Numerical simulation of the two-locus Wright-Fisher stochastic differential equation with application to approximating transition probability densities

Zhangyi He^{a,1,*}, Mark Beaumont^b, Feng Yu^{a,*}

^a*School of Mathematics, University of Bristol, Bristol BS8 1UG, United Kingdom*

^b*School of Biological Sciences, University of Bristol, Bristol BS8 1TQ, United Kingdom*

Abstract

Over the past decade there has been an increasing focus on the application of the Wright-Fisher diffusion to the inference of natural selection from genetic time series. A key ingredient for modelling the trajectory of gene frequencies through the Wright-Fisher diffusion is its transition probability density function. Recent advances in DNA sequencing techniques have made it possible to monitor genomes in great detail over time, which presents opportunities for investigating natural selection while accounting for genetic recombination and local linkage. However, most existing methods for computing the transition probability density function of the Wright-Fisher diffusion are only applicable to one-locus problems. To address two-locus problems, in this work we propose a novel numerical scheme for the Wright-Fisher stochastic differential equation of population dynamics under natural selection at two linked loci. Our key innovation is that we reformulate the stochastic differential equation in a closed form that is amenable to simulation, which enables us to avoid boundary issues and reduce computational costs. We also propose an adaptive importance sampling approach based on the proposal introduced by Fearnhead (2008) for computing the transition probability density of the Wright-Fisher diffusion between any two observed states. We show through extensive simulation studies that our approach can achieve comparable performance to the method of Fearnhead (2008) but can avoid manually tuning the parameter ρ to deliver superior performance for different observed states.

Keywords: Natural selection, Two linked loci, Wright-Fisher diffusion, Stochastic Taylor scheme, Importance sampling

*Corresponding author.

Email addresses: zh0131@bristol.ac.uk (Zhangyi He), feng.yu@bristol.ac.uk (Feng Yu)

¹Present address: Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London W2 1PG, United Kingdom & Email: z.he@imperial.ac.uk

1. Introduction

Throughout the history of population genetics, stochastic models have played a significant role in the study of the population evolving under the influences of various demographic and evolutionary forces within species such as genetic drift and natural selection. One of the most popular stochastic models is the Wright-Fisher model, introduced by Fisher (1922) and Wright (1931), which forms the basis for most theoretical and applied research in population genetics to date, including Kingman’s coalescent (Kingman, 1982), Ewens’ sampling formula (Ewens, 1972), Kimura’s work on fixation probabilities (Kimura, 1955) and statistical techniques for inferring demographic and genetic properties of biological populations (*e.g.*, see Tataru et al., 2017, and references therein for the inference of natural selection with the Wright-Fisher model). However, it is intractable to investigate the gene distribution dynamics in evolving populations through the Wright-Fisher model for large population sizes and evolutionary timescales since a prohibitively large amount of storage and computation is required. The analysis of the Wright-Fisher model today is greatly facilitated by its diffusion approximation, commonly known as the Wright-Fisher diffusion, which can be traced back to Kimura (1964). Crow & Kimura (1970), Ewens (2004) and Durrett (2008) provided an excellent theoretical introduction to the Wright-Fisher model and its diffusion approximation.

In recent years, the Wright-Fisher diffusion has already been successfully applied in the population genetic analysis of time series data of allele frequencies (*e.g.*, Bollback et al., 2008; Gutenkunst et al., 2009; Lukić & Hey, 2012; Malaspinas et al., 2012; Gautier & Vitalis, 2013; Steinrücken et al., 2014; Vitalis et al., 2014; Živković et al., 2015; Ferrer-Admetlla et al., 2016; Schraiber et al., 2016; He et al., 2019), which is mainly due to the fact that it captures the essential features of the underlying evolutionary model and provides a concise framework for characterising the gene distribution dynamics, even in complex evolutionary scenarios. The most important ingredient for the population genetic analysis of genetic time series data through the Wright-Fisher diffusion is its transition probability density function, which can completely determine the gene distribution dynamics in evolving populations. Most existing methods for computing the transition probability density function are based on solving the Kolmogorov equations associated with the Wright-Fisher diffusion, such as finite difference methods (*e.g.*, Boitard & Loisel, 2007; Bollback et al., 2008; Gutenkunst et al., 2009; He et al., 2019), finite

volume methods (*e.g.*, Zhao et al., 2013), and spectral methods (*e.g.*, Lukić et al., 2011; Song & Steinrücken, 2012; Steinrücken et al., 2013, 2014, 2015; Živković et al., 2015), most of which, however, are limited to one-locus problems.

With recent advances in DNA preparation and sequencing techniques, the increased availability of genetic time series sampled from multiple loci such as the time serial samples of ancient Eurasians (Mathieson et al., 2015; Allentoft et al., 2015) should provide new opportunities for investigating the gene distribution dynamics underlying such samples, possibly under natural selection, by accounting for the genetic variation at multiple loci jointly. However, according to Kopp et al. (2012), it is infeasible to extend most existing approaches to compute the transition probability density function by solving the Kolmogorov equations associated with the Wright-Fisher diffusion for multi-locus problems since such an approach suffers from dramatically increased computational costs and becomes increasingly inaccurate or even numerically unstable with an increase in the dimension of the problem. Boitard & Loisel (2007) employed a finite difference method to obtain a numerical solution for the probability distribution of haplotype frequencies under the Wright-Fisher model by a diffusion approximation for the two-locus problem of the neutral population. To guarantee a numerically stable computation of the solution, such a method requires an appropriate discretisation for the state space, which may depend strongly on the population genetic quantities such as the recombination rate. However, in population genetic analysis, where we have no prior knowledge of demographic and genetic properties of evolving populations, it is intractable to predict whether the results produced with a particular discretisation are accurate or not. Furthermore, such an approach can only be applied to the states in the interior of the state space, and the approximation of the boundary conditions introduce significant errors.

Fortunately, standard diffusion theory establishes an intimate link between partial differential equation (PDE) and stochastic differential equation (SDE) formulations of the Wright-Fisher diffusion (see, *e.g.*, Karlin & Taylor, 1981), which provides an opportunity for investigating the gene distribution dynamics in evolving populations under the Wright-Fisher diffusion by simulating the SDE instead of solving the PDE (*i.e.*, Kolmogorov backward equation or its adjoint). A number of specialised temporal discretisation methods have been developed for simulating the Wright-Fisher diffusion (*e.g.*, Schurz, 1996; Dangerfield et al., 2012; Schraiber et al., 2013;

Neuenkirch & Szpruch, 2014). Recently, Jenkins & Spanó (2017) proposed an approach for exact simulation of the Wright-Fisher diffusion. However, all of these methods are still limited to one-locus problems.

In this work, we introduce a Wright-Fisher SDE of population dynamics subject to natural selection at two linked loci, where the genetic recombination effect and local linkage information are explicitly modelled, and propose an efficient numerical scheme for simulating the Wright-Fisher diffusion. The key innovation of our approach is that we reformulate the Wright-Fisher SDE into a closed form that is amenable to simulation, which enables us to avoid boundary issues and reduce computational costs. Stochastic Taylor schemes are used to obtain an approximate, but often very accurate, solution of the Wright-Fisher SDE, and the associated discretisation error can be reduced by letting the discretisation become more refined. Inspired by Fearnhead (2008), we also develop an adaptive importance sampling method for computing the transition probability density of the Wright-Fisher diffusion between any two observed states with a set of simulated sample paths of the Wright-Fisher diffusion. Unlike the method of Fearnhead (2008), our approach can automatically tune the parameter ρ for different observed states and achieve comparable performance. Our method can play an important role in many existing simulation-based inference approaches such as simulated maximum-likelihood methods and Markov chain Monte Carlo methods (see Fuchs, 2013, and reference therein).

The remainder of this work is organised as follows. In Section 2, we provide a brief review of the Wright-Fisher model and its diffusion approximation for the population evolving subject to natural selection at two linked loci. In Section 3, we construct a stochastic Taylor scheme to numerically solve the Wright-Fisher SDE. In Section 4, we develop a numerical approximation of the transition probability density of the Wright-Fisher diffusion through importance sampling and compare the performance of several different importance samplers. Finally, in Section 5, we summarise our main results obtained in the previous sections and discuss possible applications and extensions.

2. Wright-Fisher diffusion

In this section, we begin with a short review of the Wright-Fisher model for a population of diploid individuals evolving subject to natural selection at a pair of linked loci. We then develop the diffusion approximation of the Wright-Fisher model in both SDE and PDE formulations.

2.1. Wright-Fisher model

Let us consider a diploid population of N randomly mating individuals at two linked autosomal loci $\mathcal{A}$ and $\mathcal{B}$ evolving under natural selection with discrete and nonoverlapping generations. We assume that the $\mathcal{A}$ and $\mathcal{B}$ loci are located on a same chromosome with recombination rate $r \in [0, 0.5]$ between them (*i.e.*, the rate that a recombinant gamete is produced at meiosis), and at each locus there are two allele types, labelled $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{B}_1$, $\mathcal{B}_2$, separately, which gives rise to four possible haplotypes $\mathcal{A}_1\mathcal{B}_1$, $\mathcal{A}_1\mathcal{B}_2$, $\mathcal{A}_2\mathcal{B}_1$ and $\mathcal{A}_2\mathcal{B}_2$, called haplotypes 1, 2, 3 and 4, respectively. We attach the symbols $\mathcal{A}_1$ and $\mathcal{B}_1$ to the mutant alleles, which are assumed to arise only once in the population, and we attach the symbols $\mathcal{A}_2$ and $\mathcal{B}_2$ to the ancestral alleles, which are assumed to originally exist in the population.

We incorporate viability selection into the population dynamics and let w_{ij} represent the viability of an individual with the genotype formed by haplotype i and haplotype j . In addition to positing absence of sex effects, *i.e.*, $w_{ij} = w_{ji}$, we assume that the viability of the genotype at the two loci is determined multiplicatively from the viabilities at individual loci and is fixed from the time that the mutant allele arises. More specifically, the viabilities of the three possible genotypes at each locus, *e.g.*, genotypes $\mathcal{A}_1\mathcal{A}_1$, $\mathcal{A}_1\mathcal{A}_2$ and $\mathcal{A}_2\mathcal{A}_2$ at a given locus $\mathcal{A}$, are taken to be 1, $1 - h_{\mathcal{A}}s_{\mathcal{A}}$ and $1 - s_{\mathcal{A}}$, respectively, where $s_{\mathcal{A}} \in [0, 1]$ is the selection coefficient and $h_{\mathcal{A}} \in [0, 1]$ is the dominance parameter. For example, the viability of an individual with the $\mathcal{A}_1\mathcal{B}_2/\mathcal{A}_2\mathcal{B}_2$ genotype is $w_{24} = (1 - h_{\mathcal{A}}s_{\mathcal{A}})(1 - s_{\mathcal{B}})$.

Let $X_i^{(N)}(k)$ denote the frequency of gamete i in a population of N individuals in generation $k \in \mathbb{N}$ for $i = 1, 2, 3, 4$, and $\mathbf{X}^{(N)}(k)$ denote the vector of frequencies of the four possible gametes. In the Wright-Fisher model, the population size N is fixed over time, and gametes are chosen randomly from an effectively infinite gamete pool reflecting the parental gamete frequencies at each generation. Under multinomial sampling, we have

$$\mathbf{X}^{(N)}(k) \mid \mathbf{X}^{(N)}(k-1) = \mathbf{x} \sim \frac{1}{2N} \text{Multinomial}(2N, \mathbf{p}), \quad (1)$$

where $\mathbf{p}$ is the vector of the parental gamete frequencies in generation $k-1$, *i.e.*, the haplotype frequencies of the effectively infinite population after a single generation of random mating, natural selection and genetic recombination. We can express the sampling probabilities $\mathbf{p}$ as a

function of the gamete frequencies $\mathbf{x}$,

$$\begin{aligned} p_1 &= (1-r)x'_1 + r(x'_1 + x'_2)(x'_1 + x'_3) \\ p_2 &= (1-r)x'_2 + r(x'_1 + x'_2)(x'_2 + x'_4) \\ p_3 &= (1-r)x'_3 + r(x'_3 + x'_4)(x'_1 + x'_3) \\ p_4 &= (1-r)x'_4 + r(x'_3 + x'_4)(x'_2 + x'_4), \end{aligned} \quad (2)$$

where

$$x'_i = \frac{\bar{w}_i}{\bar{w}} x_i = \frac{\sum_{j=1}^4 w_{ij} x_j}{\sum_{i=1}^4 \sum_{j=1}^4 w_{ij} x_j x_i} x_i \quad (3)$$

for $i = 1, 2, 3, 4$. In Eq. (3), the numerator $\bar{w}_i$ is the marginal fitness of gamete i and the denominator $\bar{w}$ is the mean fitness.

