Moment-based approximations for the Wright-Fisher model of population dynamics under natural selection at two linked loci

Zhangyi He[a,1,2,*], Wenyang Lyu[b,1], Mark Beaumont[c], Feng Yu[b,*]

[a] *Department of Statistics, University of Oxford, Oxford OX1 3LB, United Kingdom*
[b] *School of Mathematics, University of Bristol, Bristol BS8 1UG, United Kingdom*
[c] *School of Biological Sciences, University of Bristol, Bristol BS8 1TQ, United Kingdom*

Abstract

Properly modelling genetic recombination and local linkage has been shown to bring a significant improvement to the inference of natural selection from time series genetic data under a Wright-Fisher model. Existing approaches that can take genetic recombination effect and local linkage information into account are built upon either the diffusion approximation or the moment-based approximation of the Wright-Fisher model. However, such approximations are either limited to the increased computational cost like the diffusion approximation or suffer from the distribution support issue like the normal approximation, which can seriously affect computational efficiency and accuracy. In this work, we propose two novel moment-based approximations for the Wright-Fisher model of population dynamics subject to natural selection at a pair of linked loci. Our key innovation is that we extend existing approaches to calculate the mean and (co)variance of the two-locus Wright-Fisher model with selection and suggest a logistic normal distribution or a hierarchical beta distribution as a parametric continuous probability distribution to approximate the Wright-Fisher model by matching its first two moments to those of the Wright-Fisher model. Compared with the diffusion approximation, our approximations enable the computation of the transition probability distribution of the Wright-Fisher model at a far smaller computational cost and also allow us to avoid the distribution support issue found in the normal approximation.

Keywords: Natural selection, Linked loci, Wright-Fisher model, Moment approximation, Moment-based approximation

*Corresponding author.
Email addresses: z.he@imperial.ac.uk (Zhangyi He), feng.yu@bristol.ac.uk (Feng Yu)
[1] These authors contributed equally to this work.
[2] Present address: Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London W2 1PG, United Kingdom

1. Introduction

In the fields of evolutionary and population genetics, understanding the relative importance of genetic drift and natural selection as the mechanisms for maintaining polymorphisms within populations and driving divergences between populations has been a central issue. Mathematical descriptions of population dynamics under the effects of various demographic and evolutionary forces such as genetic drift and natural selection are typically built on the Wright-Fisher model, introduced by Fisher (1922) and Wright (1931). The Wright-Fisher model forms the basis for most theoretical and applied research in evolutionary and population genetics to date, *e.g.*, Ewens' sampling formula (Ewens, 1972) and Kingman's coalescent (Kingman, 1982). Moreover, the increased availability of time series genetic data in recent years enables accurate inference of demographic and genetic properties of observed populations, especially for detection and estimation of natural selection. The Wright-Fisher model is one of the most popular mathematical models that statistical techniques for inferring natural selection are built on as it provides an elegant mathematical framework for explicitly accounting for the effects of various demographic and evolutionary forces on the change in allele frequencies over time (see Tataru et al., 2017, for a review of the statistical inference in the Wright-Fisher model using allele frequency data).

The most important ingredient for carrying out statistical inference of natural selection in the Wright-Fisher model from time series allele frequency data is how to compute the probability distribution of the allele frequency at each given time point conditional on the initial allele frequency. However, to the best of our knowledge, there is no existing tractable analytical form for the probability distribution of the allele frequency under the Wright-Fisher model. Numerical evaluation of the probability distribution of the allele frequency is theoretically feasible with the transition probability distribution but computationally prohibitive for large population sizes and evolutionary timescales. Most existing approaches are therefore built on either the diffusion approximation of the Wright-Fisher model (*e.g.*, Bollback et al., 2008; Malaspinas et al., 2012; Steinrücken et al., 2014; Schraiber et al., 2016; Ferrer-Admetlla et al., 2016; He et al., 2020b,c) or the moment-based approximation of the Wright-Fisher model (*e.g.*, Lacerda & Seoighe, 2014; Terhorst et al., 2015; Paris et al., 2019). These approximations enable efficient integration over all possible allele frequency trajectories of the underlying population, therefore allowing the likelihood computation to be completed in a reasonable amount of time.

Recent advances in sequencing techniques have revolutionized the collection of time series genomic data, increasing both the volume and quality of available sequenced individuals from a large variety of populations (*e.g.*, Mathieson et al., 2015; Fages et al., 2019; Papkou et al., 2019; Barghi et al., 2019). This provides an opportunity for studying natural selection while taking into account the process of genetic recombination and the information of local linkage. Properly modelling genetic recombination and local linkage has been shown to bring a significant improvement to the inference of natural selection from time series genetic data (He et al., 2020b). However, with the exception of Terhorst et al. (2015) and He et al. (2020b), all existing methods built upon the Wright-Fisher model for estimating selection coefficients from temporally spaced genetic data are limited to either a single locus (*e.g.*, Bollback et al., 2008; Malaspinas et al., 2012; Steinrücken et al., 2014; Lacerda & Seoighe, 2014; Schraiber et al., 2016; He et al., 2020c) or multiple independent loci (*e.g.*, Ferrer-Admetlla et al., 2016; Paris et al., 2019), *i.e.*, genetic recombination and local linkage are ignored in these approaches.

Terhorst et al. (2015), one of the exceptions amongst these methods, extended the moment-based approximation of the Wright-Fisher model introduced by Feder et al. (2014) to the case of natural selection acting on multiple linked loci, where the Wright-Fisher model was approximated through a deterministic path with added Gaussian noises. Their Gaussian approximation works well for many applications when modelling the allele frequencies with intermediate values. However, as soon as the allele frequencies get close to their boundaries 0 or 1 (*i.e.*, allele loss or fixation), the Wright-Fisher model will be poorly approximated due to the infinite support of the Gaussian distribution that will result in a leak of probability mass into the values of the allele frequency that are smaller than 0 or larger than 1, which is not mathematically possible. This issue becomes more problematic in the inference of natural selection as natural selection is expected to rapidly drive the allele frequencies towards the boundaries. Most recently, He et al. (2020b) developed a Bayesian statistical framework for inferring natural selection at two linked loci while explicitly modelling genetic recombination and local linkage. Their method was built on the diffusion approximation of the Wright-Fisher model for the population evolving under natural selection at two linked loci, represented in the stochastic differential equation form introduced by He et al. (2020a). The computation of the transition probability density function by numerically solving the Kolmogorov backward equation (or its adjoint) resulting from the

diffusion approximation is computationally challenging and prohibitively expensive. Their posterior computation was therefore carried out through the particle marginal Metropolis-Hastings algorithm developed by Andrieu et al. (2010), which avoids the computation of the transition probability density function but is still time-consuming.

In this work, we introduce two novel moment-based approximations for the Wright-Fisher model of the population dynamics subject to natural selection at two linked loci. The principle of the moment-based approximation is to approximate the Wright-Fisher model with a parametric continuous probability distribution by matching its first two moments to those of the Wright-Fisher model, which could be reminiscent of Cavalli-Sforza & Edwards (1967). To calculate the first two moments of the two-locus Wright-Fisher model with selection, we extend the moment approximations developed by Lacerda & Seoighe (2014) and Paris et al. (2019), respectively. To avoid the distribution support issue found in Terhorst et al. (2015), we choose two different parametric continuous probability distributions to approximate the transition probability distribution of the Wright-Fisher model, the logistic normal distribution and the hierarchical beta distribution. Compared to the diffusion approximation of He et al. (2020b), our moment-based approximations enable the computation of the transition probability distribution of the Wright-Fisher model at a far smaller computational cost. Above all, our moment-based approximations are expected to improve the performance of the existing statistical frameworks for the inference of natural selection from time series genetic data.

The remainder of this work is organised as follows. In Section 2, we provide a brief review of the Wright-Fisher model for a population of diploid individuals evolving subject to natural selection at two linked loci. In Section 3, we introduce two moment-based approximations for the two-locus Wright-Fisher model with selection, and then we evaluate the performance of our moment-based approximations in Section 4. Finally, we summarise our main results and discuss possible applications and extensions in Section 5.

2. Wright-Fisher model

We consider a population of N randomly mating diploid individuals at a pair of linked loci, labelled $\mathcal{A}$ and $\mathcal{B}$, evolving under natural selection with discrete and nonoverlapping generations. Suppose that loci $\mathcal{A}$ and $\mathcal{B}$ 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). At each locus,

there are two possible allele types, labelled $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{B}_1$, $\mathcal{B}_2$, respectively, 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. Here 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 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} be 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 over time from the time that the mutant allele was created. 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 carrying the $\mathcal{A}_1\mathcal{B}_2/\mathcal{A}_2\mathcal{B}_1$ genotype is $w_{23} = (1-h_{\mathcal{A}}s_{\mathcal{A}})(1-h_{\mathcal{B}}s_{\mathcal{B}})$.

Let $X_i(k)$ 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 $\boldsymbol{X}(k)$ be the vector of the frequencies of the four possible gametes. In the Wright-Fisher model, the population size N is assumed to be fixed over time, and at each generation gametes are randomly sampled from an effectively infinite gamete pool reflecting the parental gamete frequencies. Under multinomial sampling, we have

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

where $\boldsymbol{p}$ 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. The sampling probabilities $\boldsymbol{p}$ can be written down as a function of the gamete frequencies $\boldsymbol{x}$,

$$\begin{aligned}
p_1 &= (1-r)q_1 + r(q_1+q_2)(q_1+q_3) \\
p_2 &= (1-r)q_2 + r(q_1+q_2)(q_2+q_4) \\
p_3 &= (1-r)q_3 + r(q_3+q_4)(q_1+q_3) \\
p_4 &= (1-r)q_4 + r(q_3+q_4)(q_2+q_4),
\end{aligned} \tag{2}$$

where

$$q_i = \frac{\bar{w}_i}{\bar{w}} x_i = \frac{\sum_{j=1}^{4} w_{ij} x_j}{\sum_{i=1}^{4} \sum_{j=1}^{4} w_{ij} x_j x_i} x_i \tag{3}$$

for $i = 1, 2, 3, 4$. In Eq. (3), the numerator $\bar{w}_i$ is the marginal fitness of haplotype 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 $\boldsymbol{X} = \{\boldsymbol{X}(k), k \in \mathbb{N}\}$ evolving in the state space

$$\Omega_{\boldsymbol{X}} = \left\{ \boldsymbol{x} \in \left\{ 0, \frac{1}{2N}, \ldots, 1 \right\}^4 : \sum_{i=1}^{4} x_i = 1 \right\}$$

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

3. Moment-based approximations

In the Wright-Fisher model, the one-step transition probability distribution can be readily obtained with Eqs. (1)-(3), but it is intractable to write down an explicit formula for the k-step transition probability distribution even though k is small. Numerical evaluation of the k-step transition probability distribution is theoretically feasible but computationally prohibitive for large population sizes and evolutionary timescales so approximations to the Wright-Fisher model based on parametric continuous probability distributions have been proposed. However, existing moment-based approximations suffer from either the distribution support issue (*e.g.*, Lacerda & Seoighe, 2014; Terhorst et al., 2015) or the limitation in applicability to the scenario of natural selection acting on multiple linked loci (*e.g.*, Lacerda & Seoighe, 2014; Paris et al., 2019). To address these issues, we introduce two moment-based approximations for the two-locus Wright-Fisher model with selection in this section.

3.1. Moment approximations

The moment-based approximation of the Wright-Fisher model can be traced back to Cavalli-Sforza & Edwards (1967). The principle of the moment-based approximation is to approximate the transition probability distribution of the Wright-Fisher model with a parametric continuous

probability distribution, matching its first two moments to those of the Wright-Fisher model. This method was first extended by Lacerda & Seoighe (2014) to incorporate natural selection and applied with a normal distribution to the inference of natural selection from time series data of allele frequencies and then employed in several other studies (*e.g.*, Terhorst et al., 2015; Paris et al., 2019). The prerequisite of this procedure is that we can calculate the mean and (co)variance of the Wright-Fisher model at any given time point. Terhorst et al. (2015) proposed a recursion to approximate the first two moments of the Wright-Fisher model, which is the only existing method can be applied to the case of natural selection acting on multiple linked loci. Here we separately extend other two moment approximations developed in Lacerda & Seoighe (2014) and Paris et al. (2019) to the two-locus Wright-Fisher model with selection.

From Eq. (1), we can formulate the mean and (co)variance of the gamete frequencies $\boldsymbol{X}(k)$ given the gamete frequencies $\boldsymbol{X}(k-1) = \boldsymbol{x}$ as

$$\begin{aligned}
\boldsymbol{\mu}(\boldsymbol{x}) &= \mathrm{E}\left[\boldsymbol{X}(k) \mid \boldsymbol{X}(k-1) = \boldsymbol{x}\right] \\
&= \boldsymbol{p} \qquad (4) \\
\boldsymbol{\Sigma}(\boldsymbol{x}) &= \mathrm{Var}\left[\boldsymbol{X}(k) \mid \boldsymbol{X}(k-1) = \boldsymbol{x}\right] \\
&= \frac{1}{2N}\left(\mathrm{diag}(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^{\intercal}\right), \qquad (5)
\end{aligned}$$

respectively, where $\mathrm{diag}(\boldsymbol{p})$ is the operator that constructs a diagonal matrix and places the elements of the sampling probability vector $\boldsymbol{p}$ on the main diagonal.

