Journal of Theoretical Biology 269 (2011) 79–87

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Comparison and content of the Wright–Fisher model of random genetic drift, the diffusion approximation, and an intermediate model

D. Waxman *

Centre for Computational Systems Biology, Fudan University, Shanghai, People's Republic of China

ARTICLE INFO

Article history:
Received 1 June 2010
Received in revised form
9 October 2010
Accepted 12 October 2010
Available online 19 October 2010

This work is dedicated to the memory of my mother

Keywords:
Random genetic drift
Diffusion analysis
Wright–Fisher model
Theory

ABSTRACT

We investigate the detailed connection between the Wright–Fisher model of random genetic drift and the diffusion approximation, under the assumption that selection and drift are weak and so cause small changes over a single generation. A representation of the mathematics underlying the Wright–Fisher model is introduced which allows the connection to be made with the corresponding mathematics underlying the diffusion approximation. Two 'hybrid' models are also introduced which lie 'between' the Wright–Fisher model and the diffusion approximation. In model 1 the relative allele frequency takes discrete values while time is continuous; in model 2 time is discrete and relative allele frequency is continuous. While both hybrid models appear to have a similar status and the same level of plausibility, the different nature of time and frequency in the two models leads to significant mathematical differences. Model 2 is mathematically inconsistent and has to be ruled out as being meaningful. Model 1 is used to clarify the content of Kimura's solution of the diffusion equation, which is shown to have the natural interpretation as describing only those populations where alleles are segregating. By contrast the Wright–Fisher model and the solution of the diffusion equation of McKane and Waxman cover populations of all categories, namely populations where alleles segregate, are lost, or fix.

1. Introduction

Random genetic drift is said to occur when, largely independent of selection, mutation and migration, there is variation in the contribution of adults of one generation to offspring of the next generation. Random genetic drift is a process associated with a finite number of individuals. It results in the number of copies of a particular allele at a locus randomly varying from generation to generation. One of the key mathematical approaches for dealing with genetic drift is the diffusion approximation. This was introduced into population genetics by Fisher (1922) and Wright (1931) and substantially extended and applied by Kimura (1955a).

Under a diffusion approximation, the relative frequency of an allele is treated as a continuous random variable whose distribution obeys a diffusion equation. This approach has been used to derive many fundamental results and insights into evolution and population genetics (Crow and Kimura, 1970).

A notable result derived by Kimura is an exact time-dependent solution of the diffusion equation for a selectively neutral population (Kimura, 1955b). However, in a paper by McKane and the author, a different solution of the diffusion equation was established which, in contrast to Kimura's result, has the property of conserving probability for all times (McKane and Waxman, 2007). McKane and Waxman solved the forward diffusion equation under a boundary condition that ensures that no probability is lost. The processes of fixation and loss then directly emerged in the guise of singular spikes, i.e., Dirac delta functions (Dirac, 1958), in the solution at the relative frequencies 0 and 1. When the total integrated probability includes the contribution of the two spikes, its value is conserved and takes the value of unity for all values of the time.

In the present work we present a clarification of the connection of the Wright–Fisher model of random genetic drift with the diffusion approximation. There are considerable mathematical subtleties associated with the diffusion approximation, and we largely avoid these by proceeding as follows.

1. We introduce a representation of the mathematics underlying the Wright–Fisher model (namely the eigenvectors associated with the model) that allows a transparently clear connection with the corresponding mathematics underlying the diffusion approximation (namely the eigenfunctions of a diffusion operator).
2. We introduce two instructive 'hybrid' models which can be viewed as lying 'between' the Wright–Fisher model and the diffusion approximation. The intermediate character of these models follows from the nature of time or allele frequency

* Correspondence address: Faculty of Science and Engineering, Cockroft Building, University of Brighton, Brighton BN2 4GJ, UK.
E-mail address: dwaxman.biology@gmail.com

0022-5193/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2010.10.014

adopted. In the Wright–Fisher model, both time and allele frequency take discrete values, while in the diffusion approximation, both of these quantities take continuous values. In one of the hybrid models introduced here, time is taken as a continuous quantity, while the relative allele frequency remains discrete; in the other hybrid model, relative frequency is taken as continuous while time remains discrete. Despite the fact that both hybrid models appear to have a similar status and the same level of plausibility, the different nature of time and frequency in the two models leads to significant mathematical differences. We find that one of the hybrid models (where the relative frequency takes continuous values but time remains discrete) is mathematically inconsistent and has to be ruled out as a meaningful model.

3. We use the remaining hybrid model (where time is continuous and relative frequency discrete) to clarify the content of Kimura's solution of the diffusion equation. Our analysis allows us to show that the solution of the diffusion equation obtained by Kimura can be naturally interpreted as describing only those populations where alleles are segregating. By contrast the Wright–Fisher model and the solution of the diffusion equation of McKane and Waxman (2007) cover all categories of populations, namely populations where alleles segregate, are lost, or fix.

2. Wright–Fisher model

We shall consider a Wright–Fisher model of genetic drift for a population of N haploid individuals that undergo random mating (i.e., that have a brief diploid stage of their predominantly haploid life cycle) and are subject to viability selection. For definiteness, all calculations in this work will be restricted to such haploid sexual populations, which cover organisms such as the green alga *Ulothrix* (Mable and Otto, 1998). However, we note that under the substitution $N\to 2N$ the same results apply to randomly mating diploid organisms that are subject to genic selection: see e.g., Gale (1990, p. 129). This applicability means the conclusions of this work also apply to 'standard diploid populations.'

To begin, consider a population of randomly mating haploid individuals with a single locus under selection. We take there to be two alleles at this locus, labelled A and a, and generations are assumed to be non-overlapping. In the adults of generation t ($=0,1,2,\ldots$) the proportion of alleles at the locus that are allele A is written $X(t)$. This is the relative frequency (henceforth termed *frequency*) of allele A in adults; the corresponding frequency of allele a in adults is $1-X(t)$.

In a very large (effectively infinite) population, the frequency $X(t)$ changes according to the deterministic rule[1]

$$X(t+1)=C(X(t)). \tag{1}$$

In this equation the function $C(x)$ generally depends on mutation, migration and selection. In the absence of these processes $C(x)=x$, in which case the allele frequency $X(t)$ takes the same value in all generations in a very large population.

In the present work we incorporate viability selection into the dynamics and assume there are no fertility differences between different genotypes. The relative viabilities of the A and a haploid genotypes are taken to be $1+s$ and 1, respectively, so the A allele confers a fitness advantage of s over the a allele. The function $C(x)$ is then given by

$$C(x)=x+\frac{sx(1-x)}{1+sx}. \tag{2}$$

[1] This rule holds unless the A allele is at very low copy numbers.

In a *finite* population, a non-selective ecological thinning process is taken to occur that ensures that N adults are maintained in a population every generation. The lifecycle adopted is

Generation t **adults (haploid)**
↓ random mating followed by the production of haploid offspring and the death of all adults
offspring
↓ viability selection
juveniles
↓ thinning (number regulation)
Generation $t+1$ **adults.**

Assuming the number of offspring in a population is very large ($\gg N$), the principal place within the lifecycle where randomness occurs is when the number of juveniles is thinned to N individuals.

Thinning effectively corresponds to randomly picking N juveniles, without replacement, to become the adults of the next generation. It results in the frequency of the A allele in adults having a hypergeometric distribution (for more information on this distribution, see e.g., the textbook by Haigh, 2002). Making the assumption that the number of adults, N, is small compared with the number of juveniles, the hypergeometric distribution can be well-approximated by a binomial distribution. Given such a binomial distribution, the relation between the frequency of the A allele in generations t and $t+1$ is

$$X(t+1)=\frac{\mathrm{Bin}(N,C(X(t)))}{N}, \tag{3}$$

where $\mathrm{Bin}(n,p)$ denotes a random number that is drawn from a binomial distribution with parameters n (number of trials) and p (probability of success on each trial). We thus have a Wright–Fisher model of random genetic drift (Fisher, 1922; Wright, 1931).

The statistical description of the allele frequency $X(t)$ in a Wright–Fisher model arises from consideration of an infinite number of replicate populations which each maintain N adults every generation. Possible values of $X(t)$ follow from Eq. (3) and are given by

$$x_n=n/N \quad \text{where } n=0,1,\ldots,N. \tag{4}$$

When $X(t)$ has the value x_n it corresponds to a population where n of the alleles at the locus in question are the A allele.

