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 significant improvement to the inference of natural selection from time series data of allele frequencies under a Wright-Fisher model. Existing approaches that can account for genetic recombination and local linkage are built on either the diffusion approximation or a moment-based approximation of the Wright-Fisher model. However, methods based on the diffusion approximation are likely to require much higher computational cost, whereas moment-based approximations may suffer from the distribution support issue: for example, the normal approximation can seriously affect computational accuracy. In the present work, we introduce two novel moment-based approximations of the Wright-Fisher model on a pair of linked loci, both subject to natural selection. Our key innovation is to extend existing methods to account for both the mean and (co)variance of the two-locus Wright-Fisher model with selection. We devise two approximation schemes, using a logistic normal distribution and a hierarchical beta distribution, respectively, by matching the first two moments of the Wright-Fisher model and the approximating model. As compared with the diffusion approximation, our approximations enable the approximate computation of the transition probabilities of the Wright-Fisher model at a far smaller computational cost. We can also avoid the distribution support issue found in the normal approximation.

Keywords: Natural selection, Linked loci, Wright-Fisher model, Moment approximation, Logistic normal approximation, Hierarchical beta 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

A central issue in evolutionary and population genetics is to understand the relative importance of genetic drift and natural selection as the mechanisms for maintaining polymorphisms within populations while at the same time driving divergence between populations. Mathematical descriptions of population dynamics under the effects of various demographic and evolutionary forces such as genetic drift and natural selection can be modelled through the Wright-Fisher model, introduced by Fisher (1922) and Wright (1931). To date, the Wright-Fisher model forms the basis of most theoretical and applied research in evolutionary and population genetics, *e.g.*, Kimura's work on fixation probabilities (Kimura, 1955), Ewens' sampling formula (Ewens, 1972) and Kingman's coalescent (Kingman, 1982). Moreover, the increased availability of time series genetic data in recent years makes it possible to accurately infer demographic and genetic properties of observed populations, especially detection and estimation of natural selection. For this reason, the Wright-Fisher model is also a popular choice as the underlying mathematical model on which statistical techniques for inferring natural selection are built (see Tataru et al., 2017, for a review of such statistical inference techniques using allele frequency data).

The most important ingredient for carrying out statistical inference of natural selection from time series data of allele frequencies is to compute, conditioning on the initial allele frequency, the probability distribution of the allele frequency at each given time point. However, to our knowledge, there is no existing general 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 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 moment-based approximations 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, thereby allowing the likelihood computation to be completed in a reasonable amount of time.

Recent advances in sequencing techniques have revolutionised the collection of time series genomic data, with increased volume and improved quality (*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 accounting for the process of genetic recombination and the information of local linkage. Properly modelling genetic recombination and local linkage has been shown to significantly improve the inference of natural selection from genetic time series (see He et al., 2020b). However, with the exception of Terhorst et al. (2015) and He et al. (2020b), all existing Wright-Fisher model based methods 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 approach their boundaries 0 or 1 (*i.e.*, allele loss or fixation), the Wright-Fisher model will be poorly approximated owing to the infinite support of the Gaussian distribution that will result in a leak of probability mass into unphysical values of the allele frequency that are smaller than 0 or larger than 1. 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 approach was built upon the diffusion approximation of the Wright-Fisher model for a population evolving under natural selection at a pair of 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 resulting Kolmogorov backward equation (or its adjoint) is computationally challenging and prohibitively expensive. Their posterior computation was therefore carried out with 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 population dynamics subject to natural selection at two linked loci. The main idea is to pick from a parametric family of probability distributions that matches the first two moments to those of the Wright-Fisher model. This is along the line of Cavalli-Sforza & Edwards (1967). To achieve the first two moments of the two-locus Wright-Fisher model with selection, we extend the moment approximations introduced by Lacerda & Seoighe (2014) and Paris et al. (2019), respectively. To avoid the distribution support issue found in the Gaussian approximation of Terhorst et al. (2015), we choose two different parametric families: logistic normal distributions and hierarchical beta distributions. As compared with the diffusion approximation of He et al. (2020b), our moment-based approximations enable us to compute the transition probabilities of the Wright-Fisher model at a much smaller computational cost. Our moment-based approximations are therefore expected to improve the performance of the existing statistical frameworks for the inference of natural selection from time series allele frequency data.

2. Wright-Fisher model

We consider a population of N randomly mating diploid individuals, consisting of 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 the 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, we assume that 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. The symbols $\mathcal{A}_2$ and $\mathcal{B}_2$ are the ancestral alleles, which are assumed to originally exist in the population.

We incorporate viability selection into population dynamics and let w_{ij} denote the viability of an individual with the genotype formed by haplotypes i and 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. More specifically, the viabilities of the three possible genotypes at each locus, *i.e.*, 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 of an effectively infinite population in generation $k-1$ and can be written down as a function of the gamete frequencies $\boldsymbol{x}$,

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

with

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

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 $\Omega_{\boldsymbol{X}} = \{\boldsymbol{x} \in \{0, 1/(2N), \ldots, 1\}^4 : \sum_{i=1}^4 x_i = 1\}$ with the multinomial one-step transition probabilities as specified in Eqs. (1)-(3).

3. Moment-based approximations

Even though the one-step transition probabilities of the Wright-Fisher model, specified in Eqs. (1)-(3), can be written down, the k-step transition probabilities are intractable for general k. Numerical evaluation of the k-step transition probabilities is theoretically feasible but computationally prohibitive for large population sizes and evolutionary timescales. For this reason,

approximations to the transition probabilities in the Wright-Fisher model based on parametric families have been developed, which could be traced back to Cavalli-Sforza & Edwards (1967). However, existing moment-based approximations either suffer from the distribution support issue (*e.g.*, Lacerda & Seoighe, 2014; Terhorst et al., 2015) or only work for the case of natural selection acting on a single locus (*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. We first derive approximations for the mean and (co)variance, and then, to approximate the transition probabilities, we describe a logistic normal approximation and a hierarchical beta approximation.

3.1. Mean and (co)variance approximations

The main idea of the moment-based approximation is to approximate the transition probabilities of the Wright-Fisher model with a continuous probability distribution chosen from a parametric family, so that the first two moments are matched. This approach was first extended by Lacerda & Seoighe (2014) to incorporate natural selection and using normal distributions, applied to the inference of natural selection from time series data of allele frequencies. Its prerequisite is to compute the mean and (co)variance of the Wright-Fisher model with selection at any given time point. Terhorst et al. (2015) proposed an iteration formula to approximate the first two moments of the Wright-Fisher model, which is the only method applicable to the case of natural selection acting on multiple linked loci. We 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), given the gamete frequencies $\boldsymbol{X}(k-1) = \boldsymbol{x}$, we can write down the mean and (co)variance of the gamete frequencies $\boldsymbol{X}(k)$ 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 $\boldsymbol{\mu}_{k-1}$. The mean $\boldsymbol{\mu}_k$ and the (co)variance $\boldsymbol{\Sigma}_k$ can then be approximated as

$$\boldsymbol{\mu}_k = \boldsymbol{\mu}(\boldsymbol{\mu}_{k-1}) \qquad (8)$$

$$\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}(\boldsymbol{\mu}_{k-1}) + \boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})\boldsymbol{\Sigma}_{k-1}\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_{k-1})^{\mathsf{T}}, \qquad (9)$$

where $\boldsymbol{J}_{\boldsymbol{\mu}}(\boldsymbol{x})$ is the Jacobian matrix defined over the function $\boldsymbol{\mu}(\boldsymbol{x})$ in Eq. (4). By substituting Eqs. (2) and (3) into Eq. (4) and taking the derivative with respect to the gamete frequencies $\boldsymbol{x}$, we have

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

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.

Alternatively, following Paris et al. (2019), we substitute Eqs. (4) and (5) into Eq. (7). Then the (co)variance $\boldsymbol{\Sigma}_k$ 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 $\boldsymbol{\Sigma}_k$ as

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

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

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 $\boldsymbol{\Sigma}_k$ as

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

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

Using the iteration formula in Eqs. (8) and (9) or in Eqs. (10) and (11) or in Eqs. (12) and (13), we can achieve the approximations of the mean and (co)variance of the two-locus Wright-Fisher model with selection, $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$, for $k \in \mathbb{N}$ with initial values of $\boldsymbol{\mu}_0 = \mathrm{E}[\boldsymbol{X}(0)]$ and $\boldsymbol{\Sigma}_0 = \mathrm{Var}[\boldsymbol{X}(0)]$. See 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

With the approximations for the first two moments of the two-locus Wright-Fisher model with selection, we now set out to find distributions to approximate the true transition probabilities. We first consider the family of normal distributions, which has already been successfully

used in population genetics studies (*e.g.*, Nicholson et al., 2002; Pickrell & Pritchard, 2012; Lacerda & Seoighe, 2014; Terhorst et al., 2015). We assume that the gamete frequencies $\boldsymbol{X}(k)$ are approximately distributed according to

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

where $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are the mean and (co)variance of the two-locus Wright-Fisher model with selection, respectively, which can be obtained by running the iteration 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 mean and (co)variance of the normal distribution to those of the Wright-Fisher model. Unfortunately, the normal approximation in Eq. (14) is able to take values outside the state space $\Omega_{\boldsymbol{X}}$, resulting in poor performance once gamete frequencies $\boldsymbol{X}(k)$ approach 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 normal distribution outside the state space of the Wright-Fisher model to the discrete fixation probabilities at boundaries 0 and 1. As an alternative, Paris et al. (2019) simply rejected the values outside the state space of the Wright-Fisher model and rescaled the normal distribution. However, both these procedures lead to the mean and (co)variance of the resulting probability distribution no longer matching those of the Wright-Fisher model.