Let $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ denote the mean and (co)variance of the gamete frequencies $\boldsymbol{X}(k)$ given the gamete frequencies $\boldsymbol{X}(0)$ for $k \in \mathbb{N}^+$, respectively. Using the law of total mean and (co)variance, we have

$$\begin{aligned}
\boldsymbol{\mu}_k &= \mathrm{E}\left[\boldsymbol{X}(k) \mid \boldsymbol{X}(0)\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[\boldsymbol{X}(k) \mid \boldsymbol{X}(k-1)\right] \mid \boldsymbol{X}(0)\right] \\
&= \mathrm{E}\left[\boldsymbol{\mu}(\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right] \qquad (6) \\
\boldsymbol{\Sigma}_k &= \mathrm{Var}\left[\boldsymbol{X}(k) \mid \boldsymbol{X}(0)\right] \\
&= \mathrm{E}\left[\mathrm{Var}\left[\boldsymbol{X}(k) \mid \boldsymbol{X}(k-1)\right] \mid \boldsymbol{X}(0)\right] + \mathrm{Var}\left[\mathrm{E}\left[\boldsymbol{X}(k) \mid \boldsymbol{X}(k-1)\right] \mid \boldsymbol{X}(0)\right] \\
&= \mathrm{E}\left[\boldsymbol{\Sigma}(\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right] + \mathrm{Var}\left[\boldsymbol{\mu}(\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right]. \qquad (7)
\end{aligned}$$

Following Lacerda & Seoighe (2014), we apply the delta method with the first-order Taylor expansion of the term $\boldsymbol{\mu}(\boldsymbol{X}(k-1))$ about the mean $\boldsymbol{\mu}_{k-1}$, and then the mean $\boldsymbol{\mu}_k$ and the (co)variance $\boldsymbol{\Sigma}_k$ of the Wright-Fisher model $\boldsymbol{X}$ can be approximated as

$$\boldsymbol{\mu}_k \approx \boldsymbol{\mu}(\boldsymbol{\mu}_{k-1}) \tag{8}$$

$$\boldsymbol{\Sigma}_k \approx \boldsymbol{\Sigma}(\boldsymbol{\mu}_{k-1}) + \boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})\boldsymbol{\Sigma}_{k-1}\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})^{\intercal}, \tag{9}$$

where $\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{x}) = (\frac{\partial f_i}{\partial x_j})$ is the Jacobian matrix defined over the vector function $\boldsymbol{f}(\boldsymbol{x})$.

Alternatively, following Paris et al. (2019), we substitute Eqs. (4) and (5) into Eq. (7), and then the (co)variance $\boldsymbol{\Sigma}_k$ of the Wright-Fisher model $\boldsymbol{X}$ can be expressed as

$$\boldsymbol{\Sigma}_k = \frac{1}{2N}\,\mathrm{E}\left[\mathrm{diag}(\boldsymbol{\mu}(\boldsymbol{X}(k-1))) - \boldsymbol{\mu}(\boldsymbol{X}(k-1))\boldsymbol{\mu}(\boldsymbol{X}(k-1))^{\intercal} \mid \boldsymbol{X}(0)\right] + \mathrm{Var}\left[\boldsymbol{\mu}(\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right].$$

With the first-order Taylor expansion of the term $\boldsymbol{\mu}(\boldsymbol{X}(k-1))$ about the mean $\boldsymbol{\mu}_{k-1}$, we can approximate the mean $\boldsymbol{\mu}_k$ and the (co)variance $\boldsymbol{\Sigma}_k$ of the Wright-Fisher model $\boldsymbol{X}$ as

$$\boldsymbol{\mu}_k \approx \boldsymbol{\mu}(\boldsymbol{\mu}_{k-1}) \tag{10}$$

$$\boldsymbol{\Sigma}_k \approx \frac{1}{2N}\left(\mathrm{diag}(\boldsymbol{\mu}_k) - \boldsymbol{\mu}_k\boldsymbol{\mu}_k^{\intercal}\right) + \left(1 - \frac{1}{2N}\right)\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})\boldsymbol{\Sigma}_{k-1}\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})^{\intercal}. \tag{11}$$

With the second-order Taylor expansion of the term $\boldsymbol{\mu}(\boldsymbol{X}(k-1))$ about the mean $\boldsymbol{\mu}_{k-1}$, we can approximate the mean $\boldsymbol{\mu}_k$ and the (co)variance $\boldsymbol{\Sigma}_k$ of the Wright-Fisher model $\boldsymbol{X}$ as

$$\boldsymbol{\mu}_k \approx \boldsymbol{\mu}(\boldsymbol{\mu}_{k-1}) + \frac{1}{2}\sum_{i,j}\frac{\partial^2\boldsymbol{\mu}}{\partial x_i \partial x_j}(\boldsymbol{\mu}_{k-1})(\boldsymbol{\Sigma}_{k-1})_{ij} \tag{12}$$

$$\boldsymbol{\Sigma}_k \approx \frac{1}{2N}\left(\mathrm{diag}(\boldsymbol{\mu}_k) - \boldsymbol{\mu}_k\boldsymbol{\mu}_k^{\intercal}\right) + \left(1 - \frac{1}{2N}\right)\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})\boldsymbol{\Sigma}_{k-1}\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})^{\intercal}. \tag{13}$$

Substituting Eqs. (2) and (3) into Eq. (4) and taking the derivative with respect to the gamete frequencies $\boldsymbol{x}$, we can write down the Jacobian matrix defined over the term $\boldsymbol{\mu}(\boldsymbol{x})$ in Eq. (4) as

$$\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{x}) = (\boldsymbol{I} + r\boldsymbol{Q})\left(\mathrm{diag}(\boldsymbol{x})\left(\frac{\boldsymbol{w}}{\bar{w}} - 2\left(\frac{\bar{\boldsymbol{w}}}{\bar{w}}\right)\left(\frac{\bar{\boldsymbol{w}}}{\bar{w}}\right)^{\intercal}\right) + \mathrm{diag}\left(\frac{\bar{\boldsymbol{w}}}{\bar{w}}\right)\right)$$

with

$$\boldsymbol{Q} = \begin{pmatrix} q_1 - q_4 & q_1 + q_3 & q_1 + q_2 & q_1 - q_1 \\ q_2 + q_4 & q_2 - q_3 & q_2 - q_2 & q_2 + q_1 \\ q_3 + q_4 & q_3 - q_3 & q_3 - q_2 & q_3 + q_1 \\ q_4 - q_4 & q_4 + q_3 & q_4 + q_2 & q_4 - q_1 \end{pmatrix},$$

where $\boldsymbol{w} = (w_{ij})$ is the viability matrix for all genotypes, and $\bar{\boldsymbol{w}} = (\bar{w}_i)$ is the marginal viability vector for all haplotypes.

Iterating the recursion formula in Eqs. (8) and (9) or in Eqs. (10) and (11) or in Eqs. (12) and (13), we can obtain the approximations of the first two moments of the two-locus Wright-Fisher model with selection, $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$, for $k \in \mathbb{N}$ with the starting values of the mean $\boldsymbol{\mu}_0 = \mathrm{E}[\boldsymbol{X}(0)]$ and the (co)variance $\boldsymbol{\Sigma}_0 = \mathrm{Var}[\boldsymbol{X}(0)]$. We present the detailed derivation of the extension of the two moment approximations developed by Lacerda & Seoighe (2014) and Paris et al. (2019) for the two-locus Wright-Fisher model with selection in Appendix A, where the detailed derivation of the moment approximation of Terhorst et al. (2015) can also be found.

3.2. Logistic normal approximation

Given the approximations for the first two moments of the two-locus Wright-Fisher model with selection, our problem reduces to how to find a proper parametric continuous probability distribution to fit the transition probability distribution of the Wright-Fisher model. The first candidate is a normal distribution, which has already been successfully applied in population genetics studies (*e.g.*, Nicholson et al., 2002; Lacerda & Seoighe, 2014; Terhorst et al., 2015).

Using a normal distribution, we have that the gamete frequencies $\boldsymbol{X}(k)$ given the gamete frequencies $\boldsymbol{X}(0)$ is approximately distributed with

$$\boldsymbol{X}(k) \mid \boldsymbol{X}(0) \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \tag{14}$$

where $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ denote the mean and (co)variance of the two-locus Wright-Fisher model with selection, respectively, which can be obtained by iterating the recursion formula in Eqs. (8) and (9) or in Eqs. (10) and (11) or in Eqs. (12) and (13) with the initial values $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$. The normal approximation is motivated by matching the first two moments of the normal distribution to those of the Wright-Fisher model. However unfortunately, the normal approximation in Eq. (14) can take the values outside the state space $\Omega_{\boldsymbol{X}}$, which makes the normal approximation poor

once the gamete frequencies $\boldsymbol{X}(k)$ get close to the boundaries of the state space $\Omega_{\boldsymbol{X}}$. The same distribution support issue can be found in Terhorst et al. (2015). In the single-locus case, to narrow the state space of the normal approximation, Nicholson et al. (2002) allocated the mass of the the normal distribution outside the state space of the Wright-Fisher model to the discrete fixation probabilities at boundaries 0 and 1. Alternatively, Paris et al. (2019) simply rejected the values outside the state space of the Wright-Fisher model and appropriately rescale the normal distribution. However, both of their procedures lead to the mean and (co)variance of the resulting probability distribution that no longer match those of the Wright-Fisher model.

To tackle the distribution support issue of the normal approximation, we construct a transformation, denoted by $\boldsymbol{\phi}$, to map the gamete frequencies $\boldsymbol{X}(k)$ to the state space $\mathbb{R}^3$:

$$\boldsymbol{\phi}(\boldsymbol{X}(k)) = \left(\log\left(\frac{X_1(k)}{X_4(k)}\right), \log\left(\frac{X_2(k)}{X_4(k)}\right), \log\left(\frac{X_3(k)}{X_4(k)}\right)\right)^{\intercal}$$

and introduce a stochastic process, denoted by $\boldsymbol{Y} = \{\boldsymbol{Y}(k), k \in \mathbb{N}\}$, to approximate the transformed Wright-Fisher model $\boldsymbol{\phi}(\boldsymbol{X})$, where we let $\boldsymbol{Y}(k) = (Y_1(k), Y_2(k), Y_3(k))^{\intercal} \in \mathbb{R}^3$ follow a normal distribution

$$\boldsymbol{Y}(k) \mid \boldsymbol{Y}(0) \sim \mathcal{N}(\boldsymbol{m}_k, \boldsymbol{V}_k), \tag{15}$$

where $\boldsymbol{m}_k$ and $\boldsymbol{V}_k$ are the mean and (co)variance of the normal distribution in Eq. (15). The unique inverse mapping given by

$$\boldsymbol{\phi}^{-1}(\boldsymbol{Y}(k)) = \left(\frac{e^{Y_1(k)}}{1+\sum_{i=1}^{3} e^{Y_i(k)}}, \frac{e^{Y_2(k)}}{1+\sum_{i=1}^{3} e^{Y_i(k)}}, \frac{e^{Y_3(k)}}{1+\sum_{i=1}^{3} e^{Y_i(k)}}, \frac{1}{1+\sum_{i=1}^{3} e^{Y_i(k)}}\right)^{\intercal}$$

yields our moment-based approximation of the Wright-Fisher model $\boldsymbol{X}$, where the inverse mapping $\boldsymbol{\phi}^{-1}$ is commonly known as the additive logistic transformation. Conditional on the gamete frequencies $\boldsymbol{X}(0)$, our moment-based approximation $\boldsymbol{\phi}^{-1}(\boldsymbol{Y}(k))$ suggests that the gamete frequencies $\boldsymbol{X}(k)$ approximately follows a logistic normal distribution

$$\boldsymbol{X}(k) \mid \boldsymbol{X}(0) \sim P(\mathcal{N}(\boldsymbol{m}_k, \boldsymbol{V}_k)), \tag{16}$$

which we refer to as the logistic normal approximation of the two-locus Wright-Fisher model with selection.

Our remaining task is to find proper values for the mean $\boldsymbol{m}_k$ and the (co)variance $\boldsymbol{V}_k$ of the normal distribution in Eq. (15) such that the logistic normal distribution in Eq. (16) matches its mean and (co)variance to those of the Wright-Fisher model $\boldsymbol{X}$. Unfortunately, to our knowledge, there is no analytical solution for the mean and (co)variance of the logistic normal distribution, but we can apply the delta method to approximate the mean $\boldsymbol{m}_k$ and the (co)variance $\boldsymbol{V}_k$ of the normal distribution with the mean $\boldsymbol{\mu}_k$ and the (co)variance $\boldsymbol{\Sigma}_k$ of the Wright-Fisher model $\boldsymbol{X}$. More specifically, using the delta method with the first-order Taylor expansion of the term $\boldsymbol{\phi}(\boldsymbol{X}(k))$ about the mean $\boldsymbol{\mu}_k$, we have

$$\begin{aligned}
\boldsymbol{m}_k &= \mathrm{E}\left[\boldsymbol{\phi}(\boldsymbol{X}(k)) \mid \boldsymbol{X}(0)\right] \\
&\approx \boldsymbol{\phi}(\boldsymbol{\mu}_k) \qquad (17) \\
\boldsymbol{V}_k &= \mathrm{Var}\left[\boldsymbol{\phi}(\boldsymbol{X}(k)) \mid \boldsymbol{X}(0)\right] \\
&\approx \boldsymbol{J}_{\boldsymbol{\phi}}(\boldsymbol{\mu}_k)\boldsymbol{\Sigma}_k\boldsymbol{J}_{\boldsymbol{\phi}}(\boldsymbol{\mu}_k)^{\intercal}, \qquad (18)
\end{aligned}$$

where

$$\boldsymbol{J}_{\boldsymbol{\phi}}(\boldsymbol{x}) = \begin{pmatrix} \frac{1}{x_1} & 0 & 0 & -\frac{1}{x_4} \\ 0 & \frac{1}{x_2} & 0 & -\frac{1}{x_4} \\ 0 & 0 & \frac{1}{x_3} & -\frac{1}{x_4} \end{pmatrix}.$$

It should be noted that the mean $\boldsymbol{\mu}_k$ and the (co)variance $\boldsymbol{\Sigma}_k$ of the Wright-Fisher model $\boldsymbol{X}$ are readily available by iterating the recursion formula in Eqs. (8) and (9) or in Eqs. (10) and (11) or in Eqs. (12) and (13) with the initial values $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$.

By substituting Eqs. (17) and (18) into Eq. (16), our logistic normal approximation of the two-locus Wright-Fisher model with selection can therefore be represented as

$$\boldsymbol{X}(k) \mid \boldsymbol{X}(0) \sim P(\mathcal{N}(\boldsymbol{\phi}(\boldsymbol{\mu}_k), \boldsymbol{J}_{\boldsymbol{\phi}}(\boldsymbol{\mu}_k)\boldsymbol{\Sigma}_k\boldsymbol{J}_{\boldsymbol{\phi}}(\boldsymbol{\mu}_k)^{\intercal})).$$

The transition probability density function of our logistic normal approximation can then be explicitly written down with the change-of-variable technique. Notice that the first two moments of our logistic normal approximation are no longer exactly matched to those of the Wright-Fisher model since the delta method is used to approximate the mean $\boldsymbol{m}_k$ and the (co)variance $\boldsymbol{V}_k$ of the normal distribution in Eqs. (17) and (18) (see Figure 1 for an illustration). In Section 4, we

will explore how such a discrepancy affect the performance of our logistic normal approximation through extensive simulations.