We write the probability that $X(t)$ has the value x_n as $F(x_n,t)$. The change of the $F(x_n,t)$ over time follows from Eq. (3) and is given by $F(x_n,t+1)=\sum_{m=0}^{N}W_{n,m}F(x_m,t)$ where

$$W_{n,m}=\binom{N}{n}[C(x_m)]^n[1-C(x_m)]^{N-n}, \quad n,m=0,1,2,\ldots,N \tag{5}$$

and in this expression $\binom{a}{b}=a!/[(a-b)!b!]$ denotes a binomial coefficient. We adopt a matrix representation where the $F(x_n,t)$ are elements of a column vector $\mathbf{F}(t)$ and the $W_{n,m}$ are elements of a matrix $\mathbf{W}$. The equation governing the change of $\mathbf{F}(t)$ can then be written as $\mathbf{F}(t+1)=\mathbf{W}\mathbf{F}(t)$.

Note that where we have written $W_{n,m}$, in Eq. (5), it is standard in the study of Markov processes to write $W_{m,n}$. Thus in the standard formulation, the matrix $\mathbf{W}$ is replaced by its transpose, which acts from the *right* on a *row* vector of probabilities. In the present work it is more natural to work with the formulation presented above, which leads to $\mathbf{F}(t+1)=\mathbf{W}\mathbf{F}(t)$. Furthermore, with $\mathbf{I}$ an identity matrix of size $(N+1)\times(N+1)$ and $\mathbf{M}$ the matrix

$$\mathbf{M}=\mathbf{I}-\mathbf{W} \tag{6}$$

we can write the equation governing the change of $\mathbf{F}(t)$ as

$$-[\mathbf{F}(t+1)-\mathbf{F}(t)]=\mathbf{M}\mathbf{F}(t). \tag{7}$$

3. Different approximations leading to two 'hybrid' models

Eq. (7) is an exact formulation for the dynamics of the distribution $\mathbf{F}(t)$ in the Wright–Fisher model. It has been adopted for the present work since this form of equation is close to that of a diffusion equation. All that it would take to obtain a diffusion equation from Eq. (7) would be to

(i) assume changes in $\mathbf{F}(t)$ from one generation to the next are small (i.e., selection and drift are weak), so time can be approximated as a continuous quantity, thereby allowing $\mathbf{F}(t+1)-\mathbf{F}(t)$ to be approximated by $d\mathbf{F}(t)/dt$;
(ii) approximate x_n as a continuous variable x and, at the same time, replace $\mathbf{M}$ (which appears on the right hand side of Eq. (7)) by a differential operator of diffusion form.

It is convenient to give here the relevant diffusion operator, which we denote by D. For weak selection ($|s| \ll 1$), we approximate $C(x)$ of Eqs. (1) and (2) by $C(x)=x+sx(1-x)$ and the action of D on a function $f(x)$ is then given by (Kimura, 1955a)

$$Df(x) \equiv -\frac{1}{2N}\frac{\partial^2[x(1-x)f(x)]}{\partial x^2}+\frac{\partial[sx(1-x)f(x)]}{\partial x}. \tag{8}$$

The approximations (i) and (ii) above can be *separately* applied to the Wright–Fisher model; they lead to two approximations that can be viewed as intermediate or 'hybrid' models that lie between the Wright–Fisher model and the diffusion approximation. The full set of models is summarised in Table 1.

Table 1
This table summarises the salient features of the Wright–Fisher model, two hybrid models that take either time or frequency as a continuous variable, and the diffusion approximation, where both time and frequency take continuous values.

Model	Time	Frequency	Dynamical equation
Wright–Fisher	Discrete	Discrete	$-[\mathbf{F}(t+1)-\mathbf{F}(t)]=\mathbf{M}\mathbf{F}(t)$
Hybrid model 1	Continuous	Discrete	$-d\mathbf{F}(t)/dt=\mathbf{M}\mathbf{F}(t)$
Hybrid model 2	Discrete	Continuous	$-[f(x,t+1)-f(x,t)]=Df(x,t)$
Diffusion	Continuous	Continuous	$-\partial f(x,t)/\partial t=Df(x,t)$

When the frequency takes discrete values the column vector $\mathbf{F}(t)$ denotes the probability distribution of the A allele frequency at time t. When the frequency takes continuous values the quantity $f(x,t)$ denotes the probability density of the A allele frequency at time t and frequency x. The matrix $\mathbf{M}$ determines the dynamics when the frequency takes discrete values, and follows from Eqs. (5) and (6). The analogue of $\mathbf{M}$, for a frequency that takes continuous values, is the diffusion operator D which is given in Eq. (8).

4. Formal solutions of the different models

In order to solve the four models given in Table 1, we shall employ the eigenvectors of the matrix $\mathbf{M}$ or the eigenfunctions of the diffusion operator D. We make the assumption that $\mathbf{M}$ is diagonalisable and so has a complete set of eigenvectors, i.e., any vector can be expressed as a linear superposition of the eigenvectors of $\mathbf{M}$. Evidence for the diagonalisability of $\mathbf{M}$ is given in the next section. The diffusion operator D is, in a very definite sense, an approximation of the matrix $\mathbf{M}$, hence we also assume the diffusion operator D has a complete set of eigenfunctions, so any function can be expressed as a linear superposition of its eigenfunctions.

For the matrix $\mathbf{M}$ we write the kth *right* eigenvector as $\mathbf{\Phi}^{(k)}$. This is a column vector with elements $\Phi_n^{(k)}$ where both n and k can take the values $0,1,2,\ldots,N$. The kth *left* eigenvector of $\mathbf{M}$ is a row vector which is written as the transpose of the column vector $\mathbf{\Psi}^{(k)}$ and whose elements are $\Psi_n^{(k)}$. With a T superscript denoting the transpose of a matrix, the eigenvalue equations for $\mathbf{M}$ are

$$\left.\begin{aligned}\mathbf{M}\mathbf{\Phi}^{(k)}&=\lambda_k\mathbf{\Phi}^{(k)}\\ \mathbf{\Psi}^{(k)T}\mathbf{M}&=\lambda_k\mathbf{\Psi}^{(k)T}\end{aligned}\right\}\quad k=0,1,2,\ldots,N \tag{9}$$

and λ_k are the eigenvalues.

For a matrix $\mathbf{M}$ that is diagonalisable, the eigenvectors $\mathbf{\Psi}^{(k)T}$ and $\mathbf{\Phi}^{(k)}$ can be chosen to be both orthogonal when the labels differ ($\mathbf{\Psi}^{(j)T}\mathbf{\Phi}^{(k)}=0$ when $j\neq k$), and normalised to unity ($\mathbf{\Psi}^{(k)T}\mathbf{\Phi}^{(k)}=1$).

The eigenvalue equations for the diffusion operator D that are analogous to Eq. (9) are the forward and backward equations

$$\left.\begin{aligned}-\frac{1}{2N}\frac{\partial^2}{\partial x^2}[x(1-x)\Phi_k(x)]+\frac{\partial}{\partial x}[sx(1-x)\Phi_k(x)]&=\tilde{\lambda}_k\Phi_k(x)\\ -\frac{1}{2N}y(1-y)\frac{\partial^2\Psi_k(y)}{\partial y^2}-sy(1-y)\frac{\partial\Psi_k(y)}{\partial y}&=\tilde{\lambda}_k\Psi_k(y)\end{aligned}\right\}\quad k=0,1,2,\ldots, \tag{10}$$

where $\tilde{\lambda}_k$ are the eigenvalues. The diffusion operator D has an infinite number of eigenfunctions hence k, in Eq. (10), can become indefinitely large. From analogy with the eigenvectors of the matrix $\mathbf{M}$, we assume the eigenfunctions of D also have the properties of orthogonality and normalisation, which in this case means $\int_0^1\Psi_j(x)\Phi_k(x)\,dx=0$ when $j\neq k$ and $\int_0^1\Psi_k(x)\Phi_k(x)\,dx=1$.

Using the properties of the various eigenvectors/eigenfunctions, it can be shown (see Appendix A) that the four models in Table 1 have the *formal* solutions given in Table 2.

5. Eigenvectors of the matrix M

In order to investigate the properties of the solutions given in Table 2 when the allele frequency takes discrete values, we shall elucidate some of the properties of the eigenvectors of the matrix $\mathbf{M}$ of Eq. (6).

