Journal of Mathematical Biology (2019) 79:2069–2110
https://doi.org/10.1007/s00285-019-01420-w

Mathematical Biology

Diffusion approximation for an age-class-structured population under viability and fertility selection with application to fixation probability of an advantageous mutant

Cíntia Dalila Soares[1] · Sabin Lessard[1]

Received: 16 October 2018 / Revised: 22 May 2019 / Published online: 29 August 2019

Abstract
In this paper, we ascertain the validity of a diffusion approximation for the frequencies of different types under recurrent mutation and frequency-dependent viability and fertility selection in a haploid population with a fixed age-class structure in the limit of a large population size. The approximation is used to study, and explain in terms of selection coefficients, reproductive values and population-structure coefficients, the differences in the effects of viability versus fertility selection on the fixation probability of an advantageous mutant.

Keywords Age-class-structured population · Leslie matrix · Frequency-dependent selection · Diffusion approximation · Fixation probability · Two timescales · Reproductive value · Population-structure coefficient

Mathematics Subject Classification Primary 92D25 · Secondary 60J70

1 Introduction

Mathematical models for biological populations often assume that individuals reproduce and survive at the same rates throughout their life. This is a very strong assumption, however. This is not what happens in general in real populations with overlapping generations in which the fertility parameters and the probabilities of survival usually depend on the age of the individuals. A review of studies on such a dependence in nature can be found, e.g., in Jones et al. (2014).

✉ Sabin Lessard
lessards@dms.umontreal.ca

1 Département de mathématiques et de statistique, Université de Montréal, C.P. 6128 Succursale Centre-ville, Montréal, QC H3C 3J7, Canada

In a neutral finite population in discrete time with a fixed age-class structure such that the number of individuals in each age class is kept constant from one time step to the next, which is an assumption in the basic model introduced in Felsenstein (1971), it is known that the frequency of a mutant type in each age class converges to a same weighted frequency of the mutant type in the whole population that does not change from one time step to the next. Can we expect a similar phenomenon over a short time scale in a population under selection if selection is weak enough, and does this make the dynamics of the mutant type over a longer time scale amenable to exact analysis under the same assumption of a fixed age-class structure? These are the questions that are addressed in this paper.

Still recently, using numerical iterations based on matrix analysis and some computer simulations for a haploid population with two age classes of fixed finite sizes and age-dependent differences in fertility parameters or probabilities of survival, Li et al. (2016) exhibited a critical point for the probability of fixation of an advantageous mutant introduced in the first age class as a function of the proportion of individuals in this age class. It was shown that the probability of fixation of the mutant increases up to the critical point in the case of a mutant that is advantageous in survival, and beyond the critical point in the case of a mutant that is advantageous in reproduction. Notice that, in the neutral model, the probability of fixation of a mutant type that is introduced in the first age class is a decreasing function with respect to the proportion of individuals in the first age class as shown in Emigh (1979a). The observed patterns of the effects of viability selection and fertility selection on the fixation probability, and their differences, appear to be puzzling and show the importance of considering life histories in selection models. Explanations are needed to understand these surprising patterns and this is a strong motivation to get further analytical results.

Let us recall that, using Felsenstein's (1971) model Emigh (1979a, b) already considered constant differences in fertility parameters, survival probabilities and mutation rates between two types in a haploid population in discrete time with a finite number of age classes. Assuming a diffusion approximation for the frequency of one of the two types in continuous time, Emigh (1979b) obtained the probability of fixation of this type in the absence of mutation and the stationary distribution of its frequency in the presence of recurrent mutation. In his approximation by a diffusion process, however, Emigh (1979b) assumes, without any verification other than numerical, that the frequency of each type tends rapidly to the same value in each age class like in the neutral model, and that this value changes slowly under the effects of selection. This two-timescale argument remains to be proved and this will be done in this paper.

For a general treatment on evolution in age-structured populations, we refer to Charlesworth (1980). A rigorous mathematical analysis of equilibrium states in infinite populations can be found in Cushing (1998). See also Kebir et al. (2010, 2015) for the role of life-history parameters on sex allocation in hermaphroditic species. Notice that aggregation methods to take into account two timescales have been applied to infinite age-structured populations in a patchy environment with a migration process assumed to be fast in comparison to the demographic process [see Auger et al. (2008) and references therein, and also Marvà and San Segundo (2018)], and in a game-theoretic framework with the game dynamics assumed to be faster than the demographic dynamics (Marvà et al. 2013). In Lessard and Soares (2018), on the other hand, a continuous-time dynamics for an infinite population with a finite number of age classes has been obtained in the limit of weak selection with the inverse of selection intensity as unit of time.