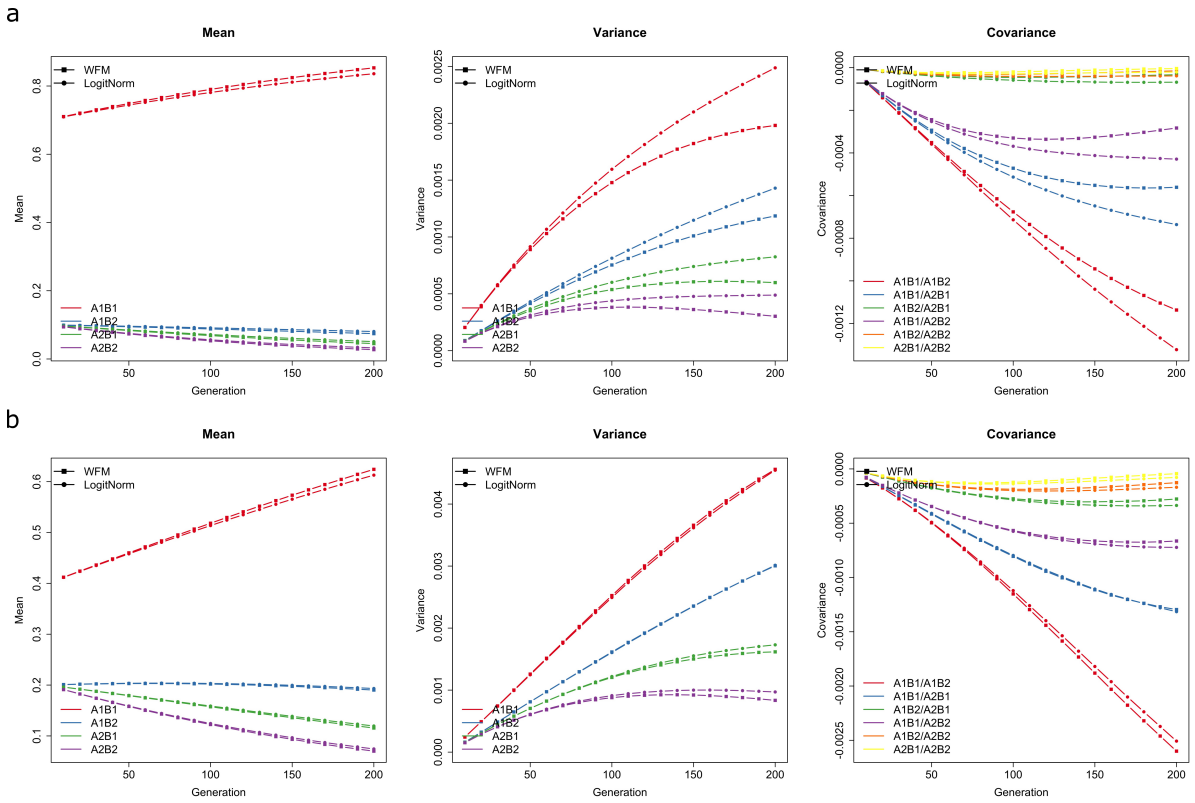

Figure 1: An illustration for the mean, variance and covariance of the Wright-Fisher model and our logistic normal approximation, where we set the selection coefficients $s_{\mathcal{A}} = 0.01$ and $s_{\mathcal{B}} = 0.005$, the dominance parameters $h_{\mathcal{A}} = 0.5$ and $s_{\mathcal{B}} = 0.5$, the recombination rate $r = 0.00001$ and the population size $N = 5000$. We pick the initial gamete frequencies (a) $\boldsymbol{X}_0 = (0.7, 0.1, 0.1, 0.1)$ and (b) $\boldsymbol{X}_0 = (0.2, 0.2, 0.3, 0.3)$, respectively. The first two moments of the Wright-Fisher model and the logistic normal approximation are estimated from 1,000,000 replicated simulations.

3.3. Hierarchical beta approximation

To our knowledge, except for the normal distribution, the beta distribution is an alternative candidate of the parametric continuous probability distribution for the moment-based approximation of the Wright-Fisher model, which has been shown to deliver superior performance in comparison to the normal distribution for single-locus problems (see Tataru et al., 2017) and has also been successfully applied in a number of population genetics studies (*e.g.*, Tataru et al., 2015; Paris et al., 2019). For two-locus problems, we need a multivariate generalisation of the beta distribution to be the parametric continuous probability distribution, but unfortunately, the multivariate beta distribution, commonly known as the Dirichlet distribution, is not appropriate since its mean and (co)variance fail to match those of the two-locus Wright-Fisher model

with selection. More specifically, the Dirichlet distribution of order four has four parameters, three of which are used to determine the mean of the Wright-Fisher model, leaving one free parameter to model the (co)variance. However, a single parameter is not enough to capture the structure of the (co)variance of the Wright-Fisher model.

Inspired by Hobolth & Siren (2016), we decompose the two-locus Wright-Fisher model with selection into two conditionally independent beta distributed levels, which is an alternative multivariate generalisation of the beta distribution, called the hierarchical beta distribution. More specifically, we construct a transformation, denoted by $\boldsymbol{\psi}$, to map the gamete frequencies $\boldsymbol{X}(k)$ to the state space $[0,1]^3$:

$$\boldsymbol{\psi}(\boldsymbol{X}(k)) = \left(X_1(k) + X_2(k), \frac{X_1(k)}{X_1(k) + X_2(k)}, \frac{X_3(k)}{X_3(k) + X_4(k)} \right)^{\intercal}$$

and introduce a stochastic process, denoted by $\boldsymbol{Z} = \{\boldsymbol{Z}(k), k \in \mathbb{N}\}$, to approximate the transformed Wright-Fisher model $\boldsymbol{\psi}(\boldsymbol{X})$, where we let $\boldsymbol{Z}(k) = (Z_1(k), Z_2(k), Z_3(k))^{\intercal} \in [0,1]^3$ follow a hierarchical beta distribution

$$Z_1(k) \mid \boldsymbol{Z}(0) \sim \text{Beta}(\alpha_{1,k}, \beta_{1,k}) \tag{19}$$

$$Z_2(k) \mid Z_1(k), \boldsymbol{Z}(0) \sim \text{Beta}(\alpha_{2,k}, \beta_{2,k}) \tag{20}$$

$$Z_3(k) \mid Z_1(k), \boldsymbol{Z}(0) \sim \text{Beta}(\alpha_{3,k}, \beta_{3,k}), \tag{21}$$

where $\alpha_{i,k}$ and $\beta_{i,k}$ are the shape parameters of each beta distribution in Eqs. (19)-(21). Notice that $Z_1(k)$ is the frequency of the $\mathcal{A}_1$ allele, $Z_2(k)$ is the frequency of the $\mathcal{A}_1\mathcal{B}_1$ haplotype in the haplotypes with the $\mathcal{A}_1$ allele, and $Z_3(k)$ is the frequency of the $\mathcal{A}_2\mathcal{B}_1$ haplotype in the haplotypes with the $\mathcal{A}_2$ allele. The unique inverse mapping given by

$$\boldsymbol{\psi}^{-1}(\boldsymbol{Z}(k)) = (Z_1(k)Z_2(k), Z_1(k)(1 - Z_2(k)), (1 - Z_1(k))Z_3(k), (1 - Z_1(k))(1 - Z_3(k)))^{\intercal}$$

yields our moment-based approximation of the Wright-Fisher model $\boldsymbol{X}$, which we refer to as the hierarchical beta approximation of the two-locus Wright-Fisher model with selection.

Our task now is to find proper values for the shape parameters $\alpha_{i,k}$ and $\beta_{i,k}$ of each beta distribution in Eqs. (19)-(21) such that the probability distribution for our hierarchical beta

approximation $\boldsymbol{\psi}^{-1}(\boldsymbol{Z}(k))$ matches its mean and (co)variance to those of the Wright-Fisher model $\boldsymbol{X}$. Since the shape parameters of the beta distribution can be expressed in terms of its mean and variance as

$$\alpha_{i,k} = \left(\frac{m_{i,k}(1 - m_{i,k})}{V_{i,k}} - 1 \right) m_{i,k} \tag{22}$$

$$\beta_{i,k} = \left(\frac{m_{i,k}(1 - m_{i,k})}{V_{i,k}} - 1 \right) (1 - m_{i,k}) \tag{23}$$

for $i = 1, 2, 3$, where $m_{i,k}$ and $V_{i,k}$ denote the mean and variance of each beta distribution in Eqs. (19)-(21), our problem reduces to the calculation of the mean $m_{i,k}$ and the variance $V_{i,k}$ of each beta distribution with the mean $\boldsymbol{\mu}_k$ and the (co)variance $\boldsymbol{\Sigma}_k$ of the Wright-Fisher model $\boldsymbol{X}$. Note that the mean $\boldsymbol{\mu}_k$ and the (co)variance $\boldsymbol{\Sigma}_k$ of the Wright-Fisher model $\boldsymbol{X}$ are readily available by iterating the recursion formula in Eqs. (8) and (9) or in Eqs. (10) and (11) or in Eqs. (12) and (13) with the initial values $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$.

Using the law of total mean, we can represent the mean of the frequency of the $\mathcal{A}_1\mathcal{B}_1$ gamete under the Wright-Fisher model $\boldsymbol{X}$ in terms of the mean $m_{i,k}$ of each beta distribution as

$$\begin{aligned}
\mu_{1,k} &= \mathrm{E}\left[Z_1(k)Z_2(k) \mid \boldsymbol{Z}(0)\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[Z_1(k)Z_2(k) \mid Z_1(k), \boldsymbol{Z}(0)\right] \mid \boldsymbol{Z}(0)\right] \\
&= \mathrm{E}\left[Z_1(k)m_{2,k} \mid \boldsymbol{Z}(0)\right] \\
&= m_{1,k}m_{2,k}.
\end{aligned} \tag{24}$$

Similarly, the mean of the other three gamete frequencies can be written down as

$$\mu_{2,k} = m_{1,k}(1 - m_{2,k}) \tag{25}$$

$$\mu_{3,k} = (1 - m_{1,k})m_{3,k} \tag{26}$$

$$\mu_{4,k} = (1 - m_{1,k})(1 - m_{3,k}). \tag{27}$$

Solving the system of coupled equations given by Eqs. (24)-(26) as the equations for the mean $\mu_{i,k}$ of each gamete frequency under the Wright-Fisher model in terms of the mean $m_{i,k}$ of each

beta distribution, we have

$$m_{1,k} = \mu_{1,k} + \mu_{2,k} \tag{28}$$

$$m_{2,k} = \frac{\mu_{1,k}}{\mu_{1,k} + \mu_{2,k}} \tag{29}$$

$$m_{3,k} = \frac{\mu_{3,k}}{\mu_{3,k} + \mu_{4,k}}. \tag{30}$$

Note that the mean $m_{i,k}$ of each beta distribution in Eqs. (28)-(30) also satisfies Eq. (27).

With the law of total variance, the variance of the frequency of the $\mathcal{A}_1\mathcal{B}_1$ gamete under the Wright-Fisher model $\boldsymbol{X}$ can be represented in terms of the mean $m_{i,k}$ and the variance $V_{i,k}$ of each beta distribution as

$$\begin{aligned}
\Sigma_{11,k} &= \mathrm{Var}\left[Z_1(k)Z_2(k) \mid \boldsymbol{Z}(0)\right] \\
&= \mathrm{Var}\left[\mathrm{E}\left[Z_1(k)Z_2(k) \mid Z_1(k), \boldsymbol{Z}(0)\right] \mid \boldsymbol{Z}(0)\right] + \mathrm{E}\left[\mathrm{Var}\left[Z_1(k)Z_2(k) \mid Z_1(k), \boldsymbol{Z}(0)\right] \mid \boldsymbol{Z}(0)\right] \\
&= \mathrm{Var}\left[Z_1(k)m_{2,k} \mid \boldsymbol{Z}(0)\right] + \mathrm{E}\left[Z_1^2(k)V_{2,k} \mid \boldsymbol{Z}(0)\right] \\
&= V_{1,k}m_{2,k}^2 + (V_{1,k} + m_{1,k}^2)V_{2,k}.
\end{aligned} \tag{31}$$

Similarly, the variance of the other three gamete frequencies can be expressed as

$$\Sigma_{22,k} = V_{1,k}(1 - m_{2,k})^2 + (V_{1,k} + m_{1,k}^2)V_{2,k} \tag{32}$$

$$\Sigma_{33,k} = V_{1,k}m_{3,k}^2 + (V_{1,k} + (1 - m_{1,k})^2)V_{3,k} \tag{33}$$

$$\Sigma_{44,k} = V_{1,k}(1 - m_{3,k})^2 + (V_{1,k} + (1 - m_{1,k})^2)V_{3,k}. \tag{34}$$

Solving the system of coupled equations given by Eqs. (31)-(33) as the equations for the variance $\Sigma_{ii,k}$ of each gamete frequency under the Wright-Fisher model in terms of the mean $m_{i,k}$ and the variance $V_{i,k}$ of each beta distribution, we have

$$V_{1,k} = \frac{\Sigma_{11,k} - \Sigma_{22,k}}{m_{2,k}^2 - (1 - m_{2,k})^2} \tag{35}$$

$$V_{2,k} = \frac{-(1 - m_{2,k})^2\Sigma_{11,k} + m_{2,k}^2\Sigma_{22,k}}{m_{1,k}^2(m_{2,k}^2 - (1 - m_{2,k})^2) + (\Sigma_{11,k} - \Sigma_{22,k})} \tag{36}$$

$$V_{3,k} = \frac{-m_{3,k}^2\Sigma_{11,k} + m_{3,k}^2\Sigma_{22,k} + (m_{2,k}^2 - (1 - m_{2,k})^2)\Sigma_{33,k}}{(1 - m_{1,k})^2(m_{2,k}^2 - (1 - m_{2,k})^2) + (\Sigma_{11,k} - \Sigma_{22,k})}, \tag{37}$$

where $m_{i,k}$ is the mean of each beta distribution given by Eqs. (28)-(30). Note that the variance $V_{i,k}$ of each beta distribution in Eqs. (35)-(37) are not quite guaranteed to satisfy Eq. (34).