The matrix $\mathbf{W}$ of Eq. (5) has the form

$$\mathbf{W}=\begin{pmatrix}W_{0,0} & W_{0,1} & \cdots\\ W_{1,0} & \cdots & \\ \vdots & & W_{N,N}\end{pmatrix}=\begin{pmatrix}1 & \mathbf{v}^T & 0\\ \mathbf{0} & \mathbf{w} & \mathbf{0}\\ 0 & \mathbf{u}^T & 1\end{pmatrix}, \tag{11}$$

where a T superscript denotes the transpose of a matrix and (i) $\mathbf{u}$, $\mathbf{v}$ are column vectors of length $N-1$ with all elements non-zero, (ii) $\mathbf{0}$ is a column vector of length $N-1$ with all elements zero and (iii) $\mathbf{w}$ is an $(N-1)\times(N-1)$ matrix.

It follows that the matrix $\mathbf{M}$ of Eq. (6) has the form

$$\mathbf{M}=\mathbf{I}-\mathbf{W}=\begin{pmatrix}M_{0,0} & M_{0,1} & \cdots\\ M_{1,0} & \cdots & \\ \vdots & & M_{N,N}\end{pmatrix}=\begin{pmatrix}0 & -\mathbf{v}^T & 0\\ \mathbf{0} & \mathbf{m} & \mathbf{0}\\ 0 & -\mathbf{u}^T & 0\end{pmatrix}, \tag{12}$$

where, with $\mathbf{i}$ an identity matrix of size $(N-1)\times(N-1)$, the quantity $\mathbf{m}$ is the matrix

$$\mathbf{m}=\mathbf{i}-\mathbf{w}. \tag{13}$$

The matrix $\mathbf{M}$ only appears in Table 1 in those models where frequency is discrete. In the Wright–Fisher model (where both time and frequency take discrete values), elements of $\mathbf{u}$ and $\mathbf{v}$ are probabilities of transition from states of a population where allele A is segregating, to states where this allele is fixed or lost, while the matrix $\mathbf{w}$ contains transition probabilities between pairs of states of a population that each have allele A segregating. In hybrid model 1 (where time is continuous, but frequency is discrete), elements of $\mathbf{u}$ and $\mathbf{v}$ are probabilities/unit time of transitions from states of a population where allele A is segregating, to states where this allele is fixed or lost, while elements of $\mathbf{m}$ are transition probabilities/unit time between pairs of states that each have allele A segregating.

Table 2
This table gives the formal solutions of the four models of Table 1 in terms of the eigenvectors of the matrix $\mathbf{M}$ or the eigenfunctions of the diffusion operator D, along with the corresponding eigenvalues.

Model	Time	Frequency	Formal solution
Wright–Fisher	Discrete	Discrete	$\mathbf{F}(t)=\sum_{k=0}^{N}(1-\lambda_k)^t\boldsymbol{\Phi}^{(k)}\boldsymbol{\Psi}^{(k)T}\mathbf{F}(0)$
Hybrid model 1	Continuous	Discrete	$\mathbf{F}(t)=\sum_{k=0}^{N}\exp(-\lambda_k t)\boldsymbol{\Phi}^{(k)}\boldsymbol{\Psi}^{(k)T}\mathbf{F}(0)$
Hybrid model 2	Discrete	Continuous	$f(x,t)=\int_0^1\sum_{k=0}^{\infty}(1-\tilde{\lambda}_k)^t\Phi_k(x)\Psi_k(y)f(y,0)\,dy$
Diffusion	Continuous	Continuous	$f(x,t)=\int_0^1\sum_{k=0}^{\infty}\exp(-\tilde{\lambda}_k t)\Phi_k(x)\Psi_k(y)f(y,0)\,dy$

We work under the assumption that the matrix $\mathbf{m}$ has all eigenvalues distinct and hence is diagonalisable. For evidence of this and other properties of the matrix $\mathbf{m}$, see Appendix B.

For reasons that shall shortly become clear, we label the eigenvalues and eigenvectors of the matrix $\mathbf{m}$ with an index that ranges 2 to N. We write the right eigenvectors of this matrix as the column vector $\boldsymbol{\phi}^{(k)}$ and the left eigenvectors as $\boldsymbol{\psi}^{(k)T}$ (i.e., as the transpose of a column vector $\boldsymbol{\psi}^{(k)}$). With λ_k the eigenvalue associated with both $\boldsymbol{\phi}^{(k)}$ and $\boldsymbol{\psi}^{(k)T}$ the eigenvalue equations of matrix $\mathbf{m}$ are

$$\left.\begin{aligned}\mathbf{m}\boldsymbol{\phi}^{(k)}&=\lambda_k\boldsymbol{\phi}^{(k)}\\ \boldsymbol{\psi}^{(k)T}\mathbf{m}&=\lambda_k\boldsymbol{\psi}^{(k)T}\end{aligned}\right\}\quad k=2,3,\ldots,N. \tag{14}$$

Distinctness of the eigenvalues of $\mathbf{m}$ (i.e., $\lambda_j\neq\lambda_k$ if $j\neq k$) results in the eigenvectors $\boldsymbol{\psi}^{(j)T}$ and $\boldsymbol{\phi}^{(k)}$ with different labels being orthogonal ($\boldsymbol{\psi}^{(j)T}\boldsymbol{\phi}^{(k)}=0$ when $j\neq k$). We take the eigenvectors to be chosen so they are normalised to unity ($\boldsymbol{\psi}^{(k)T}\boldsymbol{\phi}^{(k)}=1$).

Let us now return to the matrix $\mathbf{M}$. Because this matrix has the form given in Eq. (12), and hence has a relationship to the matrix $\mathbf{m}$, it inherits from $\mathbf{m}$ the property of being diagonalisable. We can directly construct the eigenvectors of the matrix $\mathbf{M}$ according to Table 3.

Table 3
This table gives the eigenvalues of the matrix $\mathbf{M}$ along with its right eigenvectors, $\boldsymbol{\Phi}^{(k)}$, and left eigenvectors, $\boldsymbol{\Psi}^{(k)T}$.

Label, k	Eigenvalue	$\boldsymbol{\Phi}^{(k)}$	$\boldsymbol{\Psi}^{(k)}$
0	0	$\begin{pmatrix}1\\ \mathbf{0}\\ 0\end{pmatrix}$	$\begin{pmatrix}1\\ (\mathbf{v}^T\mathbf{m}^{-1})^T\\ 0\end{pmatrix}$
1	0	$\begin{pmatrix}0\\ \mathbf{0}\\ 1\end{pmatrix}$	$\begin{pmatrix}0\\ (\mathbf{u}^T\mathbf{m}^{-1})^T\\ 1\end{pmatrix}$
2,3,…,N	λ_k	$\begin{pmatrix}-\mathbf{v}^T\boldsymbol{\phi}^{(k)}/\lambda_k\\ \boldsymbol{\phi}^{(k)}\\ -\mathbf{u}^T\boldsymbol{\phi}^{(k)}/\lambda_k\end{pmatrix}$	$\begin{pmatrix}0\\ \boldsymbol{\psi}^{(k)}\\ 0\end{pmatrix}$

The matrix $\mathbf{M}$ has two zero eigenvalues associated with the values 0 and 1 of the label k. The remaining eigenvalues of the matrix $\mathbf{M}$, associated with $k\geq 2$, coincide with the eigenvalues λ_k of the matrix $\mathbf{m}$.