To tackle the distribution support issue of the normal approximation, we consider the logistic normal distribution with the support $\Delta_{\boldsymbol{X}} = \{\boldsymbol{x} \in [0,1]^4 : \sum_{i=1}^4 x_i = 1\}$. More specifically, we construct a transformation, denoted by $\boldsymbol{\phi}$, to map the gamete frequencies $\boldsymbol{X}(k)$ to the state space $\mathbb{R}^3$:

$$\boldsymbol{\phi}(\boldsymbol{X}(k)) = \left(\log\left(\frac{X_1(k)}{X_4(k)}\right), \log\left(\frac{X_2(k)}{X_4(k)}\right), \log\left(\frac{X_3(k)}{X_4(k)}\right)\right)^\intercal \tag{15}$$

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

$$\boldsymbol{Y}(k) \mid \boldsymbol{Y}(0) \sim \text{Normal}(\boldsymbol{m}_k, \boldsymbol{V}_k), \tag{16}$$

where $\boldsymbol{m}_k$ and $\boldsymbol{V}_k$ are the mean and (co)variance of the normal distribution. 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. Given the initial 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

$$\boldsymbol{X}(k) \mid \boldsymbol{X}(0) \sim \text{Logistic-normal}(\boldsymbol{m}_k, \boldsymbol{V}_k), \tag{17}$$

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

Our task now is to find proper values for the mean and (co)variance of the normal distribution in Eq. (16) such that the logistic normal distribution in Eq. (17) matches its mean and (co)variance to those of the Wright-Fisher model $\boldsymbol{X}$. To the best of our knowledge, there is no analytical solution for the mean and (co)variance of the logistic normal distribution, but 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 can 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 $\boldsymbol{X}$,

$$\begin{aligned} \boldsymbol{m}_k &= \mathrm{E}\left[\boldsymbol{\phi}(\boldsymbol{X}(k)) \mid \boldsymbol{X}(0)\right] \\ &= \boldsymbol{\phi}(\boldsymbol{\mu}_k) \end{aligned} \tag{18}$$

$$\begin{aligned} \boldsymbol{V}_k &= \mathrm{Var}\left[\boldsymbol{\phi}(\boldsymbol{X}(k)) \mid \boldsymbol{X}(0)\right] \\ &= \boldsymbol{J}_{\boldsymbol{\phi}}(\boldsymbol{\mu}_k)\boldsymbol{\Sigma}_k\boldsymbol{J}_{\boldsymbol{\phi}}(\boldsymbol{\mu}_k)^{\intercal}, \end{aligned} \tag{19}$$

where

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

is the Jacobian matrix defined over the function $\boldsymbol{\phi}(\boldsymbol{x})$ in Eq. (15). 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 running the iteration 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. (18) and (19) into Eq. (17), our logistic normal approximation of the two-locus Wright-Fisher model with selection can therefore be represented as

$$\boldsymbol{X}(k) \mid \boldsymbol{X}(0) \sim \text{Logistic-normal}(\boldsymbol{\phi}(\boldsymbol{\mu}_k), \boldsymbol{J}_{\boldsymbol{\phi}}(\boldsymbol{\mu}_k)\boldsymbol{\Sigma}_k\boldsymbol{J}_{\boldsymbol{\phi}}(\boldsymbol{\mu}_k)^{\intercal}).$$

The transition probability density function of our logistic normal approximation can be explicitly written down by using 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. (18) and (19) (see Figure 1 for an illustration). In Section 4, we will explore how such a discrepancy affects the performance of our logistic normal approximation through extensive simulations.

3.3. Hierarchical beta approximation

In addition to normal distributions, the beta family of distributions is an alternative candidate of the parametric family for the moment-based approximation of the Wright-Fisher model. This has been shown to deliver superior performance in comparison with normal distributions for single-locus problems (see Tataru et al., 2017; Paris et al., 2019) and successfully applied in a number of population genetics studies (*e.g.*, Sirén et al., 2011; Hui & Burt, 2015; Tataru et al., 2015; Paris et al., 2019). For two-locus problems, we need a multivariate generalisation of beta distributions, but unfortunately, the multivariate beta distribution, commonly known as the Dirichlet distribution, is not appropriate since its mean and (co)variance fail to fully 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 completely capture the (co)variance structure of the Wright-Fisher model.

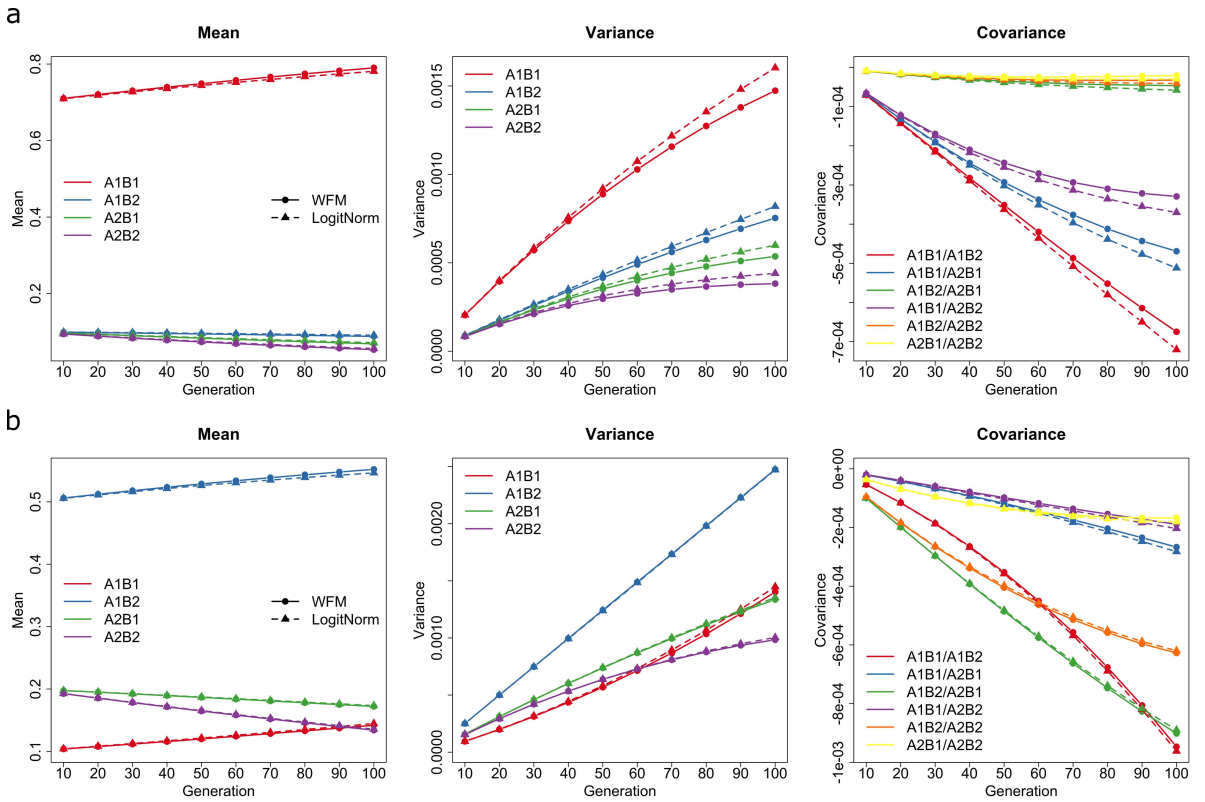

Figure 1: An illustration for the mean, variance and covariance of the Wright-Fisher model and our logistic normal approximation, where we set the selection coefficients $s_{\mathcal{A}} = 0.01$ and $s_{\mathcal{B}} = 0.005$, the dominance parameters $h_{\mathcal{A}} = 0.5$ and $s_{\mathcal{B}} = 0.5$, 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) $\boldsymbol{X}_0 = (0.1, 0.5, 0.2, 0.2)$, respectively. The first two moments of the Wright-Fisher model and the logistic normal approximation are estimated from 1,000,000 replicated simulations.