Using Eqs. (19)-(21), we can therefore represent our hierarchical beta approximation of the two-locus Wright-Fisher model with selection as

$$
\begin{aligned}
X_1(k) + X_2(k) \mid \boldsymbol{X}(0) &\sim \mathrm{Beta}(\alpha_{1,k}, \beta_{1,k}) \\
X_1(k) \mid X_1(k) + X_2(k), \boldsymbol{X}(0) &\sim (X_1(k) + X_2(k))\,\mathrm{Beta}(\alpha_{2,k}, \beta_{2,k}) \\
X_3(k) \mid X_1(k) + X_2(k), \boldsymbol{X}(0) &\sim (1 - X_1(k) - X_2(k))\,\mathrm{Beta}(\alpha_{3,k}, \beta_{3,k}),
\end{aligned}
$$

where the shape parameters $\alpha_{i,k}$ and $\beta_{i,k}$ of each beta distribution can be obtained by substituting Eqs. (28)-(30) and (32)-(34) into Eqs. (22) and (23). The transition probability density function of our hierarchical beta approximation can then be explicitly written down through the change-of-variable technique. Notice that in our hierarchical beta approximation, its mean exactly fit those of the Wright-Fisher model, but only three of its variance are exactly matched (see Figure 2 for an illustration). This is because in our hierarchical beta approximation there is no extra free parameter leaving for fitting the one remaining variance and all covariance of the two-locus Wright-Fisher model with selection. With the law of total covariance, the covariance between four gamete frequencies under our hierarchical beta approximation can be represented in terms of the mean $m_{i,k}$ and the variance $V_{i,k}$ of each beta distribution as

$$
\begin{aligned}
\mathrm{Cov}\left[X_1(k), X_2(k) \mid \boldsymbol{X}(0)\right] &= -(V_{1,k} + m_{1,k}^2)V_{2,k} + V_{1,k}m_{2,k}(1 - m_{2,k}) \\
\mathrm{Cov}\left[X_1(k), X_3(k) \mid \boldsymbol{X}(0)\right] &= -V_{1,k}m_{2,k}m_{3,k} \\
\mathrm{Cov}\left[X_1(k), X_4(k) \mid \boldsymbol{X}(0)\right] &= -V_{1,k}m_{2,k}(1 - m_{3,k}) \\
\mathrm{Cov}\left[X_2(k), X_3(k) \mid \boldsymbol{X}(0)\right] &= -V_{1,k}(1 - m_{2,k})m_{3,k} \\
\mathrm{Cov}\left[X_2(k), X_4(k) \mid \boldsymbol{X}(0)\right] &= -V_{1,k}(1 - m_{2,k})(1 - m_{3,k}) \\
\mathrm{Cov}\left[X_3(k), X_4(k) \mid \boldsymbol{X}(0)\right] &= -(V_{1,k} + (1 - m_{1,k})^2)V_{3,k} + V_{1,k}m_{3,k}(1 - m_{3,k}).
\end{aligned}
$$

We will discuss how such a discrepancy affect the performance of our hierarchical beta approximation through extensive simulations in Section 4.

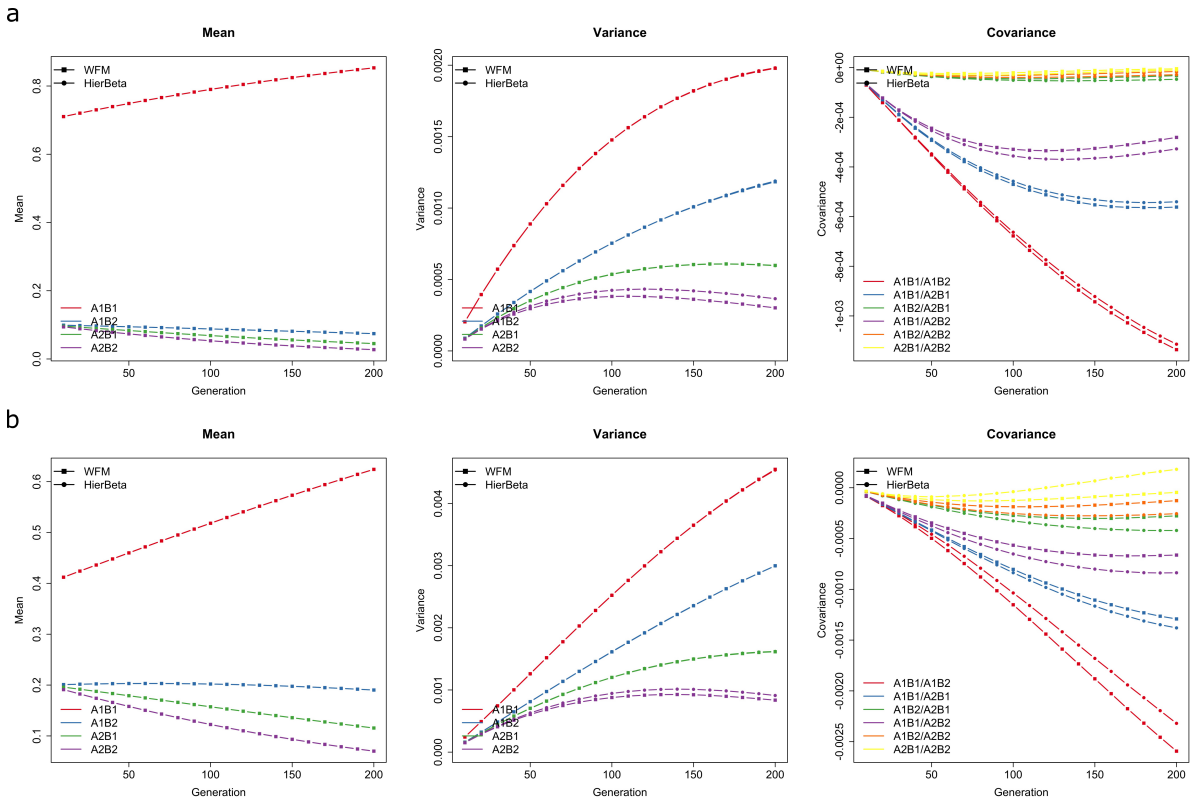

Figure 2: An illustration for the mean, variance and covariance of the Wright-Fisher model and our hierarchical beta approximation, where set the selection coefficients $s_{\mathcal{A}} = 0.01$ and $s_{\mathcal{B}} = 0.005$, the dominance parameters $h_{\mathcal{A}} = 0.5$ and $s_{\mathcal{B}} = 0.5$, the recombination rate $r = 0.00001$ and the population size $N = 5000$. We pick the initial gamete frequencies (a) $\boldsymbol{X}_0 = (0.7, 0.1, 0.1, 0.1)$ and (b) $\boldsymbol{X}_0 = (0.2, 0.2, 0.3, 0.3)$, respectively. The first two moments of the Wright-Fisher model are estimated from 1,000,000 replicated simulations.

4. Moment-based approximation performance

In this section, we use simulation studies to evaluate the performance of our approximations for the mean and (co)variance of the Wright-Fisher model and assess the performance of our logistic normal and hierarchical beta approximations for the Wright-Fisher model, including the comparison with the diffusion and normal approximations. In the following, we fix the selection coefficient $s_{\mathcal{A}} = 0.005$ and vary the selection coefficient $s_{\mathcal{B}} \in \{0, 0.001, 0.005, 0.01\}$, the recombination rate $r \in \{0.00000001, 0.00001, 0.01\}$ and the population size $N \in \{500, 5000, 50000\}$. We adopt the dominance parameters $h_{\mathcal{A}} = 0.5$ and $h_{\mathcal{B}} = 0.5$ (*i.e.*, the heterozygous fitness is the arithmetic average of the homozygous fitness, called genic selection) unless otherwise noted. In principle, the conclusions hold for any other values of the dominance parameters $h_{\mathcal{A}}$ and $h_{\mathcal{B}} \in [0, 1]$. Source code implementing all approximations and simulations described in this work is written in R with C++ through the Rcpp and RcppArmadillo packages, and available at `https://github.com/zhangyi-he/WFM-2L-Approx`.

4.1. Performance of moment approximations

To evaluate our approximations for the mean and (co)variance of the Wright-Fisher model, we perform 200 replicated runs for each of the 36 possible combinations of the selection coefficient, the recombination rate and the population size. For each run, we randomly choose the initial gamete frequencies $\boldsymbol{X}_0$ from a flat Dirichlet distribution on the three-simplex $\Omega_{\boldsymbol{X}}$ and simulate 1,000,000 gamete frequency trajectories throughout 200 generations using the Wright-Fisher model. We calculate the mean and (co)variance of the Wright-Fisher model from these 1,000,000 replicates at every 10 generations through Monte Carlo techniques. We also approximate its mean and (co)variance at every 10 generations using the approach of Terhorst et al. (2015) and our extensions of Lacerda & Seoighe (2014) and Paris et al. (2019). We apply the Euclidean vector norm to measure the difference between the mean of the Wright-Fisher model and their approximations and the Frobenius matrix norm to measure the difference between the (co)variance of the Wright-Fisher model and their approximations.

We illustrate the difference between the mean of the Wright-Fisher model and their approximations in Figure 3 and the difference between the (co)variance of the Wright-Fisher model and their approximations in Figure 4, respectively, for the case of the population size $N = 5000$. From Figures 3 and 4, we see that the error of our approximations gets smaller with the decrease in the selection coefficient and/or the recombination rate. We also observe that the error of our approximations becomes larger with longer times, *i.e.*, the larger mean error with the larger error bar. This is because the error caused by our approximations gets accumulated over time. A similar conclusion can be reached with different population sizes. See Supplemental Material, Figures S1-S4 for the comparison between the first two moments of the Wright-Fisher model and their approximations for the cases of the population size $N = 500$ and $N = 50000$.

As shown in Figure 3, we get the same performance for the mean approximation with our extensions of Lacerda & Seoighe (2014) and Paris et al. (2019) up to the first order since they are of the same form. Compared to our extensions of Lacerda & Seoighe (2014) and Paris et al. (2019) up to the first order, the method of Terhorst et al. (2015) and our extension of Paris et al. (2019) up to the second order have better performance across all 12 possible combinations of the selection coefficient and the recombination rate, which however is narrowing with the increase in the selection coefficient and/or the recombination rate. More specifically, our extension of

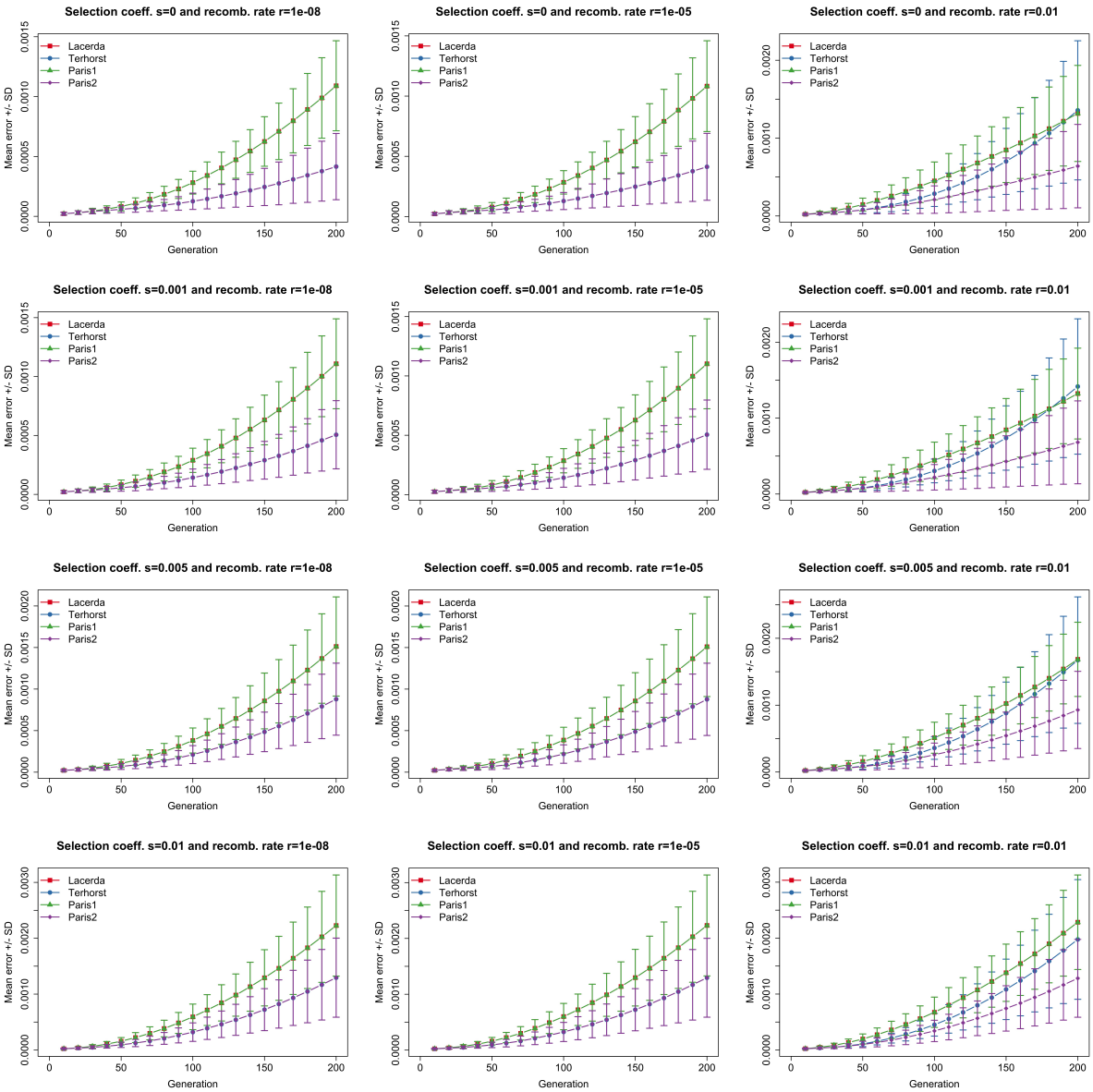

Figure 3: Comparison between the mean of the Wright-Fisher model and their approximations for the case of the population size $N = 5000$, where we fix the selection coefficient $s_{\mathcal{A}} = 0.005$ and vary the selection coefficient $s_{\mathcal{B}} \in \{0, 0.001, 0.005, 0.01\}$ and the recombination rate $r \in \{0.00000001, 0.00001, 0.01\}$.