In this paper, we are interested in a continuous-time limit of a finite discrete-time population with a fixed age-class structure. In a population of size N with discrete, non-overlapping generations, the classical limit as $N \to +\infty$ with N generations as unit of time and N^{-1} as intensity of selection and mutation in a multi-type setting is a Wright–Fisher diffusion as shown in Kimura (1964). Notice that, as shown in Chalub and Souza (2009) for the Moran model and Chalub and Souza (2014) for the Wright–Fisher model with weak frequency-dependent selection between two types of individuals, the limiting probability density of one type, if it exists, depends on the relationship between the intensity of selection and the time step as the population size goes to infinity: the diffusion equation, the replicator-diffusion equation and the partial differential version of the replicator equation. Notice also that when modeling a population according to a stochastic point process with birth, mutation, and death occurring at continuous random times, large population approximations can be deterministic, in the form of ordinary, integro-, or partial differential equations, or probabilistic, in the form of stochastic partial differential equations or superprocesses depending on the scalings of the parameters (Champagnat et al. 2006, 2008).

In the case of weak selection and mutation in a finite age-class-structured population in discrete time, we can resort to a diffusion approximation for a Markov chain with two timescales as established in Ethier and Nagylaki (1980). This has been used, for instance, to ascertain a strong-migration limit in geographically structured populations as the population size goes to infinity (Nagylaki 1980) and in group-structured populations as the number of groups goes to infinity (Lessard 2009).

The first objective of this paper is to ascertain the validity of a diffusion approximation for the frequencies of different types under weak selection and mutation in a haploid population with a fixed age-class structure by checking the conditions in Ethier and Nagylaki (1980). This is done in Sect. 4 in the general case of n types with frequency-dependent viability and fertility selection parameters and general mutation probabilities, following the presentation of the model in Sects. 2 and 3. The proof relies on rigorous approximations of the moments of the multivariate Wallenius' noncentral hypergeometric distribution by the moments of a multinomial distribution.

The second objective of this paper is to provide explanations for the results obtained in Li et al. (2016). This is discussed in Sect. 6 after applying the diffusion approximation to the situations studied in the aforementioned paper in Sect. 5.

Technical details are all relegated to Appendices.

2 Selection model with age-class structure

Consider a finite population of size N with $n \geq 1$ types of individuals distributed in $d \geq 1$ age classes. Time is assumed to be discrete. The number of individuals in each age class remains constant from one time step to the next. Let the size of age class k be

$$N_k = f_k N \tag{1}$$

for $k = 1, \ldots, d$, with $f_1 \geq f_2 \geq \cdots \geq f_d$ and

$$\sum_{k=1}^{d} f_k = 1. \tag{2}$$

The population is assumed to be haploid so that each offspring that is produced from one time step to the next comes from only one parent.

From one time step to the next, N_1 offspring are produced independently from one another to form the next cohort of individuals in age class 1, while N_{k+1} individuals are chosen without replacement among the N_k individuals in age class k to form the next cohort of individuals in age class $k+1$ for $k = 1, \ldots, d-1$.

Let the number of individuals of type $i = 1, \ldots, n$ in age class $k = 1, \ldots, d$ at time step $\tau \geq 0$ be $c_{i,k}(\tau)$. The array $\mathbf{c}(\tau) = (c_{i,k}(\tau))$ gives the population state at time step $\tau \geq 0$. Let

$$x_{i,k}(\tau) = \frac{c_{i,k}(\tau)}{N_k} \tag{3}$$

be the frequency of type $i = 1, \ldots, n$ in age class $k = 1, \ldots, d$ at time step $\tau \geq 0$, so that the population state at this time step can be represented by the array of frequencies

$$\mathbf{x}(\tau) = (x_{i,k}(\tau)). \tag{4}$$

In the following, we assume $\mathbf{c}(0) = \mathbf{c}$ and $\mathbf{x}(0) = \mathbf{x}$.