We define the two-locus Wright-Fisher model with selection to be the stochastic process $\mathbf{X}^{(N)} = \{\mathbf{X}^{(N)}(k), k \in \mathbb{N}\}$ evolving in the state space

$$\Omega_{\mathbf{X}^{(N)}} = \left\{ \mathbf{x} \in \left\{ 0, \frac{1}{2N}, \dots, 1 \right\}^4 : \sum_{i=1}^4 x_i = 1 \right\}$$

with a multinomial one-step transition probability matrix. From Eqs. (1)-(3), we observe that the transition probabilities of the gamete frequencies from generation k to the next are only dependent on the gamete frequencies in generation k but are independent of generation k , which implies that the Wright-Fisher model is a time-homogeneous Markov process.

2.2. Diffusion approximation

Owing to the interplay of stochastic and deterministic forces, the two-locus Wright-Fisher model with selection presents analytical challenges far beyond the Wright-Fisher model for neutral evolution at a single locus. Moreover, it requires a prohibitively large amount of storage and computation to explore the gene distribution dynamics in evolving populations under the Wright-Fisher model when population sizes and evolutionary timescales become large. However, fortunately, we can find an accurate diffusion approximation of the two-locus Wright-Fisher model with selection by running time at rate $2N$, *i.e.*, $t = k/(2N)$, and holding the scaled selection coefficients $\alpha_A = 2Ns_A$, $\alpha_B = 2Ns_B$ and the scaled recombination rate $\rho = 4Nr$ fixed while taking the population size $N \rightarrow \infty$.

Theorem 1. *If we let the population size N go to infinity and hold the scaled parameters $\alpha_{\mathcal{A}}$, $\alpha_{\mathcal{B}}$ and R constant, the two-locus Wright-Fisher model with selection $\mathbf{X}^{(N)}$ converges to a diffusion process, denoted by $\mathbf{X} = \{\mathbf{X}(t), t \geq 0\}$, satisfying the stochastic differential equation of the form*

$$d\mathbf{X}(t) = \boldsymbol{\mu}(\mathbf{X}(t))dt + \boldsymbol{\nu}(\mathbf{X}(t))d\mathbf{W}(t), \quad (4)$$

where the drift term $\boldsymbol{\mu}(\mathbf{x})$ is a four-dimensional vector being

$$\begin{aligned} \mu_1(\mathbf{x}) &= \alpha_{\mathcal{A}}x_1(x_3 + x_4)[(x_1 + x_2)h_{\mathcal{A}} + (x_3 + x_4)(1 - h_{\mathcal{A}})] \\ &\quad + \alpha_{\mathcal{B}}x_1(x_2 + x_4)[(x_1 + x_3)h_{\mathcal{B}} + (x_2 + x_4)(1 - h_{\mathcal{B}})] - \frac{\rho}{2}(x_1x_4 - x_2x_3) \\ \mu_2(\mathbf{x}) &= \alpha_{\mathcal{A}}x_2(x_3 + x_4)[(x_1 + x_2)h_{\mathcal{A}} + (x_3 + x_4)(1 - h_{\mathcal{A}})] \\ &\quad - \alpha_{\mathcal{B}}x_2(x_1 + x_3)[(x_1 + x_3)h_{\mathcal{B}} + (x_2 + x_4)(1 - h_{\mathcal{B}})] + \frac{\rho}{2}(x_1x_4 - x_2x_3) \\ \mu_3(\mathbf{x}) &= -\alpha_{\mathcal{A}}x_3(x_1 + x_2)[(x_1 + x_2)h_{\mathcal{A}} + (x_3 + x_4)(1 - h_{\mathcal{A}})] \\ &\quad + \alpha_{\mathcal{B}}x_3(x_2 + x_4)[(x_1 + x_3)h_{\mathcal{B}} + (x_2 + x_4)(1 - h_{\mathcal{B}})] + \frac{\rho}{2}(x_1x_4 - x_2x_3) \\ \mu_4(\mathbf{x}) &= -\alpha_{\mathcal{A}}x_4(x_1 + x_2)[(x_1 + x_2)h_{\mathcal{A}} + (x_3 + x_4)(1 - h_{\mathcal{A}})] \\ &\quad - \alpha_{\mathcal{B}}x_4(x_1 + x_3)[(x_1 + x_3)h_{\mathcal{B}} + (x_2 + x_4)(1 - h_{\mathcal{B}})] - \frac{\rho}{2}(x_1x_4 - x_2x_3), \end{aligned} \quad (5)$$

the diffusion term $\boldsymbol{\nu}(\mathbf{x})$ is a four by three matrix satisfying

$$\boldsymbol{\nu}(\mathbf{x})\boldsymbol{\nu}^{\top}(\mathbf{x}) = \Sigma(\mathbf{x}), \quad (6)$$

where

$$\Sigma_{ij}(\mathbf{x}) = x_i(\delta_{ij} - x_j) \quad (7)$$

for $i, j = 1, 2, 3, 4$, and $\mathbf{W}(t)$ is a three-dimensional standard Brownian motion.

The proof of Theorem 1 follows in the similar manner to that employed for neutral populations by Durrett (2008, p. 323), and the major change being the substitution of the gamete frequencies of an effectively infinite population after a single generation of random mating, natural selection and genetic recombination for the sampling probabilities (see He et al., 2017). The term $x_1x_4 - x_2x_3$ in Eq. (3) is a measure of the linkage disequilibrium between loci $\mathcal{A}$ and $\mathcal{B}$, which quantifies the non-random association of the alleles at the two loci. We refer to the

diffusion process $\mathbf{X}$ satisfying the SDE in Eq. (4) as the two-locus Wright-Fisher diffusion with selection, which evolves in the state space

$$\Omega_{\mathbf{X}} = \left\{ \mathbf{x} \in [0, 1]^4 : \sum_{i=1}^4 x_i = 1 \right\}. \quad (8)$$

Moreover, for any $\mathbf{x}$ in the interior of the state space $\Omega_{\mathbf{X}}$, the infinitesimal mean vector $\boldsymbol{\mu}(\mathbf{x})$ and the infinitesimal covariance matrix $\boldsymbol{\Sigma}(\mathbf{x})$ are twice continuously differentiable with respect to $\mathbf{x}$ with bounded derivatives, and the infinitesimal covariance matrix $\boldsymbol{\Sigma}(\mathbf{x})$ is also positive definite. Therefore, for any $\mathbf{x}$ and $\mathbf{x}'$ in the interior of the state space $\Omega_{\mathbf{X}}$, the transition probability distribution function $P(t, \mathbf{x}, \mathbf{x}')$ of the Wright-Fisher diffusion $\mathbf{X}$ is strongly continuous with respect to Lebesgue's measure (Stroock & Varadhan, 1979), and the transition probability density function $p(t, \mathbf{x}, \mathbf{x}')$ satisfies the Kolmogorov backward equation (KBE)

$$\frac{\partial}{\partial t} p(t, \mathbf{x}, \mathbf{x}') = \frac{1}{2} \sum_{i=1}^4 \sum_{j=1}^4 \Sigma_{ij}(\mathbf{x}) \frac{\partial^2 p(t, \mathbf{x}, \mathbf{x}')}{\partial x_i \partial x_j} + \sum_{i=1}^4 \mu_i(\mathbf{x}) \frac{\partial p(t, \mathbf{x}, \mathbf{x}')}{\partial x_i} \quad (9)$$

for fixed $\mathbf{x}'$ and the Kolmogorov forward equation (KFE)

$$\frac{\partial}{\partial t} p(t, \mathbf{x}, \mathbf{x}') = \frac{1}{2} \sum_{i=1}^4 \sum_{j=1}^4 \frac{\partial^2 (\Sigma_{ij}(\mathbf{x}') p(t, \mathbf{x}, \mathbf{x}'))}{\partial x'_i \partial x'_j} - \sum_{i=1}^4 \frac{\partial (\mu_i(\mathbf{x}') p(t, \mathbf{x}, \mathbf{x}'))}{\partial x'_i} \quad (10)$$

for fixed $\mathbf{x}$, where the infinitesimal mean vector $\boldsymbol{\mu}(\mathbf{x})$ and the infinitesimal covariance matrix $\boldsymbol{\Sigma}(\mathbf{x})$ are given by Eqs. (5) and (7), respectively.

To solve the KBE in Eq. (9) or the KFE in Eq. (10), we need to specify certain boundary conditions. At the boundary of the state space $\Omega_{\mathbf{X}}$, the Wright-Fisher diffusion $\mathbf{X}$ can have different behaviours depending on the type of the boundary: corners, edges and faces. Indeed, its boundary behaviour corresponds to a regular boundary except in the case that fixation occurs at least at one locus (*i.e.*, corners and edges). Since fixation at one locus is non-reversible, this kind of boundary behaviour is characteristic of an exit boundary. When fixation occurs at one locus (*i.e.*, edges), the two-locus Wright-Fisher diffusion with selection reduces to the one-locus Wright-Fisher diffusion with selection, for the population evolving at the other locus. Notice that the classification of boundaries here is extended from Feller's boundary classification (see Karlin & Taylor, 1981, for more details). Even though the boundary behaviour of the Wright-

Fisher diffusion $\mathbf{X}$ is known, it is still intractable to formulate the boundary conditions with exact equations, which could cause a serious problem in solving the KBE and the KFE, either analytically or numerically (see Boitard & Loisel, 2007).

3. Stochastic Taylor scheme

Although the Kolmogorov equations satisfied by the transition probability density function of the Wright-Fisher diffusion $\mathbf{X}$ can be readily obtained in Section 2.2, unfortunately, to the best of our knowledge, there are still no known analytical solutions to the Kolmogorov equations, not even numerical solutions. In this section, we resort to a numerical simulation scheme to solve the Wright-Fisher SDE in Eq. (4) rather than the resulting Kolmogorov equations in Eq. (9) or (10).

3.1. Equivalent SDE models

To find a numerical solution of the Wright-Fisher SDE in Eq. (4), we need to compute the diffusion term $\sigma(\mathbf{x})$, which we have to perform at every time step in most existing numerical simulation schemes. The diffusion term $\sigma(\mathbf{x})$ can be derived using the Cholesky decomposition. This has been given analytically by Eq. (5.3) in Sato (1976). However, many coefficients in this decomposition explodes at the boundaries. For example,

$$\sigma_{22}(x) \propto x_2^{1/2}(1 - x_1 - x_2)^{1/2}(1 - x_1)^{-1/2}, \quad (11)$$

which will be numerically unstable when x_1 is close to 1. More precisely, even though in Eq. (11) $(1 - x_1 - x_2)^{1/2}$ will be tiny, which cancels out the magnitude of $(1 - x_1)^{-1/2}$, any round-off error will be magnified when computing $(1 - x_1)^{-1/2}$. There exist other matrix decompositions capable of computing the diffusion term $\sigma(\mathbf{x})$ such as spectral decomposition, which are valid for positive semi-definite matrices, typically at the expense of either additional numerical errors and computational costs, or limitations in applicability to the infinitesimal covariance matrix $\Sigma(\mathbf{x})$ of the form in Eq. (7).

Inspired by Allen et al. (2008), we construct a structurally different SDE from the population evolving subject to natural selection at two linked loci, which is equivalent to the Wright-Fisher SDE in Eq. (4) but enables us to avoid the problem with the decomposition of the infinitesimal

200 covariance matrix $\Sigma(\mathbf{x})$. Such an SDE can be formulated in the form:

$$d\mathbf{X}(t) = \boldsymbol{\mu}(\mathbf{X}(t))dt + \boldsymbol{\sigma}(\mathbf{X}(t))d\mathbf{W}(t), \quad (12)$$

201 where the diffusion term $\boldsymbol{\sigma}(\mathbf{x})$ can be explicitly written down as

$$\boldsymbol{\sigma}(\mathbf{x}) = \begin{pmatrix} \sqrt{x_1 x_2} & \sqrt{x_1 x_3} & \sqrt{x_1 x_4} & 0 & 0 & 0 \\ -\sqrt{x_2 x_1} & 0 & 0 & \sqrt{x_2 x_3} & \sqrt{x_2 x_4} & 0 \\ 0 & -\sqrt{x_3 x_1} & 0 & -\sqrt{x_3 x_2} & 0 & \sqrt{x_3 x_4} \\ 0 & 0 & -\sqrt{x_4 x_1} & 0 & -\sqrt{x_4 x_2} & -\sqrt{x_4 x_3} \end{pmatrix}, \quad (13)$$

202 and $\mathbf{W}(t)$ is a six-dimensional standard Brownian motion. We observe that unlike the Cholesky
203 decomposition found in Eq. (5.3) of Sato (1976), there is no division or negative power in the
204 diffusion term $\boldsymbol{\sigma}(\mathbf{x})$ in Eq. (13), which can lead to numerical instability.