Paris et al. (2019) up to the second order matches the performance of Terhorst et al. (2015) for tightly linked loci (see the left two panels of Figure 3) and becomes better for loosely linked loci (see the right panel of Figure 3). Although the error of all mean approximations gets larger over time, the approach of Terhorst et al. (2015) sees a much faster increase in the approximation error than other three methods when the two loci are loosely linked. Our findings are consistent with that observed in Supplemental Material, Figures S1 and S3 for the cases of the population size $N = 500$ and $N = 50000$.

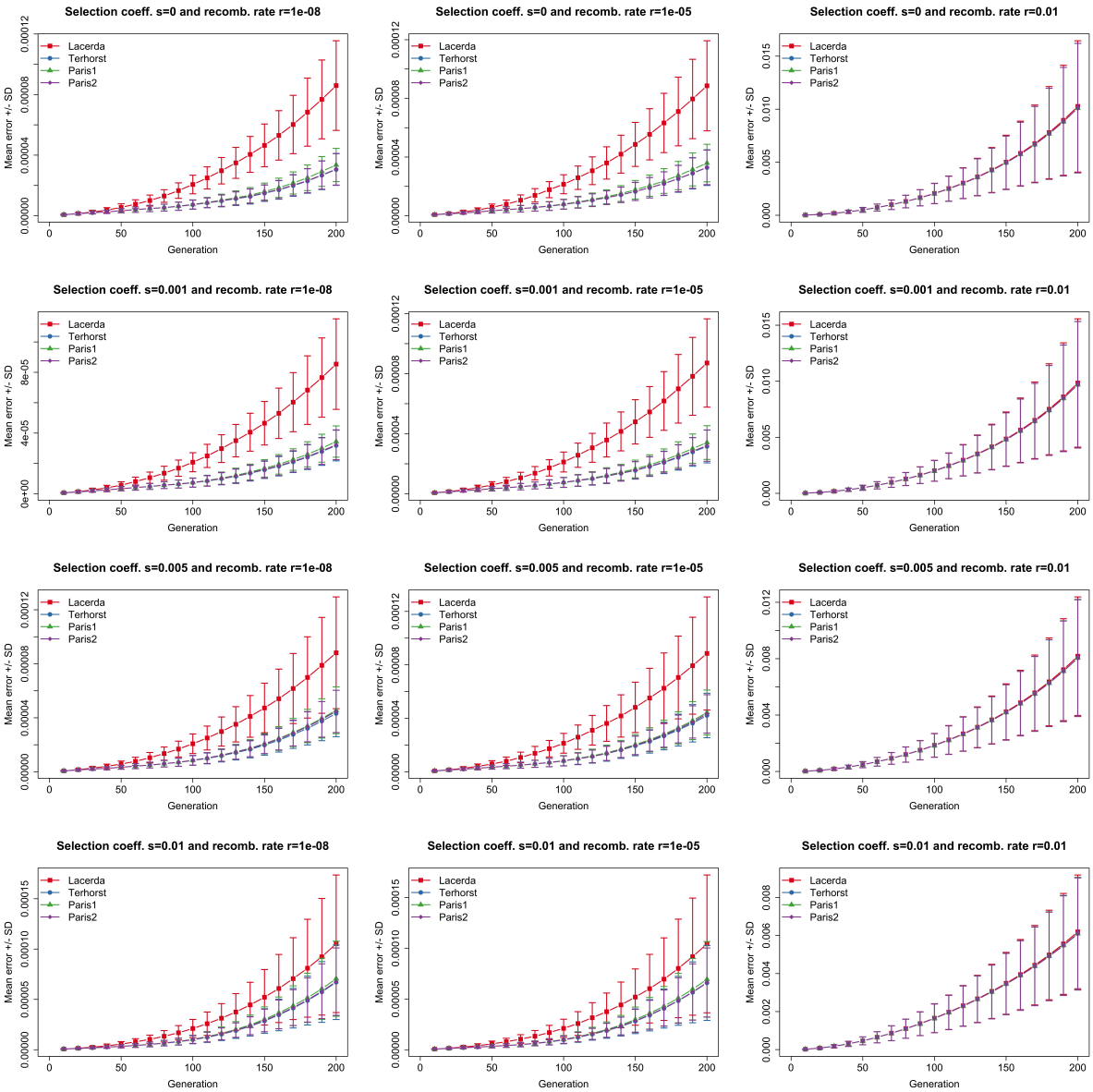

Figure 4: Comparison between the (co)variance of the Wright-Fisher model and their approximations for the case of the population size $N = 5000$, where we fix the selection coefficient $s_{\mathcal{A}} = 0.005$ and vary the selection coefficient $s_{\mathcal{B}} \in \{0, 0.001, 0.005, 0.01\}$ and the recombination rate $r \in \{0.00000001, 0.00001, 0.01\}$.

We find from Figure 4 that the approximation of Terhorst et al. (2015) and our extensions of Paris et al. (2019) have better performance than our extension of Lacerda & Seoighe (2014), but such an advantage narrows with the decrease in the selection coefficient and/or the recombination rate or even nearly diminishes for loosely linked loci (see the right panel of Figure 4). Compared with the approach of Terhorst et al. (2015), our extension of Paris et al. (2019) up to the second order matches its performance, but our extension of Paris et al. (2019) up to the first order perform slightly worse. Combining with the comparison between the (co)variance of the

Wright-Fisher model and their approximations for the cases of the population size $N = 500$ and $N = 50000$ (see Supplemental Material, Figures S2 and S4), we observe that the performance advantage of the approximation of Terhorst et al. (2015) and our extensions of Paris et al. (2019) weakens continually with the increase of the population size or even nearly diminishes.

In summary, moment approximations overall work well. Their performance gets better with the decrease in the selection coefficient and/or recombination rate but becomes worse with time. In the four moment approximations, our extension of Paris et al. (2019) up to the second order achieves the best performance across all 36 possible combinations of the selection coefficient, the recombination rate and the population size. More accurate approximations for the mean and (co)variance of the Wright-Fisher model are available if we keep third-order or even higher-order terms in the Taylor expansion in our extension of Paris et al. (2019).

4.2. Performance of moment-based approximations

To evaluate the performance of our logistic normal and hierarchical beta approximations for the Wright-Fisher model, following Hobolth & Siren (2016) we simulate the Wright-Fisher model and its logistic normal and hierarchical beta approximations and compare the generated samples. We also simulate the diffusion and normal approximations of the Wright-Fisher model to check whether our logistic normal and hierarchical beta approximations can match the performance of these two widely used approximations. More specifically, we perform 200 replicated runs for each of the 36 possible combinations of the selection coefficient, the recombination rate and the population size. For each run, we randomly generate the initial gamete frequencies $\boldsymbol{X}_0$ from a flat Dirichlet distribution over the three-simplex $\Omega_{\boldsymbol{X}}$ and simulate 1,000,000 gamete frequency trajectories throughout 200 generations using the Wright-Fisher model and its approximations, where the mean and (co)variance required in the moment-based approximation are calculated with our extension of Paris et al. (2019) up to the second order, and the diffusion approximation is simulated with the approach of He et al. (2020a). We then randomly pick 5000 vectors of the gamete frequencies from a flat Dirichlet distribution on the three-simplex $\Omega_{\boldsymbol{X}}$ and the empirical cumulative distribution function (CDF) is evaluated at these sampled gamete frequencies at every 10 generations from the 1,000,000 simulated gamete frequency trajectories. We measure the difference between the Wright-Fisher model and its approximation by the root mean squared difference (RMSD) between the two empirical CDFs.

We summarise the RMSD between the Wright-Fisher model and its approximations in Figures 5-7 for all 36 possible combinations of the selection coefficient, the recombination rate and the population size, respectively. Compared with the three moment-based approximations, the diffusion approximation shows better overall performance across all 36 possible combinations in terms of computational accuracy, *i.e.*, the lowest mean error with the smallest error bar. Note that even though some combinations of the selection coefficient, the recombination rate and the population size do not completely satisfy the assumption required in the diffusion approximation (see, *e.g.*, He et al., 2020a), the diffusion approximation still works well and outperforms the moment-based approximations. However, in terms of computational efficiency, the diffusion approximation performs significantly worse than the moment-based approximations. For example, to construct the empirical CDFs at every 10 generations in the top left plot of Figures 5, it takes approximately 392.018 seconds for each sampled initial gamete frequencies through the diffusion approximation with 1,000,000 realisations on a single core of an Intel Core i7 processor at 4.2 GHz, which only costs around 0.001 seconds to achieve the required parameters for the moment-based approximations. Additionally, at every 10 generations 1,000,000 realisations are required to be stored to compute the empirical CDFs through the diffusion approximation, but we only need the mean and (co)variance in the moment-based approximations.

As shown in Figures 5-7, we find that the moment-based approximations get better overall performance with the increase in the population size but become worse with longer times. This is consistent with the conclusion for the single locus presented in Lacerda & Seoighe (2014). In the three moment-based approximations, our logistic normal approximation shows the worst overall performance except for small population sizes. More specifically, we see from Figure 5 that our logistic normal approximation achieves the worst performance at the initial tens of generations and then gets better performance than the normal approximation with longer times for small population sizes. In general, alleles drift to loss or fixation is much faster in smaller populations, which will result in a leak of probability mass into the values of the gamete frequencies that are out of the state space $\Omega_{\boldsymbol{X}}$ due to the infinite support of the normal distribution. That is, only when the Wright-Fisher model gets close to the boundaries, our logistic normal approximation works better than the normal approximation.

Figures 5-7 also illustrate that our hierarchical beta approximation yields the best overall

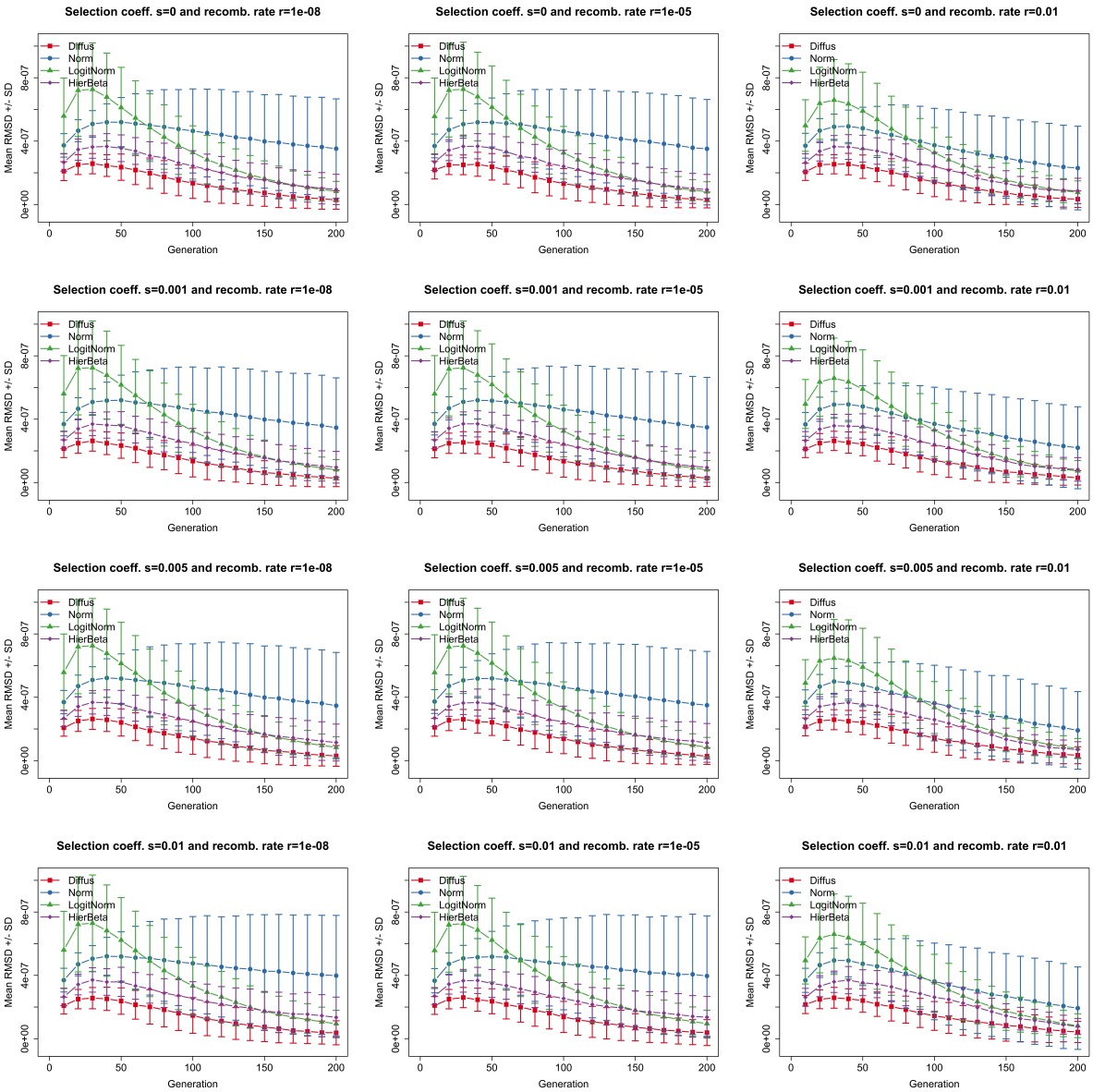

Figure 5: Comparison between the empirical CDFs of the Wright-Fisher model and their approximations for the case of the population size $N = 500$, where we fix the selection coefficient $s_{\mathcal{A}} = 0.005$ and vary the selection coefficient $s_{\mathcal{B}} \in \{0, 0.001, 0.005, 0.01\}$ and the recombination rate $r \in \{0.00000001, 0.00001, 0.01\}$.

performance among the three moment-based approximations across all 36 possible combinations of the selection coefficient, the recombination rate and the population size, which even matches the performance of the diffusion approximation for small recombination rates and large population sizes (see the left two panels of Figure 7). However, for some mean and (co)variance of the Wright-Fisher model can be fitted with a hierarchical beta distribution, *i.e.*, negative shape parameters will be produced in our hierarchical beta approximation from some mean and (co)variance of the Wright-Fisher model, which is not mathematically possible. We summarise