Given a population state $\mathbf{x}$, let $r_{j,k}(\mathbf{x})$ and $s_{j,k}(\mathbf{x})$ be the selective advantages in reproduction and survival, respectively, for an individual of type $j = 1, \ldots, n$ in age class $k = 1, \ldots, d$, with $s_{j,d}(\mathbf{x}) = 0$ so that the last age class is d. Moreover, let $u_{j,i,k}(\mathbf{x})$ be the probability for an individual of type $j = 1, \ldots, n$ in age class $k = 1, \ldots, d$ to produce an offspring of type $i = 1, \ldots, n$ with

$$\sum_{i=1}^{n} u_{j,i,k}(\mathbf{x}) = 1. \tag{5}$$

Weak mutation and weak selection for N large enough are modeled by assuming mutation and selection parameters of order N^{-1} in the form

$$u_{j,i,k}(\mathbf{x}) = \begin{cases} \frac{1}{N}\mu_{j,i,k}(\mathbf{x}) & \text{for} \quad i \neq j, \\ 1 - \frac{1}{N}\sum_{l=1, l\neq j}^{n} \mu_{j,l,k}(\mathbf{x}) & \text{for} \quad i = j, \end{cases} \tag{6}$$

$$r_{j,k}(\mathbf{x}) = \frac{1}{N}\rho_{j,k}(\mathbf{x}) \tag{7}$$

and

$$s_{j,k}(\mathbf{x}) = \frac{1}{N}\sigma_{j,k}(\mathbf{x}) \tag{8}$$

for some continuous functions $\mu_{j,i,k}(\mathbf{x})$, $\rho_{j,k}(\mathbf{x})$ and $\sigma_{j,k}(\mathbf{x})$ for $i, j = 1, \ldots, n$ and $k = 1, \ldots, d$. Actually (see remark following the statement of theorem 1 in Sect. 4), these functions are assumed to have continuous partial derivatives of all orders.

The cohort of individuals in age class $k+1$ at time step 1 for $k = 1, \ldots, d-1$ is obtained by sampling without replacement N_{k+1} individuals among the N_k individuals in age class k at time step 0 with weight $1 + s_{i,k}(\mathbf{x})$ given to individuals of type i for $i = 1, \ldots, n$. Let

$$\begin{aligned}\mathbf{c}_{k+1}(1) &= (c_{1,k+1}(1), \ldots, c_{n,k+1}(1))^T,\\ \mathbf{c}_k(0) &= \mathbf{c}_k = (c_{1,k}, \ldots, c_{n,k})^T,\\ \mathbf{s}_k(\mathbf{x}) &= (s_{1,k}(\mathbf{x}), \ldots, s_{n,k}(\mathbf{x}))^T,\end{aligned}$$

where T stands for transpose, and $\mathbf{1}$ be the n-dimensional vector with all entries equal to 1. Then

$$\mathbf{c}_{k+1}(1) \mid \mathbf{c}(0) = \mathbf{c} \sim mwnchypg(N_{k+1}, N_k, \mathbf{c}_k, \mathbf{1} + \mathbf{s}_k(\mathbf{x})).$$

This is a multivariate Wallenius' noncentral hypergeometric (*mwnchypg*) distribution (Wallenius 1963; Chesson 1976). See "Appendix B" for more details and some approximations.

As for reproduction, selection can be soft with competition within each age class (and each age class contributing a fixed expected proportion of offspring), or hard with competition in the whole population (see, e.g., Karlin 1982). In the following, we consider the case of soft selection. See "Appendix A" for the case of hard selection.

With soft selection taking place and p_k being the expected proportion of offspring coming from age class k for $k = 1, \ldots, d$, the probability for an offspring produced in a population in state $\mathbf{x}$ to be of type i is given by

$$P_i(\mathbf{x}) = \sum_{k=1}^{d} p_k \sum_{j=1}^{n} \frac{x_{j,k}(1 + r_{j,k}(\mathbf{x}))u_{j,i,k}(\mathbf{x})}{1 + \sum_{j'=1}^{n} x_{j',k} r_{j',k}(\mathbf{x})} \tag{9}$$

for $i = 1, \ldots, n$. Using (6) and (7), this probability can be written as

$$P_i(\mathbf{x}) = \sum_{k=1}^{d} p_k x_{i,k} + \frac{1}{N}\sum_{k=1}^{d} p_k \left(\tilde{\mu}_{i,k}(\mathbf{x}) + x_{i,k}\tilde{\rho}_{i,k}(\mathbf{x})\right) + o(N^{-1}) \tag{10}$$