Inspired by Hobolth & Siren (2016), we employ the hierarchical beta distribution with the support $\Delta_{\boldsymbol{X}} = \{\boldsymbol{x} \in [0,1]^4 : \sum_{i=1}^4 x_i = 1\}$, which is an alternative multivariate generalisation of the beta distribution. The main idea is to decompose the two-locus Wright-Fisher model with selection into two conditionally independent beta distributed levels. More specifically, we consider any three linearly independent gamete frequencies out of the four, which is illustrated below for $X_1(k)$, $X_2(k)$ and $X_3(k)$, and construct a transformation, denoted by $\boldsymbol{\psi}$, to map the gamete frequencies $\boldsymbol{X}(k)$ to the state space $[0,1]^3$:

$$\boldsymbol{\psi}(\boldsymbol{X}(k)) = \left(X_1(k) + X_2(k), \frac{X_1(k)}{X_1(k) + X_2(k)}, \frac{X_3(k)}{1 - X_1(k) - X_2(k)} \right)^{\intercal}.$$

We introduce a stochastic process, denoted by $\boldsymbol{Z} = \{\boldsymbol{Z}(k), k \in \mathbb{N}\}$, to approximate the transformed Wright-Fisher model $\boldsymbol{\psi}(\boldsymbol{X})$. More concretely, let $\boldsymbol{Z}(k) = (Z_1(k), Z_2(k), Z_3(k))^{\intercal} \in [0,1]^3$

follow a hierarchical beta distribution

$$Z_1(k) \mid \boldsymbol{Z}(0) \sim \text{Beta}(\alpha_{1,k}, \beta_{1,k}) \tag{20}$$

$$Z_2(k) \mid Z_1(k), \boldsymbol{Z}(0) \sim \text{Beta}(\alpha_{2,k}, \beta_{2,k}) \tag{21}$$

$$Z_3(k) \mid Z_1(k), \boldsymbol{Z}(0) \sim \text{Beta}(\alpha_{3,k}, \beta_{3,k}), \tag{22}$$

where $\alpha_{i,k}$ and $\beta_{i,k}$ are the shape parameters of each beta distribution in Eqs. (20)-(22). 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 $\boldsymbol{X}$. We call the procedure we just outlined 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. (20)-(22) such that the probability distribution for our hierarchical beta approximation $\boldsymbol{\psi}^{-1}(\boldsymbol{Z}(k))$ has its mean and (co)variance matching those of the Wright-Fisher model $\boldsymbol{X}$. The shape parameters of the beta distribution are related its mean and variance as follows:

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

$$\beta_{i,k} = \left(\frac{m_{i,k}(1 - m_{i,k})}{V_{i,k}} - 1\right) (1 - m_{i,k}) \tag{24}$$

for $i = 1, 2, 3$, where $m_{i,k}$ and $V_{i,k}$ denote the mean and variance of each beta distribution in Eqs. (20)-(22). Our problem then 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 $\boldsymbol{X}$.

Using the law of total mean, we can now write down 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

$$
\begin{aligned}
\mu_{1,k} &= \mathrm{E}\left[Z_1(k)Z_2(k) \mid \boldsymbol{Z}(0)\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[Z_1(k)Z_2(k) \mid Z_1(k), \boldsymbol{Z}(0)\right] \mid \boldsymbol{Z}(0)\right] \\
&= \mathrm{E}\left[Z_1(k)m_{2,k} \mid \boldsymbol{Z}(0)\right] \\
&= m_{1,k}m_{2,k}. \qquad (25)
\end{aligned}
$$

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

$$
\mu_{2,k} = m_{1,k}(1 - m_{2,k}) \qquad (26)
$$

$$
\mu_{3,k} = (1 - m_{1,k})m_{3,k} \qquad (27)
$$

$$
\mu_{4,k} = (1 - m_{1,k})(1 - m_{3,k}). \qquad (28)
$$

Solving the system of coupled equations given by Eqs. (25)-(27) yields

$$
m_{1,k} = \mu_{1,k} + \mu_{2,k} \qquad (29)
$$

$$
m_{2,k} = \frac{\mu_{1,k}}{\mu_{1,k} + \mu_{2,k}} \qquad (30)
$$

$$
m_{3,k} = \frac{\mu_{3,k}}{\mu_{3,k} + \mu_{4,k}}. \qquad (31)
$$

Note that the mean $m_{i,k}$ given in Eqs. (29)-(31) also satisfies Eq. (28).

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

$$
\begin{aligned}
\Sigma_{11,k} &= \mathrm{Var}\left[Z_1(k)Z_2(k) \mid \boldsymbol{Z}(0)\right] \\
&= \mathrm{Var}\left[\mathrm{E}\left[Z_1(k)Z_2(k) \mid Z_1(k), \boldsymbol{Z}(0)\right] \mid \boldsymbol{Z}(0)\right] \\
&\quad + \mathrm{E}\left[\mathrm{Var}\left[Z_1(k)Z_2(k) \mid Z_1(k), \boldsymbol{Z}(0)\right] \mid \boldsymbol{Z}(0)\right] \\
&= \mathrm{Var}\left[Z_1(k)m_{2,k} \mid \boldsymbol{Z}(0)\right] + \mathrm{E}\left[Z_1^2(k)V_{2,k} \mid \boldsymbol{Z}(0)\right] \\
&= V_{1,k}m_{2,k}^2 + (V_{1,k} + m_{1,k}^2)V_{2,k}. \qquad (32)
\end{aligned}
$$

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

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

$$\Sigma_{33,k} = V_{1,k}m_{3,k}^2 + (V_{1,k} + (1 - m_{1,k})^2)V_{3,k} \tag{34}$$

$$\Sigma_{44,k} = V_{1,k}(1 - m_{3,k})^2 + (V_{1,k} + (1 - m_{1,k})^2)V_{3,k}. \tag{35}$$

Solving the system of coupled equations given by Eqs. (32)-(34) yields

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

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

$$V_{3,k} = \frac{-m_{3,k}^2\Sigma_{11,k} + m_{3,k}^2\Sigma_{22,k} + (m_{2,k}^2 - (1 - m_{2,k})^2)\Sigma_{33,k}}{(1 - m_{1,k})^2(m_{2,k}^2 - (1 - m_{2,k})^2) + (\Sigma_{11,k} - \Sigma_{22,k})}, \tag{38}$$

where $m_{i,k}$ is the mean of each beta distribution given by Eqs. (29)-(31). Note that the variance $V_{i,k}$ given in Eqs. (36)-(38) is no longer guaranteed to satisfy Eq. (35).

Recall that the mean $\boldsymbol{\mu}_k$ and the (co)variance $\boldsymbol{\Sigma}_k$ of the Wright-Fisher model $\boldsymbol{X}$ are readily available by running the iteration 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 Eqs. (20)-(22), 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. (29)-(31) and (33)-(35) into Eqs. (23) and (24). The transition probability density function of our hierarchical beta approximation can be explicitly written down by the change-of-variable technique. Note that in our hierarchical beta approximation, the mean of our hierarchical beta approximation exactly matches those of the Wright-Fisher model, but only three of four variances are exactly matched (see Figure 2 for an illustration). This is because in our hi-

erarchical beta approximation there is no extra free parameter left for fitting the one remaining variance and all covariance of the two-locus Wright-Fisher model with selection.

Using the law of total covariance, we can write down the covariance between four gamete frequencies under our hierarchical beta approximation in terms of the mean $m_{i,k}$ and the variance $V_{i,k}$ of each beta distribution. For example, the covariance between the frequencies of the $\mathcal{A}_1\mathcal{B}_1$ gamete and the $\mathcal{A}_1\mathcal{B}_2$ gamete under our hierarchical beta approximation can be represented as

$$\begin{aligned}
&\mathrm{Cov}\left[X_1(k), X_2(k) \mid \boldsymbol{X}(0)\right] \\
&\quad = \mathrm{Cov}\left[Z_1(k)Z_2(k), Z_1(k)(1-Z_2(k)) \mid \boldsymbol{Z}(0)\right] \\
&\quad = \mathrm{E}\left[\mathrm{Cov}\left[Z_1(k)Z_2(k), Z_1(k)(1-Z_2(k)) \mid Z_1(k), \boldsymbol{Z}(0)\right] \mid \boldsymbol{Z}(0)\right] \\
&\qquad + \mathrm{Cov}\left[\mathrm{E}\left[Z_1(k)Z_2(k) \mid Z_1(k), \boldsymbol{Z}(0)\right], \mathrm{E}\left[Z_1(k)(1-Z_2(k)) \mid Z_1(k), \boldsymbol{Z}(0)\right] \mid \boldsymbol{Z}(0)\right] \\
&\quad = -\mathrm{E}\left[Z_1(k)^2 V_{2,k} \mid \boldsymbol{Z}(0)\right] + \mathrm{Cov}\left[Z_1(k)m_{2,k}, Z_1(k)(1-m_{2,k}) \mid \boldsymbol{Z}(0)\right] \\
&\quad = -(V_{1,k} - m_{1,k}^2)V_{2,k} + V_{1,k}m_{2,k}(1-m_{2,k}).
\end{aligned}$$

Similarly, we can write down the other covariance between four gamete frequencies under our hierarchical beta approximation as

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

We will discuss how the discrepancy of the (co)variance structure affects the performance of our hierarchical beta approximation through extensive simulations in Section 4.