205 From Eqs. (5) and (13), we find that the drift term $\boldsymbol{\mu}(\mathbf{x})$ and the diffusion term $\boldsymbol{\sigma}(\mathbf{x})$ do not
206 satisfy the local Lipschitz conditions (see Karlin & Taylor, 1981), but do satisfy the following
207 conditions arising in the study of the Brownian motion on the group of diffeomorphisms of the
208 circle (*e.g.*, Malliavin, 1999; Airault & Ren, 2002):

$$|\boldsymbol{\sigma}(\mathbf{x}) - \boldsymbol{\sigma}(\mathbf{x}')|^2 \leq |\mathbf{x} - \mathbf{x}'|^2 \cdot g(|\mathbf{x} - \mathbf{x}'|) \quad (14)$$

$$|\boldsymbol{\mu}(\mathbf{x}) - \boldsymbol{\mu}(\mathbf{x}')| \leq |\mathbf{x} - \mathbf{x}'| \cdot g(|\mathbf{x} - \mathbf{x}'|) \quad (15)$$

209 for all $\mathbf{x}, \mathbf{x}' \in \Omega_{\mathbf{X}}$, where $|\cdot|$ denotes the Euclidean vector norm or the Frobenius matrix norm
210 as appropriate and

$$g(x) = C \left(\log \left(\frac{1}{x} \right) \vee K \right)^{\frac{1}{\beta}} \quad (16)$$

211 for some $C > 0$, $K > 0$ and $\beta > 1$. Here we can take $C = 1$, $K = 1$ and $\beta = 2$, respectively. Ac-
212 cording to Yamada & Watanabe (1971) and Taniguchi (1992), the non-Lipschitz conditions (14)
213 and (15) are sufficient to ensure the existence and uniqueness of the solution for the SDE in
214 Eq. (12). The equivalence of the Wright-Fisher SDE in Eq. (4) and (12) is given by the following
215 theorem, which guarantees that we can obtain the numerical solution of the SDE in Eq. (4) by
216 solving the SDE in Eq. (12) with numerical simulation schemes.

Theorem 2. *Solutions to the Wright-Fisher SDE in Eq. (4) and (12) possess the same probability law.*

The proof of Theorem 2 can be found in Appendix A.

3.2. Euler-Maruyama scheme

We now turn to the development of a numerical simulation scheme for the TLWFDS-SDE with the decomposition of the infinitesimal covariance matrix $\Sigma(\mathbf{x})$ we proposed in Section 3.1. Assuming that the closed-form expression of the diffusion coefficient matrix $\sigma(\mathbf{x})$ is available, we consider the construction of stochastic Taylor schemes here. Without loss of generality, we take the initial time point to be 0 and focus our attention on the TLWFDS* model for time t on a interval $[0, T]$ in the following due to the fact that the Wright-Fisher diffusion $\mathbf{X}$ is time-homogeneous.

To obtain the stochastic Itô-Taylor expansion for the solution of the SDE in Eqs. (11)-(13), we first represent the Wright-Fisher SDE in the integral component form as

$$X_i(t) = X_i(0) + \int_0^t \mu_i(\mathbf{X}(s)) ds + \sum_{j=1}^6 \int_0^t \sigma_{ij}(\mathbf{X}(s)) dW_j(s), \quad (17)$$

for $i = 1, 2, 3, 4$. Using Itô's formula, we then have

$$\mu_i(\mathbf{X}(t)) = \mu_i(\mathbf{X}(0)) + \int_0^t \mathcal{L}_0 \mu_i(\mathbf{X}(s)) ds + \sum_{k=1}^6 \int_0^t \mathcal{L}_k \mu_i(\mathbf{X}(s)) dW_k(s) \quad (18)$$

$$\sigma_{ij}(\mathbf{X}(t)) = \sigma_{ij}(\mathbf{X}(0)) + \int_0^t \mathcal{L}_0 \sigma_{ij}(\mathbf{X}(s)) ds + \sum_{k=1}^6 \int_0^t \mathcal{L}_k \sigma_{ij}(\mathbf{X}(s)) dW_k(s), \quad (19)$$

where $\mathcal{L}_0$ and $\mathcal{L}_k$ are differential operators defined as

$$\begin{aligned} \mathcal{L}_0 &= \sum_{i=1}^4 \mu_i(\mathbf{x}) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i=1}^4 \sum_{j=1}^4 \Sigma_{ij}(\mathbf{x}) \frac{\partial^2}{\partial x_i \partial x_j} \\ \mathcal{L}_k &= \sum_{i=1}^4 \sigma_{ik}(\mathbf{x}) \frac{\partial}{\partial x_i}, \end{aligned}$$

for $k = 1, 2, \dots, 6$. Substituting Eqs. (18) and (19) into Eq. (17) and separating the constant terms out from the integrand, we can formulate the stochastic Itô-Taylor expansion for the

234 solution of the Wright-Fisher SDE as

$$X_i(t) = X_i(0) + \mu_i(\mathbf{X}(0)) \int_0^t ds_1 + \sum_{j=1}^6 \sigma_{ij}(\mathbf{X}(0)) \int_0^t dW_j(s_1) + \epsilon_i, \quad (20)$$

235 for $i = 1, 2, 3, 4$, where

$$\begin{aligned} \epsilon_i = & \int_0^t \int_0^{s_1} \mathcal{L}_0 \mu_i(\mathbf{X}(s_2)) ds_2 ds_1 \\ & + \sum_{k=1}^6 \int_0^t \int_0^{s_1} \mathcal{L}_k \mu_i(\mathbf{X}(s_2)) dW_k(s_2) ds_1 \\ & + \sum_{j=1}^6 \int_0^t \int_0^{s_1} \mathcal{L}_0 \sigma_{ij}(\mathbf{X}(s_2)) ds_2 dW_j(s_1) \\ & + \sum_{j=1}^6 \sum_{k=1}^6 \int_0^t \int_0^{s_1} \mathcal{L}_k \sigma_{ij}(\mathbf{X}(s_2)) dW_k(s_2) dW_j(s_1). \end{aligned}$$

236 We can get a truncated stochastic Itô-Taylor expansion by keeping the first three constant
237 terms in Eq. (20) or continue applying Itô's formula to other terms in ϵ_i to include further
238 higher-order constant terms, which enables us to get truncated stochastic Itô-Taylor expansions
239 of higher order but requires more derivatives and integrals. That is, in theory, we can construct
240 the stochastic Taylor scheme of arbitrarily high order by including further components of the
241 stochastic Itô-Taylor expansion for the solution of the Wright-Fisher SDE in Eqs. (11)-(13).

242 We introduce a partition of the time interval $[0, T]$, defined as

$$\Delta_{[0,T]}^{(U)} = \left\{ \tau_u : \tau_u = \frac{T}{U}u \text{ for } u = 0, 1, \dots, U \right\}.$$

243 Keeping the first three constant terms in Eq. (20), we can obtain the stochastic Taylor scheme
244 of the form

$$\hat{X}_i(\tau_u) = \hat{X}_i(\tau_{u-1}) + \mu_i(\hat{\mathbf{X}}(\tau_{u-1}))\Delta\tau_{u-1} + \sum_{j=1}^6 \sigma_{ij}(\hat{\mathbf{X}}(\tau_{u-1}))\Delta W_j(\tau_{u-1}),$$

245 for $i = 1, 2, 3, 4$, where $\Delta\tau_{u-1} = \tau_u - \tau_{u-1}$ and $\Delta W_j(\tau_{u-1}) = W_j(\tau_u) - W_j(\tau_{u-1})$ are independent
246 and normally distributed with mean 0 and variance $\Delta\tau_{u-1}$ for $j = 1, 2, \dots, 6$.

247 The stochastic Taylor scheme we adopt here is rather simple, commonly known as the Euler-
248 Maruyama scheme, which is one of the most popular stochastic Taylor schemes in practice due

to its high efficiency and low complexity. As demonstrated in Kloeden & Platen (1992), the Euler-Maruyama scheme is numerically stable and strongly consistent, and the convergence of the Euler-Maruyama approximation $\hat{\mathbf{X}}$ of the Wright-Fisher diffusion is guaranteed by the non-Lipschitz conditions (14) and (15) (see Zhang, 2006, for more details). If one desires a stochastic Taylor scheme of higher order, an efficient systematic procedure proposed by Tocino & Vigo-Aguiar (2003) can be applied to obtain truncated stochastic Itô-Taylor expansions of higher order for the Wright-Fisher diffusion $\mathbf{X}$ expressed in terms of powers of the increments Δt and $\Delta \mathbf{X}(t)$.

Figure 1: Empirical cumulative distribution functions for each gamete frequency in generation 200 under the Wright-Fisher diffusion $\mathbf{X}$ for different interval partitions. We partition the time interval into $U = 10, 50, 100$ equidistant subintervals for the Euler-Maruyama scheme, respectively, and we construct the empirical cumulative distribution functions with $V = 100000$ independent realizations of the Wright-Fisher diffusion. In addition, we provide the empirical cumulative distribution functions for each haplotype frequency under the Wright-Fisher model $\mathbf{X}^{(N)}$ for reference. We adopt $\mathbf{x}_0 = (0.3, 0.2, 0.2, 0.3)$, $N = 1000$, $s_A = 0.01$, $h_A = 0.5$, $s_B = 0.02$, $h_B = 0.5$ and $r = 0.05$.

We illustrate the empirical cumulative distribution functions of the Wright-Fisher diffusion $\mathbf{X}$ for each gamete frequency at the given time point in Figures 1 and 2, which are obtained with a sample of the realizations of the Wright-Fisher diffusion simulated with the Euler-Maruyama scheme. By comparing with the empirical cumulative distribution functions of the Wright-Fisher model $\mathbf{X}^{(N)}$ for each gamete frequency at the same time point, we find that the Euler-Maruyama

Figure 2: Empirical cumulative distribution functions for each gamete frequency in generation 200 under the Wright-Fisher diffusion $\mathbf{X}$ for different sample sizes. We partition the time interval into $U = 100$ equidistant subintervals for the Euler-Maruyama scheme, and we construct the empirical cumulative distribution functions with $V = 1000, 10000, 100000$ independent realizations of the Wright-Fisher diffusion, respectively. In addition, we provide the empirical cumulative distribution functions for each haplotype frequency under the Wright-Fisher model $\mathbf{X}^{(N)}$ for reference. We adopt $\mathbf{x}_0 = (0.3, 0.2, 0.2, 0.3)$, $N = 1000$, $s_A = 0.01$, $h_A = 0.5$, $s_B = 0.02$, $h_B = 0.5$ and $r = 0.05$.

approximation $\hat{\mathbf{X}}$ of the Wright-Fisher diffusion is sufficiently accurate on the time interval $[0, T]$ for a sufficiently fine partition $\Delta_{[0, T]}^{(U)}$ with enough realizations in the sample.

4. Transition probability densities

As we have stated in Section 1, the most important ingredient in parametric inference procedures for the time series data of allele frequencies under the Wright-Fisher diffusion is the transition probability density function, which can completely characterise the gene distribution dynamics in evolving populations. Most existing approaches for computing the transition probability density function are limited to one-locus problems, and it becomes intractable to extend them to multi-locus problems due to dramatically increased computational costs associated with an increase in the dimensionality of the problem. However, not all existing parametric inference procedures require the transition probability density function (*i.e.*, transition probability densities for all possible states in the state space). For instance, simulated maximum-likelihood

methods (*e.g.*, Pedersen, 1995a,b; Brandt & Santa-Clara, 2002; Durham & Gallant, 2002) just need to evaluate the transition probability density of the Wright-Fisher diffusion between two observed states. In this section, we resort to importance sampling techniques to construct an efficient numerical approximation of the transition probability density of the Wright-Fisher diffusion between two observed states, which may introduce some statistical errors but can be reduced by higher-order numerical simulation schemes and larger numbers of simulated sample paths. For simplicity, we demonstrate the construction of the importance sampling approximation of the transition probability density with the Euler-Maruyama scheme in the following.