with $No(N^{-1}) \to 0$ uniformly in $\mathbf{x}$ as $N \to +\infty$, where

$$\tilde{\mu}_{i,k}(\mathbf{x}) = \sum_{j=1, j\neq i}^{n} \left(x_{j,k}\mu_{j,i,k}(\mathbf{x}) - x_{i,k}\mu_{i,j,k}(\mathbf{x}) \right) \tag{11}$$

and

$$\tilde{\rho}_{i,k}(\mathbf{x}) = \rho_{i,k}(\mathbf{x}) - \sum_{j=1}^{n} x_{j,k}\rho_{j,k}(\mathbf{x}) \tag{12}$$

for $i = 1, \ldots, n$ and $k = 1, \ldots, d$. Finally, assuming N_1 independent trials to form the next cohort of individuals in age class 1, we have

$$\mathbf{c}_1(1) \mid \mathbf{c}(0) = \mathbf{c} \sim multinomial(N_1, P_1(\mathbf{x}), \ldots, P_n(\mathbf{x})). \tag{13}$$

In the rest of this section, we study the expected frequencies of types in the next cohorts of individuals.

Let us define $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,d})^T$ for $i = 1, \ldots, n$, and

$$M = \begin{pmatrix} p_1 & p_2 & p_3 & \ldots & p_d \\ 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & 1 & 0 \end{pmatrix}. \tag{14}$$

Using properties of the multinomial distribution for the type frequencies in age class 1 (with parameters as above so that $E_{\mathbf{x}}[x_{i,1}(1)] = P_i(\mathbf{x})$ for $1, \ldots, n$) and properties of the *mwnchypg* distribution for the type frequencies in the other age classes [see (141) in "Appendix B" with parameters $r = N_k$, $R = N_{k-1}$ and $f = f_{k-1}$], the conditional expected frequency of type i in the cohort of individuals in age class k at time step 1 takes the form

$$E_{\mathbf{x}}[x_{i,k}(1)] = (M\mathbf{x}_i)_k + \frac{1}{N}\phi_{i,k}(\mathbf{x}) + o(N^{-1}) \tag{15}$$

with $No(N^{-1}) \to 0$ uniformly in $\mathbf{x}$ as $N \to +\infty$. Here, we have

$$\phi_{i,1}(\mathbf{x}) = \sum_{k=1}^{d} p_k \left(\tilde{\mu}_{i,k}(\mathbf{x}) + x_{i,k}\tilde{\rho}_{i,k}(\mathbf{x}) \right) \tag{16}$$

and

$$\phi_{i,k}(\mathbf{x}) = x_{i,k-1}C_{k-1}\tilde{\sigma}_{i,k-1}(\mathbf{x}) \tag{17}$$

with

$$C_{k-1} = \frac{f_{k-1} - f_k}{f_k} \ln\left(\frac{f_{k-1}}{f_{k-1} - f_k}\right) \tag{18}$$

and

$$\tilde{\sigma}_{i,k}(\mathbf{x}) = \sigma_{i,k}(\mathbf{x}) - \sum_{j=1}^{n} \sigma_{j,k}(\mathbf{x}) x_{j,k} \tag{19}$$

for $k = 2, \ldots, d$ and $i = 1, \ldots, n$. The coefficient C_{k-1} that distinguishes viability selection from fertility selection depends on the population age-class structure, more precisely on the relative sizes of the age classes $k-1$ and k. It modulates the strength of viability selection from age class $k-1$ to age class k.

3 Neutral model with age-class structure

The stochastic matrix $M = (m_{kl})$ is the backward transition matrix under neutrality, with m_{kl} being the expected proportion of individuals in age class k coming from age class l one time step back in the absence of selection for $k, l = 1, \ldots, d$. Under the additional assumptions that $p_d > 0$ and $\gcd\{k : 1 \leq k \leq d, p_k > 0\} = 1$ (which is the case if $p_{d-1} > 0$ with $p_d > 0$), the stochastic matrix M is necessarily irreducible and aperiodic. Its stationary probability distribution

$$\mathbf{w}^T = (w_1, \ldots, w_d) = \mathbf{w}^T M \tag{20}$$

is given by

$$w_k = \frac{1}{\bar{k}} \sum_{m=k}^{d} p_m \tag{21}$$

for $k = 1, \ldots, d$, with

$$\bar{k} = \sum_{k=1}^{d} k p_k = \frac{1}{w_1} \tag{22}$$

being the mean age of reproduction in the neutral model.