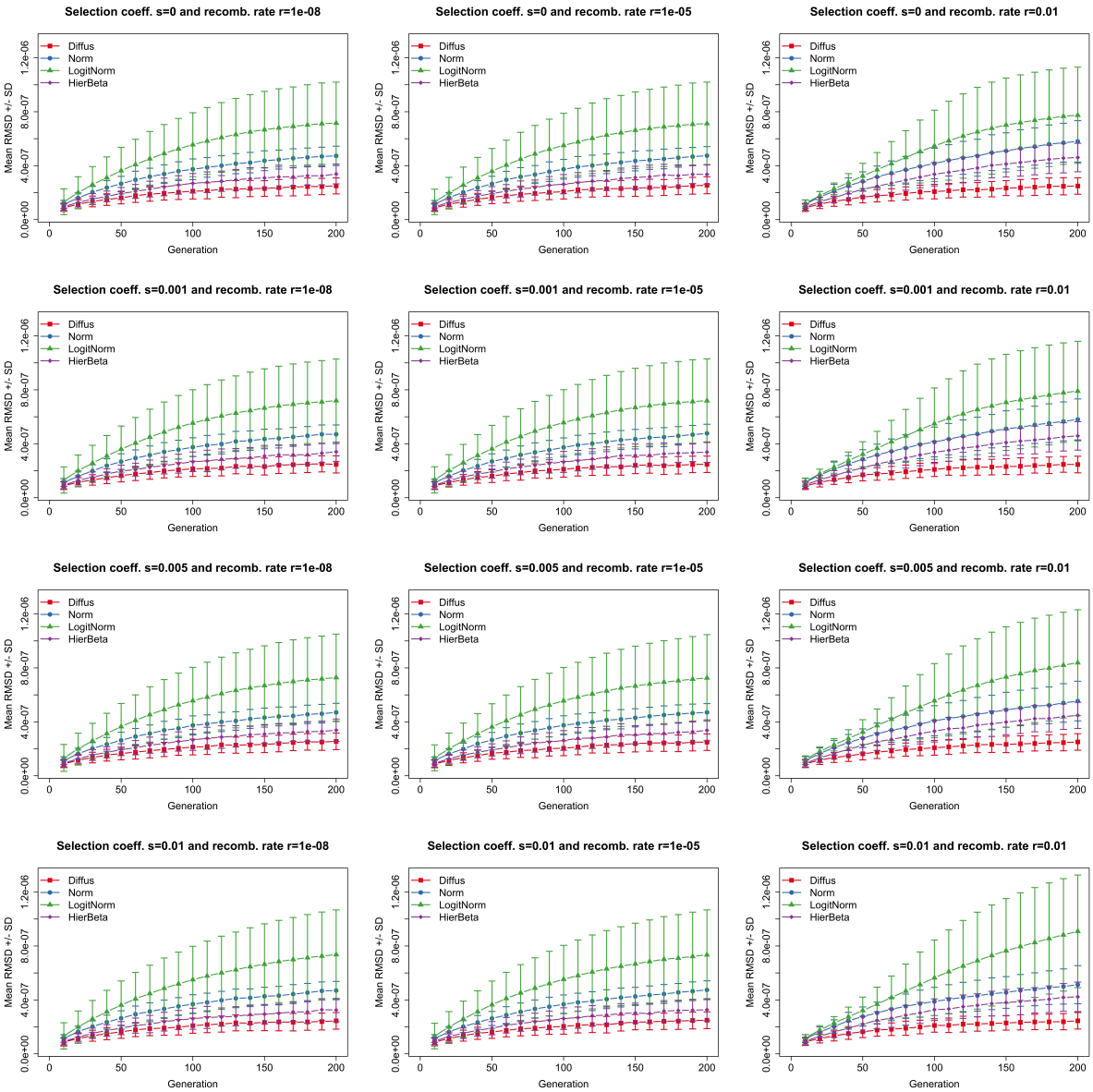

Figure 6: Comparison between the empirical CDFs of the Wright-Fisher model and their approximations for the case of the population size $N = 5000$, where we fix the selection coefficient $s_{\mathcal{A}} = 0.005$ and vary the selection coefficient $s_{\mathcal{B}} \in \{0, 0.001, 0.005, 0.01\}$ and the recombination rate $r \in \{0.00000001, 0.00001, 0.01\}$.

the proportion of the cases that the Wright-Fisher model cannot be matched by our hierarchical beta approximation at every 10 generations for each possible combination in Figure 8. We find from Figure 8 that the failure proportion has a trend of slow increase over time for small recombination rates but shows a sharp increase for large recombination rates. We also see that the failure proportion slightly increases with the decrease in the population size, which becomes significant for large recombination rates. This implies that our hierarchical beta approximation is not suitable for the analysis of the genetic data with large evolutionary timescales such as

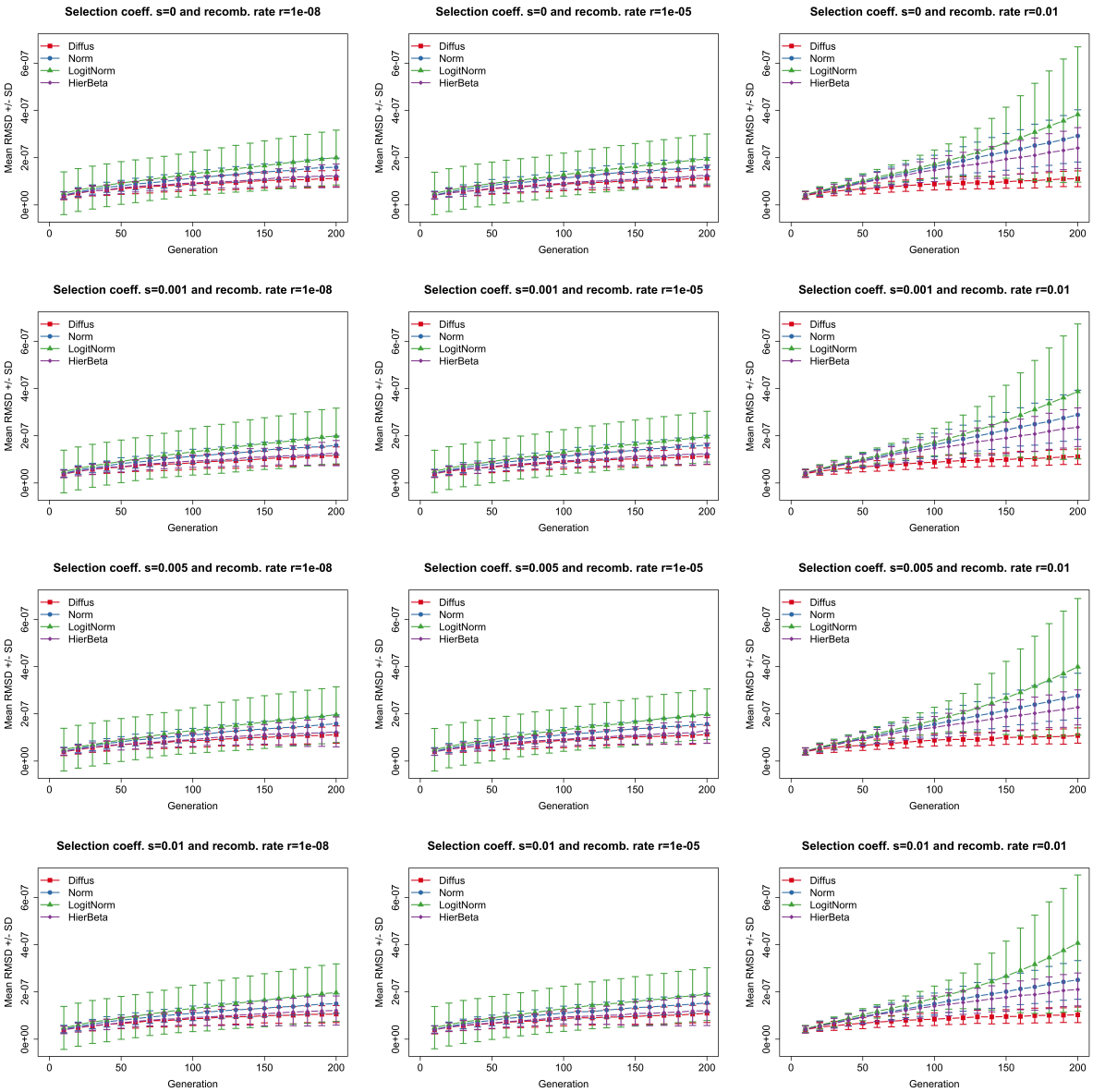

Figure 7: Comparison between the empirical CDFs of the Wright-Fisher model and their approximations for the case of the population size $N = 50000$, where we fix the selection coefficient $s_{\mathcal{A}} = 0.005$ and vary the selection coefficient $s_{\mathcal{B}} \in \{0, 0.001, 0.005, 0.01\}$ and the recombination rate $r \in \{0.00000001, 0.00001, 0.01\}$.

the ancient DNA (aDNA) data.

In conclusion, the moment-based approximations enable us to compute the transition probability distribution of the Wright-Fisher model at a far smaller computational cost than the diffusion approximation, but the diffusion approximation yields higher computational accuracy. Compared with the normal approximation of Terhorst et al. (2015), our logistic normal approximation yields higher computational accuracy only when the Wright-Fisher model gets close to the boundaries. Our hierarchical beta approximation shows the highest computational accuracy

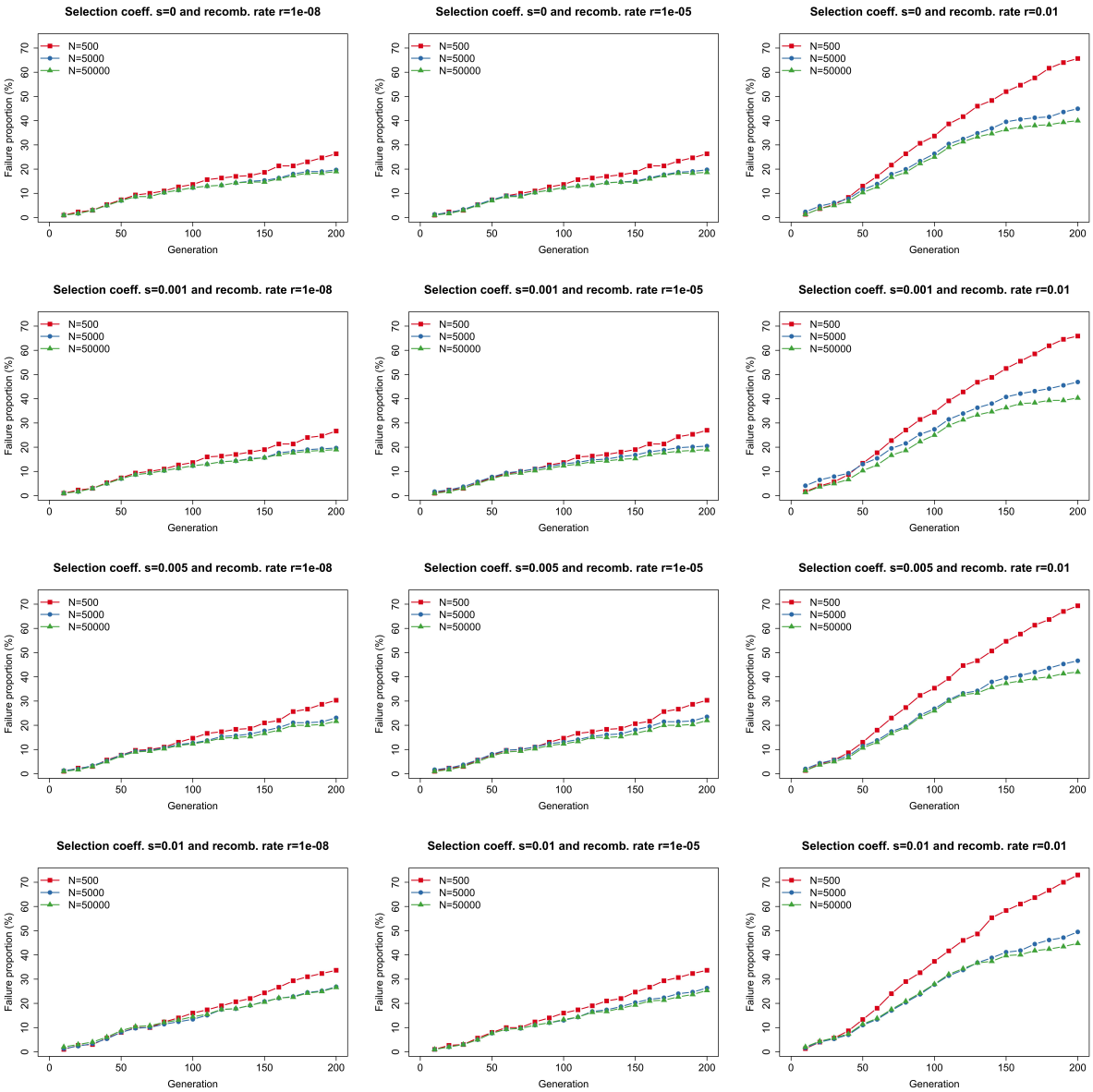

Figure 8: Proportions of the cases that the Wright-Fisher model cannot be fitted through our hierarchical beta approximation across all 36 possible combinations of the selection coefficient, the recombination rate and the population size.

among the three moment-based approximations but suffers from the computational instability issue, especially for large evolutionary timescales and/or recombination rates.

5. Discussion

In this work, we proposed two moment-based approximations for the two-locus Wright-Fisher model with selection, the logistic normal approximation and the hierarchical beta approximation. The main idea is to match the true moments of the Wright-Fisher model with the fitted

moments of the selected parametric continuous probability distribution, which could date back to Cavalli-Sforza & Edwards (1967). A prerequisite for the moment-based approximation is that we are able to calculate the mean and (co)variance of the two-locus Wright-Fisher model with selection at any given time point, so we extended the methods of Lacerda & Seoighe (2014) and Paris et al. (2019) that were limited to the Wright-Fisher model of the population dynamics under natural selection at a single locus. Compared to the existing moment-based approximation introduced by Terhorst et al. (2015), our logistic normal and hierarchical beta approximations enable us to avoid the distribution support issue found in their approximation.