In particular, it may be directly verified, using Eqs. (12) and (14) and Table 3, that for $k\geq 2$: $\mathbf{M}\boldsymbol{\Phi}^{(k)}=\lambda_k\boldsymbol{\Phi}^{(k)}$ and $\boldsymbol{\Psi}^{(k)T}\mathbf{M}=\lambda_k\boldsymbol{\Psi}^{(k)T}$. Accordingly, all eigenvalues of the matrix $\mathbf{m}$ are also eigenvalues of the larger matrix $\mathbf{M}$. In addition, Table 3 contains two right eigenvectors, $\boldsymbol{\Phi}^{(0)}$ and $\boldsymbol{\Phi}^{(1)}$, and two left eigenvectors, $\boldsymbol{\Psi}^{(0)T}$ and $\boldsymbol{\Psi}^{(1)T}$, which have the properties $\mathbf{M}\boldsymbol{\Phi}^{(0)}=\mathbf{0}$, $\mathbf{M}\boldsymbol{\Phi}^{(1)}=\mathbf{0}$, $\boldsymbol{\Psi}^{(0)T}\mathbf{M}=\mathbf{0}$, $\boldsymbol{\Psi}^{(1)T}\mathbf{M}=\mathbf{0}$. Thus the matrix $\mathbf{M}$ also has two eigenvalues which vanish and which are not shared with the matrix $\mathbf{m}$. We have attributed the labels $k=0$ and 1 to these eigenvalues and the associated eigenvectors, hence $\lambda_0=0$ and $\lambda_1=0$. For $k\geq 2$ orthogonality and normalisation of the $\boldsymbol{\Phi}^{(k)}$ and $\boldsymbol{\Psi}^{(k)T}$ follow directly from orthogonality and normalisation of the $\boldsymbol{\phi}^{(k)}$ and $\boldsymbol{\psi}^{(j)T}$.

For future use, we note that the form of the right eigenvectors $\boldsymbol{\Phi}^{(k)}$ given in Table 3 can be written as

$$\Phi_n^{(0)}=\delta_{n,0}, \tag{15}$$

$$\Phi_n^{(1)}=\delta_{n,N}, \tag{16}$$

$$\Phi_n^{(k)}=\phi_n^{(k)}\Delta_n-\frac{\mathbf{v}^T\boldsymbol{\phi}^{(k)}}{\lambda_k}\delta_{n,0}-\frac{\mathbf{u}^T\boldsymbol{\phi}^{(k)}}{\lambda_k}\delta_{n,N},\quad k\geq 2, \tag{17}$$

where $\delta_{a,b}$ is a Kronecker delta ($\delta_{a,b}$ equals 1 if $a=b$ and is zero otherwise) and Δ_n equals 1 for $1\leq n\leq N-1$ and is zero for both $n=0$ and N.

6. Eigenfunctions of the diffusion operator, D

To investigate properties of the solutions given in Table 2 when the allele frequency takes continuous values, we shall elucidate some of the properties of the eigenfunctions of the diffusion operator D of Eq. (8) and which appears in the diffusion equation $-\partial f(x,t)/\partial t=Df(x,t)$.

To start, we note that the diffusion operator D of Eq. (8) is, as already stated, some sort of an approximation of the matrix $\mathbf{M}$ and hence apparently has a set of (forward) eigenfunctions, $\Phi_k(x)$, which are directly analogous to the right eigenvectors $\boldsymbol{\Phi}^{(k)}$ of the matrix $\mathbf{M}$. This, however, is true *only* when the eigenfunctions of D satisfy the appropriate boundary conditions. McKane and Waxman (2007) required that probability is conserved in the diffusion equation and accordingly imposed the condition that the eigenfunctions of D have no probability current flowing at $x=0$ and 1. This leads to the eigenfunctions of D containing singularities, namely Dirac delta functions (Dirac, 1958), which are 'located' at $x=0$ and 1.

For the case of selective neutrality ($s=0$) it is possible to explicitly determine the full form for the eigenfunctions of D under the 'zero current' boundary condition of McKane and Waxman (2007). Inspection of the forms given in McKane and Waxman (2007) shows that for the haploid case the eigenfunctions can be written as[2]

$$\Phi_0(x)=\delta(x), \tag{18}$$

$$\Phi_1(x)=\delta(1-x), \tag{19}$$

[2] There are two eigenfunctions associated with zero eigenvalue. Different linear combinations of these eigenfunctions also have zero eigenvalue. In the present work we have made a particular choice of these eigenfunctions; an alternative, but equivalent choice of these eigenfunctions has been made by McKane and Waxman (2007).

$$\Phi_k(x) = \phi_k(x)\Lambda(x) - \frac{\phi_k(0)}{2N\tilde{\lambda}_k}\delta(x) - \frac{\phi_k(1)}{2N\tilde{\lambda}_k}\delta(1-x), \quad k \geq 2. \tag{20}$$

Here $\delta(x)$ is a Dirac delta function, $\phi_k(x)$ is a non-singular function satisfying $-(2N)^{-1}\partial^2[x(1-x)\phi_k(x)]/\partial x^2 = \tilde{\lambda}_k\phi_k(x)$, and $\Lambda(x)$ is a box shaped function which equals 1 for $0 < x < 1$ and is zero elsewhere. The form of the eigenfunctions in Eqs. (18), (19) and (20) are strongly analogous to the eigenvectors in Eqs. (15), (16) and (17), with discrete quantities simply being replaced by continuous analogues. In particular, the Dirac delta function $\delta(x)$ becomes the replacement of the Kronecker delta $\delta_{n,0}$ and it appears natural to identify $\phi_k(0)/(2N\tilde{\lambda}_k)$ as the analogue of $\mathbf{v}^T\boldsymbol{\phi}^{(k)}/\lambda_k$. In this last case, the identification makes sense when it is appreciated that the elements, v_n, of $\mathbf{v}$ rapidly decay with n (we have $v_n = (1-x_n)^N$) so that $\mathbf{v}^T\boldsymbol{\phi}^{(k)}$ obtains the largest contribution from the elements $\phi_n^{(k)}$ of $\boldsymbol{\phi}^{(k)}$ with small n. Thus $\mathbf{v}^T$, in a discrete frequency problem, plays a similar role to a function proportional to $\delta(x)$ in a continuous frequency model. Similarly $\mathbf{u}^T$ plays a similar role to a function proportional to $\delta(1-x)$.

In Appendix C we show that the general form of the eigenfunctions given in Eqs. (18)–(20) continues to apply when there is a non-zero level of selection.

6.1. Singular spikes

We note that singular spikes (Dirac delta functions) are present in all of the eigenfunctions, $\Phi_k(x)$, of the diffusion operator D (see Eqs. (18)–(20)), and we have made the very natural identification of these singularities with related features of the right eigenvectors of the Wright–Fisher model.

Generally, in a Wright–Fisher model, the probability distribution of the allele frequency and the 'building blocks' from which the probability distribution can be composed, namely the right eigenvectors of the matrix $\mathbf{M}$, can be represented as a set of $N+1$ bars that make up a histogram. We can take these bars to be of width $1/N$ and centered at the frequencies $0/N, 1/N, \ldots, N/N$. Thus the bar corresponding to frequency $x_n = n/N$ covers the frequency interval $x_n - 1/(2N)$ to $x_n + 1/(2N)$, with $n = 0,1,\ldots,N$. What happens, under the diffusion approximation, is that (i) the bars at the terminal frequencies 0 and 1 shrink to zero width, but have a finite area, i.e., become proportional to Dirac delta functions that are located at the frequencies 0 and 1, and (ii) the contributions of all other bars are interpolated into a continuous function that covers the frequency range $0 < x < 1$, which excludes the end points $x = 0$ and 1. Thus the diffusion approximation shrinks the domain of states, corresponding to loss or fixation of the A allele, to zero width, while it simultaneously stretches the frequencies of the segregating states, from infinitesimally above 0 to infinitesimally below 1. See Fig. 1 for an illustration of this.

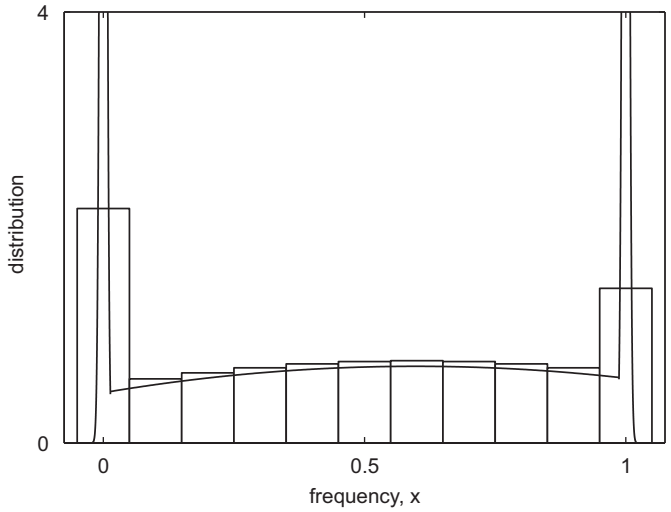

Fig. 1. This figure illustrates how the distribution of the diffusion approximation is related to the distribution of the Wright–Fisher model. The histogram reflects the discrete distribution of the Wright–Fisher model for a population of size $N = 10$. Under a diffusion approximation, the distribution in the frequency range $0 < x < 1$ becomes continuous, while at the boundaries $x = 0$ and 1 it possesses two spikes of finite area but zero width (Dirac delta functions). We have broadened the width of the Dirac delta functions for the purpose of visualisation. We note that numerical computations of the 'weights' associated with the spikes that arise in a diffusion approximation (and which are associated with fixation and loss probabilities) can be compared with the numerical results following from the exact model of the problem (the Wright–Fisher model). Such a comparison was made in Fig. 1 of the paper of McKane and Waxman (2007); see also Fig. 2 of the present paper.