In the absence of mutation and selection, that is

$$\mu_{j,i,k}(\mathbf{x}) = \rho_{j,k}(\mathbf{x}) = \sigma_{j,k}(\mathbf{x}) = 0 \tag{23}$$

for $k = 1, \ldots, d$ and $i, j = 1, \ldots, n$ with $i \neq j$, the expected frequencies of type i in the different age classes from time step $\tau - 1$ to time step τ, according to (15), are given by

$$E[x_{i,1}(\tau) \mid \mathbf{x}_i(\tau-1)] = \sum_{k=1}^{d} p_k x_{i,k}(\tau-1) \tag{24}$$

and

$$E[x_{i,k}(\tau) \mid \mathbf{x}_i(\tau-1)] = x_{i,k-1}(\tau-1) \tag{25}$$

for $k = 2, \ldots, d$ and $i = 1, \ldots, n$. This yields

$$E[\mathbf{x}_i(\tau) \mid \mathbf{x}_i(\tau-1)] = M\mathbf{x}_i(\tau-1), \tag{26}$$

from which

$$E[\mathbf{x}_i(\tau)] = E[E[\mathbf{x}_i(\tau) \mid \mathbf{x}_i(\tau-1)]] = ME[\mathbf{x}_i(\tau-1)] = M^\tau E[\mathbf{x}_i(0)] = M^\tau \mathbf{x}_i, \tag{27}$$

by conditional expectation and mathematical induction, for $\tau \geq 0$ and $i = 1, \ldots, n$. The Perron-Frobenius theory for an irreducible aperiodic stochastic matrix (see, e.g., Karlin and Taylor 1975) guarantees that

$$\lim_{\tau \to +\infty} M^\tau = \mathbf{1}\mathbf{w}^T. \tag{28}$$

Thus, as time goes to infinity, the array of frequencies for the different types in the different age classes satisfies

$$\lim_{\tau \to +\infty} E[\mathbf{x}_i(\tau)] = \lim_{\tau \to +\infty} M^\tau \mathbf{x}_i = \left(\mathbf{w}^T \mathbf{x}_i\right) \mathbf{1} \tag{29}$$

for $i = 1, \ldots, n$.

Another consequence of neutrality is that the frequencies of the different types in the different age classes weighted by the entries of the stationary probability distribution of the backward transition matrix M, namely, the vector $\mathbf{z}(\tau) = (z_1(\tau), \ldots, z_n(\tau))^T$ with

$$z_i(\tau) = \mathbf{w}^T \mathbf{x}_i(\tau) \tag{30}$$

for $i = 1, \ldots, n$ and $\tau \geq 0$, is a martingale. As a matter of fact, we have

$$E[z_i(\tau)|\mathbf{x}_i(\tau-1)] = E[\mathbf{w}^T \mathbf{x}_i(\tau)|\mathbf{x}_i(\tau-1)] = \mathbf{w}^T M\mathbf{x}_i(\tau-1) = z_i(\tau-1) \tag{31}$$

for $i = 1, \ldots, n$ and $\tau \geq 0$.

The stationary probability distribution of the backward transition matrix M has an interpretation in terms of reproductive values. Note that

$$l_k = \frac{N_2}{N_1}\frac{N_3}{N_2}\cdots\frac{N_k}{N_{k-1}} = \frac{N_k}{N_1} = \frac{f_k}{f_1} \tag{32}$$

is the probability for an individual in age class 1 to survive up to age class k for $k = 1, \ldots, d$. The notion of reproductive value, introduced in Fisher (1930), represents the expected contribution to all future generations. In our neutral model, the average number of individuals produced by an individual in age class k is

$$b_k = p_k \frac{N_1}{N_k} = \frac{p_k}{l_k} \tag{33}$$

for $k = 1, \ldots, d$. Taking $v_1 = 1$ as the reproductive value of an individual in age class 1, the reproductive value of an individual in age class k is given by

$$v_k = \sum_{m=k}^{d} b_m \frac{l_m}{l_k} = \frac{1}{l_k} \sum_{m=k}^{d} p_m = T \frac{w_k}{l_k}, \tag{34}$$

which means

$$w_k = \frac{l_k v_k}{T} = \frac{N_k v_k}{N_1 T}, \tag{35}$$

for $k = 1, \ldots, d$. Since N_k is the size of age class k and v_k is the reproductive value of an individual in this age class in a neutral population, while $N_1 T$ is a normalizing factor, then w_k is the relative reproductive value of age class k in a neutral population.

