1-z_2}\right)(1-x_1-x_2)\right.$$

$$\left. + \sum_{j=1}^{2}\left(\frac{|\theta|}{\theta_j}z_j \vee \frac{2(1+|\theta|)}{z_j}\right)x_j\right]\mathbb{E}[\mathcal{D}_{\theta+L_m}(z)],$$

where, in the first inequality, terms starting at $l_j = \lfloor mz_j \rfloor + 1$ are compared one-to-one to terms starting at $l_j = \lfloor mz_j \rfloor$, and the last inequality holds after taking common factors and noting that the terms for $l_j = \lfloor mz_j \rfloor$ are bounded by both

$$\left(\frac{|\theta|}{\theta_3}(1-z_1-z_2) \vee \frac{2(1+|\theta|)}{1-z_1-z_2}\right)(1-x_1-x_2)$$

and

$$\sum_{j=1}^{2}\left(\frac{|\theta|}{\theta_j}z_j \vee \frac{2(1+|\theta|)}{z_j}\right)x_j.$$

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

$$\mathbb{E}[\mathcal{D}_{\theta+L_m}(z)] = \sum_{l_1=0}^{m} \sum_{l_2=0}^{m-l_1} \cdots \sum_{l_{d-1}=0}^{m-|l|_{d-2}} \mathbb{P}(L_m = l)\mathcal{D}_{\theta+l}(z),$$

where $|l|_{d-2} = \sum_{j=1}^{d-2} l_j$.

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

(a) $l_j \le \lfloor mz_j \rfloor$ for all j;

(b) $\lfloor mz_j \rfloor \le l_j$ for all $j \ne 1$, and $l_i \le \lfloor mz_i \rfloor$ for all $i \ne j$; or

(c) $\lfloor mz_1 \rfloor \le l_1$, and l_i is free for all $i \ne 1$.

Note that this partition covers all (non-exclusive) combinations and includes all the elements of the sum. Now, comparing $\mathbb{E}[\mathcal{D}_{\theta+L_{m+1}}(z)]$ with $\mathbb{E}[\mathcal{D}_{\theta+L_m}(z)]$, the bounding constants for each case are as follows:

(a)
$$\left(\frac{|\theta|}{\theta_d} \left(1 - \sum_{j=1}^{d-1} z_j \right) \vee \frac{2(1+|\theta|)}{1 - \sum_{j=1}^{d-1} z_j} \right) \left(1 - \sum_{j=1}^{d-1} x_j \right),$$

(b)
$$\left(\frac{|\theta|}{\theta_j} z_j \vee \frac{2(1+|\theta|)}{z_j} \right) x_j,$$

(c)
$$\left(\frac{|\theta|}{\theta_1} z_1 \vee \frac{2(1+|\theta|)}{z_1} \right) x_1.$$

This yields

$$\mathbb{E}[\mathcal{D}_{\theta+L_{m+1}}(z)] \le \tilde{K}^{(\theta,x,z)} \mathbb{E}[\mathcal{D}_{\theta+L_m}(z)],$$

with

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

Proof of Proposition 6.2. The proof follows from that of [32, Proposition 3], which is reproduced here for completeness.

The inequality $d_{2m+1} < d_{2m}$ follows because if $m \ge E^{(t,\theta)}$ then $2j \ge C_{m-j}^{(t,\theta)}$ for all $j = 0, \dots, m$, which, by Proposition 3.1, implies

$$b_{m+j+1}^{(t,\theta)}(m-j) < b_{m+j}^{(t,\theta)}(m-j)).$$

Multiplying by $\mathbb{E}[\mathcal{D}_{\theta+L_{m-j}}(z)]$ and then summing over $j = 0, \dots, m$ gives

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

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

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

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

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

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

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

Note now that

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

where

$$h_m(k) := \frac{|\theta| + m + k - 1}{k - m + 1}\frac{|\theta| + 2k + 1}{|\theta| + 2k - 1},$$

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 [32, Proposition 1].