7. Implications of the range of the eigenvalues

The eigenvalues of the matrix $\mathbf{M}$ or the diffusion operator D are important in determining the existence or non-existence of the discrete time models that appear in Table 1 and whose formal solutions are given in Table 2.

We have already established that the matrix $\mathbf{M}$ has two zero eigenvalues and that the remainder of the eigenvalues coincide with the eigenvalues of the matrix $\mathbf{m}$. In Appendix B we provide evidence that all eigenvalues of the matrix $\mathbf{m}$ are real and positive but are no larger than unity. These properties of the eigenvalues ensure that the time-dependent factor $(1-\lambda_k)^t$, that is present in the solution to the Wright–Fisher model in Table 2, never goes negative and remains in the range 0–1. By contrast, in the case of selective neutrality, the *complete set* of the eigenvalues of the diffusion operator D under the boundary conditions of McKane and Waxman (2007) and hence including the two vanishing eigenvalues, can be written as $\tilde{\lambda}_k = k(k-1)/2N$ where $k = 0,1,2,\ldots$ and k can become arbitrarily large. The eigenvalues of D thus have the property that they also can become arbitrarily large. Numerical considerations (not presented here) indicate that when there is selection ($s \neq 0$) the eigenvalues of the diffusion operator deviate from the neutral form, but still remain real and can become arbitrarily large. Thus generally, for sufficiently large k the eigenvalues $\tilde{\lambda}_k$ will be larger than unity. Under a diffusion approximation, where time is continuous, this feature is not problematic, since the time-dependent factor in the solution of the diffusion equation in Table 2, namely $\exp(-\tilde{\lambda}_k t)$, simply becomes very small for sufficiently large k. By contrast, in hybrid model 2 in Table 2, the time-dependent factor is $(1-\tilde{\lambda}_k)^t$ and for those eigenvalues with $\tilde{\lambda}_k > 1$ this factor alternates in sign over adjacent generations and for $\tilde{\lambda}_k > 2$ the *magnitude* of the time-dependent factor grows with k. Thus if, as is natural, we approximate the solution to hybrid model 2 by truncating the sum in Table 2 at a large value of k which has $\tilde{\lambda}_k > 2$, then as time increases, the approximation increasingly wildly oscillates between negative and positive values. This feature indicates that despite its reasonable motivation, hybrid model 2 suffers from the severe mathematical problem of convergence of the probability density $f(x,t)$. We thus rule out hybrid model 2 as being meaningful and conclude that there is no 'natural' model that is intermediate between the Wright–Fisher model and the diffusion approximation, where time is discrete but frequency continuous.

8. Lessons that can be learnt from continuous time dynamics

The dynamics of the Wright–Fisher model has, in past studies, been numerically investigated and the results directly compared with the diffusion approximation (Ewens, 1963). Analytical estimates have also been made (Ethier and Norman, 1977). Here we

shall investigate the continuous time dynamics of hybrid model 1 and compare its results with solutions of the diffusion equation. We shall use the representation of the eigenvectors in Table 3 and the close connection we have established between these eigenvectors and the eigenfunctions of the diffusion operator in Section 6. In this way we can establish some of the implications of the dynamics while avoiding mathematical subtleties associated with singularities.

To proceed, we write the solution to hybrid model 1 (in Table 2) as $\mathbf{F}(t)=\mathbf{K}(t)\mathbf{F}(0)$ where

$$\mathbf{K}(t)=\sum_{k=0}^{N} e^{-\lambda_k t}\boldsymbol{\Phi}^{(k)}\boldsymbol{\Psi}^{(k)T}. \tag{21}$$

Using the eigenvectors in Table 3 it is shown in Appendix D that we can also write

$$\mathbf{K}(t)=\begin{pmatrix} 1 & \int_0^t \mathbf{v}^T\mathbf{K}_{\rm seg}(r)\,dr & 0 \\ \mathbf{0} & \mathbf{K}_{\rm seg}(t) & \mathbf{0} \\ 0 & \int_0^t \mathbf{u}^T\mathbf{K}_{\rm seg}(r)\,dr & 1 \end{pmatrix}, \tag{22}$$

where $\mathbf{0}$ is a column vector of length $N-1$ with all elements zero and

$$\mathbf{K}_{\rm seg}(t)=\sum_{k=2}^{N} e^{-\lambda_k t}\phi^{(k)}\psi^{(k)T}. \tag{23}$$

The matrix $\mathbf{K}_{\rm seg}(t)$ in Eq. (22) is associated with the indices $1,2,\ldots,N-1$ of the matrix $\mathbf{K}(t)$. It therefore only describes states of a population where allele A is segregating, i.e., does not have a frequency of 0 or 1. Additionally, the matrix $\mathbf{K}_{\rm seg}(t)$ has the property that as the time t increases, all elements decay to zero, because it depends on $e^{-\lambda_2 t}$, $e^{-\lambda_3 t},\ldots$ and the eigenvalues λ_2, $\lambda_3,\ldots$ are all positive, and not zero.

When Kimura determined his solution of the diffusion equation he implicitly imposed the requirements that the eigenfunctions of the diffusion operator are normalisable (i.e., have a non-infinite integral) and *non-singular*. The condition of being non-singular is generally incompatible with conservation of probability in the diffusion equation. It also leads to differences in the eigenvalues and the eigenfunctions, compared with Eqs. (18)–(20). Kimura's (1955b) solution, given an initial frequency of y, has the form $K_{\rm Kimura}(x,t;y)=\sum_k e^{-\tilde{\lambda}_k t}\phi_k(x)\psi_k(y)$ where $\phi_k(x)$ ($\psi_k(y)$) are eigenfunctions of the forward (backward) diffusion operator associated with eigenvalue $\tilde{\lambda}_k$. A key observation is that the eigenvalues, $\tilde{\lambda}_k$, in Kimura's expression are all non-zero, and for the haploid case, Kimura's smallest eigenvalue has the value $1/N$. This indicates that Kimura's solution cannot be identified with $\mathbf{K}(t)$ of Eq. (21), which includes contributions associated with the two vanishing eigenvalues. Rather, Kimura's solution is very naturally associated with the solution $\mathbf{K}_{\rm seg}(t)$ of Eq. (23), which describes purely segregating alleles and obtains contributions from terms associated with explicitly non-zero eigenvalues. Indeed, in the selectively neutral case ($s=0$) it may be verified that the smallest eigenvalue of the matrix $\mathbf{m}$ is precisely $1/N$, in full agreement with Kimura's smallest eigenvalue (other eigenvalues of $\mathbf{m}$ do not precisely coincide with Kimura's eigenvalues).