Note that the vector $\mathbf{v} = (v_1, v_2, \ldots, v_d)^T$ is a left eigenvector associated with the leading eigenvalue 1 of the Leslie matrix (Leslie 1945) given by

$$A = \begin{pmatrix} p_1/l_1 & p_2/l_2 & p_3/l_3 & \ldots & p_D/l_D \\ l_2/l_1 & 0 & 0 & \ldots & 0 \\ 0 & l_3/l_2 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & l_D/l_{D-1} & 0 \end{pmatrix}. \tag{36}$$

The matrix $A = (a_{ij})$ is a forward or projection matrix, with a_{ij} being the expected number of individuals left in age class i by an individual in age class j. On the other hand, the stationary age distribution is given by the corresponding right eigenvector $\mathbf{u} = (l_1, l_2, \ldots, l_d)^T$. Therefore, we have

$$A\mathbf{u} = \mathbf{u} \quad \text{and} \quad \mathbf{v}^T A = \mathbf{v}^T \text{ with } \mathbf{v}^T \mathbf{u} = 1 \tag{37}$$

owing to (21) and (35).

4 Diffusion approximation for the selection model with age-class structure

In this section, we shall apply a diffusion approximation theorem for Markov chains with two time scales that is due to Ethier and Nagylaki (1980). The model considered is the general selection model with fixed age-class structure described in Sect. 2.

The population age-class structure

$$\mathbf{f} = (f_1, f_2, \dots, f_d)^T = \left(\frac{N_1}{N}, \frac{N_2}{N}, \dots, \frac{N_d}{N}\right)^T \tag{38}$$

as defined in (1) and (2) is kept constant, while we let the population size N go to plus infinity. The array of frequencies $\mathbf{x}(\tau) = (x_{i,k}(\tau))$ as defined in (3) and (4) for $\tau \geq 0$ is a Markov chain on the state space

$$S = \left\{ \mathbf{x} = (x_{i,k}) : x_{i,k} = \frac{c_{i,k}}{N_k} \text{ with } \sum_{i=1}^{n} c_{i,k} = N_k \text{ for } i = 1, \dots, n \text{ and } k = 1, \dots, d \right\}. \tag{39}$$

This Markov chain actually depends on the population size $N = \sum_{k=1}^{d} N_k$. Consider the weighted frequency of type i in the whole population at time step $\tau \geq 0$ defined as

$$z_i(\tau) = \mathbf{w}^T \mathbf{x}_i(\tau) = \sum_{k=1}^{d} w_k x_{i,k}(\tau), \tag{40}$$

where $\mathbf{w}^T = (w_1, \dots, w_d)$ is the stationary probability distribution of the backward transition matrix under neutrality M defined in Sect. 4, and the difference between this weighted frequency and the frequency of type i in age class k at the same time step

$$y_{i,k}(\tau) = x_{i,k}(\tau) - z_i(\tau), \tag{41}$$

for $i = 1, \dots, n$ and $k = 1, \dots, d$. Let the initial population state be $\mathbf{x}(0) = \mathbf{x} = (x_{i,k})$, which corresponds to

$$\mathbf{z}(0) = (z_1(0), \dots, z_n(0))^T = (z_1, \dots, z_n)^T = \mathbf{z} \tag{42}$$

and

$$\mathbf{y}(0) = (y_{i,k}(0)) = (y_{i,k}) = \mathbf{y}. \tag{43}$$

Note that

$$\mathbf{x} = \mathbf{z}\mathbf{1}^T + \mathbf{y}, \tag{44}$$

where **1** denotes here the d-dimensional row vector with all entries equal to 1. Define the changes over one time step

$$\Delta z_i = z_i(1) - z_i \tag{45}$$

and

$$\Delta x_{i,k} = x_{i,k}(1) - x_{i,k}, \tag{46}$$

so that

$$\Delta y_{i,k} = \Delta x_{i,k} - \Delta z_i, \tag{47}$$

for $i = 1, \ldots, n$ and $k = 1, \ldots, d$.