4.1. Euler-Maruyama approximation

Given that the Wright-Fisher diffusion $\mathbf{X}$ is time-homogeneous, we let $p(T, \mathbf{x}_0, \mathbf{x}_T)$ denote the probability density of the transition of the Wright-Fisher diffusion $\mathbf{X}$ from state $\mathbf{x}_0$ to state $\mathbf{x}_T$, where $\mathbf{x}_0$ and $\mathbf{x}_T$ are two observed states at the time points 0 and T , respectively. If the length of the time interval T is sufficiently small, the Euler-Maruyama approximation of the Wright-Fisher diffusion $\hat{\mathbf{X}}$ has the desired accuracy on the time interval $[0, T]$. Therefore, we can accurately approximate the transition probability density $p(T, \mathbf{x}_0, \mathbf{x}_T)$ through the Euler-Maruyama approximation of the Wright-Fisher diffusion $\hat{\mathbf{X}}$ as

$$\hat{p}_{EM}^{(1)}(T, \mathbf{x}_0, \mathbf{x}_T) = \phi(\mathbf{x}_T; \mathbf{x}_0 + \boldsymbol{\mu}(\mathbf{x}_0)T, \boldsymbol{\Sigma}(\mathbf{x}_0)T),$$

where $\phi(\cdot; \mathbf{x}_0 + \boldsymbol{\mu}(\mathbf{x}_0)T, \boldsymbol{\Sigma}(\mathbf{x}_0)T)$ is the probability density function of the multivariate normal distribution with the mean vector $\mathbf{x}_0 + \boldsymbol{\mu}(\mathbf{x}_0)T$ and the covariance matrix $\boldsymbol{\Sigma}(\mathbf{x}_0)T$. Otherwise, we can use the partition $\Delta_{[0,T]}^{(U)}$ such that the Euler-Maruyama approximation of the Wright-Fisher diffusion $\hat{\mathbf{X}}$ is sufficiently accurate on each subinterval $[\tau_{u-1}, \tau_u]$. Combining the Markov property of the Wright-Fisher diffusion and the Chapman-Kolmogorov equation, we can accurately approximate the transition probability density $p(T, \mathbf{x}_0, \mathbf{x}_T)$ through the Euler-Maruyama approximation of the Wright-Fisher diffusion $\hat{\mathbf{X}}$ as

$$\hat{p}_{EM}^{(U)}(T, \mathbf{x}_0, \mathbf{x}_T) = \int_{\Omega_{\mathbf{X}}} \hat{p}_{EM}^{(1)}(\Delta\tau_{U-1}, \boldsymbol{\xi}_{U-1}, \boldsymbol{\xi}_U) \hat{p}_{EM}^{(U-1)}(\tau_{U-1} - \tau_0, \boldsymbol{\xi}_0, \boldsymbol{\xi}_{U-1}) d\boldsymbol{\xi}_{U-1} \quad (21)$$

$$= \int_{\Omega_{\mathbf{X}}} \prod_{u=1}^U \phi(\boldsymbol{\xi}_u; \boldsymbol{\xi}_{u-1} + \boldsymbol{\mu}(\boldsymbol{\xi}_{u-1})\Delta\tau_{u-1}, \boldsymbol{\Sigma}(\boldsymbol{\xi}_{u-1})\Delta\tau_{u-1}) d\boldsymbol{\xi}_1 \cdots d\boldsymbol{\xi}_{U-1}, \quad (22)$$

where we define $\hat{\mathbf{X}}(\tau_u) = \boldsymbol{\xi}_u$ for $u = 0, 1, \dots, U$, and use the convention $\boldsymbol{\xi}_0 = \mathbf{x}_0$ and $\boldsymbol{\xi}_U = \mathbf{x}_T$ to conserve notation.

From Theorem 2 in Pedersen (1995b), we have

$$\lim_{U \rightarrow \infty} \hat{p}_{EM}^{(U)}(T, \mathbf{x}_0, \mathbf{x}_T) = p(T, \mathbf{x}_0, \mathbf{x}_T). \quad (23)$$

However, it is infeasible to evaluate the integrals in Eq. (22) analytically, and unfortunately, quadrature-based numerical integration techniques become computationally infeasible with a very fine partition of the time interval $[0, T]$.

4.2. Monte Carlo integration

Inspired by Brandt & Santa-Clara (2002), Monte Carlo integration is an alternative approach of evaluating the integrals in the computation of the Euler-Maruyama approximation $\hat{p}_{EM}^{(U)}(T, \mathbf{x}_0, \mathbf{x}_T)$. From Eq. (21), using Monte Carlo integration, we can approximate the Euler-Maruyama approximation $\hat{p}_{EM}^{(U)}(T, \mathbf{x}_0, \mathbf{x}_T)$ by

$$\hat{p}_{EM,MC}^{(U,V)}(T, \mathbf{x}_0, \mathbf{x}_T) = \frac{1}{V} \sum_{v=1}^V \phi(\boldsymbol{\xi}_U; \boldsymbol{\xi}_{U-1}^{(v)} + \boldsymbol{\mu}(\boldsymbol{\xi}_{U-1}^{(v)})\Delta\tau_{U-1}, \boldsymbol{\Sigma}(\boldsymbol{\xi}_{U-1}^{(v)})\Delta\tau_{U-1}), \quad (24)$$

where $\{\boldsymbol{\xi}_{U-1}^{(v)}\}_{v=1}^V$ is the sample of the realisations generated by setting $\boldsymbol{\xi}_0 = \mathbf{x}_0$ and $\boldsymbol{\xi}_U = \mathbf{x}_T$, and successively drawing

$$\boldsymbol{\xi}_u \sim \mathcal{N}(\boldsymbol{\xi}_{u-1} + \boldsymbol{\mu}(\boldsymbol{\xi}_{u-1})\Delta\tau_{u-1}, \boldsymbol{\Sigma}(\boldsymbol{\xi}_{u-1})\Delta\tau_{u-1}) \quad (25)$$

for $u = 1, 2, \dots, U-1$, and V is the total number of the realisations in the sample. In Figure 3, we illustrate a set of the simulated sample paths for the Wright-Fisher diffusion $\mathbf{X}$ with two observed states $\mathbf{x}_0$ and $\mathbf{x}_T$ generated by using Eq. (25) as an example. However, the simulated sample paths in Figure 3 do not look likely to start at state $\mathbf{x}_0$ and end at state $\mathbf{x}_T$.

The strong law of large numbers guarantees the convergence of the Monte Carlo approximation $\hat{p}_{EM,MC}^{(U,V)}(T, \mathbf{x}_0, \mathbf{x}_T)$ to the Euler-Maruyama approximation $\hat{p}_{EM}^{(U)}(T, \mathbf{x}_0, \mathbf{x}_T)$ as the sample size V goes to infinity. Moreover, combining with Eq. (23), we have

$$\lim_{V \rightarrow \infty} \lim_{U \rightarrow \infty} \hat{p}_{EM,MC}^{(U,V)}(T, \mathbf{x}_0, \mathbf{x}_T) = p(T, \mathbf{x}_0, \mathbf{x}_T).$$

Figure 3: Simulated sample paths of each gamete evolving from generation 0 to 10 drawn from the the Wright-Fisher diffusion. In this illustration, we partition the time interval $[0, 0.001]$ into 100 equidistant subintervals for the Euler-Maruyama scheme and generate 10 simulated sample paths with two observed states $\mathbf{x}_0 = (0.3, 0.2, 0.2, 0.3)$ and $\mathbf{x}_{0.001} = (0.4, 0.2, 0.2, 0.2)$. Other parameters $N = 5000$, $s_A = 0.01$, $h_A = 0.5$, $s_B = 0.05$, $h_B = 0.5$ and $r = 0.05$.

317 We show the convergence of the Monte Carlo approximation $\hat{p}_{EM,MC}^{(U,V)}(T, \mathbf{x}_0, \mathbf{x}_T)$ to the transition
318 probability density $p(T, \mathbf{x}_0, \mathbf{x}_T)$ on a logarithmic scale in Figure 4 and find that the Monte Carlo
319 approximation $\hat{p}_{EM,MC}^{(U,V)}(T, \mathbf{x}_0, \mathbf{x}_T)$ given by Eq. (24) suffers from poor convergence largely due to
320 the enormous variance resulting from Monte Carlo integration. As shown in Figure 3, most of the
321 realisations $\xi_{U-1}^{(v)}$ drawn from the Wright-Fisher diffusion $\mathbf{X}$ with the Euler-Maruyama scheme
322 are in regions where the integrand $\hat{p}_{EM}^{(1)}(\Delta\tau_{U-1}, \xi_{U-1}, \xi_U)$ is small, which leads to prohibitive
323 Monte Carlo variability (Durham & Gallant, 2002).

Figure 4: Convergence of the Monte Carlo approximation of the transition probability density $\hat{p}_{EM,MC}^{(U,V)}(T, \mathbf{x}_0, \mathbf{x}_T)$. In this illustration, we partition the time interval $[0, 0.001]$ into 100 equidistant subintervals for the Euler-Maruyama scheme and take $\mathbf{x}_0 = (0.3, 0.2, 0.2, 0.3)$, $\mathbf{x}_{0.001} = (0.4, 0.2, 0.2, 0.2)$, $N = 5000$, $s_A = 0.01$, $h_A = 0.5$, $s_B = 0.05$, $h_B = 0.5$ and $r = 0.05$.

4.3. Monte Carlo integration with importance sampling

Now we pursue a more efficient Monte Carlo approximation of the transition probability density $\hat{p}_{EM,MC}^{(U,V)}(T, \mathbf{x}_0, \mathbf{x}_T)$ through importance sampling, whose efficiency depends largely on the choice of the importance sampler. Fuchs (2013) provided a number of importance samplers for Monte Carlo variance reduction. In this section, we consider the importance samplers based on the guided process according to the terminology suggested in Papaspiliopoulos & Roberts (2012), which can force the simulated sample paths generated from the importance sampler to terminate at state $\mathbf{x}_T$ almost surely through the so-called guiding drift term defined by

$$\boldsymbol{\mu}_G(t, \mathbf{x}) = \frac{\mathbf{x}_T - \mathbf{x}}{T - t}.$$

Note that we use G to refer to a generic guided process in the following.

The first G -based importance sampler we consider was developed by Delyon & Hu (2006), which suggests generating the simulated sample paths from the guided process, denoted by G_I , satisfying the following proposal SDE of the form

$$d\mathbf{X}(t) = \boldsymbol{\mu}_{G_I}(t, \mathbf{X}(t))dt + \boldsymbol{\sigma}(\mathbf{X}(t))d\mathbf{W}(t),$$

where the drift term

$$\boldsymbol{\mu}_{G_I}(t, \mathbf{x}) = \boldsymbol{\mu}(\mathbf{x}) + \boldsymbol{\mu}_G(t, \mathbf{x}), \quad (26)$$

the diffusion term is given by Eq. (13), and $\mathbf{W}(t)$ is a six-dimensional standard Brownian motion. From Eq. (26), the drift term $\boldsymbol{\mu}_{G_I}(t, \mathbf{x})$ of the G_I is composed of the drift term $\boldsymbol{\mu}(\mathbf{x})$ of the underlying Wright-Fisher diffusion and the guiding drift term $\boldsymbol{\mu}_G(t, \mathbf{x})$, which enables the resulting evolution of the G_I -SDE to be influenced by both the underlying Wright-Fisher diffusion and the observed state $\mathbf{x}_T$.

More specifically, using the Euler-Maruyama scheme, we set $\boldsymbol{\xi}_0 = \mathbf{x}_0$ and $\boldsymbol{\xi}_U = \mathbf{x}_T$, and successively draw

$$\boldsymbol{\xi}_u \sim \mathcal{N}(\boldsymbol{\xi}_{u-1} + \boldsymbol{\mu}_{G_I}(\tau_{u-1}, \boldsymbol{\xi}_{u-1})\Delta\tau_{u-1}, \boldsymbol{\Sigma}(\boldsymbol{\xi}_{u-1})\Delta\tau_{u-1})$$

for $u = 1, 2, \dots, U - 1$. We show a set of simulated sample paths generated from the G_I -based

importance sampler with two observed states $\mathbf{x}_0$ and $\mathbf{x}_T$ as an example in Figure 5.

Figure 5: Simulated sample paths of each gamete evolving from generation 0 to 10 drawn from the G_I -based importance sampler. In this illustration, we partition the time interval $[0, 0.001]$ into 100 equidistant subintervals for the Euler-Maruyama scheme and generate 10 simulated sample paths with two observed states $\mathbf{x}_0 = (0.3, 0.2, 0.2, 0.3)$ and $\mathbf{x}_{0.001} = (0.4, 0.2, 0.2, 0.2)$. Other parameters $N = 5000$, $s_A = 0.01$, $h_A = 0.5$, $s_B = 0.05$, $h_B = 0.5$ and $r = 0.05$.

The second G -based importance sampler we consider was proposed by Fearnhead (2008), which suggests generating the simulated sample paths from the guided process, denoted by G_2 , satisfying the following proposal SDE of the form

$$d\mathbf{X}(t) = \boldsymbol{\mu}_{G_2}(t, \mathbf{X}(t))dt + \boldsymbol{\sigma}(\mathbf{X}(t))d\mathbf{W}(t),$$

where the drift term

$$\boldsymbol{\mu}_{G_2}(t, \mathbf{x}) = (1 - e^{-\rho(T-t)})\boldsymbol{\mu}(\mathbf{x}) + e^{-\rho(T-t)}\boldsymbol{\mu}_G(t, \mathbf{x})$$

for a predetermined non-negative constant ρ , the diffusion term is given by Eq. (13), and $\mathbf{W}(t)$ is a six-dimensional standard Brownian motion. Compared to the drift term $\boldsymbol{\mu}_{G_1}(t, \mathbf{x})$ of the G_I , the construction of the drift term $\boldsymbol{\mu}_{G_2}(t, \mathbf{x})$ of the G_2 increases the contribution of the drift term $\boldsymbol{\mu}(\mathbf{x})$ of the underlying Wright-Fisher diffusion for time t close to 0, whereas for time t close to T , it increases the contribution of the guiding drift term $\boldsymbol{\mu}_G(t, \mathbf{x})$.