4. Moment-based approximation performance

We evaluate the performance of both our logistic normal and hierarchical beta approximations for the two-locus Wright-Fisher model with selection through extensive simulation studies, comparing them to the diffusion and normal approximations. In what follows, we fix the selec-

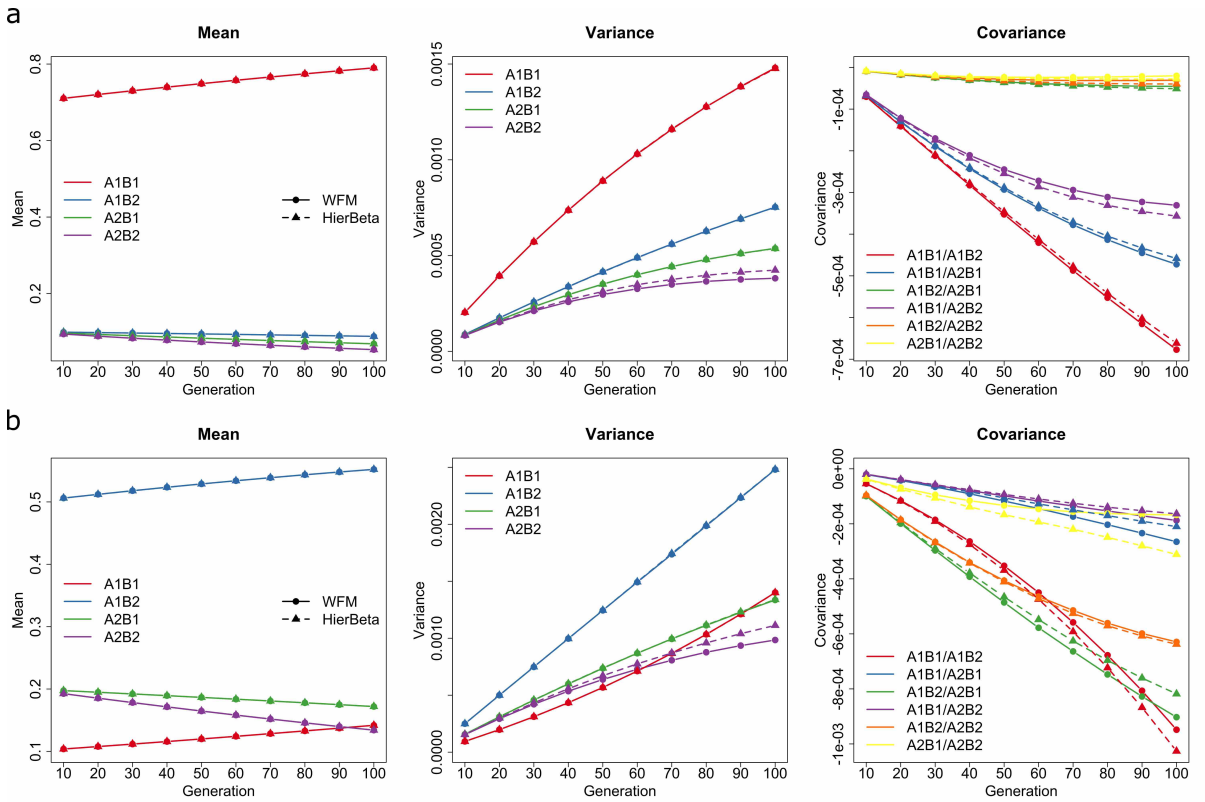

Figure 2: An illustration for the mean, variance and covariance of the Wright-Fisher model and our hierarchical beta approximation, where we set the selection coefficients $s_{\mathcal{A}} = 0.01$ and $s_{\mathcal{B}} = 0.005$, the dominance parameters $h_{\mathcal{A}} = 0.5$ and $s_{\mathcal{B}} = 0.5$, 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) $\boldsymbol{X}_0 = (0.1, 0.5, 0.2, 0.2)$, respectively. The first two moments of the Wright-Fisher model are estimated from 1,000,000 replicated simulations.

tion coefficient $s_{\mathcal{A}} = 0.005$ and vary the second selection coefficient $s_{\mathcal{B}} \in \{0, 0.001, 0.005, 0.01\}$, recombination rate $r \in \{0.00000001, 0.00001, 0.01\}$ and population size $N \in \{500, 5000, 50000\}$. We take 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 $h_{\mathcal{B}} \in [0, 1]$. All approximations and simulations described in this work were implemented in R with C++ through the Rcpp and RcppArmadillo packages, and the source code is available at `https://github.com/zhangyi-he/WFM-2L-Approx`.

4.1. Performance of mean and (co)variance approximations

We perform 300 replicated runs for each of the 36 possible combinations of the selection coefficient, recombination rate and population size. For each run, we randomly select the initial gamete frequencies $\boldsymbol{X}_0$ from a flat Dirichlet distribution on the state space $\Omega_{\boldsymbol{X}}$ and simulate

1,000,000 gamete frequency trajectories for 100 generations under the Wright-Fisher model. We compute the empirical mean and (co)variance of the Wright-Fisher model from these 1,000,000 replicates every 10 generations. To find out which method gives the best moment approximation, we approximate the mean and (co)variance every 10 generations using the approach of Terhorst et al. (2015) and our extensions of Lacerda & Seoighe (2014) and Paris et al. (2019). In the sequel, we refer to our three approximation methods as follows:

- Method EL: a scheme using the approximation of mean and (co)variance by our extension of Lacerda & Seoighe (2014).
- Method EP1: a scheme using the approximation of mean and (co)variance by our extension of Paris et al. (2019) up to first order.
- Method EP2: a scheme using the approximation of mean and (co)variance by our extension of Paris et al. (2019) up to second order.

We use the Euclidean vector norm to measure the approximation error of the mean vector, *i.e.*, the difference between the mean under the Wright-Fisher model and their approximations, and the Frobenius matrix norm to measure the approximation error of the (co)variance matrix.

For the case of the population size $N = 5000$, we plot the approximation errors of the mean vector in Figure 3 and the approximation errors of the (co)variance matrix in Figure 4. From Figures 3 and 4, we observe that the error of our approximations increases, *i.e.*, the larger mean error with the larger error bar, with the increase in the selection coefficient and/or recombination rate, and also accumulates with time. A similar conclusion can be obtained for other different population sizes. See Supplemental Material, Figures S1-S4 for the comparison between the mean vector and (co)variance matrix of the Wright-Fisher model and their approximations for the cases of the population size $N = 500$ and $N = 50000$.

As illustrated in Figure 3, we get the same performance with methods EL and EP1. Compared to these two methods, the approach of Terhorst et al. (2015) and our method EP2 have better performance across all 12 possible combinations of the selection coefficient and recombination rate. However, the performance advantage of the latter two methods narrows with the increase in the selection coefficient and/or recombination rate. More specifically, our method EP2 matches the performance of the approach 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