As rigorously shown in "Appendix C", there exist continuous functions $b_i(\mathbf{z}, \mathbf{y})$, $a_{i,j}(\mathbf{z}, \mathbf{y})$ and $c_{i,k}(\mathbf{z}, \mathbf{y})$ such that the following conditions on conditional expected values hold:

$$I.\ E_{\mathbf{x}}(\Delta z_i) = b_i(\mathbf{z}, \mathbf{y})N^{-1} + o\left(N^{-1}\right), \tag{48a}$$

$$II.\ E_{\mathbf{x}}((\Delta z_i)(\Delta z_j)) = a_{i,j}(\mathbf{z}, \mathbf{y})N^{-1} + o\left(N^{-1}\right), \tag{48b}$$

$$III.\ E_{\mathbf{x}}((\Delta z_i)^4) = o(N^{-1}), \tag{48c}$$

$$IV.\ E_{\mathbf{x}}(\Delta y_{i,k}) = c_{i,k}(\mathbf{z}, \mathbf{y}) + o(1), \tag{48d}$$

$$V.\ Var_{\mathbf{x}}(\Delta y_{i,k}) = o(1). \tag{48e}$$

Here, $o(1) \to 0$ and $No(N^{-1}) \to 0$ uniformly for $\mathbf{x}$ in S as $N \to +\infty$. Moreover, the continuous functions are given by

$$b_i(\mathbf{z}, \mathbf{y}) = \sum_{k=1}^{d} w_k \phi_{i,k}(\mathbf{z}\mathbf{1}^T + \mathbf{y}), \tag{49a}$$

$$a_{i,j}(\mathbf{z}, \mathbf{y}) = \sum_{k=1}^{d} w_k^2((M\mathbf{y}_i)_k + z_i)(\delta_{ij} - (M\mathbf{y}_j)_k - z_j)\left(\frac{1}{f_k} - \frac{1}{f_{k-1}}\right), \tag{49b}$$

$$c_{i,k}(\mathbf{z}, \mathbf{y}) = (M\mathbf{y}_i)_k - y_{i,k}, \tag{49c}$$

with $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,d})^T$ and $\phi_{i,k}$ defined in (16) and (17), for $i = 1, \ldots, n$ and $k = 1, \ldots, d$. Here, we use the convention that $1/f_0 = 0$.

Furthermore, the recurrence system of equations

$$Y_{i,k}(t+1, \mathbf{z}, \mathbf{y}) - Y_{i,k}(t, \mathbf{z}, \mathbf{y}) = \mathbf{c}_{i,k}(\mathbf{z}, \mathbf{Y}(t, \mathbf{z}, \mathbf{y})) \tag{50}$$

for all integers $t \geq 0$ with the initial condition

$$Y_{i,k}(0, \mathbf{z}, \mathbf{y}) = y_{i,k}, \tag{51}$$

for $k = 1, \ldots, d$ and $i = 1, \ldots, n$, has a solution that goes to 0 uniformly for $\mathbf{z} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n \times \mathbb{R}^d$ as $t \to +\infty$. This is referred to as condition VI, which is also shown in "Appendix C".

According to Ethier and Nagylaki (1980), we have proved the following result:

Theorem 1 *The frequency process* $\left(x_{i,k}(\tau)\right)$ *for* $\tau \geq 0$ *with* N *time steps as unit of time as* $N \to +\infty$ *converges in law to a diffusion with infinitesimal means given by*

$$b_i(\mathbf{z}, \mathbf{0}) = \sum_{k=1}^{d} w_k \phi_{i,k}(\mathbf{z}\mathbf{1}^T) \tag{52}$$

and infinitesimal covariances given by

$$a_{i,j}(\mathbf{z}, \mathbf{0}) = z_i(\delta_{ij} - z_j)\alpha \tag{53}$$

for $i, j = 1, \ldots, n$*, with*

$$\alpha = \sum_{k=1}^{d} w_k^2 \left(\frac{1}{f_k} - \frac{1}{f_{k-1}} \right). \tag{54}$$

Remark Under our assumptions, the function $b_i(\mathbf{z}, \mathbf{0})$ has continuous partial derivatives of all orders on the simplex of n-dimensional frequency vectors and satisfies $b_i(\mathbf{z}, \mathbf{0}) \geq 0$ when $z_i = 0$, for $i = 1, \ldots, n$. Then, it can be checked that the closure of the diffusive operator defined by (52) and (53) generates a strongly continuous semigroup (see Ethier 1976), which is a supplementary technical condition for the above result to be ascertained.