More specifically, using the Euler-Maruyama scheme, we set $\boldsymbol{\xi}_0 = \mathbf{x}_0$ and $\boldsymbol{\xi}_U = \mathbf{x}_T$, and successively draw

$$\boldsymbol{\xi}_u \sim \mathcal{N}(\boldsymbol{\xi}_{u-1} + \boldsymbol{\mu}_{G_2}(\tau_{u-1}, \boldsymbol{\xi}_{u-1})\Delta\tau_{u-1}, \boldsymbol{\Sigma}(\boldsymbol{\xi}_{u-1})\Delta\tau_{u-1}) \quad (27)$$

for $u = 1, 2, \dots, U - 1$. In particular, when we set $\rho = 0$ in Eq. (27), the G_2 -based importance

Figure 6: Simulated sample paths of each gamete evolving from generation 0 to 10 drawn from the G_2 -based importance sampler. In this illustration, we partition the time interval $[0, 0.001]$ into 100 equidistant subintervals for the Euler-Maruyama scheme and generate 10 simulated sample paths with two observed states $\mathbf{x}_0 = (0.3, 0.2, 0.2, 0.3)$ and $\mathbf{x}_{0.001} = (0.4, 0.2, 0.2, 0.2)$. Other parameters $N = 5000$, $s_A = 0.01$, $h_A = 0.5$, $s_B = 0.05$, $h_B = 0.5$ and $r = 0.05$.

sampler is commonly known as the diffusion bridge proposal. In Figure 7, we illustrate several sets of simulated sample paths generated from the G_2 -based importance sampler with two observed states $\mathbf{x}_0$ and $\mathbf{x}_T$ for different values of the constant ρ as examples.

Fearnhead (2008) also provided a modified G_2 -based importance sampler, which suggests successively drawing

$$\xi_u \sim \mathcal{N}\left(\xi_{u-1} + \mu_{G_2}(\tau_{u-1}, \xi_{u-1})\Delta\tau_{u-1}, \left(1 - e^{-\rho\Delta\tau_{u-1}} \frac{\Delta\tau_{u-1}}{\tau_U - \tau_{u-1}}\right) \Sigma(\xi_{u-1})\Delta\tau_{u-1}\right) \quad (28)$$

for $u = 1, 2, \dots, U - 1$. When we take $\rho = 0$ in Eq. (28), the modified G_2 -based importance

sampler reduces to the modified diffusion bridge proposal discussed in Durham & Gallant (2002), in which they empirically demonstrate that compared with the diffusion bridge proposal, the modification leads to much better performance in their applications. We illustrate several sets of simulated sample paths generated from the modified G_2 -based importance sampler with two observed states $\mathbf{x}_0$ and $\mathbf{x}_T$ for different values of the constant ρ as examples in Figure 7.

Figure 7: Simulated sample paths of each gamete evolving from generation 0 to 10 drawn from the modified G_2 -based importance sampler. In this illustration, we partition the time interval $[0, 0.001]$ into 100 equidistant subintervals for the Euler-Maruyama scheme and generate 10 simulated sample paths with two observed states $\mathbf{x}_0 = (0.3, 0.2, 0.2, 0.3)$ and $\mathbf{x}_{0.001} = (0.4, 0.2, 0.2, 0.2)$. Other parameters $N = 5000$, $s_A = 0.01$, $h_A = 0.5$, $s_B = 0.05$, $h_B = 0.5$ and $r = 0.05$.

We propose an adaptive G_2 -based importance sampler, which suggests generating the simulated sample paths from the guided process, denoted by G_3 , satisfying the following proposal SDE of the form

$$d\mathbf{X}(t) = \boldsymbol{\mu}_{G_3}(t, \mathbf{X}(t))dt + \boldsymbol{\sigma}(\mathbf{X}(t))d\mathbf{W}(t),$$

where the drift term

$$\mu_{G_3}(t, \mathbf{x}) = (1 - e^{-\rho(\mathbf{x}_T)(T-t)})\mu(\mathbf{x}) + e^{-\rho(\mathbf{x}_T)(T-t)}\mu_G(t, \mathbf{x})$$

for a predetermined non-negative function $\rho(\mathbf{x}_T)$, the diffusion term is given by Eq. (13), and $\mathbf{W}(t)$ is a six-dimensional standard Brownian motion. Compared to the drift term $\mu_{G_2}(t, \mathbf{x})$ of the G_2 , the ρ -term is a function of the observed state $\mathbf{x}_T$ in the drift term $\mu_{G_3}(t, \mathbf{x})$ of the G_3 , which is defined as

$$\rho(\mathbf{x}_T) = \frac{1}{\sqrt{(\mathbf{x}_T - \bar{\mathbf{x}}_T)^\top \Sigma^{-1}(\bar{\mathbf{x}}_T)(\mathbf{x}_T - \bar{\mathbf{x}}_T)}},$$

where $\bar{\mathbf{x}}_T$ is the value at time T of the solution to the following ordinary differential equation (ODE) of the form

$$\frac{d\bar{\mathbf{x}}(t)}{dt} = \mu(\bar{\mathbf{x}}(t)) \quad (29)$$

with the initial condition $\bar{\mathbf{x}}(0) = \mathbf{x}_0$. The solution to the ODE (29) characterises an effectively infinite population evolving under natural selection at two linked loci, and therefore $\bar{\mathbf{x}}_T$ can be considered to be the effectively infinite population limit of $\mathbf{X}(T)$, where the influence of genetic drift on the changes in gamete frequencies is completely removed. The advantage of constructing such a ρ -term is to enable the resulting G_3 -SDE to evolve largely under the influence of the underlying Wright-Fisher diffusion if its sample path is already close to the observed state $\mathbf{x}_T$ with high probability, but allows a more significant contribution from the guiding drift term if its sample path gets further away from the observed state $\mathbf{x}_T$.

Figure 8: Simulated sample paths of each gamete evolving from generation 0 to 10 drawn from the G_3 -based importance sampler. In this illustration, we partition the time interval $[0, 0.001]$ into 100 equidistant subintervals for the Euler-Maruyama scheme and generate 10 simulated sample paths with two observed states $\mathbf{x}_0 = (0.3, 0.2, 0.2, 0.3)$ and $\mathbf{x}_{0.001} = (0.4, 0.2, 0.2, 0.2)$. Other parameters $N = 5000$, $s_A = 0.01$, $h_A = 0.5$, $s_B = 0.05$, $h_B = 0.5$ and $r = 0.05$.

More specifically, using the Euler-Maruyama scheme, we set $\xi_0 = \mathbf{x}_0$ and $\xi_U = \mathbf{x}_T$, and successively draw

$$\xi_u \sim \mathcal{N}(\xi_{u-1} + \mu_{G_3}(\tau_{u-1}, \xi_{u-1})\Delta\tau_{u-1}, \Sigma(\xi_{u-1})\Delta\tau_{u-1})$$

for $u = 1, 2, \dots, U - 1$. Figure 8 illustrates a set of simulated sample paths generated from the G_3 -based importance sampler with two observed states $\mathbf{x}_0$ and $\mathbf{x}_T$ as an example.

Furthermore, the modification made in the G_2 -based importance sampler can be naturally applied to the G_3 -based importance sampler, which suggests setting $\xi_0 = \mathbf{x}_0$ and $\xi_U = \mathbf{x}_T$, and successively drawing

$$\xi_u \sim \mathcal{N}\left(\xi_{u-1} + \mu_{G_3}(\tau_{u-1}, \xi_{u-1})\Delta\tau_{u-1}, \left(1 - e^{-\rho(\mathbf{x}_T)\Delta\tau_{u-1}} \frac{\Delta\tau_{u-1}}{\tau_U - \tau_{u-1}}\right) \Sigma(\xi_{u-1})\Delta\tau_{u-1}\right)$$

for $u = 1, 2, \dots, U - 1$. In Figure 9, we illustrate a set of simulated sample paths generated from the modified G_3 -based importance sampler with two observed states $\mathbf{x}_0$ and $\mathbf{x}_T$ as an example.

Figure 9: Simulated sample paths of each gamete evolving from generation 0 to 10 drawn from the modified G_3 -based importance sampler. In this illustration, we partition the time interval $[0, 0.001]$ into 100 equidistant subintervals for the Euler-Maruyama scheme and generate 10 simulated sample paths with two observed states $\mathbf{x}_0 = (0.3, 0.2, 0.2, 0.3)$ and $\mathbf{x}_{0.001} = (0.4, 0.2, 0.2, 0.2)$. Other parameters $N = 5000$, $s_A = 0.01$, $h_A = 0.5$, $s_B = 0.05$, $h_B = 0.5$ and $r = 0.05$.

Compared to the simulated sample paths generated from the Wright-Fisher diffusion shown in Figure 3, the simulated sample paths generated from the G -based importance samplers shown in Figures 5-9 more closely resemble typical paths that start at state $\mathbf{x}_0$ and end at state $\mathbf{x}_T$ since it is unlikely for the gamete frequency to undergo a drastic change at a specific time. This is due to the guiding drift term $\mu_G(t, \mathbf{x})$ driving the simulated sample paths to terminate at the observed state $\mathbf{x}_T$, which enables most of the realisations $\xi_{U-1}^{(v)}$ drawn from the G -based

importance samplers to be in regions where the integrand $\hat{p}_{EM}^{(1)}(\Delta\tau_{U-1}, \boldsymbol{\xi}_{U-1}, \boldsymbol{\xi}_U)$ is large. The drift term $\boldsymbol{\mu}(\mathbf{x})$ of the Wright-Fisher diffusion incorporated into the G enables the simulated sample paths generated from the G -based importance samplers to imitate the local behaviour of the Wright-Fisher diffusion, which provides a sufficiently good approximation of the target. Hence, intuitively, the construction of the G -based importance samplers may result in much better performance.

Let $\hat{q}_{EM}^{(1)}(\Delta\tau_{u-1}, \boldsymbol{\xi}_{u-1}, \boldsymbol{\xi}_u)$ denote the probability density function of the multivariate normal distribution given in the G -based importance sampler and $w(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_{U-1})$ denote the importance sampling weight given by

$$w(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_{U-1}) = \prod_{u=1}^{U-1} \frac{\hat{p}_{EM}^{(1)}(\Delta\tau_{u-1}, \boldsymbol{\xi}_{u-1}, \boldsymbol{\xi}_u)}{\hat{q}_{EM}^{(1)}(\Delta\tau_{u-1}, \boldsymbol{\xi}_{u-1}, \boldsymbol{\xi}_u)}.$$

We can thereby approximate the Euler-Maruyama approximation $\hat{p}_{EM}^{(U)}(T, \mathbf{x}_0, \mathbf{x}_T)$ through importance sampling as

$$\hat{p}_{EM,IS}^{(U,V)}(T, \mathbf{x}_0, \mathbf{x}_T) = \frac{1}{V} \sum_{v=1}^V \phi(\mathbf{x}_T; \boldsymbol{\xi}_{U-1}^{(v)} + \boldsymbol{\mu}(\boldsymbol{\xi}_{U-1}^{(v)})\Delta\tau_{U-1}, \boldsymbol{\Sigma}(\boldsymbol{\xi}_{U-1}^{(v)})\Delta\tau_{U-1}) w(\boldsymbol{\xi}_1^{(v)}, \boldsymbol{\xi}_2^{(v)}, \dots, \boldsymbol{\xi}_{U-1}^{(v)}),$$

where $\{\boldsymbol{\xi}_1^{(v)}, \boldsymbol{\xi}_2^{(v)}, \dots, \boldsymbol{\xi}_{U-1}^{(v)}\}_{v=1}^V$ is the sample of the realisations successively generated from the G -based importance sampler, and V is the total number of the realisations in the sample. Similarly, we have

$$\lim_{V \rightarrow \infty} \lim_{U \rightarrow \infty} \hat{p}_{EM,IS}^{(U,V)}(T, \mathbf{x}_0, \mathbf{x}_T) = p(T, \mathbf{x}_0, \mathbf{x}_T).$$

We illustrate the convergence of the importance sampling approximation $\hat{p}_{EM,IS}^{(U,V)}(T, \mathbf{x}_0, \mathbf{x}_T)$ to the transition probability density $p(T, \mathbf{x}_0, \mathbf{x}_T)$ on a logarithmic scale in Figure 10. Compared to the Monte Carlo approximation $\hat{p}_{EM,MC}^{(U,V)}(T, \mathbf{x}_0, \mathbf{x}_T)$ demonstrated in Figure 4, the variance resulting from Monte Carlo integration is largely removed through importance sampling with the G_3 -based importance sampler, especially for the modified G_3 importance sampler.