There is additional evidence of the identification of $\mathbf{K}_{\rm seg}(t)$ with $K_{\rm Kimura}(x,t;y)$. Let us consider an initial distribution $\mathbf{F}(0)$ in the equation $\mathbf{F}(t)=\mathbf{K}(t)\mathbf{F}(0)$ in which the pth element is unity and all other elements are zero (i.e., $F_n(0)=\delta_{n,p}$) corresponding to the A allele having an initial frequency of x_p. This leads to the interpretation that the (n,p) element of $\mathbf{K}(t)$, namely $K_{n,p}(t)$, is the probability that $X(t)$ has the value x_n given $X(0)$ had the value x_p. Thus the upper row of $\mathbf{K}(t)$ in Eq. (22), namely $(1,\int_0^t \mathbf{v}^T\mathbf{K}_{\rm seg}(r)\,dr,0)$, has elements $K_{0,p}(t)$ which give the probability of loss of the A allele by time t from an initial frequency of x_p. Similarly, the bottom row of $\mathbf{K}(t)$, namely $(0,\int_0^t \mathbf{u}^T\mathbf{K}_{\rm seg}(r)\,dr,1)$ has elements $K_{N,p}(t)$ which give the time-dependent probability of fixation of the A allele from an initial frequency of x_p. We note that from considerations of current at the boundary $x=1$, Kimura established a result that is equivalent, for a haploid population, to $(2N)^{-1}\int_0^t K_{\rm Kimura}(1,r;y)\,dr$ for the probability of fixation of an allele by time t, given that the allele had a frequency of y, at time 0. This is very similar to the result $\int_0^t \mathbf{u}^T\mathbf{K}_{\rm seg}(r)\,dr$. The time dependence of the two different expressions are essentially identical: $(2N)^{-1}\int_0^t K_{\rm Kimura}(1,r;y)\,dr$ involves a sum of terms involving time-dependent factors of the form $1-e^{-\tilde{\lambda}_k t}$ where all eigenvalues ($\tilde{\lambda}_k$) occurring in the sum are non-zero, while $\int_0^t \mathbf{u}^T\mathbf{K}_{\rm seg}(r)\,dr$ involves a sum over terms involving time-dependent factors of the form $1-e^{-\lambda_k t}$ where, again, all eigenvalues (λ_k) occurring in the sum are non-zero.

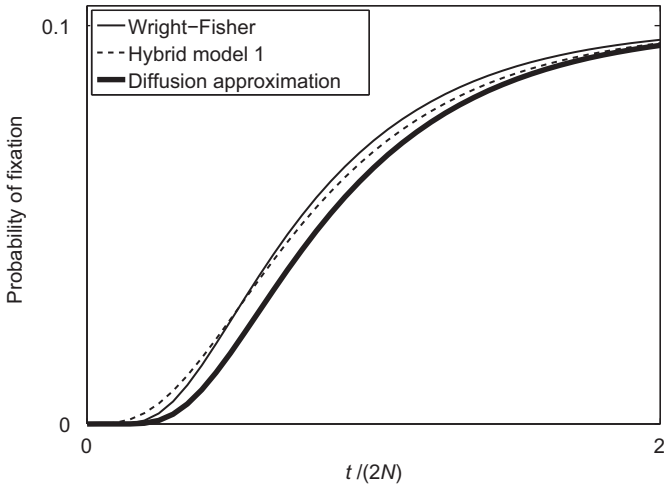

Fig. 2. This figure illustrates the probability that fixation has occurred by time t for: hybrid model 1, the diffusion approximation and the Wright–Fisher model. The figure is calculated for a population of size $N=10$ and an initial allele frequency of 0.1, for the case of selective neutrality ($s=0$). We note that the fixation probability of hybrid model 1 is not always smaller than the fixation probability of the Wright–Fisher model.

Given the differences in the expressions for the probability of fixation that arise from hybrid model 1, the diffusion approximation and, indeed, the Wright–Fisher model, it is interesting to see the actual similarities in the results, and we illustrate this with an example in Fig. 2.

9. Discussion

In this work we have investigated the content of the Wright–Fisher model of random genetic drift and its diffusion approximation. The diffusion approximation of the Wright–Fisher model is not a single approximation but several different approximations. These involve

(i) replacing discrete time by continuous time,
(ii) replacing discrete allele frequencies by continuous allele frequencies and simultaneously determining the appropriate diffusion operator that leads to closely equivalent dynamics to the Wright–Fisher model.

The complications and obscurities associated with approximation (ii) make it hard to form a detailed assessment of what is included in the diffusion approximation, and, indeed, what *can* be included in it. For example, in the past, there has been discussion on whether the forward diffusion equation can directly deal with gene loss and gene fixation (see e.g., Gale, 1990 and references therein). To make some progress with the clarification of these and related

mathematical issues, we have introduced a representation of the eigenvectors associated with the Wright–Fisher model that can be seen to have extremely close connections with the corresponding eigenfunctions of the diffusion operator. This representation allows us to see the relation of the solution of the Wright–Fisher model (which includes gene loss and fixation) to the solution of the diffusion equation of McKane and Waxman (2007) (which also includes gene loss and fixation).

The solution of the diffusion equation of McKane and Waxman (2007), which includes singular delta function spikes at the frequency boundaries $x=0$ and 1, differs from the earlier solution of Kimura (1955b), which has no such singularities. To probe this difference, two hybrid models were introduced that are intermediate (in the character of frequency or time) between the Wright–Fisher model and the diffusion approximation. One of the hybrid models (with discrete time and continuous frequency) is not mathematically well behaved and was ruled out of consideration, while the other (with discrete frequency and continuous time) avoids mathematical subtleties associated with continuous allele frequencies, yet provides clear insight into the phenomena occurring. Analysis of this model convincingly establishes that the solution of Kimura (1955b) only covers populations where alleles are segregating, but not fixing or being lost. This is in contrast to the solution of McKane and Waxman (2007), which includes the processes of loss and fixation, and is in close correspondence with the solution of the Wright–Fisher model.

In the investigations carried out here, it was also established that the probability distribution $\mathbf{F}(t)$ of the continuous time, discrete frequency, hybrid model can be written as $\mathbf{F}(t)=\mathbf{K}(t)\mathbf{F}(0)$ where $\mathbf{K}(t)$ is given in Eq. (22). However, $\mathbf{K}(t)$ is determined from the quantity $\mathbf{K}_{\rm seg}(t)$ of Eq. (23), along with the vectors $\mathbf{u}$ and $\mathbf{v}$ that appear in Eqs. (11) and (12). Under a diffusion approximation, $\mathbf{K}(t)$ has a direct analogue in the function $K(x,t;y)$ that was found by McKane and Waxman (2007), while $\mathbf{K}_{\rm seg}(t)$ has a direct analogue with the function $K_{\rm Kimura}(x,t;y)$ found by Kimura (1955b). Additionally the action of $\mathbf{u}^T$ or $\mathbf{v}^T$ becomes equivalent, under a diffusion approximation, to acting as a constant times either $\delta(1-x)$ or $\delta(x)$, followed by integration over all x. Thus under a diffusion approximation, we only require knowledge of the 'inner solution' $K_{\rm Kimura}(x,t;y)$ that covers only populations where alleles segregate, to construct a full solution $K(x,t;y)$ that conserves probability and so also covers the processes of fixation and loss. However, from a computational point of view, it is not obvious what precise boundary conditions should be imposed on the x dependence of $K_{\rm Kimura}(x,t;y)$, since being normalisable and singularity-free are not conditions that have an obvious numerical implementation. It is also unclear how to carry out numerical solution of the forward diffusion equation when probability is conserved (McKane and Waxman, 2007), since this automatically entails informative, but numerically intractable, Dirac delta functions in the solution. It could be argued that the diffusion approach is an approximation, thus when a numerical result is required, one should simply return to the underlying exact theory and numerically solve that. However, dealing with large matrices, possibly with time-dependent parameters, is not always feasible or convenient, and it would be useful, and also important, to have a principled approach for solving the forward diffusion equation that can be numerically implemented, or capable of being analytically solved/approximated.

To conclude this work, let us summarise some of the different issues touched upon here. From a fundamental point of view, we started with a Markov stochastic process in discrete time and with a discrete state space that exhibits irreversibility, due to the presence of absorbing states. Its diffusion approximation—where time and 'space' are continuous, results in the absorbing states non-trivially becoming singularities (Dirac delta functions), while the remaining states constitute a continuum. These singularities are present in the solution of the diffusion equation only if the appropriate boundary conditions, that conserve probability, are imposed. Conservation of probability means that the 'weights' associated with the singularities, combined with the probability associated with the continuum states, add to unity (for all times). It is possible to consider intermediate models where only time or space is continuous—the other independent variable remaining discrete, and so probe the nature of the continuum approximations leading to the diffusion approximation. In the particular problem (of random genetic drift) considered here, it was found that only of one of these intermediate models exists, due to the unbounded range of the eigenvalues. It seems possible that results and observations found in this work may have broader implications than just for the process of random genetic drift.

Acknowledgements

I thank Professor Jianfeng Feng and Professor Wei Lin for the invitation to visit Fudan University and the kind hospitality extended to me. I thank two anonymous reviewers for their constructive comments that have greatly improved this work.