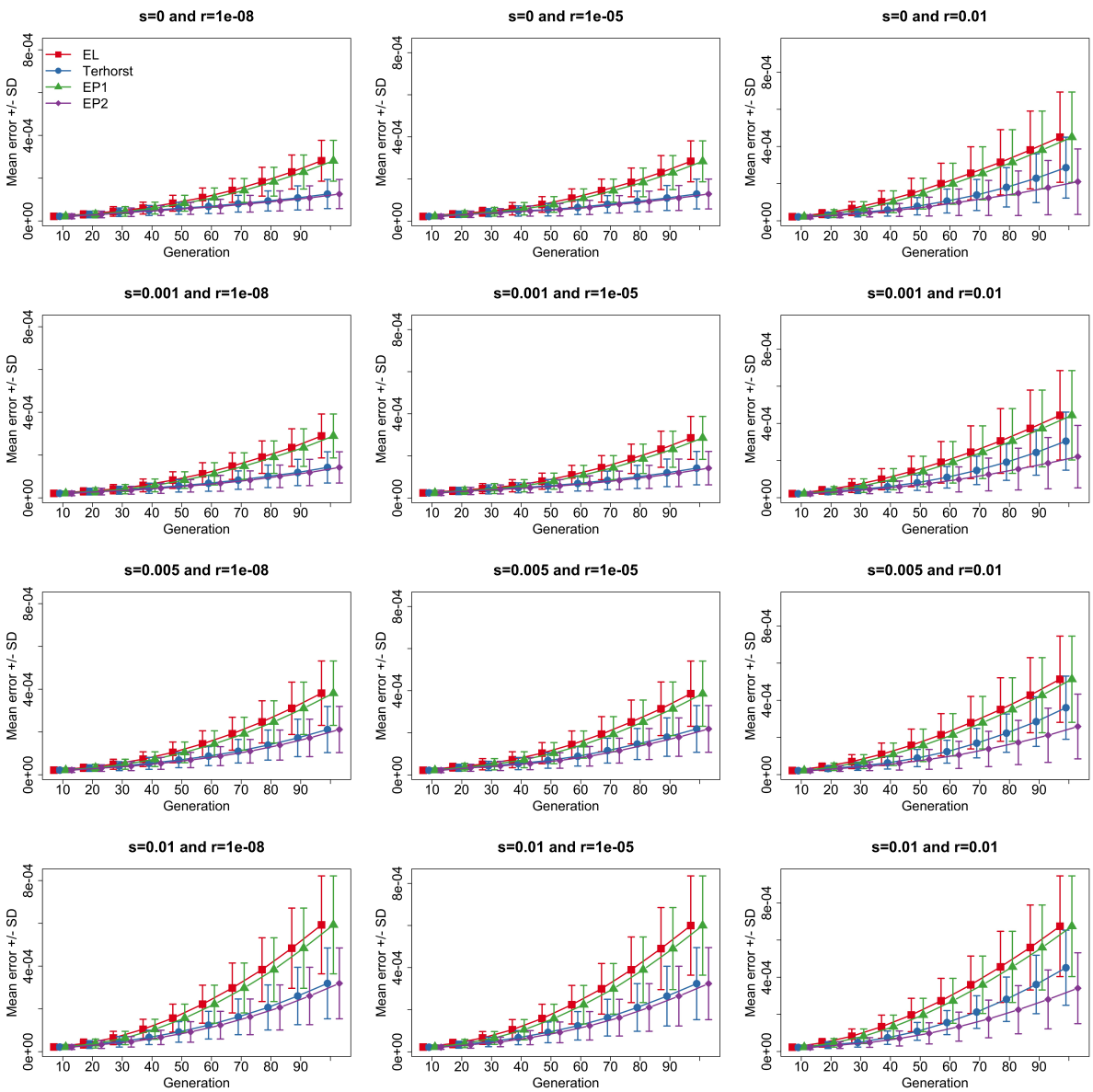

Figure 3: Comparison between the mean vector of the Wright-Fisher model and their approximations for the case of the population size $N = 5000$, where we fix the selection coefficient $s_{\mathcal{A}} = 0.005$ and vary the selection coefficient $s_{\mathcal{B}} \in \{0, 0.001, 0.005, 0.01\}$ and the recombination rate $r \in \{0.00000001, 0.00001, 0.01\}$. To aid visual comparison of different approximation schemes, we plot the approximation error for each scheme using slightly different x coordinate ticks even though the x coordinates are exactly the same. We adopt similar practice for all subsequent figures.

panel of Figure 3). Although the error of all mean approximations increases over time, the approach of Terhorst et al. (2015) sees a much faster increase in the approximation error than the other three methods when the two loci are loosely linked. In Supplemental Material, Figures S1 and S3, we illustrate the results for the cases of the population size $N = 500$ and $N = 50000$. They reveal similar behaviour in approximation errors.

For the approximation error of the (co)variance matrix, we observe from Figure 4 that the

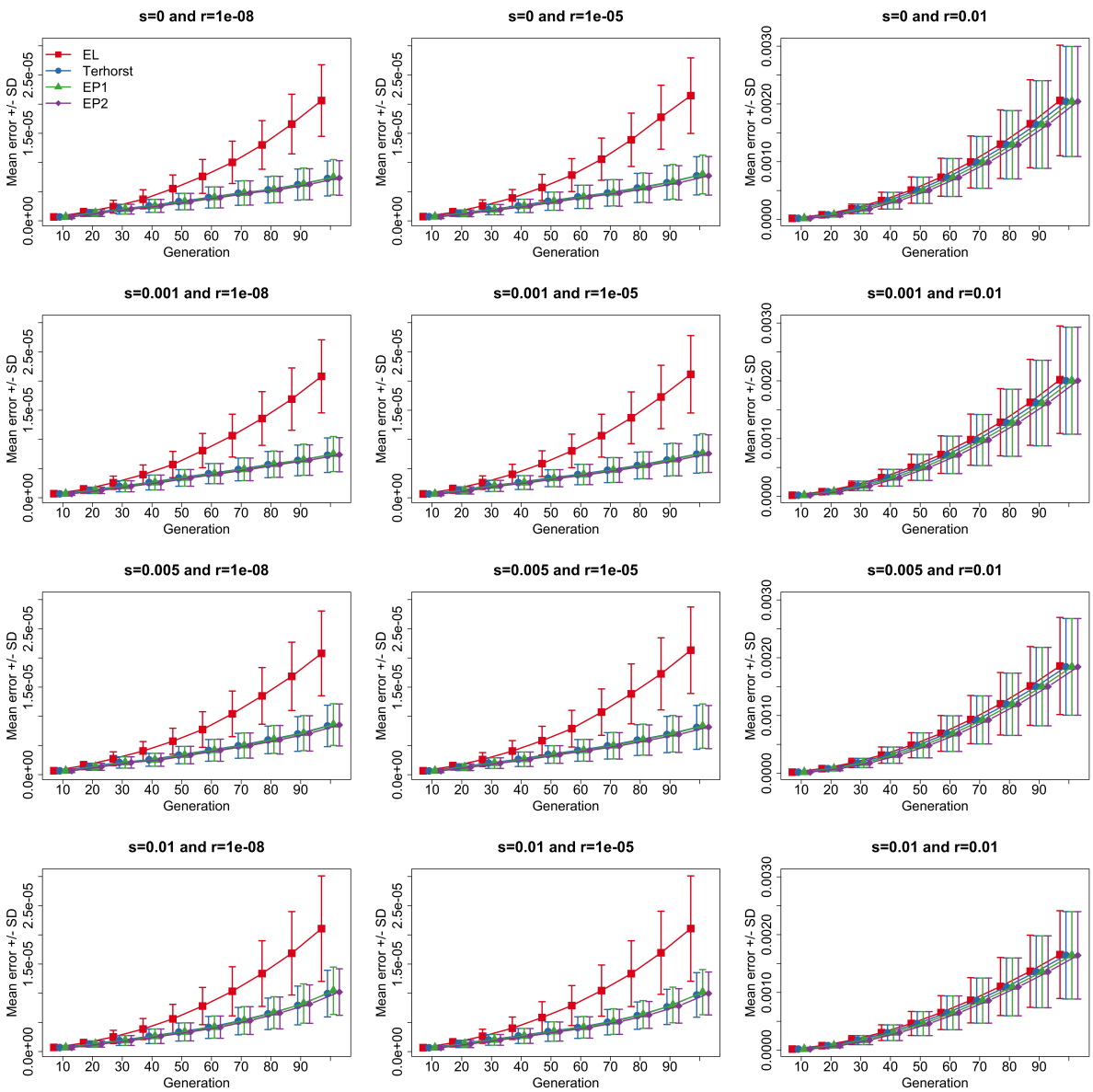

Figure 4: Comparison between the (co)variance matrix of the Wright-Fisher model and their approximations for the case of the population size $N = 5000$, where we fix the selection coefficient $s_{\mathcal{A}} = 0.005$ and vary the selection coefficient $s_{\mathcal{B}} \in \{0, 0.001, 0.005, 0.01\}$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

approach of Terhorst et al. (2015) and our methods EP1 and EP2 have better performance than our method EL. This advantage is diminished with the decrease in the selection coefficient and/or recombination rate, to nearly non-existent for loosely linked loci (see the right panel of Figure 4). Compared to the approach of Terhorst et al. (2015), our method EP2 matches its performance, but our method EP1 performs slightly worse. Combining with their performance for the cases of population sizes $N = 500$ and $N = 50000$ (see Supplemental Material, Figures S2 and S4), we find that the performance advantage of the approximation of Terhorst et al. (2015)

and our methods EP1 and EP2 weakens with the increase in the population size.

In conclusion, moment approximations overall work well. Their performance becomes better with the decrease in the selection coefficient and/or recombination rate, but the approximation error increases with time. Amongst the four moment approximation schemes, our extension of Paris et al. (2019) up to the second order (method EP2) achieves the best performance across all 36 possible combinations of the selection coefficient, recombination rate and population size. More accurate approximations for the mean and (co)variance of the Wright-Fisher model are available if we keep third- or even higher-order terms in the Taylor expansion in our extension of Paris et al. (2019).

4.2. Performance of transition probability density approximations

To evaluate the performance of our logistic normal and hierarchical beta approximations for the Wright-Fisher model, we simulate them alongside the diffusion and normal approximations of the Wright-Fisher model, so that we can compare the approximation schemes we introduce with these two widely used methods. More specifically, we perform 300 replicated runs for each of the 36 possible combinations of the selection coefficient, recombination rate and population size. For each run, we randomly draw the starting gamete frequencies $\boldsymbol{X}_0$ from a flat Dirichlet distribution over the state space $\Omega_{\boldsymbol{X}}$ and simulate 1,000,000 gamete frequency trajectories for 100 generations according to the Wright-Fisher model and its approximations. The mean vector and (co)variance matrix of the Wright-Fisher model required in the moment-based approximation are calculated with our extension of Paris et al. (2019) up to the second order, *i.e.*, method EP2 from Section 4.1. The diffusion approximation is simulated with the procedure of He et al. (2020a). Evaluation of the quality of the approximations is carried out by comparing the empirical cumulative distribution functions (CDFs) computed from these simulated gamete frequency trajectories. We randomly choose 5000 vectors of gamete frequencies from a flat Dirichlet distribution on the state space $\Omega_{\boldsymbol{X}}$, and the empirical 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 using 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, recombination rate and pop-

ulation size. Compared to 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, even though some combinations of the selection coefficient, recombination rate and population size do not completely satisfy the assumption required in the diffusion approximation (see, *e.g.*, He et al., 2020a). 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, we generate 1,000,000 realisations. This takes about 181.358 seconds on a single core of an Intel Core i7 processor at 4.2 GHz with the diffusion approximation. In contrast, this only takes approximately 0.001 seconds with the moment-based approximations. Additionally, using diffusion approximation, at every 10 generations 1,000,000 realisations are required to be stored for the computation of the empirical CDFs, 341.386 MB in total, whereas in moment-based approximations we only need to store the relevant parameters, *e.g.*, the shape parameters for the hierarchical beta distribution at every 10 generations, 523.2 bytes in total.