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

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

Finally,

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

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

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

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

which concludes the proof.   □

Acknowledgements

References

[1] AURELL, E., EKEBERG, M. AND KOSKI, T. (2019). On a multilocus Wright–Fisher model with mutation and a Svirezhev–Shahshahani gradient-like selection dynamics. Preprint. Available at https://arxiv.org/abs/1906.00716.

[2] BEAUMONT, M. A., ZHANG, W. AND BALDING, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025–2035.

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

[4] BESKOS, A., PAPASPILIOPOULOS, O. AND ROBERTS, G. O. (2008). A factorisation of diffusion measure and finite sample path constructions. *Methodology Comput. Appl. Prob.* **10**, 85–104.

[5] BESKOS, A. AND ROBERTS, G. O. (2005). Exact simulation of diffusions. *Ann. Appl. Prob.* **15**, 2422–2444.

[6] BLANCHET, J. AND ZHANG, F. (2020). Exact simulation for multivariate Itô diffusions. *Advances in Applied Probability*, **52**(4), 1003–1034.

[7] BÜRGER, R. (2000). *The Mathematical Theory of Selection, Recombination, and Mutation*. John Wiley, Chichester.

[8] BUZBAS, E. O., JOYCE, P. AND ROSENBERG, N. A. (2011). Inference on the strength of balancing selection for epistatically interacting loci. *Theoret. Pop. Biol.* **79**, 102–113.

[9] CASELLA, B. AND ROBERTS, G. O. (2008). Exact Monte Carlo simulation of killed diffusions. *Adv. Appl. Prob.* **40**, 273–291.

[10] CHEN, N. AND HUANG, Z. (2013). Localization and exact simulation of Brownian motion-driven stochastic differential equations. *Math. Operat. Res.* **38**, 591–616.

[11] CORANDER, J. *et al.* (2017). Frequency-dependent selection in vaccine-associated pneumococcal population dynamics. *Nature Ecol. Evol.* **1**, 1950–1960.

[12] CROW, J. F. AND KIMURA, M. (1970). *An Introduction to Population Genetics Theory.* Harper & Row, New York, Evanston, London.

[13] DEVROYE, L. (2006). Nonuniform random variate generation. *Handbooks Operat. Res. Manag. Sci.* **13**, 83–121.

[14] EKEBERG, M. *et al.* (2013). Improved contact prediction in proteins: using pseudolikelihoods to infer Potts models. *Phys. Rev. E* **87**, 012707.

[15] ETHERIDGE, A. M. AND GRIFFITHS, R. C. (2009). A coalescent dual process in a Moran model with genic selection. *Theoret. Pop. Biol.* **75**, 320–330.

[16] ETHIER, S. N. AND KURTZ, T. G. (2009). *Markov Processes: Characterization and Convergence*. John Wiley, New York.

[17] ETHIER, S. N. AND NAGYLAKI, T. (1989). Diffusion approximations of the two-locus Wright–Fisher model. *J. Math. Biol.* **27**, 17–28.

[18] EWENS, W. J. (2004). *Mathematical Population Genetics*, 2nd edn. Springer, New York.

[19] FAVERO, M., HULT, H. AND KOSKI, T. (2019). A dual process for the coupled Wright–Fisher diffusion. To appear in *J. Math. Biol.*

[20] FEARNHEAD, P. (2006). The stationary distribution of allele frequencies when selection acts at unlinked loci. *Theoret. Pop. Biol.* **70**, 376–386.

[21] FEARNHEAD, P. AND PRANGLE, D. (2012). Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. *J. R. Statist. Soc. B [Statist. Methodology]* **74**, 419–474.

[22] FITZSIMMONS, P., PITMAN, J. AND YOR, M. (1993). Markovian bridges: construction, Palm interpretation, and splicing. In *Seminar on Stochastic Processes, 1992*, Birkhäuser, Boston, pp. 101–134.

[23] GAO, C.-Y., ZHOU, H.-J. AND AURELL, E. (2018). Correlation-compressed direct-coupling analysis. *Phys. Rev. E* **98**, 032407.

[24] GRIFFITHS, R. C. (1980). Lines of descent in the diffusion approximation of neutral Wright–Fisher models. *Theoret. Pop. Biol.* **17**, 37–50.