Appendix A. Formal solutions of the models in Table 1

In this Appendix, we determine the formal solutions of the four models given in Table 1.

The procedure is similar in all four models.

For the Wright–Fisher model we use completeness of the right eigenvectors of the matrix $\mathbf{M}$, to allow us to write

$$\mathbf{F}(t)=\sum_{k=0}^{N}c_k(t)\mathbf{\Phi}^{(k)} \tag{24}$$

for some set of time-dependent coefficients $c_k(t)$. Substituting this solution into the dynamical equation for the model (see Table 1) and using the eigenvalue equation for $\mathbf{M}$ gives $-\sum_{k=0}^{N}[c_k(t+1)-c_k(t)]\mathbf{\Phi}^{(k)}=\sum_{k=0}^{N}\lambda_k c_k(t)\mathbf{\Phi}^{(k)}$. Multiplying this equation from the left with $\mathbf{\Psi}^{(j)T}$ and using orthogonality and normalisation of the eigenvectors yields $-[c_j(t+1)-c_j(t)]=\lambda_j c_j(t)$. This has the solution $c_j(t)=(1-\lambda_j)^t c_j(0)$, which on substituting into Eq. (24) yields

$$\mathbf{F}(t)=\sum_{k=0}^{N}(1-\lambda_k)^t\mathbf{\Phi}^{(k)}c_k(0). \tag{25}$$

To determine $c_k(0)$, we multiple the equation $\mathbf{F}(0)=\sum_{k=0}^{N}c_k(0)\mathbf{\Phi}^{(k)}$ from the left with $\mathbf{\Psi}^{(j)T}$ to obtain

$$c_j(0)=\mathbf{\Psi}^{(j)T}\mathbf{F}(0). \tag{26}$$

Using this result in Eq. (25) yields $\mathbf{F}(t)=\sum_{k=0}^{N}(1-\lambda_k)^t\mathbf{\Phi}^{(k)}\mathbf{\Psi}^{(k)T}\mathbf{F}(0)$ which is the solution in Table 2.

For hybrid model 1 we again write $\mathbf{F}(t)$ in the form of Eq. (24) and quickly arrive at $-dc_j(t)/dt=\lambda_j c_j(t)$, with solution $c_j(t)=e^{-\lambda_j t}c_j(0)$. Using Eq. (26) yields the solution in Table 2 for this model.

For hybrid model 2, the analogue of Eq. (24) is $f(x,t)=\sum_{k=0}^{\infty}c_k(t)\Phi_k(x)$ and using the dynamical equation and the eigenvalue equation for D gives $-\sum_{k=0}^{\infty}[c_k(t+1)-c_k(t)]\Phi_k(x)=\sum_{k=0}^{N}\tilde{\lambda}_k c_k(t)\Phi_k(x)$. On multiplying this equation by $\Psi_j(x)$, integrating from $x=0$ to 1, and using orthogonality and normalisation of the eigenfunctions yields $-[c_j(t+1)-c_j(t)]=\tilde{\lambda}_j c_j(t)$. This equation and the analogue of Eq. (26), namely $c_j(0)=\int_0^1\Psi_j(y)f(y,0)\,dy$, yield the solution for hybrid model 2 given in Table 2.

The diffusion approximation requires the same steps as hybrid model 2; it produces to $-dc_j(t)/dt=\tilde{\lambda}_j c_j(t)$ and leads to the solution given in Table 2.

Appendix B. Properties of eigenvalues of the matrix m

In this Appendix we provide evidence that the matrix $\mathbf{m}$, which was introduced in Eq. (12), has eigenvalues that are real, distinct, positive and do not exceed 1. Since the matrix $\mathbf{m}$ is related to the matrix $\mathbf{w}$ by $\mathbf{m}=\mathbf{i}-\mathbf{w}$ (Eq. (13)) the corresponding properties of the matrix $\mathbf{w}$ are that it has eigenvalues which are real, distinct, positive and do not exceed 1.

We begin, assuming $s>-1$, so all elements of $\mathbf{w}$ are positive. This is the simplest situation where the Perron Frobenius Theorem applies (Seneta, 1981). It leads to the eigenvalue of $\mathbf{w}$ of largest magnitude being real and positive. There are bounds on this eigenvalue, which we denote μ (Seneta, 1981). In terms of the matrix $\mathbf{w}$, with elements $w_{i,j}$ and i,j running from 1 to $N-1$, we have

$$\min_j\left(\sum_{i=1}^{N-1}w_{i,j}\right)\le\mu\le\max_j\left(\sum_{i=1}^{N-1}w_{i,j}\right). \tag{27}$$

A direct calculation yields $\sum_{i=1}^{N-1}w_{i,j}=1-[C(x_j)]^N-[1-C(x_j)]^N$. The sum is maximised at $C(x_j)=1/2$ thus $\mu\le 1-2^{-(N-1)}$ and is minimised when $C(x_j)$ takes its smallest value (which is positive). Thus $0<\mu<1$. The remaining eigenvalues of $\mathbf{w}$ have a magnitude which is smaller than μ and hence less than 1. We have numerically investigated the properties of these eigenvalues and have extensive numerical evidence, covering s in the range -0.1 to 0.1 and N up to 100, that all remaining eigenvalues are real, distinct, positive and do not exceed 1.

Appendix C. Form of the eigenfunctions of D with non-zero selection

In this Appendix we show that the general form of the eigenfunctions given in Eqs. (18)–(20) also applies when there is non-zero selection.

We start with the diffusion equation which reads

$$\begin{aligned}-\frac{\partial f(x,t)}{\partial t}&=-\frac{1}{2N}\frac{\partial^2}{\partial x^2}[x(1-x)f(x,t)]+\frac{\partial}{\partial x}[sx(1-x)f(x,t)]\\&=Df(x,t).\end{aligned} \tag{28}$$

The probability current of the diffusion equation associated with a function $f(x)$ is

$$j(x;f)=-\frac{1}{2N}\frac{\partial}{\partial x}[x(1-x)f(x)]+sx(1-x)f(x). \tag{29}$$

The property of Dirac delta functions that

$$x\delta(x)=0,\quad (1-x)\delta(1-x)=0 \tag{30}$$

means that the eigenfunctions $\Phi_0(x)=\delta(x)$ and $\Phi_1(x)=\delta(1-x)$ of the non-selective ($s=0$) problem remain eigenfunctions of the diffusion operator D in the presence of selection, and continue to have vanishing eigenvalues: $D\Phi_0(x)=0$ and $D\Phi_1(x)=0$.

Eq. (30) also results in $\Phi_0(x)$ and $\Phi_1(x)$ having zero probability current ($j(x;\Phi_0)=0$ and $j(x;\Phi_1)=0$) and as a consequence they satisfy the 'zero current' boundary condition of McKane and Waxman (2007).

For the remaining eigenfunctions, with assumed non-zero eigenvalues, we take them to have the form

$$\Phi(x)=\phi(x)\Delta(x)-a\delta(x)-b\delta(1-x), \tag{31}$$

where $\phi(x)$ is normalisable and does not contain any Dirac delta functions, the function $\Delta(x)$ is 1 for $0<x<1$ and vanishes outside this interval, while a and b are constants.

A property of $\Delta(x)$ we shall make use of is

$$\frac{d}{dx}\Delta(x)=\delta(x)-\delta(1-x). \tag{32}$$

We note that the probability current associated with Eq. (31) is, using Eqs. (30) and (32),

$$\begin{aligned}j(x;\Phi)&=-\frac{1}{2N}\frac{\partial}{\partial x}[x(1-x)\phi(x)\Delta(x)]+sx(1-x)\phi(x)\Delta(x)\\&=\left(-\frac{1}{2N}\frac{\partial}{\partial x}[x(1-x)\phi(x)]+sx(1-x)\phi(x)\right)\Delta(x).\end{aligned} \tag{33}$$

Because of the overall factor of $\Delta(x)$ on the right hand side, the probability current $j(x;\Phi)$ vanishes for both $x\ge 1$ and $x\le 0$ and hence obeys the 'zero current' boundary condition of McKane and Waxman (2007).