Through extensive simulations studies, we demonstrated that our approximations for the mean and (co)variance had better performance with the decrease in the selection coefficient and/or the recombination rate, which got continuously worse over time due to the cumulative error. Our extension of Paris et al. (2019) up to the second order showed the best performance in the four moment approximations, which could be further improved by keeping third-order or even higher-order terms in the Taylor expansion. In our simulation studies, we also illustrated that the performance of the moment-based approximations became better with the increase in the population size but became worse over time due to the cumulative error, which is consistent with the conclusion for the single locus presented in Lacerda & Seoighe (2014). Our hierarchical beta approximation outperformed the normal approximation of Terhorst et al. (2015) but suffered from the issue of computational instability, especially for large evolutionary timescales and/or recombination rates. By contrast, the performance of our logistic normal approximation was not as good as the normal approximation of Terhorst et al. (2015) except when the Wright-Fisher model began to be absorbed at the boundaries. Compared to the diffusion approximation, our moment-based approximations could not yield the same computational accuracy but enabled the computation of the transition probability distribution of the Wright-Fisher model at a far smaller computational cost.

In recent years, the Wright-Fisher model has already been successfully applied in the population genetic analysis of time series allele frequency data. However, 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, which may lead to incorrect natural selection inferences, especially for the scenario of tightly linked loci (He et al., 2020b). Existing

methods that allow us to account for genetic recombination and local linkage suffer from either the distribution support issue in Terhorst et al. (2015) or the prohibitive computational cost in He et al. (2020b). Our moment-based approximations provide an alternative way to carry out the likelihood computation, especially our hierarchical beta approximation that has been shown to balance the computational accuracy and efficiency in our simulation studies. Our hierarchical beta approximation is therefore expected to improve the performance of the existing statistical frameworks for the inference of natural selection from time series genetic data. Moreover, our hierarchical beta approximation will be likely used by the community developing further statistical inference methods, *e.g.*, it provides the possibility of the genome-wide inference of natural selection from time series genomic data while modelling genetic recombination and local linkage through a maximum composite likelihood framework. It should be noted that due to the issue of computational instability, our hierarchical beta approximation is better suited to be used for the analysis of the genetic data of short evolutionary time scales such as experimental evolution studies (*e.g.*, Lang et al., 2013; Wiser et al., 2013; Burke et al., 2014; Le Bihan-Duval et al., 2018; Papkou et al., 2019; Barghi et al., 2019).

A number of interesting aspects arise from the presented material in this work. First of all, our moment-based approximations cannot completely capture the (co)variance structure of the two-locus Wright-Fisher model with selection although their performance has been shown to be acceptable. Recently, Speed et al. (2019) introduced the pyramid hierarchical beta approximation for the Wright-Fisher model for evolution of the relative frequencies of four alleles at a single locus, which has enough free parameters to capture the entire (co)variance structure. However, their pyramid hierarchical beta approximation cannot be extended to the two-locus Wright-Fisher model with selection since only a small subset of the state space can guarantee positive shape parameters for each beta distribution. So is there any other parametric continuous probability distribution that enables us to completely capture the (co)variance structure of the two-locus Wright-Fisher model with selection? Second, our moment-based approximations avoid the distribution support issue found in Terhorst et al. (2015) but fail to model fixation or loss of alleles. Tataru et al. (2015) addressed this issue and introduced the beta with spikes approximation for the Wright-Fisher model for evolution of the relative frequencies of two alleles at a single locus, and Paris et al. (2019) extended it to incorporate natural selection and applied

it to the inference of natural selection from time series data of allele frequencies. In their work, modelling fixation or loss of alleles has been illustrated to improve the performance of the beta approximation. So is it possible to capture the behaviour of the two-locus Wright-Fisher model with selection at the boundaries?

Acknowledgements

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

References

Andrieu, C., Doucet, A., & Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*, 269–342.

Barghi, N., Tobler, R., Nolte, V., Jakšić, A. M., Mallard, F. et al. (2019). Genetic redundancy fuels polygenic adaptation in Drosophila. *PLoS Biology*, *17*, e3000128.

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

Burke, M. K., Liti, G., & Long, A. D. (2014). Standing genetic variation drives repeatable experimental evolution in outcrossing populations of Saccharomyces cerevisiae. *Molecular Biology and Evolution*, *31*, 3228–3239.

Cavalli-Sforza, L. L., & Edwards, A. W. F. (1967). Phylogenetic analysis: models and estimation procedures. *Evolution*, *21*, 550–570.

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

Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A. et al. (2019). Tracking five millennia of horse management with extensive ancient genome time series. *Cell*, *177*, 1419–1435.

Feder, A. F., Kryazhimskiy, S., & Plotkin, J. B. (2014). Identifying signatures of selection in genetic time series. *Genetics*, *196*, 509–522.

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

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

He, Z., Beaumont, M. A., & Yu, F. (2020a). Numerical simulation of the two-locus Wright-Fisher stochastic differential equation with application to approximating transition probability densities. *bioRxiv*, (p. 213769).

He, Z., Dai, X., Beaumont, M. A., & Yu, F. (2020b). Detecting and quantifying natural selection at two linked loci from time series data of allele frequencies with forward-in-time simulations. *Genetics*, *216*, 521–541.

He, Z., Dai, X., Beaumont, M. A., & Yu, F. (2020c). Estimation of natural selection and allele age from time series allele frequency data using a novel likelihood-based approach. *Genetics*, *216*, 463–480.

Hobolth, A., & Siren, J. (2016). The multivariate Wright-Fisher process with mutation: Moment-based analysis and inference using a hierarchical beta model. *Theoretical Population Biology*, *108*, 36–50.

Kingman, J. F. C. (1982). The coalescent. *Stochastic Processes and their Applications*, *13*, 235–248.

Lacerda, M., & Seoighe, C. (2014). Population genetics inference for longitudinally-sampled mutants under strong selection. *Genetics*, *198*, 1237–1250.

Lang, G. I., Rice, D. P., Hickman, M. J., Sodergren, E., Weinstock, G. M. et al. (2013). Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature*, *500*, 571–574.

Le Bihan-Duval, E., Hennequet-Antier, C., Berri, C., Beauclercq, S. A., Bourin, M. C. et al. (2018). Identification of genomic regions and candidate genes for chicken meat ultimate pH by combined detection of selection signatures and QTL. *BMC Genomics*, *19*, 294.

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.

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N. et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, *528*, 499–503.

Nicholson, G., Smith, A. V., Jónsson, F., Gústafsson, Ó., Stefánsson, K. et al. (2002). Assessing population differentiation and isolation from single-nucleotide polymorphism data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*, 695–715.

Oehlert, G. W. (1992). A note on the delta method. *The American Statistician*, *46*, 27–29.

Papkou, A., Guzella, T., Yang, W., Koepper, S., Pees, B. et al. (2019). The genomic basis of Red Queen dynamics during rapid reciprocal host-pathogen coevolution. *Proceedings of the National Academy of Sciences*, *116*, 923–928.

Paris, C., Servin, B., & Boitard, S. (2019). Inference of selection from genetic time series using various parametric approximations to the Wright-Fisher model. *G3: Genes, Genomes, Genetics*, *9*, 4073–4086.

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

Speed, M. S., Balding, D. J., & Hobolth, A. (2019). A general framework for moment-based analysis of genetic data. *Journal of Mathematical Biology*, *78*, 1727–1769.

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.

Tataru, P., Bataillon, T., & Hobolth, A. (2015). Inference under a Wright-Fisher model using an accurate beta approximation. *Genetics*, *201*, 1133–1141.

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.

Terhorst, J., Schlötterer, C., & Song, Y. S. (2015). Multi-locus analysis of genomic time series data from experimental evolution. *PLoS Genetics*, *11*, e1005069.

Wiser, M. J., Ribeck, N., & Lenski, R. E. (2013). Long-term dynamics of adaptation in asexual populations. *Science*, *342*, 1364–1367.

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

Appendix A. Derivation of the moment approximations

We demonstrate how to extend the moment approximations developed in Lacerda & Seoighe (2014) and Paris et al. (2019) to the two-locus Wright-Fisher model with selection in this section, including the first- and second-order approximations proposed by Paris et al. (2019). We also introduce the moment approximations developed in Terhorst et al. (2015).

Appendix A.1. Extension of Lacerda & Seoighe (2014)

In this section, we provide the detailed derivation of the extension of the moment approximation proposed in Lacerda & Seoighe (2014) to the two-locus Wright-Fisher model with selection. Using the law of total mean and (co)variance, we can formulate the mean and (co)variance of the gamete frequencies $\boldsymbol{X}(k)$ given the gamete frequencies $\boldsymbol{X}(0)$ as

$$\begin{aligned}
\boldsymbol{\mu}_k &= \mathrm{E}\left[\mathrm{E}\left[\boldsymbol{X}(k) \mid \boldsymbol{X}(k-1)\right] \mid \boldsymbol{X}(0)\right] \\
&= \mathrm{E}\left[\boldsymbol{\mu}(\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right] && \text{(A.1)} \\
\boldsymbol{\Sigma}_k &= \mathrm{E}\left[\mathrm{Var}\left[\boldsymbol{X}(k) \mid \boldsymbol{X}(k-1)\right] \mid \boldsymbol{X}(0)\right] + \mathrm{Var}\left[\mathrm{E}\left[\boldsymbol{X}(k) \mid \boldsymbol{X}(k-1)\right] \mid \boldsymbol{X}(0)\right] \\
&= \mathrm{E}\left[\boldsymbol{\Sigma}(\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right] + \mathrm{Var}\left[\boldsymbol{\mu}(\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right]. && \text{(A.2)}
\end{aligned}$$

Following Lacerda & Seoighe (2014), we approximate the three terms on the right-hand side of Eqs. (A.1) and (A.2) with the multivariate delta method (see Oehlert, 1992, for more details), thereby obtaining

$$\mathrm{E}\left[\boldsymbol{\mu}(\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right] \approx \boldsymbol{\mu}(\boldsymbol{\mu}_{k-1}) \tag{A.3}$$

$$\mathrm{E}\left[\boldsymbol{\Sigma}(\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right] \approx \boldsymbol{\Sigma}(\boldsymbol{\mu}_{k-1}) \tag{A.4}$$

$$\mathrm{Var}\left[\boldsymbol{\mu}(\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right] \approx \boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})\boldsymbol{\Sigma}_{k-1}\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})^{\intercal}. \tag{A.5}$$

Substituting Eqs. (A.3)-(A.5) into Eqs. (A.1) and (A.2), we have

$$\begin{aligned}
\boldsymbol{\mu}_k &= \boldsymbol{\mu}(\boldsymbol{\mu}_{k-1}) \\
\boldsymbol{\Sigma}_k &= \boldsymbol{\Sigma}(\boldsymbol{\mu}_{k-1}) + \boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})\boldsymbol{\Sigma}_{k-1}\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})^{\intercal},
\end{aligned}$$

which is the recursion formula for the approximation of the mean and (co)variance of the two-

locus Wright-Fisher model with selection extended from Lacerda & Seoighe (2014). The initial mean and (co)variance are $\boldsymbol{\mu}_0 = \mathrm{E}[\boldsymbol{X}(0)]$ and $\boldsymbol{\Sigma}_0 = \mathrm{Var}[\boldsymbol{X}(0)]$, respectively.

Appendix A.2. Extension of Paris et al. (2019)

In this section, we show how to extend the moment approximation of Paris et al. (2019) to the two-locus Wright-Fisher model with selection. Following Paris et al. (2019), we substitute Eqs. (4) and (5) into Eq. (7), and then the (co)variance of the gamete frequencies $\boldsymbol{X}(k)$ given the gamete frequencies $\boldsymbol{X}(0)$ can be formulated as

$$\begin{aligned}
\boldsymbol{\Sigma}_k &= \mathrm{E}\left[\mathrm{Var}\left[\boldsymbol{X}(k) \mid \boldsymbol{X}(k-1)\right] \mid \boldsymbol{X}(0)\right] + \mathrm{Var}\left[\mathrm{E}\left[\boldsymbol{X}(k) \mid \boldsymbol{X}(k-1)\right] \mid \boldsymbol{X}(0)\right] \\
&= \frac{1}{2N}\,\mathrm{E}\left[\mathrm{diag}(\boldsymbol{\mu}(\boldsymbol{X}(k-1))) - \boldsymbol{\mu}(\boldsymbol{X}(k-1))\boldsymbol{\mu}(\boldsymbol{X}(k-1))^{\intercal} \mid \boldsymbol{X}(0)\right] \\
&\quad + \mathrm{Var}\left[\boldsymbol{\mu}(\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right] \\
&= \frac{1}{2N}\,\mathrm{diag}(\mathrm{E}\left[\boldsymbol{\mu}(\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right]) - \frac{1}{2N}\,\mathrm{E}\left[\boldsymbol{\mu}(\boldsymbol{X}(k-1))\boldsymbol{\mu}(\boldsymbol{X}(k-1))^{\intercal} \mid \boldsymbol{X}(0)\right] \\
&\quad + \mathrm{Var}\left[\boldsymbol{\mu}(\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right] \\
&= \frac{1}{2N}\left(\mathrm{diag}(\mathrm{E}\left[\boldsymbol{\mu}(\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right]) - \mathrm{E}\left[\boldsymbol{\mu}(\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right]\mathrm{E}\left[\boldsymbol{\mu}(\boldsymbol{X}(k-1))^{\intercal} \mid \boldsymbol{X}(0)\right]\right) \\
&\quad + \left(1 - \frac{1}{2N}\right)\mathrm{Var}\left[\boldsymbol{\mu}(\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right]. \qquad \text{(A.6)}
\end{aligned}$$