4.4. Performance of the G -based Importance Samplers

We now evaluate the performance of the G -based importance samplers in the importance sampling approximation of the transition probability density $\hat{p}_{EM,IS}^{(U,V)}(T, \mathbf{x}_0, \mathbf{x}_T)$, in particular for the G_3 -based importance sampler. Given that the explicit analytical transition density function

Figure 10: Convergence of the importance sampling approximation of the transition probability density $\hat{p}_{EM,IS}^{(U,V)}(T, \mathbf{x}_0, \mathbf{x}_T)$ with the G_3 -based importance sampler. In this illustration, we partition the time interval $[0, 0.001]$ into 100 equidistant subintervals for the Euler-Maruyama scheme and take $\mathbf{x}_0 = (0.3, 0.2, 0.2, 0.3)$, $\mathbf{x}_{0.001} = (0.4, 0.2, 0.2, 0.2)$, $N = 5000$, $s_A = 0.01$, $h_A = 0.5$, $s_B = 0.05$, $h_B = 0.5$ and $r = 0.05$.

of the Wright-Fisher diffusion is still unknown, we illustrate the convergence of the importance sampling approximation $\hat{p}_{EM,IS}^{(U,V)}(T, \mathbf{x}_0, \mathbf{x}_T)$ to the transition probability density $p(T, \mathbf{x}_0, \mathbf{x}_T)$ with different G -importance samplers and evaluate their performance by looking at the effective sample size, denoted by V_E , which means that the importance sampling approximation $\hat{p}_{EM,IS}^{(U,V)}(T, \mathbf{x}_0, \mathbf{x}_T)$ based on the weighted sample of size V is approximately as accurate as one based on an independent sample of size V_E . From Liu (1996), the effective sample size is defined as

$$V_E = \frac{(\sum_{v=1}^V w(\xi_1^{(v)}, \xi_2^{(v)}, \dots, \xi_{U-1}^{(v)}))^2}{\sum_{v=1}^V (w(\xi_1^{(v)}, \xi_2^{(v)}, \dots, \xi_{U-1}^{(v)}))^2},$$

where $\{\xi_1^{(v)}, \xi_2^{(v)}, \dots, \xi_{U-1}^{(v)}\}_{v=1}^V$ is the sample of the realisations generated from the G -based importance sampler, and V is the total number of the realisations in the sample.

We fix the observed initial state $\mathbf{x}_0 = (0.3, 0.2, 0.2, 0.3)$ and consider the observed terminal state $\mathbf{x}_T$ for two cases: one is in the region where the underlying Wright-Fisher diffusion terminates at the time point T with a high probability, and the other is in the region where the underlying Wright-Fisher diffusion terminates at the time point T with a low probability. We take $T = 0.001$ and illustrate the transition probability density function for each gamete computed by kernel density estimation in Figure 3, which enable us to find high probability

Figure 11: Transition probability density function for each gamete computed by kernel density estimation provided by a standard routine in the `ks` package of R. In this illustration, we partition the time interval $[0, 0.001]$ into 100 equidistant subintervals for the Euler-Maruyama scheme and generate 10^6 simulated sample paths starting from the state $\mathbf{x}_0 = (0.3, 0.2, 0.2, 0.3)$. Other parameters $N = 5000$, $s_A = 0.01$, $h_A = 0.5$, $s_B = 0.05$, $h_B = 0.5$ and $r = 0.05$.

regions and low probability density regions where the underlying Wright-Fisher diffusion terminates at the time point T . From Figure 3, we pick $\mathbf{x}_T = (0.32, 0.20, 0.24, 0.24)$ as the observed terminal state in a high probability region and $\mathbf{x}_T = (0.38, 0.20, 0.24, 0.18)$ as the observed terminal state in a low probability region. We take use of a sample of the simulated sample paths generated from the proposal SDE of size $V = 1000$ to evaluate the importance sampling approximation $\hat{p}_{EM,IS}^{(U,V)}(T, \mathbf{x}_0, \mathbf{x}_T)$ with different G -based importance samplers for these two observed terminal states, and show the convergence of the the importance sampling approximation $\hat{p}_{EM,IS}^{(U,V)}(T, \mathbf{x}_0, \mathbf{x}_T)$ to the transition probability $p(T, \mathbf{x}_0, \mathbf{x}_T)$ on a logarithmic scale in Figures 12 and 13, respectively. As shown in Figures 12 and 13, the G_1 -based importance sampler has the worst performance of all G -based importance samplers, and the G_2 - and G_3 -based importance samplers have quite similar performance.

For further comparison of the performance of the G -based importance samplers, we replicate the importance sampling procedure 100 times and evaluate the average effective sample size for importance sampling based on different G -based importance samplers in Table 1. As shown in Table 1, the G_1 -based importance sampler has the smallest effective sample size for importance sampling of all G -based importance samplers, which confirms the results in Figures 12 and

Figure 12: Comparison of the convergence of the importance sampling approximation of the transition probability density $\hat{p}_{EM,IS}^{(U,V)}(T, \mathbf{x}_0, \mathbf{x}_T)$ with different G -based importance samplers for the observed terminal state in a high probability region. In this illustration, we partition the time interval $[0, 0.001]$ into 100 equidistant subintervals in the Euler-Maruyama scheme and generate 10^3 simulated sample paths with two observed states $\mathbf{x}_0 = (0.3, 0.2, 0.2, 0.3)$ and $\mathbf{x}_{0.001} = (0.32, 0.20, 0.24, 0.24)$. Other parameters $N = 5000$, $s_A = 0.01$, $h_A = 0.5$, $s_B = 0.05$, $h_B = 0.5$ and $r = 0.05$.

13. Overall, the effective sample size is larger for importance sampling based on the G_3 -based importance sampler than other G_2 -based importance samplers, especially for the modified G_3 -based importance sampler. For the observed state $\mathbf{x}_T$ in a high probability region, the effective sample size is larger for importance sampling based on the G_2 -based importance sampler with small values of the ρ -term than large values. In contrast, the effective sample size is larger for importance sampling based on the G_2 -based importance sampler with large values of the ρ -term than small values for the observed state $\mathbf{x}_T$ in a low probability region. These results confirm our intuition in the design of the ρ -term to improve the performance of the G_3 -based importance sampler, *i.e.*, it adjusts the contribution of the drift term of the underlying Wright-Fisher diffusion and the guiding drift term on the proposal SDE according to the observed terminal state $\mathbf{x}_T$. Moreover, from Table 1, we find that the effective sample size is larger for importance sampling based on the modified G_3 -based importance sampler than the original one, which implies that the modification suggested by Fearnhead (2008) works for the G_3 -based importance sampler. In summary, our results indicate that the modified G_3 -based importance

Figure 13: Comparison of the convergence of the importance sampling approximation of the transition probability density $\hat{p}_{EM,IS}^{(U,V)}(T, \mathbf{x}_0, \mathbf{x}_T)$ with different G -based importance samplers for the observed terminal state in a low probability region. In this illustration, we partition the time interval $[0, 0.001]$ into 100 equidistant subintervals in the Euler-Maruyama scheme and generate 10^3 simulated sample paths with two observed states $\mathbf{x}_0 = (0.3, 0.2, 0.2, 0.3)$ and $\mathbf{x}_{0.001} = (0.38, 0.20, 0.24, 0.18)$. Other parameters $N = 5000$, $s_A = 0.01$, $h_A = 0.5$, $s_B = 0.05$, $h_B = 0.5$ and $r = 0.05$.

470 sampler has the best performance of all G -based importance samplers.

Importance sampler	Average effective sample size	
	$\mathbf{x}_T = (0.32, 0.20, 0.24, 0.24)$	$\mathbf{x}_T = (0.38, 0.20, 0.24, 0.18)$
G_1	4.727507	5.510372
G_2 with $\rho = 0$	17.262143	13.079748
G_2 with $\rho = 1$	19.262232	12.439503
G_2 with $\rho = 10$	20.237910	11.992605
G_2 with $\rho = 100$	19.078593	10.780667
G_2 (modified) with $\rho = 0$	18.211303	14.126935
G_2 (modified) with $\rho = 1$	20.677225	12.980928
G_2 (modified) with $\rho = 10$	21.696237	12.597953
G_2 (modified) with $\rho = 100$	19.456650	11.723928
G_3	19.440782	13.301470
G_3 (modified)	20.459668	15.731508

Table 1: Average effective sample size for the importance sampling approximation of the transition probability density $\hat{p}_{EM,IS}^{(U,V)}(T, \mathbf{x}_0, \mathbf{x}_T)$ based on different G -based importance samplers.

471 To further improve the importance sampling approximation, we can adopt the asymptoti-
472 cally optimal choice of the total number of the realisations in the sample $V = \mathcal{O}(U^2)$ to balance

the trade-off between reducing the bias caused by the Euler-Maruyama scheme and reducing the variance resulting from Monte Carlo integration (Stramer & Yan, 2007), and take the use of antithetic variates for variance reduction, which is applied in Brandt & Santa-Clara (2002). There are other acceleration techniques, bias reduction techniques and variance reduction techniques discussed in Durham & Gallant (2002).

5. Discussion

In this work, we constructed two structurally different Wright-Fisher diffusion driven SDEs to characterise the population dynamics under natural selection at two linked loci in terms of the changes in gamete frequencies over time. One of these two SDEs has closed-form expressions for both of the drift term and the diffusion term, which can contribute to an efficient numerical sample path solution for the SDE satisfied by the two-locus Wright-Fisher diffusion with selection. Alternatively, we can describe the population evolving under natural selection at two linked loci in terms of the changes in allele frequencies and linkage disequilibrium over time with the Wright-Fisher diffusion driven SDE constructed in Appendix B, which however is of a more complicated form. We therefore only provided the derivation of the Wright-Fisher diffusion in Appendix B.

We developed an efficient importance sampling approximation of the transition probability density of the Wright-Fisher diffusion between two observed states. We evaluated the performance of several guided process based importance samplers. The main idea behind the design of the guiding term is to enable the sample path of the proposal SDE to hit the observed states while at the same time imitate the local behaviour of the underlying Wright-Fisher diffusion. Our importance sampler is an extension of the proposal introduced by Fearnhead (2008), where we take the ρ -term in the proposal of Fearnhead (2008) to be a function of the observed state $\mathbf{x}_T$. Such a modification is used to adjust the contributions of the drift term of the underlying Wright-Fisher diffusion and the guiding drift term on the proposal SDE according to the observed state $\mathbf{x}_T$, so that the resulting proposal SDE can evolve largely under the influence of the underlying Wright-Fisher diffusion if its sample path is already close to the given state $\mathbf{x}_T$ with a high probability, but allow a more significant contribution from the guiding term if its sample path gets further away from the given state $\mathbf{x}_T$ with a high probability. A simple comparison suggests that our importance sampler may have better performance, which can

be further improved through various acceleration techniques, bias reduction techniques and variance reduction techniques proposed by Durham & Gallant (2002).

In recent years, the Wright-Fisher diffusion has already been successfully applied in the population genetic analysis of time series data of allele frequencies, but most of these methods are limited to either a single locus or multiple independent loci, *i.e.*, genetic recombination effect and local linkage information are ignored in these methods. However, in some cases accounting genetic recombination and linkage information is essential for the inference of natural selection, especially when two loci are tightly linked. Otherwise incorrect results might be obtained with existing single-locus methods (see Figure 14 and 15). Our method can play an important role in the development of the

many existing simulation-based inference approaches such as simulated maximum-likelihood methods and Markov chain Monte Carlo methods for the inference of natural selection from generic time series data

Figure 14: A example of a positively selected locus $\mathcal{A}$ tightly linked with a neutral locus $\mathcal{B}$. The initial haplotype frequencies of the population are $\mathbf{x}_0 = (0.2, 0.1, 0.3, 0.4)$, and the selection coefficients are $s_A = 0.01$ and $s_B = 0$. Dark grey denotes the one-locus Wright-Fisher diffusion, and light grey denotes the the two-locus Wright-Fisher diffusion.

Given that many phenotypes are known to show variation caused by multiple linked loci, there is strong motivation to consider natural selection at more than one locus. However, theoretical studies of natural selection simultaneously acting on multiple linked loci are much more challenging. Even for the case of two linked loci, the population evolving subject to natural selection is complicated and has not been completely understood. According to Hamilton (2009), genetic recombination promotes linkage equilibrium whereas natural selection maximises the mean fitness, and the equilibrium is determined by the relative strength of these two mecha-

Figure 15: A example of two tightly linked loci $\mathcal{A}$ and $\mathcal{B}$ under natural selection. The initial haplotype frequencies of the population are $\mathbf{x}_0 = (0.05, 0.05, 0.7, 0.2)$, and the selection coefficients are $s_{\mathcal{A}} = 0.01$ and $s_{\mathcal{B}} = 0.005$. Dark grey denotes the one-locus Wright-Fisher diffusion, and light grey denotes the the two-locus Wright-Fisher diffusion.

When the two processes are of approximately equal strength, the result is a compromise that may produce an equilibrium that is neither in linkage equilibrium nor at maximum mean fitness. In addition, the equilibrium also depends strongly on gene actions (*i.e.*, the dominance parameters $h_{\mathcal{A}}$ and $h_{\mathcal{B}}$) and the initial state of the population. However, all of these conclusions have been reached under the assumption of the infinite effective population size, which is quite restrictive and is never met in natural populations. In this section, we investigate the effect of natural selection acting on the population dynamics at two linked loci based on the simulated Wright-Fisher diffusion sample paths.