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

[26] GRIFFITHS, R. C. (2006). Coalescent lineage distributions. *Adv. Appl. Prob.* **38**, 405–429.

[27] GRIFFITHS, R. C., JENKINS, P. A. AND SPANÒ, D. (2018). Wright–Fisher diffusion bridges. *Theoret. Pop. Biol.* **122**, 67–77.

[28] GRIFFITHS, R. C. AND LI, W.-H. (1983). Simulating allele frequencies in a population and the genetic differentiation of populations under mutation pressure. *Theoret. Pop. Biol.* **23**, 19–33.

[29] GRIFFITHS, R. C. AND SPANÒ, D. (2010). Diffusion processes and coalescent trees. In *Probability and Mathematical Genetics, Papers in Honour of Sir John Kingman*, eds N. H. BINGHAM AND C. M. GOLDIE, Cambridge University Press, pp. 358–375.

[30] HÖSSJER, O., TYVAND, P. A. AND MILOH, T. (2016). Exact Markov chain and approximate diffusion solution for haploid genetic drift with one-way mutation. *Math. Biosci.* **272**, 100–112.

[31] JENKINS, P. A. (2013). Exact simulation of the sample paths of a diffusion with a finite entrance boundary. Preprint. Available at https://arxiv.org/abs/1311.5777.

[32] JENKINS, P. A. AND SPANÒ, D. (2017). Exact simulation of the Wright–Fisher diffusion. *Ann. Appl. Prob.* **27**, 1478–1509.

[33] JEWETT, E. M. AND ROSENBERG, N. A. (2014). Theory and applications of a deterministic approximation to the coalescent model. *Theoret. Pop. Biol.* **93**, 14–29.

[34] KARATZAS, I. AND SHREVE, S. E. (1998). *Brownian Motion and Stochastic Calculus*, 2nd edn. Springer, New York.

[35] KIMURA, M. (1955). Stochastic processes and distribution of gene frequencies under natural selection. *Cold Spring Harb. Symp. Quant. Biol.* **20**, 33–53.

[36] KIMURA, M. (1964). Diffusion models in population genetics. *J. Appl. Prob.* **1**, 177–232.

[37] KLOEDEN, P. E. AND PLATEN, E. (1992). *Numerical Solution of Stochastic Differential Equations*. Springer, New York.

[38] NEHER, R. A. AND SHRAIMAN, B. I. (2011). Statistical genetics and evolution of quantitative traits. *Rev. Mod. Phys.* **83**, 1283–1300.

[39] NENÉ, N. R., MUSTONEN, V. AND ILLINGWORTH, C. J. R. (2018). Evaluating genetic drift in time-series evolutionary analysis. *J. Theoret. Biol.* **437**, 51–57.

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

[41] ROHLFS, R. V., SWANSON, W. J. AND WEIR, B. S. (2010). Detecting coevolution through allelic association between physically unlinked loci. *Amer. J. Human Genet.* **86**, 674–685.

[42] SCHUBERT, B., MADDAMSETTI, R., NYMAN, J., FARHAT, M. R. AND MARKS, D. S. (2019). Genome-wide discovery of epistatic loci affecting antibiotic resistance in Neisseria gonorrhoeae using evolutionary couplings. *Nature Microbiol.* 4, 328–338.

[43] SKWARK, M. J. *et al.* (2017). Interacting networks of resistance, virulence and core machinery genes identified by genome-wide epistasis analysis. *PLoS Genet.* **13**, e1006508.

[44] STEINRÜCKEN, M., BHASKAR, A. AND SONG, Y. S. (2014). A novel spectral method for inferring general diploid selection from time series genetic data. *Ann. Appl. Statist.* **8**, 2203–2222.

[45] TATARU, P., SIMONSEN, M., BATAILLON, T. AND HOBOLTH, A. (2017). Statistical inference in the Wright–Fisher model using allele frequency data. *Systematic Biol.* **66**, e30–e46.

[46] TAVARÉ, S., BALDING, D. J., GRIFFITHS, R. C. AND DONNELLY, P. (1997). Inferring coalescence times from DNA sequence data. *Genetics* **145**, 505–518.