Using the Taylor expansion of $\boldsymbol{\mu}(\boldsymbol{X}(k-1))$ about $\boldsymbol{\mu}_{k-1}$ up to the first order, we have

$$\boldsymbol{\mu}(\boldsymbol{X}(k-1)) \approx \boldsymbol{\mu}(\boldsymbol{\mu}_{k-1}) + \boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})(\boldsymbol{X}(k-1) - \boldsymbol{\mu}_{k-1}),$$

thereby obtaining

$$\begin{aligned}
\mathrm{E}\left[\boldsymbol{\mu}(\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right] &\approx \mathrm{E}\left[\boldsymbol{\mu}(\boldsymbol{\mu}_{k-1}) + \boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})(\boldsymbol{X}(k-1) - \boldsymbol{\mu}_{k-1}) \mid \boldsymbol{X}(0)\right] \\
&= \boldsymbol{\mu}(\boldsymbol{\mu}_{k-1}) + \boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})(\mathrm{E}\left[\boldsymbol{X}(k-1) \mid \boldsymbol{X}(0)\right] - \boldsymbol{\mu}_{k-1}) \\
&= \boldsymbol{\mu}(\boldsymbol{\mu}_{k-1}) \qquad \text{(A.7)} \\
\mathrm{Var}\left[\boldsymbol{\mu}(\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right] &\approx \mathrm{Var}\left[\boldsymbol{\mu}(\boldsymbol{\mu}_{k-1}) + \boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})(\boldsymbol{X}(k-1) - \boldsymbol{\mu}_{k-1}) \mid \boldsymbol{X}(0)\right] \\
&= \boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})\,\mathrm{Var}\left[\boldsymbol{X}(k-1) \mid \boldsymbol{X}(0)\right]\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})^{\intercal} \\
&= \boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})\boldsymbol{\Sigma}_{k-1}\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})^{\intercal}. \qquad \text{(A.8)}
\end{aligned}$$

Substituting Eqs. (A.7) and (A.8) into Eqs. (A.1) and (A.6), we have

$$\begin{aligned}
\boldsymbol{\mu}_k &= \mathrm{E}\left[\boldsymbol{\mu}(\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right] \\
&\approx \boldsymbol{\mu}(\boldsymbol{\mu}_{k-1}) \\
\boldsymbol{\Sigma}_k &= \mathrm{E}\left[\boldsymbol{\Sigma}(\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right] + \left(1 - \frac{1}{2N}\right) \mathrm{Var}\left[\boldsymbol{\mu}(\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right] \\
&\approx \frac{1}{2N}\left(\mathrm{diag}(\boldsymbol{\mu}_k) - \boldsymbol{\mu}_k\boldsymbol{\mu}_k^\intercal\right) + \left(1 - \frac{1}{2N}\right) \boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})\boldsymbol{\Sigma}_{k-1}\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})^\intercal.
\end{aligned}$$

Using Eqs. (4) and (5), we can formulate the recursion formula for the first-order approximation of the mean and (co)variance of the two-locus Wright-Fisher model with selection extended from Paris et al. (2019) as

$$\begin{aligned}
\boldsymbol{\mu}_k &= \boldsymbol{\mu}(\boldsymbol{\mu}_{k-1}) \\
\boldsymbol{\Sigma}_k &= \frac{1}{2N}\left(\mathrm{diag}(\boldsymbol{\mu}_k) - \boldsymbol{\mu}_k\boldsymbol{\mu}_k^\intercal\right) + \left(1 - \frac{1}{2N}\right) \boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})\boldsymbol{\Sigma}_{k-1}\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})^\intercal
\end{aligned}$$

with the initial values $\boldsymbol{\mu}_0 = \mathrm{E}[\boldsymbol{X}(0)]$ and $\boldsymbol{\Sigma}_0 = \mathrm{Var}[\boldsymbol{X}(0)]$.

Similarly, with the Taylor expansion of $\boldsymbol{\mu}(\boldsymbol{X}(k-1))$ about $\boldsymbol{\mu}_{k-1}$ up to the second order

$$\begin{aligned}
\boldsymbol{\mu}(\boldsymbol{X}(k-1)) \approx \boldsymbol{\mu}(\boldsymbol{\mu}_{k-1}) &+ \boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})(\boldsymbol{X}(k-1) - \boldsymbol{\mu}_{k-1}) \\
&+ \frac{1}{2}\sum_{i,j} \frac{\partial^2 \boldsymbol{\mu}}{\partial x_i \partial x_j}(\boldsymbol{\mu}_{k-1})(X_i(k-1) - \mu_{i,k-1})(X_j(k-1) - \mu_{j,k-1}),
\end{aligned}$$

the recursion formula for the second-order approximation of the mean and (co)variance of the two-locus Wright-Fisher model with selection can be written down as

$$\begin{aligned}
\boldsymbol{\mu}_k &= \boldsymbol{\mu}(\boldsymbol{\mu}_{k-1}) + \frac{1}{2}\sum_{i,j} \frac{\partial^2 \boldsymbol{\mu}}{\partial x_i \partial x_j}(\boldsymbol{\mu}_{k-1})(\boldsymbol{\Sigma}_{k-1})_{ij} \\
\boldsymbol{\Sigma}_k &= \frac{1}{2N}\left(\mathrm{diag}(\boldsymbol{\mu}_k) - \boldsymbol{\mu}_k\boldsymbol{\mu}_k^\intercal\right) + \left(1 - \frac{1}{2N}\right) \boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})\boldsymbol{\Sigma}_{k-1}\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})^\intercal
\end{aligned}$$

with the initial values $\boldsymbol{\mu}_0 = \mathrm{E}[\boldsymbol{X}(0)]$ and $\boldsymbol{\Sigma}_0 = \mathrm{Var}[\boldsymbol{X}(0)]$.

Appendix A.3. Introduction of Terhorst et al. (2015)

In this section, we show how to apply the moment approximation of Terhorst et al. (2015) to the two-locus Wright-Fisher model with selection. Following Terhorst et al. (2015), we write

the gamete frequencies $\boldsymbol{X}(k) = \boldsymbol{x}_k + \delta\boldsymbol{X}(k)$, where $\boldsymbol{x}_0 = \mathrm{E}[\boldsymbol{X}(0)]$ and $\boldsymbol{x}_k = \boldsymbol{\mu}(\boldsymbol{x}_{k-1})$ for $k \in \mathbb{N}^+$, and introduce two correction terms $\boldsymbol{\epsilon}_k = \mathrm{E}\left[\delta\boldsymbol{X}(k) \mid \boldsymbol{X}(0)\right]$ and $\boldsymbol{\xi}_k = \mathrm{E}\left[\delta\boldsymbol{X}(k)\delta\boldsymbol{X}(k)^\intercal \mid \boldsymbol{X}(0)\right]$.

Unlike Paris et al. (2019), we use the Taylor expansion of $\boldsymbol{\mu}(\boldsymbol{X}(k-1))$ about $\boldsymbol{x}_{k-1}$ up to the second order rather than $\boldsymbol{\mu}_{k-1}$, and then we have

$$\boldsymbol{\mu}(\boldsymbol{X}(k-1)) \approx \boldsymbol{\mu}(\boldsymbol{x}_{k-1}) + \boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{x}_{k-1})\delta\boldsymbol{X}(k-1) + \frac{1}{2}\sum_{i,j}\frac{\partial^2\boldsymbol{\mu}}{\partial x_i \partial x_j}(\boldsymbol{x}_{k-1})\delta X_i(k-1)\delta X_j(k-1). \quad \text{(A.9)}$$

With Eq. (A.1), the mean of the gamete frequencies $\boldsymbol{X}(k)$ given the gamete frequencies $\boldsymbol{X}(0)$ can be represented as

$$\boldsymbol{\mu}_k = \mathrm{E}\left[\boldsymbol{\mu}(\boldsymbol{x}_{k-1} + \delta\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right]. \quad \text{(A.10)}$$

Substituting Eq. (A.9) into (A.10), we have

$$\boldsymbol{\mu}_k \approx \boldsymbol{\mu}(\boldsymbol{x}_{k-1}) + \boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{x}_{k-1})\boldsymbol{\epsilon}_{k-1} + \frac{1}{2}\sum_{i,j}\frac{\partial^2\boldsymbol{\mu}}{\partial x_i \partial x_j}(\boldsymbol{x}_{k-1})(\boldsymbol{\xi}_{k-1})_{ij}. \quad \text{(A.11)}$$

Similarly, with Eq. (A.2), the (co)variance of the gamete frequencies $\boldsymbol{X}(k)$ given the gamete frequencies $\boldsymbol{X}(0)$ can be expressed as

$$\begin{aligned}
\boldsymbol{\Sigma}_k &= \frac{1}{2N}\mathrm{E}\left[\mathrm{diag}(\boldsymbol{\mu}(\boldsymbol{X}(k-1))) - \boldsymbol{\mu}(\boldsymbol{X}(k-1))\boldsymbol{\mu}(\boldsymbol{X}(k-1))^\intercal \mid \boldsymbol{X}(0)\right] \\
&\quad + \mathrm{Var}\left[\boldsymbol{\mu}(\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right] \\
&= \frac{1}{2N}\mathrm{E}\left[\mathrm{diag}(\boldsymbol{\mu}(\boldsymbol{x}_{k-1} + \delta\boldsymbol{X}(k-1))) - \boldsymbol{\mu}(\boldsymbol{x}_{k-1} + \delta\boldsymbol{X}(k-1))\boldsymbol{\mu}(\boldsymbol{x}_{k-1} + \delta\boldsymbol{X}(k-1))^\intercal \mid \boldsymbol{X}(0)\right] \\
&\quad + \mathrm{Var}\left[\boldsymbol{\mu}(\boldsymbol{x}_{k-1} + \delta\boldsymbol{X}(k-1)) \mid \boldsymbol{X}(0)\right]. \qquad \text{(A.12)}
\end{aligned}$$

Substituting Eq. (A.9) into (A.12), we have

$$\begin{aligned}
\boldsymbol{\Sigma}_k &\approx \frac{1}{2N}\mathrm{diag}\left(\boldsymbol{\mu}_k\right) - \frac{1}{2N}\boldsymbol{\mu}(\boldsymbol{x}_{k-1})\boldsymbol{\mu}(\boldsymbol{x}_{k-1})^\intercal \\
&\quad - \frac{1}{2N}\left(\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{x}_{k-1})\boldsymbol{\epsilon}_{k-1}\right)\boldsymbol{\mu}(\boldsymbol{x}_{k-1})^\intercal - \frac{1}{2N}\boldsymbol{\mu}(\boldsymbol{x}_{k-1})\left(\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{x}_{k-1})\boldsymbol{\epsilon}_{k-1}\right)^\intercal \\
&\quad - \frac{1}{2N}\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{x}_{k-1})\boldsymbol{\xi}_{k-1}\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{x}_{k-1})^\intercal + \boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{x}_{k-1})\boldsymbol{\Sigma}_{k-1}\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{x}_{k-1})^\intercal \\
&= \frac{1}{2N}\mathrm{diag}\left(\boldsymbol{\mu}_k\right) - \frac{1}{2N}\left(\boldsymbol{\mu}(\boldsymbol{x}_{k-1}) + \boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{x}_{k-1})\boldsymbol{\epsilon}_{k-1}\right)\left(\boldsymbol{\mu}(\boldsymbol{x}_{k-1}) + \boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{x}_{k-1})\boldsymbol{\epsilon}_{k-1}\right)^\intercal \\
&\quad + \left(1 - \frac{1}{2N}\right)\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{x}_{k-1})\boldsymbol{\Sigma}_{k-1}\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{x}_{k-1})^\intercal. \qquad \text{(A.13)}
\end{aligned}$$

By definition, we can formulate the mean and (co)variance of the gamete frequencies $\boldsymbol{X}(k)$ given the gamete frequencies $\boldsymbol{X}(0)$ as

$$\begin{aligned}
\boldsymbol{\mu}_k &= \mathrm{E}\left[\boldsymbol{x}_k + \delta\boldsymbol{X}(k) \mid \boldsymbol{X}(0)\right] \\
&= \boldsymbol{x}_k + \mathrm{E}\left[\delta\boldsymbol{X}(k) \mid \boldsymbol{X}(0)\right] \\
&= \boldsymbol{x}_k + \boldsymbol{\epsilon}_k \qquad \text{(A.14)} \\
\boldsymbol{\Sigma}_k &= \mathrm{Var}\left[\boldsymbol{x}_k + \delta\boldsymbol{X}(k) \mid \boldsymbol{X}(0)\right] \\
&= \mathrm{Var}\left[\delta\boldsymbol{X}(k) \mid \boldsymbol{X}(0)\right] \\
&= \boldsymbol{\xi}_k - \boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^{\mathsf{T}}. \qquad \text{(A.15)}
\end{aligned}$$

Combining Eqs. (A.11), (A.13), (A.14) and (A.15), we can write down the recursion formula for the approximation of the mean and (co)variance of the two-locus Wright-Fisher model with selection introduced from Terhorst et al. (2015) as

$$\begin{aligned}
\boldsymbol{x}_k &= \boldsymbol{\mu}(\boldsymbol{x}_{k-1}) \\
\boldsymbol{\mu}_k &= \boldsymbol{\mu}(\boldsymbol{x}_{k-1}) + \boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{x}_{k-1})\boldsymbol{\epsilon}_{k-1} + \frac{1}{2}\sum_{i,j}\frac{\partial^2\boldsymbol{\mu}}{\partial x_i \partial x_j}(\boldsymbol{x}_{k-1})(\boldsymbol{\xi}_{k-1})_{ij} \\
\boldsymbol{\Sigma}_k &= \frac{1}{2N}\,\mathrm{diag}\left(\boldsymbol{\mu}_k\right) - \frac{1}{2N}\left(\boldsymbol{\mu}(\boldsymbol{x}_{k-1}) + \boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{x}_{k-1})\boldsymbol{\epsilon}_{k-1}\right)\left(\boldsymbol{\mu}(\boldsymbol{x}_{k-1}) + \boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{x}_{k-1})\boldsymbol{\epsilon}_{k-1}\right)^{\mathsf{T}} \\
&\quad + \left(1 - \frac{1}{2N}\right)\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{x}_{k-1})\boldsymbol{\Sigma}_{k-1}\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{x}_{k-1})^{\mathsf{T}} \\
\boldsymbol{\epsilon}_k &= \boldsymbol{\mu}_k - \boldsymbol{x}_k \\
\boldsymbol{\xi}_k &= \boldsymbol{\Sigma}_k + \boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^{\mathsf{T}}
\end{aligned}$$

with the initial values $\boldsymbol{x}_0 = \boldsymbol{\mu}_0 = \mathrm{E}[\boldsymbol{X}(0)]$, $\boldsymbol{\Sigma}_0 = \mathrm{Var}[\boldsymbol{X}(0)]$, $\boldsymbol{\epsilon}_0 = \mathbf{0}$ and $\boldsymbol{\xi}_0 = \mathbf{0}$. Note that $\mathbf{0}$ denotes the four-dimensional zero vector or the four by four zero matrix as appropriate.