Figure 16: Dynamics of the empirical transition probabilities for the $\mathcal{A}_1$ allele, the $\mathcal{B}_1$ allele and linkage disequilibrium over 500 generations for the population initially in positive linkage disequilibrium. We take other parameters to be $\mathbf{x}_0 = (0.3, 0.2, 0.2, 0.3)$, $N = 5000$, $s_{\mathcal{A}} = 0.01$, $h_{\mathcal{A}} = 0.5$, $s_{\mathcal{B}} = 0.05$, $h_{\mathcal{B}} = 0.5$ and $r = 0.05$.

We illustrate several examples of the population dynamics subject to natural selection at two linked loci in Figures 16-20, where we partition the time interval into 5,000 equidistant

Figure 17: Dynamics of the empirical transition probabilities for the $\mathcal{A}_1$ allele, the $\mathcal{B}_1$ allele and linkage disequilibrium over 500 generations for the population initially in negative linkage disequilibrium. We take other parameters to be $\mathbf{x}_0 = (0.2, 0.3, 0.3, 0.2)$, $N = 5000$, $s_A = 0.01$, $h_A = 0.5$, $s_B = 0.05$, $h_B = 0.5$ and $r = 0.05$.

Figure 18: Dynamics of the empirical transition probabilities for the $\mathcal{A}_1$ allele, the $\mathcal{B}_1$ allele and linkage disequilibrium over 500 generations for the allele favoured by natural selection initially in low frequency. We take other parameters to be $\mathbf{x}_0 = (0.03, 0.02, 0.48, 0.47)$, $N = 5000$, $s_A = 0.01$, $h_A = 0.5$, $s_B = 0.05$, $h_B = 0.5$ and $r = 0.05$.

Figure 19: Dynamics of the empirical transition probabilities for the $\mathcal{A}_1$ allele, the $\mathcal{B}_1$ allele and linkage disequilibrium over 500 generations for non-additive gene actions. We take other parameters to be $\mathbf{x}_0 = (0.3, 0.2, 0.2, 0.3)$, $N = 5000$, $s_A = 0.01$, $h_A = 0.5$, $s_B = 0.05$, $h_B = 0.9$ and $r = 0.05$.

Figure 20: Dynamics of the empirical transition probabilities for the $\mathcal{A}_1$ allele, the $\mathcal{B}_1$ allele and linkage disequilibrium over 500 generations for neutral locus $\mathcal{A}$. We partition the time interval into 5000 equidistant subintervals in the Euler-Maruyama scheme and evaluate the empirical transition probabilities with 10^6 independent realisations of the Wright-Fisher diffusion. We take other parameters to be $\mathbf{x}_0 = (0.3, 0.2, 0.2, 0.3)$, $N = 5000$, $s_{\mathcal{A}} = 0$, $h_{\mathcal{A}} = 0.5$, $s_{\mathcal{B}} = 0.05$, $h_{\mathcal{B}} = 0.5$ and $r = 0.05$.

subintervals in the Euler-Maruyama scheme and evaluate the empirical transition probabilities with 1,000,000 independent realisations of the Wright-Fisher diffusion. Figures 16 and 17 demonstrate the cases of the population initially in positive and negative linkage disequilibrium, respectively. Figure 18 gives an example when the positively selected allele is initially in low frequency. Figure 19 shows an example of non-additive gene action. Figure 20 demonstrates the case when only locus $\mathcal{B}$ is subject to natural selection. As illustrated in Figures 16-20, we find that genetic recombination promotes linkage equilibrium whereas natural selection promotes the fixation of the gamete type favoured by natural selection, $\mathcal{A}_1\mathcal{B}_1$, *i.e.*, the type with the highest mean fitness, which confirms the results stated in Hamilton (2009). However, in finite populations, linkage disequilibrium always decays to zero, *i.e.*, linkage equilibrium is eventually reached. When both of the two loci are positively selected, natural selection leads eventually to the fixation of a certain gamete type $\mathcal{A}_i\mathcal{B}_j$, but not necessarily the $\mathcal{A}_1\mathcal{B}_1$ gamete. In fact, which gamete type becomes fixed is a random event with a corresponding fixation probability determined by the population size, the strength of natural selection and genetic recombination, gene actions and the initial state of the population (see Figures 16-19). By contrast, when at least one of the two loci is neutral, the fixation of a certain gamete type $\mathcal{A}_i\mathcal{B}_j$ may not occur (see Figure 20).

The scope of our results in the present work can be extended in various directions. A direct extension is to certain multi-locus problems. As we have stated above, in theory, our method

can be naturally extended to high-dimensional problems, but may suffer from the potential issue of computational efficiency in practice. We have two potential directions to improve the performance of our method for multi-locus problems. One direction is to consider how to exactly simulate the sample path of the Wright-Fisher diffusion to reduce the bias caused by numerical simulation schemes in this work. A series of recent papers (*e.g.*, Jenkins & Spanó, 2017) may provide some new inspirations. The other direction is to consider how to construct a more proper proposal SDE in importance sampling. In this work, we simply apply the Mahalanobis distance (De Maesschalck et al., 2000) to construct the ρ -term in our importance sampler, which roughly measures the distance from the observed state $\mathbf{x}_T$ to the regions where the underlying Wright-Fisher diffusion sample path terminates at time T with a high probability. A better construction of the ρ -term may improve the performance of the importance sampling approximation of the transition probability density.

Acknowledgements

This work was funded in part by the Engineering and Physical Sciences Research Council (EPSRC) Grant EP/I028498/1 to F.Y.

References

- Airault, H., & Ren, J. (2002). Modulus of continuity of the canonic Brownian motion “on the group of diffeomorphisms of the circle”. *Journal of Functional Analysis*, 196, 395–426.
- Allen, E. J., Allen, L. J. S., Arciniega, A., & Greenwood, P. E. (2008). Construction of equivalent stochastic differential equation models. *Stochastic Analysis and Applications*, 26, 274–297.
- Allentoft, M. E., Sikora, M., Sjogren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P. B., Schroeder, H., Ahlstrom, T., Vinner, L., Malaspinas, A.-S., Margaryan, A., Higham, T., Chivall, D., Lynnerup, N., Harvig, L., Baron, J., Casa, P. D., Dabrowski, P., Duffy, P. R., Ebel, A. V., Epimakhov, A., Frei, K., Furmanek, M., Gralak, T., Gromov, A., Gronkiewicz, S., Grupe, G., Hajdu, T., Jarysz, R., Khartanovich, V., Khokhlov, A., Kiss, V., Kolar, J., Kriiska, A., Lasak, I., Longhi, C., McGlynn, G., Merkevicius, A., Merkyte, I., Metspalu, M., Mkrtychyan, R., Moiseyev, V., Paja, L., Palfi, G., Pokutta, D., Pospieszny, L., Price, T. D., Saag, L., Sablin, M., Shishlina, N., Smrcka, V., Soenov, V. I., Szeverenyi, V.,

580 Toth, G., Trifanova, S. V., Varul, L., Vicze, M., Yepiskoposyan, L., Zhitenev, V., Orlando,
581 L., Sicheritz-Ponten, T., Brunak, S., Nielsen, R., Kristiansen, K., & Willerslev, E. (2015).
582 Population genomics of Bronze Age Eurasia. *Nature*, 522, 167–172.

583 Boitard, S., & Loisel, P. (2007). Probability distribution of haplotype frequencies under the
584 two-locus Wright-Fisher model by diffusion approximation. *Theoretical Population Biology*,
585 71, 380–391.

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

588 Brandt, M. W., & Santa-Clara, P. (2002). Simulated likelihood estimation of diffusions with
589 an application to exchange rate dynamics in incomplete markets. *Journal of Financial Eco-*
590 *nomics*, 63, 161–210.

591 Crow, J. F., & Kimura, M. (1970). *An Introduction to Population Genetics Theory*. New York:
592 Harper & Row.

593 Dangerfield, C. E., Kay, D., MacNamara, S., & Burrage, K. (2012). A boundary preserving nu-
594 merical algorithm for the Wright-Fisher model with mutation. *BIT Numerical Mathematics*,
595 52, 283–304.

596 De Maesschalck, R., Jouan-Rimbaud, D., & Massart, D. L. (2000). The mahalanobis distance.
597 *Chemometrics and Intelligent Laboratory Systems*, 50, 1–18.

598 Delyon, B., & Hu, Y. (2006). Simulation of conditioned diffusion and application to parameter
599 estimation. *Stochastic Processes and their Applications*, 116, 1660–1675.

600 Durham, G. B., & Gallant, A. R. (2002). Numerical techniques for maximum likelihood esti-
601 mation of continuous-time diffusion processes. *Journal of Business & Economic Statistics*,
602 20, 297–338.

603 Durrett, R. (2008). *Probability Models for DNA Sequence Evolution*. New York: Springer-
604 Verlag.

605 Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoretical Population*
606 *Biology*, 3, 87–112.

607 Ewens, W. J. (2004). *Mathematical Population Genetics 1: Theoretical Introduction*. New York:
608 Springer-Verlag.

609 Fearnhead, P. (2008). Computational methods for complex stochastic systems: a review of some
610 alternatives to MCMC. *Statistics and Computing*, 18, 151–171.

611 Ferrer-Admetlla, A., Leuenberger, C., Jensen, J. D., & Wegmann, D. (2016). An approximate
612 Markov model for the Wright-Fisher diffusion and its application to time series data. *Genetics*,
613 203, 831–846.

614 Fisher, R. A. (1922). On the dominance ratio. *Proceedings of the Royal Society of Edinburgh*,
615 42, 321–341.

616 Fuchs, C. (2013). *Inference for Diffusion Processes: With Applications in Life Sciences*. Berlin:
617 Springer-Verlag.

618 Gautier, M., & Vitalis, R. (2013). Inferring population histories using genome-wide allele
619 frequency data. *Molecular Biology and Evolution*, 30, 654–668.

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

623 Hamilton, M. (2009). *Population Genetics*. Chichester: Wiley-Blackwell.

624 He, Z., Beaumont, M., & Yu, F. (2017). Effects of the ordering of natural selection and popu-
625 lation regulation mechanisms on Wright-Fisher models. *G3: Genes, Genomes, Genetics*, 7,
626 2095–2106.

627 He, Z., Dai, X., Beaumont, M. A., & Yu, F. (2019). Maximum likelihood estimation of natural
628 selection and allele age from time series data of allele frequencies. *bioRxiv*, (p. 837310).

629 Jenkins, P. A., & Spanó, D. (2017). Exact simulation of the Wright-Fisher diffusion. *The Annals*
630 *of Applied Probability*, 27, 1478–1509.

631 Karlin, S., & Taylor, H. E. (1981). *A Second Course in Stochastic Processes*. London: Academic
632 Press.

- Kimura, M. (1955). Solution of a process of random genetic drift with a continuous model. *Proceedings of the National Academy of Sciences of the United States of America*, 41, 144–150.
- Kimura, M. (1964). Diffusion models in population genetics. *Journal of Applied Probability*, 1, 177–232.
- Kingman, J. F. C. (1982). The coalescent. *Stochastic Processes and their Applications*, 13, 235–248.
- Kloeden, P. E., & Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*. Berlin: Springer-Verlag.
- Kopp, A., Büsching, I., Strauss, R. D., & Potgieter, M. S. (2012). A stochastic differential equation code for multidimensional Fokker-Planck type problems. *Computer Physics Communications*, 183, 530–542.
- Liu, J. S. (1996). Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and Computing*, 6, 113–119.
- Lukić, S., & Hey, J. (2012). Demographic inference using spectral methods on SNP data, with an analysis of the human out-of-Africa expansion. *Genetics*, 192, 619–639.
- Lukić, S., Hey, J., & Chen, K. (2011). Non-equilibrium allele frequency spectra via spectral methods. *Theoretical Population Biology*, 79, 203–219.
- Malaspinas, A.-S., Malaspinas, O., Evans, S. N., & Slatkin, M. (2012). Estimating allele age and selection coefficient from time-serial data. *Genetics*, 192, 599–607.
- Malliavin, P. (1999). The canonic diffusion above the diffeomorphism group of the circle. *Comptes Rendus de l'Académie des Sciences-Series I-Mathematics*, 329, 325–329.
- Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., Sirak, K., Gamba, C., Jones, E. R., Llamas, B., Dryomov, S., Pickrell, J., Arsuaga, J. L., de Castro, J. B., Carbonell, E., Gerritsen, F., Khokhlov, A., Kuznetsov, P., Lozano, M., Meller, H., Mochalov, O., Moiseyev, V., Guerra, M. A. R., Roodenberg, J., Vergés, J. M., Krause, J., Cooper, A., Alt, K. W., Brown, D.,

Anthony, D., Lalueza-Fox, C., Haak, W., Pinhasi, R., & Reich, D. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, 528, 499–503.