As illustrated in Figures 5-7, moment-based approximations have better overall performance as the population size increases and become worse with longer times. This is consistent with the conclusion for the single-locus case presented in Lacerda & Seoighe (2014). In the three moment-based approximations, our logistic normal approximation achieves the worst overall performance except for small population sizes. We find from Figure 5 that our logistic normal approximation achieves the worst performance during the initial tens of generations and then achieves better performance than the normal approximation with longer times for small population sizes. In general, alleles drift toward loss or fixation 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 $\Omega_{\boldsymbol{X}}$ owing to the infinite support of the normal distribution. Therefore, our logistic normal approximation works better than the normal approximation only when the allele frequencies approach the boundaries.

Figures 5-7 also show that our hierarchical beta approximation yields the best overall performance amongst the three moment-based approximations across all 36 possible combinations of the selection coefficient, recombination rate and population size and even matches the perfor-

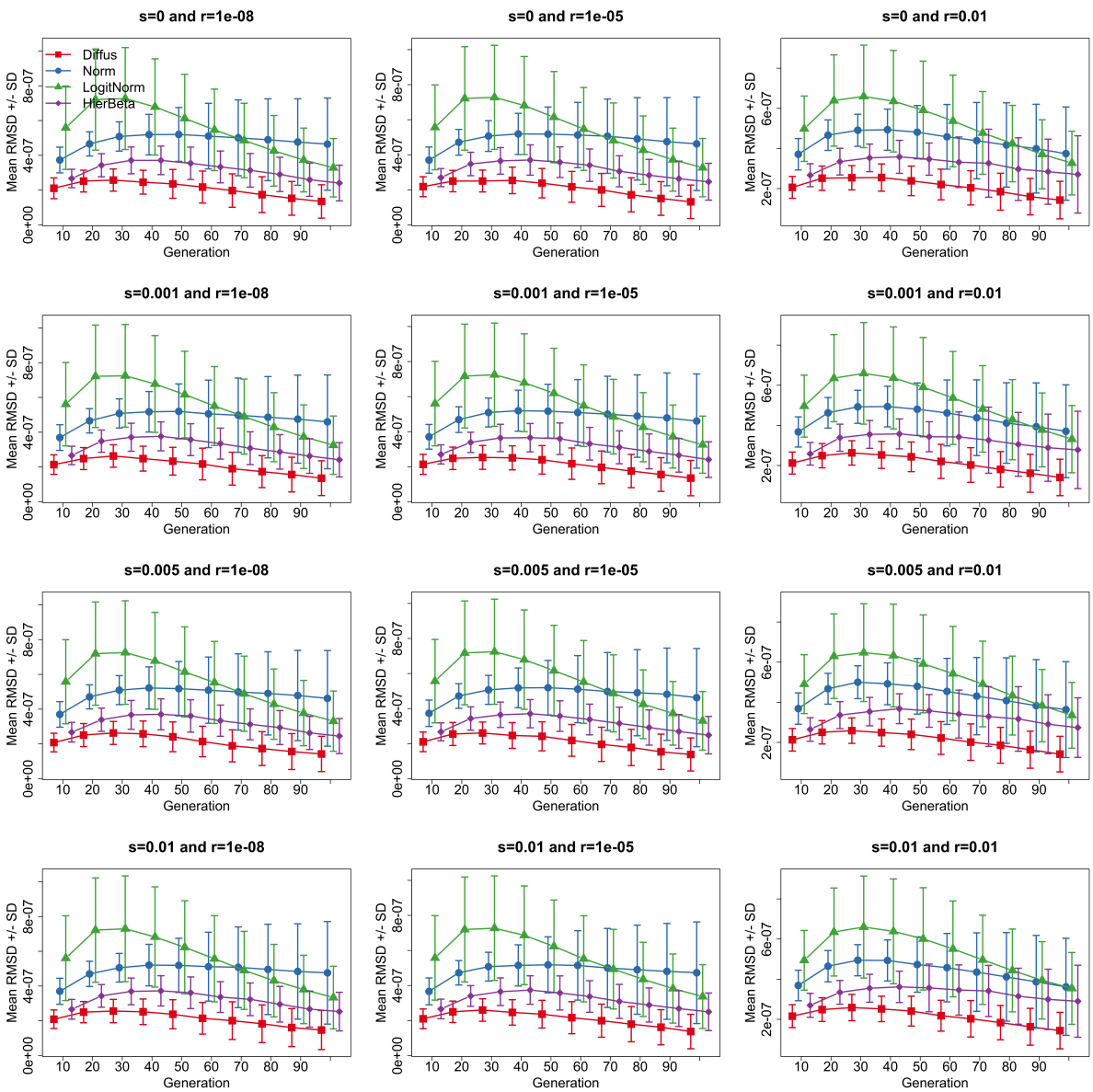

Figure 5: Comparison between the empirical CDFs of the Wright-Fisher model and their approximations for the case of the population size $N = 500$, where we fix the selection coefficient $s_{\mathcal{A}} = 0.005$ and vary the selection coefficient $s_{\mathcal{B}} \in \{0, 0.001, 0.005, 0.01\}$ and the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

mance of the diffusion approximation for small recombination rates and large population sizes (see the left two panels of Figure 7). However, it should be noted that in practice non-positive 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 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 observe that the failure proportion has a trend of slow increase over time for small recombina-

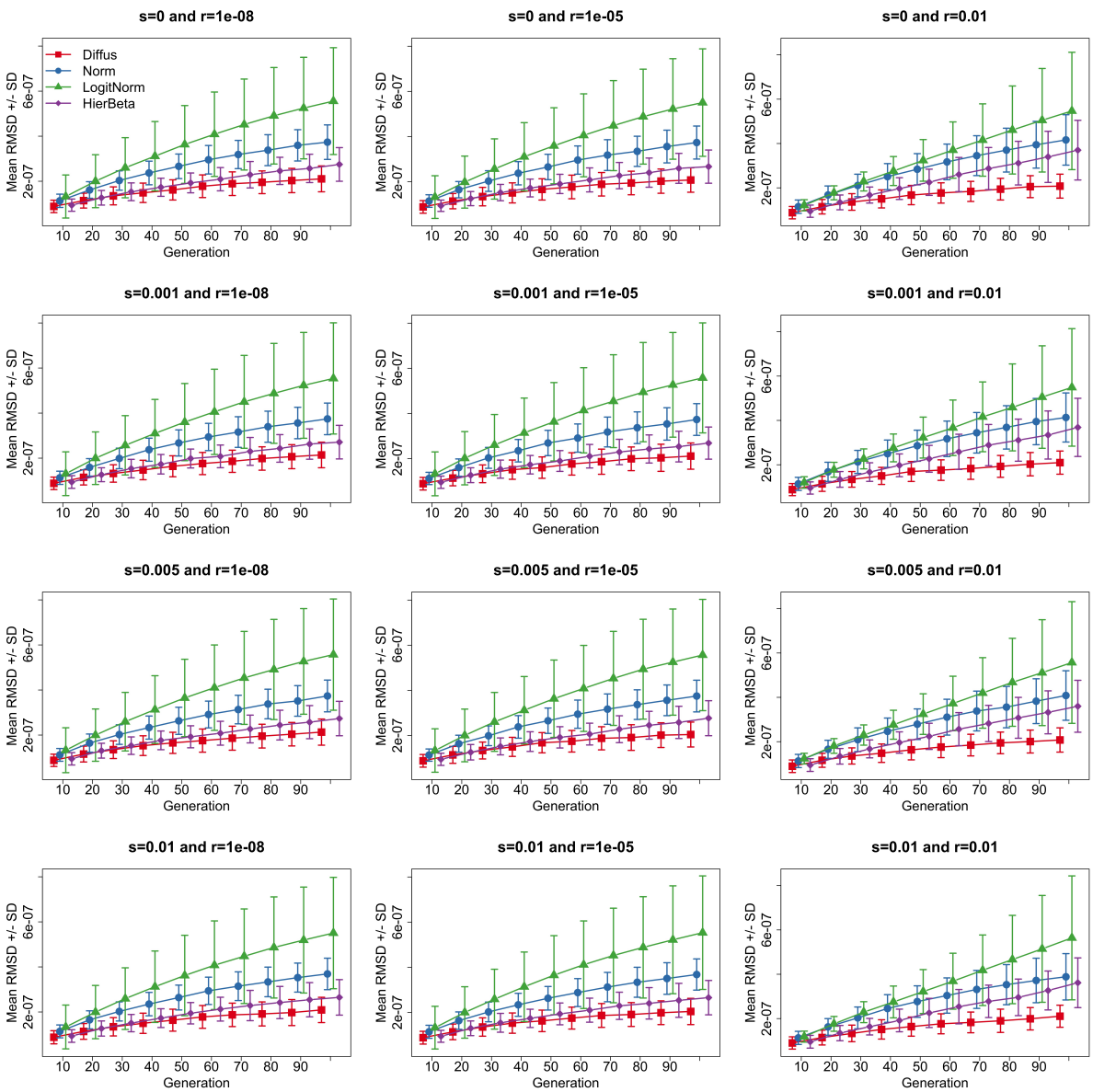

Figure 6: Comparison between the empirical CDFs of the Wright-Fisher model and their approximations for the case of the population size $N = 5000$, where we fix the selection coefficient $s_{\mathcal{A}} = 0.005$ and vary the selection coefficient $s_{\mathcal{B}} \in \{0, 0.001, 0.005, 0.01\}$ and the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

tion rates but shows a sharp increase for large recombination rates. We also see from Figure 8 that the failure proportion slightly increases with the decrease in the population size, which becomes significant for large recombination rates. Such a computational instability issue could be caused by that our moment approximation gets less accurate with time and our hierarchical beta approximation fails to model fixation or loss of alleles.