We proceed by substituting Eq. (31) into the eigenvalue equation $D\Phi(x)=\tilde{\lambda}_k\Phi_k(x)$. Making use of Eqs. (30), (32) and (33) allows us to write

$$\begin{aligned}D\Phi(x)&=\frac{\partial}{\partial x}j(x;\Phi)\\&=\left[\frac{\partial}{\partial x}\left(-\frac{1}{2N}\frac{\partial}{\partial x}[x(1-x)\phi(x)]+sx(1-x)\phi(x)\right)\right]\Delta(x)\\&\quad+\left(-\frac{1}{2N}\frac{\partial}{\partial x}[x(1-x)\phi(x)]+sx(1-x)\phi(x)\right)[\delta(x)-\delta(1-x)]\\&=[D\phi(x)]\Delta(x)+\left(-\frac{1}{2N}(1-2x)\phi(x)\right)[\delta(x)-\delta(1-x)]\\&=[D\phi(x)]\Delta(x)-\frac{\phi(0)}{2N}\delta(x)-\frac{\phi(1)}{2N}\delta(1-x).\end{aligned} \tag{34}$$

Thus the eigenvalue equation $D\Phi(x)=\tilde{\lambda}_k\Phi_k(x)$ for an eigenfunction of the form of Eq. (31) is

$$[D\phi(x)]\Delta(x)-\frac{\phi(0)}{2N}\delta(x)-\frac{\phi(1)}{2N}\delta(1-x)=\tilde{\lambda}_k[\phi(x)\Delta(x)-a\delta(x)-b\delta(1-x)]. \tag{35}$$

This equation is consistent in the sense that there are the same types of terms on the left and right sides. On comparing like terms we arrive at $D\phi(x)=\tilde{\lambda}_k\phi(x)$, $a=\phi(0)/(2N\tilde{\lambda}_k)$ and $b=\phi(1)/(2N\tilde{\lambda}_k)$. Thus the eigenfunctions associated with non-zero eigenvalues take the general form given in Eq. (20), involving singular and non-singular terms. Of course the precise form for $\phi_k(x)$ will depend on the strength of selection.

Appendix D. Solution of hybrid model 1

In this Appendix we show how the eigenvectors in Table 3 can be used to explicitly construct the solution of hybrid model 1 given in Eq. (22).

We start from Eq. (21), $\mathbf{K}(t)=\sum_{k=0}^{N}e^{-\lambda_k t}\mathbf{\Phi}^{(k)}\mathbf{\Psi}^{(k)T}$, and note that the contributions from $k\ge 2$ to the sum are

$$\begin{aligned}\sum_{k=2}^{N}e^{-\lambda_k t}\mathbf{\Phi}^{(k)}\mathbf{\Psi}^{(k)T}&=\sum_{k=2}^{N}e^{-\lambda_k t}\begin{pmatrix}-\mathbf{v}^T\boldsymbol{\phi}^{(k)}/\lambda_k\\ \boldsymbol{\phi}^{(k)}\\ -\mathbf{u}^T\boldsymbol{\phi}^{(k)}/\lambda_k\end{pmatrix}(0,\boldsymbol{\psi}^{(k)T},0)\\&=\begin{pmatrix}0 & -\sum_{k=2}^{N}e^{-\lambda_k t}\mathbf{v}^T\boldsymbol{\phi}^{(k)}\boldsymbol{\psi}^{(k)T}/\lambda_k & 0\\ \mathbf{0} & \sum_{k=2}^{N}e^{-\lambda_k t}\boldsymbol{\phi}^{(k)}\boldsymbol{\psi}^{(k)T} & \mathbf{0}\\ 0 & -\sum_{k=2}^{N}e^{-\lambda_k t}\mathbf{u}^T\boldsymbol{\phi}^{(k)}\boldsymbol{\psi}^{(k)T}/\lambda_k & 0\end{pmatrix}.\end{aligned}$$

Adding the contributions from the $k=0$ and 1 terms yields

$$\sum_{k=0}^{N}e^{-\lambda_k t}\mathbf{\Phi}^{(k)}\mathbf{\Psi}^{(k)T}=\begin{pmatrix}1 & \mathbf{v}^T\mathbf{m}^{-1}-\sum_{k=2}^{N}e^{-\lambda_k t}\mathbf{v}^T\boldsymbol{\phi}^{(k)}\boldsymbol{\psi}^{(k)T}/\lambda_k & 0\\ \mathbf{0} & \sum_{k=2}^{N}e^{-\lambda_k t}\boldsymbol{\phi}^{(k)}\boldsymbol{\psi}^{(k)T} & \mathbf{0}\\ 0 & \mathbf{u}^T\mathbf{m}^{-1}-\sum_{k=2}^{N}e^{-\lambda_k t}\mathbf{u}^T\boldsymbol{\phi}^{(k)}\boldsymbol{\psi}^{(k)T}/\lambda_k & 1\end{pmatrix}. \tag{36}$$

We use the spectral decomposition, $\mathbf{m}^{-1}=\sum_{k=2}^{N}\boldsymbol{\phi}^{(k)}\boldsymbol{\psi}^{(k)T}/\lambda_k$ to obtain

$$\sum_{k=0}^{N}e^{-\lambda_k t}\boldsymbol{\Phi}^{(k)}\boldsymbol{\Psi}^{(k)T}=\begin{pmatrix}1 & \mathbf{v}^T\sum_{k=2}^{N}(1-e^{-\lambda_k t})\boldsymbol{\phi}^{(k)}\boldsymbol{\psi}^{(k)T}/\lambda_k & 0\\ \mathbf{0} & \sum_{k=2}^{N}e^{-\lambda_k t}\boldsymbol{\phi}^{(k)}\boldsymbol{\psi}^{(k)T} & \mathbf{0}\\ 0 & \mathbf{u}^T\sum_{k=2}^{N}(1-e^{-\lambda_k t})\boldsymbol{\phi}^{(k)}\boldsymbol{\psi}^{(k)T}/\lambda_k & 1\end{pmatrix}. \tag{37}$$

Lastly, we write $(1-e^{-\lambda_k t})/\lambda_k$ as $\int_0^t e^{-\lambda_k r}\,dr$ and obtain

$$\sum_{k=0}^{N}e^{-\lambda_k t}\boldsymbol{\Phi}^{(k)}\boldsymbol{\Psi}^{(k)T}=\begin{pmatrix}1 & \mathbf{v}^T\int_0^t\sum_{k=2}^{N}e^{-\lambda_k r}\boldsymbol{\phi}^{(k)}\boldsymbol{\psi}^{(k)T}dr & 0\\ \mathbf{0} & \sum_{k=2}^{N}e^{-\lambda_k t}\boldsymbol{\phi}^{(k)}\boldsymbol{\psi}^{(k)T} & \mathbf{0}\\ 0 & \mathbf{u}^T\int_0^t\sum_{k=2}^{N}e^{-\lambda_k r}\boldsymbol{\phi}^{(k)}\boldsymbol{\psi}^{(k)T} & 1\end{pmatrix} \tag{38}$$

which is equivalent to Eq. (22) of the main text.

References

Crow, J.F., Kimura, M., 1970. An Introduction to Population Genetics Theory. Harper & Row, New York.

Dirac, P.A.M., 1958. Quantum Mechanics, fourth ed. Oxford University Press, Oxford.

Ethier, S.N., Norman, M.F., 1977. Error estimate for the diffusion approximation for the Wright–Fisher model. Proc. Natl. Acad. Sci. 74, 5096–5098.

Ewens, W.J., 1963. Numerical results and diffusion approximations in a genetic process. Biometrika 50, 241–249.

Fisher, R.A., 1922. On the dominance ratio. Proc. R. Soc. Edinburg 42, 321–431.

Gale, J.S., 1990. Theoretical Population Genetics. Unwin Hyman, London.

Haigh, J., 2002. Probability Models. Springer-Verlag, London.

Kimura, M., 1955a. Stochastic processes and distribution of gene frequencies under natural selection. Cold Spring Harbour Symp. Quant. Biol. 20, 33–53.

Kimura, M., 1955b. Solution of a process of random genetic drift with a continuous model. Proc. Natl. Acad. Sci. 41, 141–150.

Mable, B.K., Otto, S.P., 1998. The evolution of life cycles with haploid and diploid phases. BioEssays 20, 453–462.

McKane, A.J., Waxman, D., 2007. Singular solutions of the diffusion equation of population genetics. J. Theor. Biol. 247, 849–858.

Seneta, E., 1981. Non-Negative Matrices and Markov Chains, second ed. Springer-Verlag, New York.

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