Neuenkirch, A., & Szpruch, L. (2014). First order strong approximations of scalar SDEs defined in a domain. *Numerische Mathematik*, 128, 103–136.

Papaspiliopoulos, O., & Roberts, G. (2012). Importance sampling techniques for estimation of diffusion models. In *Statistical Methods for Stochastic Differential Equations* (pp. 311–340). New York: CRC Press.

Pedersen, A. R. (1995a). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics*, 22, 55–71.

Pedersen, A. R. (1995b). Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli*, 1, 257–279.

Sato, K.-I. (1976). Diffusion processes and a class of Markov chains related to population genetics. *Osaka Journal of Mathematics*, 13, 631–659.

Schraiber, J. G., Evans, S. N., & Slatkin, M. (2016). Bayesian inference of natural selection from allele frequency time series. *Genetics*, 203, 493–511.

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

Schurz, H. (1996). Numerical regularization for SDEs: Construction of nonnegative solutions. *Dynamic Systems and Applications*, 5, 323–352.

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

Steinrücken, M., Bhaskar, A., & Song, Y. S. (2014). A novel spectral method for inferring general diploid selection from time series genetic data. *The Annals of Applied Statistics*, 8, 2203–2222.

Steinrücken, M., Jewett, E. M., & Song, Y. S. (2015). SpectralTDF: transition densities of diffusion processes with time-varying selection parameters, mutation rates and effective population sizes. *Bioinformatics*, *32*, 795–797.

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

Stramer, O., & Yan, J. (2007). Asymptotics of an efficient Monte Carlo estimation for the transition density of diffusion processes. *Methodology and Computing in Applied Probability*, *9*, 483–496.

Stroock, D. W., & Varadhan, S. R. S. (1979). *Multidimensional Diffusion Processes*. Berlin: Springer-Verlag.

Taniguchi, T. (1992). Successive approximations to solutions of stochastic differential equations. *Journal of Differential Equations*, *96*, 152–169.

Tataru, P., Simonsen, M., Bataillon, T., & Hobolth, A. (2017). Statistical inference in the Wright-Fisher model using allele frequency data. *Systematic Biology*, *66*, e30–e46.

Tocino, A., & Vigo-Aguiar, J. (2003). New Itô-Taylor expansions. *Journal of Computational and Applied Mathematics*, *158*, 169–185.

Vitalis, R., Gautier, M., Dawson, K. J., & Beaumont, M. A. (2014). Detecting and measuring selection from gene frequency data. *Genetics*, *196*, 799–817.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*, 97–159.

Yamada, T., & Watanabe, S. (1971). On the uniqueness of solutions of stochastic differential equations. *Journal of Mathematics of Kyoto University*, *11*, 155–167.

Zhang, X. (2006). Euler-Maruyama approximations for SDEs with non-Lipschitz coefficients and applications. *Journal of Mathematical Analysis and Applications*, *316*, 447–458.

Zhao, L., Yue, X., & Waxman, D. (2013). Complete numerical solution of the diffusion equation of random genetic drift. *Genetics*, *194*, 973–985.

711 Živković, D., Steinrücken, M., Song, Y. S., & Stephan, W. (2015). Transition densities and
 712 sample frequency spectra of diffusion processes with selection and variable population size.
 713 *Genetics*, 200, 601–617.

714 Appendix A. Proof of Theorem 2

715 Let $\mathcal{A}$ denote the generator associated with the SDE in Eq. (4), then

$$\mathcal{A}f(x) = \sum_{i=1}^4 \mu_i(x) \frac{\partial}{\partial x_i} f(x) + \frac{1}{2} \sum_{i,j=1}^4 \Sigma_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} f(x).$$

716 Let $\mathcal{A}^*$ be the generator associated with the SDE in Eq. (12), then

$$\mathcal{A}^*f(x) = \sum_{i=1}^4 \mu_i(x) \frac{\partial}{\partial x_i} f(x) + \frac{1}{2} \sum_{i,j=1}^4 a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} f(x),$$

717 where

$$a_{ij}(x) = \sum_{k=1}^6 \sigma_{ik}(x) \sigma_{kj}(x).$$

718 From (13), we have

$$\boldsymbol{\sigma}(\mathbf{x}) \boldsymbol{\sigma}^\top(\mathbf{x}) = \begin{pmatrix} x_1(x_2 + x_3 + x_4) & -x_1x_2 & -x_1x_3 & -x_1x_4 \\ -x_1x_2 & x_2(x_1 + x_3 + x_4) & -x_2x_3 & -x_2x_4 \\ -x_1x_3 & -x_2x_3 & x_3(x_1 + x_2 + x_4) & -x_3x_4 \\ -x_1x_4 & -x_2x_4 & -x_3x_4 & x_4(x_1 + x_2 + x_3) \end{pmatrix},$$

719 which is the same as $\Sigma_{ij}(\mathbf{x}) = x_i(\delta_{ij} - x_j)$ given in (7). This shows the two SDE's in Eqs. (4)

720 and (12) have the same weak solutions. $\square$

721 Appendix B. An alternative formulation of the Wright-Fisher diffusion

722 We provide an alternative formulation of the Wright-Fisher diffusion for the population dy-
 723 namics under natural selection at two linked loci. From Eq. (8), it is clear that the population
 724 evolving according to the Wright-Fisher diffusion $\mathbf{X}$ can be characterised by considering any
 725 three linearly independent gamete frequencies out of the four, *e.g.*, X_1 , X_2 and X_3 . However,
 726 there is an obvious asymmetry in doing so, and based on Ewens (2004), we find it more conve-
 727 nient to work with a vector of new independent variables $\mathbf{Z}^{(N)}(k) = (Z_1^{(N)}(k), Z_2^{(N)}(k), Z_3^{(N)}(k))$,
 728 which is related to $\mathbf{X}^{(N)}(k)$ in the following way:

$$\begin{aligned} Z_1^{(N)}(k) &= X_1^{(N)}(k) + X_2^{(N)}(k) \\ Z_2^{(N)}(k) &= X_1^{(N)}(k) + X_3^{(N)}(k) \\ Z_3^{(N)}(k) &= X_1^{(N)}(k)X_4^{(N)}(k) - X_2^{(N)}(k)X_3^{(N)}(k). \end{aligned} \quad (\text{B.1})$$

729 Note that $Z_1^{(N)}$ is the frequency of the $\mathcal{A}_1$ allele, $Z_2^{(N)}$ is the frequency of the $\mathcal{B}_1$ allele, and
 730 $Z_3^{(N)}$ is the standard measure of linkage disequilibrium.

731 **Theorem 3.** *If we let the population size N go to infinity and hold the scaled parameters $\alpha_{\mathcal{A}}$, $\alpha_{\mathcal{B}}$
 732 and R constant, the Markov process of mutant allele frequencies and linkage disequilibrium $\mathbf{Z}^{(N)}$
 733 converges to a time-homogeneous diffusion process, denoted by $\mathbf{Z} = \{\mathbf{Z}(t), t \geq 0\}$, satisfying the
 734 following stochastic differential equations of the form*

$$d\mathbf{Z}(t) = \boldsymbol{\mu}(\mathbf{Z}(t))dt + \boldsymbol{\sigma}(\mathbf{Z}(t))d\mathbf{W}(t), \quad (\text{B.2})$$

735 where the drift term $\boldsymbol{\mu}(\mathbf{z})$ is a three-dimensional vector being

$$\begin{aligned} \mu_1(\mathbf{z}) &= \alpha_{\mathcal{A}}z_1(1 - z_1)[h_{\mathcal{A}}z_1 + (1 - h_{\mathcal{A}})(1 - z_1)] \\ &\quad + \alpha_{\mathcal{B}}[h_{\mathcal{B}}z_2 + (1 - h_{\mathcal{B}})(1 - z_2)]z_3 \\ \mu_2(\mathbf{z}) &= \alpha_{\mathcal{A}}[h_{\mathcal{A}}z_1 + (1 - h_{\mathcal{A}})(1 - z_1)]z_3 \\ &\quad + \alpha_{\mathcal{B}}z_2(1 - z_2)[h_{\mathcal{B}}z_2 + (1 - h_{\mathcal{B}})(1 - z_2)] \\ \mu_3(\mathbf{z}) &= \alpha_{\mathcal{A}}(1 - 2z_1)[h_{\mathcal{A}}z_1 + (1 - h_{\mathcal{A}})(1 - z_1)]z_3 \\ &\quad + \alpha_{\mathcal{B}}(1 - 2z_2)[h_{\mathcal{B}}z_2 + (1 - h_{\mathcal{B}})(1 - z_2)]z_3 - (1 + R)z_3, \end{aligned} \quad (\text{B.3})$$

the diffusion term $\sigma(\mathbf{z})$ is three by three matrix being the square root of the infinitesimal covariance matrix $\Sigma(\mathbf{z})$ given by

$$\begin{aligned}\Sigma_{11}(\mathbf{z}) &= z_1(1 - z_1) \\ \Sigma_{12}(\mathbf{z}) &= z_3 \\ \Sigma_{13}(\mathbf{z}) &= (1 - 2z_1)z_3 \\ \Sigma_{22}(\mathbf{z}) &= z_2(1 - z_2) \\ \Sigma_{23}(\mathbf{z}) &= (1 - 2z_2)z_3 \\ \Sigma_{33}(\mathbf{z}) &= z_1(1 - z_1)z_2(1 - z_2) + (1 - 2z_1)(1 - 2z_2)z_3 - z_3^2,\end{aligned}\tag{B.4}$$

and $\mathbf{W}(t)$ is a three-dimensional standard Brownian motion.

The proof of Theorem 3 is almost identical to the proof for neutral populations in Durrett (2008, p. 324) by using a multivariate extension of the change of variables formula. We refer to the process of mutant allele frequencies and linkage disequilibrium $\mathbf{Z}$ solving the SDE in Eq. (B.2) as the two-locus Wright-Fisher diffusion with selection, which evolves in the state space

$$\begin{aligned}\Omega_{\mathbf{Z}} &= \{\mathbf{z} \in [0, 1] \times [0, 1] \times [-0.25, 0.25] : \\ &\max\{-z_1z_2, -(1 - z_1)(1 - z_2)\} \leq z_3 \leq \min\{z_1(1 - z_2), (1 - z_1)z_2\}\}.\end{aligned}$$

From Eqs. (B.3) and (B.4), we find that the drift term $\mu(\mathbf{z})$ and the diffusion term $\sigma(\mathbf{z})$ satisfy both global Lipschitz and linear growth conditions (8) and (9) for all $\mathbf{z} \in \Omega_{\mathbf{Z}}$, which guarantees the existence and uniqueness of a non-explosive solution of the Wright-Fisher diffusion driven SDE for any given $\mathbf{Z}(0)$ independent of $\mathbf{W}(t)$ and satisfying $E(|\mathbf{Z}(0)|^2) < \infty$.

Moreover, for any $\mathbf{z}$ in the interior of the state space $\Omega_{\mathbf{Z}}$, the infinitesimal mean vector $\mu(\mathbf{z})$ and the infinitesimal covariance matrix $\Sigma(\mathbf{z})$ are twice continuously differentiable with respect to $\mathbf{z}$ with bounded derivatives, and the infinitesimal covariance matrix $\Sigma(\mathbf{z})$ is also positive definite, which implies that there exists a well-defined and smooth transition probability density function of the Wright-Fisher diffusion $\mathbf{Z}$ satisfying the Kolmogorov backward equation

$$\frac{\partial}{\partial t}p(t, \mathbf{z}, \mathbf{z}') = \frac{1}{2} \sum_{i=1}^3 \sum_{j=1}^3 \Sigma_{ij}(\mathbf{z}) \frac{\partial^2 p(t, \mathbf{z}, \mathbf{z}')}{\partial z_i \partial z_j} + \sum_{i=1}^3 \mu_i(\mathbf{z}) \frac{\partial p(t, \mathbf{z}, \mathbf{z}')}{\partial z_i}\tag{B.5}$$

753 for fixed $\mathbf{z}'$ and the Kolmogorov forward equation

$$\frac{\partial}{\partial t}p(t, \mathbf{z}, \mathbf{z}') = \frac{1}{2} \sum_{i=1}^3 \sum_{j=1}^3 \frac{\partial^2 (\Sigma_{ij}(\mathbf{z}')p(t, \mathbf{z}, \mathbf{z}'))}{\partial z'_i \partial z'_j} - \sum_{i=1}^3 \frac{\partial (\mu_i(\mathbf{z}')p(t, \mathbf{z}, \mathbf{z}'))}{\partial z'_i} \quad (\text{B.6})$$

754 for fixed $\mathbf{z}$, where the infinitesimal mean vector $\boldsymbol{\mu}(\mathbf{z})$ and the infinitesimal covariance matrix
755 $\boldsymbol{\Sigma}(\mathbf{z})$ are separately given by Eqs. (B.3) and (B.4).