To resolve the computational instability issue of our hierarchical beta approximation, if the three selected gamete frequencies, *e.g.*, $X_1(k)$, $X_2(k)$ and $X_3(k)$, fail to produce positive shape

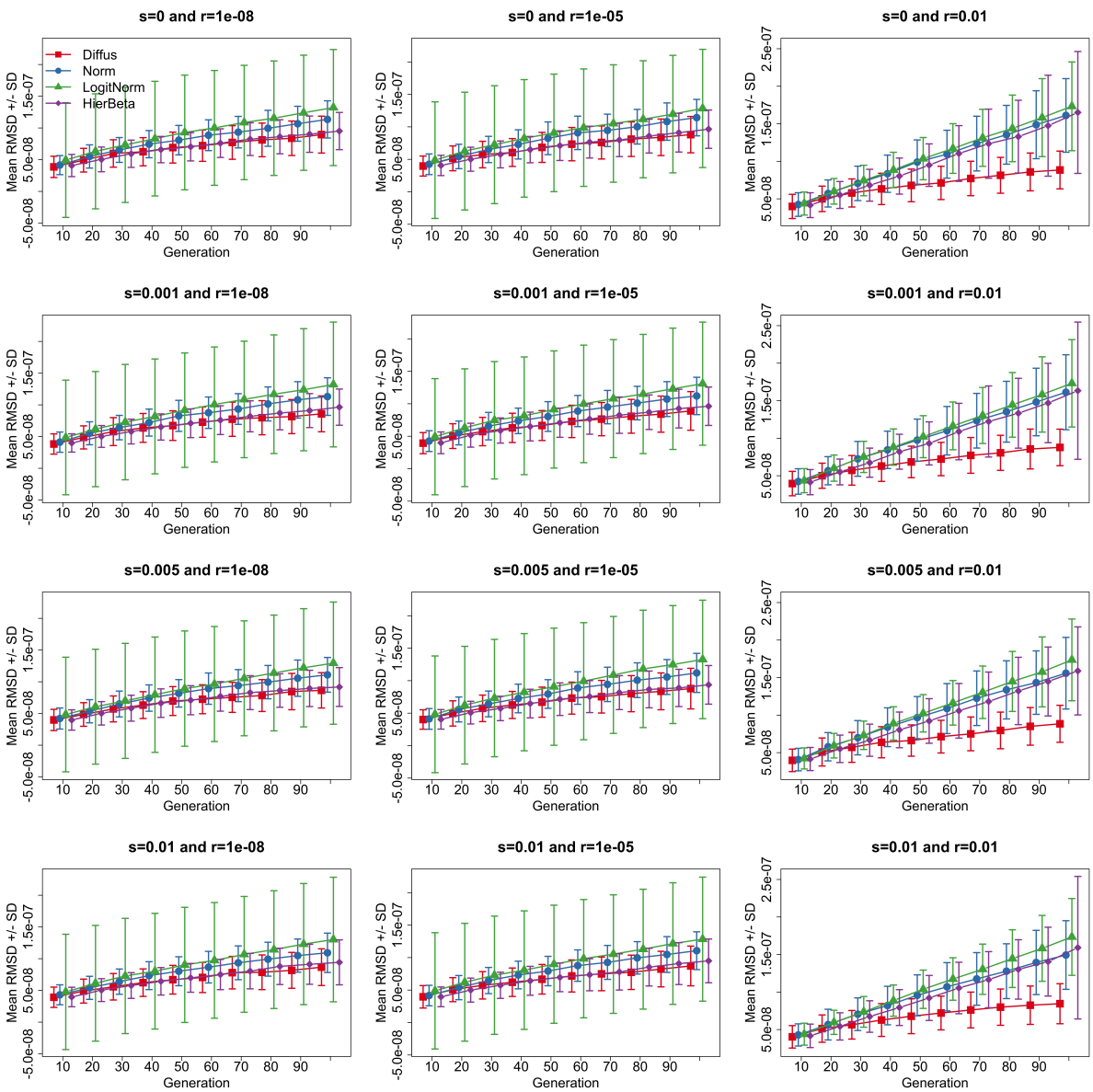

Figure 7: Comparison between the empirical CDFs of the Wright-Fisher model and their approximations for the case of the population size $N = 50000$, where we fix the selection coefficient $s_{\mathcal{A}} = 0.005$ and vary the selection coefficient $s_{\mathcal{B}} \in \{0, 0.001, 0.005, 0.01\}$ and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

parameters for each beta distribution, we choose other three gamete frequencies to construct our hierarchical beta approximation instead. Permuting the indices of our hierarchical beta approximation can completely resolve the failure cases for small recombination rates (see the left two panels of Figure 9) and significantly reduce the failure proportions for large recombination rates (see the right panel of Figure 9). For the remaining cases of failure, *i.e.*, those which permutation of the indices of the hierarchical beta approximation cannot resolve, we search for nearby mean and variance values that do yield hierarchical beta approximation with positive

shape parameters. Such a procedure completely eliminates the computational instability issue but would result in additional errors. In the simulation studies we have performed, we have used the procedure outlined in this paragraph. As shown in Figures 5-7, our hierarchical beta approximation achieves the best overall performance among the three moment-based approximations even though it needs these failure prevention steps.

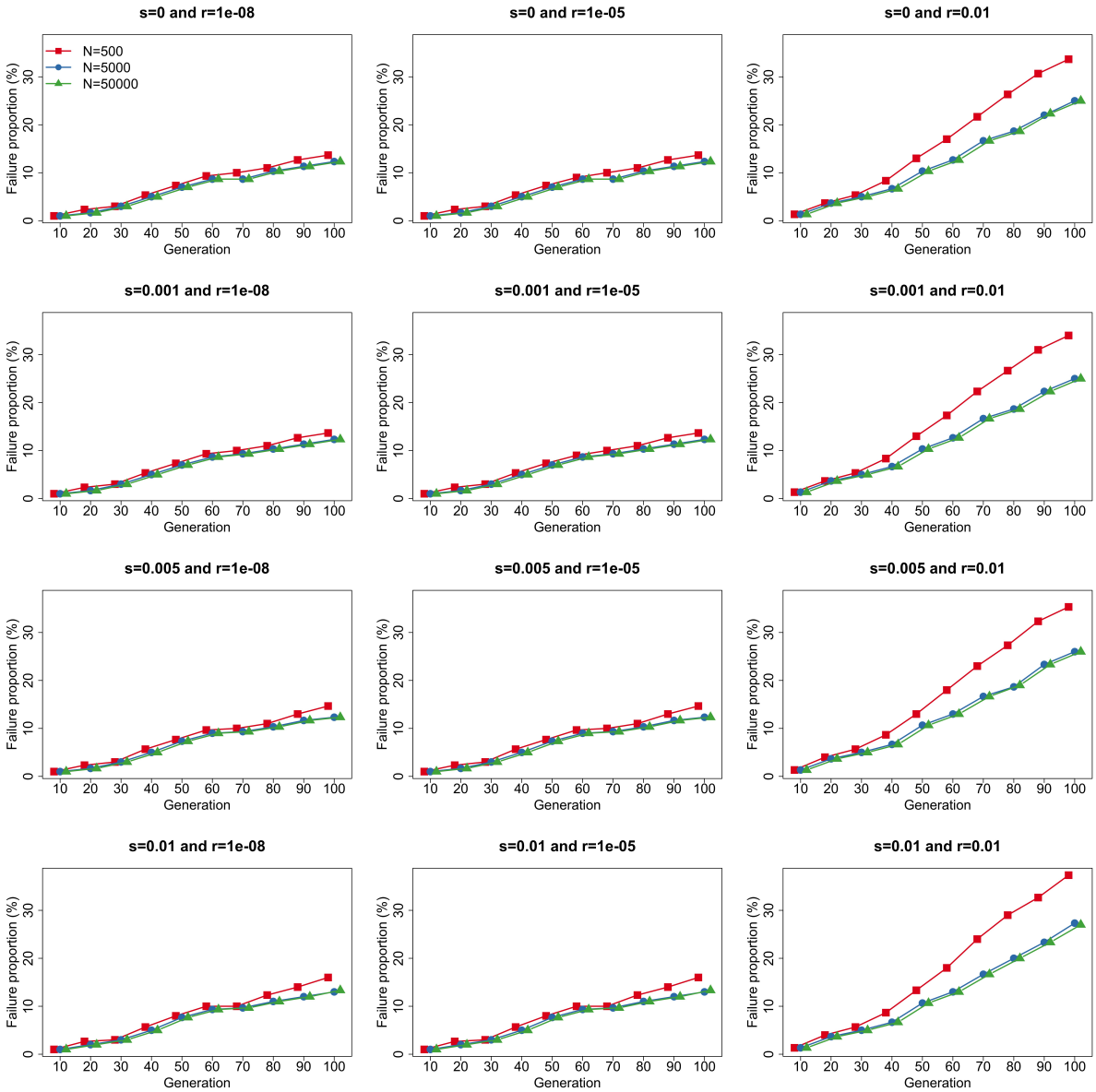

Figure 8: Proportions of the cases that the Wright-Fisher model cannot be fitted through our hierarchical beta approximation across all 36 possible combinations of the selection coefficient, recombination rate and population size.

In summary, the moment-based approximations enable us to compute the approximate transition probabilities of the Wright-Fisher model at a much smaller computational cost than the

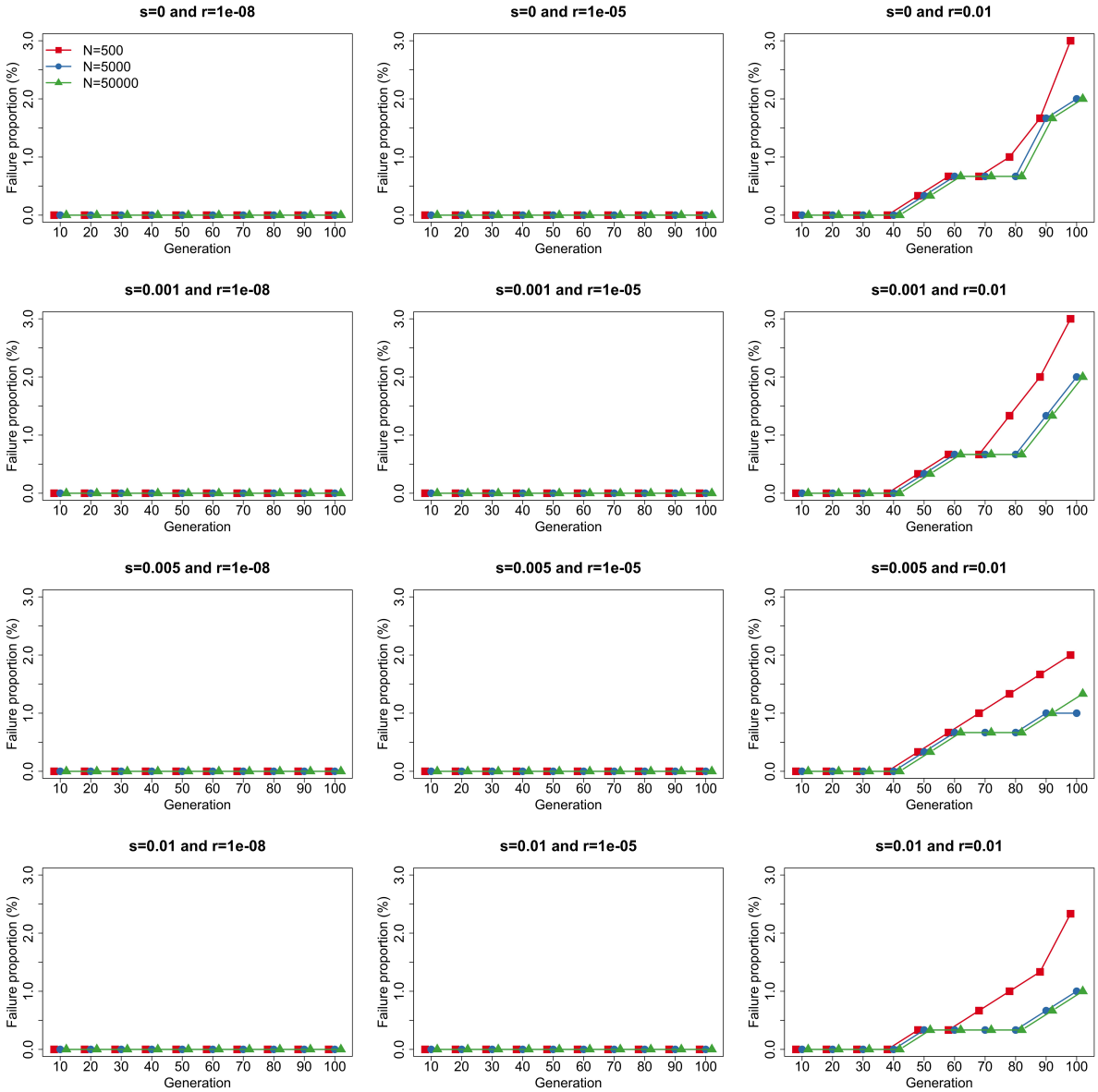

Figure 9: Proportions of the cases that the Wright-Fisher model cannot be fitted through our hierarchical beta approximation (with permutation) across all 36 possible combinations of the selection coefficient, recombination rate and population size. The remaining failure cases are resolved by the failure prevention steps as explained above.

diffusion approximation, even though the diffusion approximation achieves higher computational accuracy. Compared to the normal approximation of Terhorst et al. (2015), our logistic normal approximation achieves higher computational accuracy only when the Wright-Fisher model gets close to the boundaries. Our hierarchical beta approximation achieves the highest computational accuracy among the three moment-based approximations and even matches the performance of the diffusion approximation for small recombination rates and large population sizes. For large recombination rates, the performance advantage of our hierarchical beta approx-

imation is diminished. However, the two loosely linked loci are approximately independent with each other due to the large number of recombinations, therefore can typically be characterised through the single-locus Wright-Fisher model in practice.

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 a parametric family of probability distributions, dating back to Cavalli-Sforza & Edwards (1967). A prerequisite for the moment-based approximation is that we can calculate the mean and (co)variance of the two-locus Wright-Fisher model with selection at any given time point, so we extended the approaches of Lacerda & Seoighe (2014) and Paris et al. (2019), which in their original form were limited to the Wright-Fisher model of population dynamics under natural selection at a single locus. A major advantage of our logistic normal and hierarchical beta approximations, compared to *e.g.*, the Gaussian approximation of Terhorst et al. (2015), is that we avoid the distribution support issue.

Through extensive simulations studies, we illustrated that our approximations for the mean and (co)variance achieved better performance with smaller selection coefficients and/or recombination rates, but has deteriorating performance with the increase of time horizon, owing to the accumulation of approximation errors. Our extension of Paris et al. (2019) up to the second order achieved the best performance in the four moment approximations we compared. In our simulation studies, our moment-based approximations could not yield the same computational accuracy as the diffusion approximation but enabled the computation of the transition probabilities of the Wright-Fisher model at a fraction of the computational cost. The performance of the moment-based approximations was better for larger population sizes but deteriorated with the increase of time horizon. This is consistent with the conclusion stated in Lacerda & Seoighe (2014) for the single-locus case. The performance of our logistic normal approximation was not as good as the Gaussian approximation of Terhorst et al. (2015) except when the Wright-Fisher model approached the boundaries and began to be absorbed. In contrast, our hierarchical beta approximation outperformed the Gaussian approximation and even matched the performance of the diffusion approximation for small recombination rates and large population sizes.

In recent years, the Wright-Fisher model has been successfully applied to the 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. This may lead to a poor inference of natural selection, especially for the case of tightly linked loci (He et al., 2020b). Existing approaches that allow us to account for genetic recombination and local linkage suffer from either distribution support issues in Terhorst et al. (2015) or prohibitive computational costs in He et al. (2020b). Our moment-based approximations provide an alternative way to carry out the likelihood computation. For example, our hierarchical beta approximation has been shown to balance computational accuracy and efficiency in our simulation studies. It can therefore be expected to improve the performance of existing statistical frameworks for the inference of natural selection from time series genetic data. Moreover, our hierarchical beta approximation makes it possible to conduct genome-wide inference of natural selection from time series genomic data, modelling genetic recombination and local linkage in a maximum composite likelihood framework. Since the computational accuracy of our hierarchical beta approximation becomes worse with the increase of time horizon, it is better suited to 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 improvements to our moment-based approximations may be possible. 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) developed the pyramid hierarchical beta approximation of the Wright-Fisher model for the evolution of relative frequencies of four alleles at a single locus, which has enough free parameters to characterise the entire (co)variance structure. However, it cannot be extended to the two-locus Wright-Fisher model with selection since only a small subset of the state space can guarantee the existence of positive shape parameters for each beta distribution in their pyramid hierarchical beta approximation. It would be fruitful to find other parametric families of probability distributions that enable the complete capture of 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 the evolution of relative frequencies of two alleles at a single locus. 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 shown to improve the performance of the beta approximation. It would be interesting to find ways 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.

Hui, T.-Y. J., & Burt, A. (2015). Estimating effective population size from temporally spaced samples with a novel, efficient maximum-likelihood algorithm. *Genetics*, *200*, 285–293.

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.

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.

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genetics*, *8*, e1002967.

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

Sirén, J., Marttinen, P., & Corander, J. (2011). Reconstructing population histories from single nucleotide polymorphism data. *Molecular Biology and Evolution*, *28*, 673–683.

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.