Notice that the quantity

$$N_e = \frac{N}{\bar{k}\alpha} \tag{55}$$

is the effective size of a large neutral population taking into account that one time step corresponds to $1/\bar{k}$ generations as shown in Felsenstein (1971) (see also Charlesworth 2009). The infinitesimal covariances in the diffusion approximation for the age-class-structured population of size N with N_e time steps as unit of time are the same as in the diffusion approximation for a Wright–Fisher population of size $N/\bar{k}$ with non-overlapping generations and $N/\bar{k}$ generations as unit of time as $N \to +\infty$.

As for the infinitesimal means, all calculations are made by weighing the age classes with the entries of the stationary distribution of the backward transition matrix under neutrality.

5 Selection model with two types in two age classes

In the case of two types ($n = 2$), let the weighted frequencies of types 1 and 2 be given by $\mathbf{z} = (z_1, z_2)^T = (z, 1 - z)^T$. Then the infinitesimal mean for type 1 in the diffusion approximation stated in Theorem 1 takes the form

$$b_1(z) = (1-z)\mu_{2,1}(z) - z\mu_{1,2}(z) + z(1-z)(\rho_1(z) - \rho_2(z) + \sigma_1(z) - \sigma_2(z)), \tag{56}$$

where

$$\mu_{i,j}(z) = w_1 \sum_{k=1}^{d} p_k \mu_{i,j,k}(\mathbf{z}\mathbf{1}^{\mathbf{T}}), \tag{57a}$$

$$\rho_i(z) = w_1 \sum_{k=1}^{d} p_k \rho_{i,k}(\mathbf{z}\mathbf{1}^{T}), \tag{57b}$$

$$\sigma_i(z) = \sum_{k=2}^{d} w_k C_{k-1} \sigma_{i,k-1}(\mathbf{z}\mathbf{1}^{T}), \tag{57c}$$

with

$$C_{k-1} = \frac{f_{k-1} - f_k}{f_k} \ln\left(\frac{f_{k-1}}{f_{k-1} - f_k}\right) \tag{58}$$

for $k = 2, \ldots, d$, while the infinitesimal covariances are given by

$$a_{i,j}(z) = (-1)^{i+j} z(1-z)\alpha \tag{59}$$

for $i, j = 1, 2$.

In the absence of mutation, that is, $\mu_{1,2}(z) = \mu_{2,1}(z) = 0$, and constant selection parameters independent of the population state, that is,

$$\rho_i(z) = w_1 \sum_{k=1}^{d} p_k \rho_{i,k} = \rho_i, \tag{60a}$$

$$\sigma_i(z) = \sum_{k=2}^{d} w_k C_{k-1} \sigma_{i,k-1} = \sigma_i, \tag{60b}$$

the infinitesimal mean and variance for type 1 are given by

$$b_1(z) = z(1-z)(\rho_1 - \rho_2 + \sigma_1 - \sigma_2) \tag{61}$$

and

$$v_1(z) = a_{1,1}(z) = z(1-z)\alpha, \tag{62}$$

respectively. Then, the probability of fixation of type 1 starting from an initial weighted frequency $z_0 = \mathbf{w}^T \mathbf{x}_1(0)$ and taking N time steps as unit of time as $N \to +\infty$ is given by

$$u(z_0) = \frac{\int_0^{z_0} G(x)dx}{\int_0^1 G(x)dx}, \tag{63}$$

where

$$G(x) = \exp\left\{-2\int_0^x \frac{b_1(z)}{v_1(z)}dz\right\} \tag{64}$$

(see, e.g., Ewens 2004, pp. 136–140; Etheridge 2011, pp. 39–41; or Crow and Kimura 1970, pp. 367–382). In the case at hand, we get

$$G(x) = \exp\left\{-\frac{2(\rho_1 - \rho_2 + \sigma_1 - \sigma_2)x}{\alpha}\right\}, \tag{65}$$

from which

$$u(z_0) = \frac{1 - \exp\{-2(\rho_1 - \rho_2 + \sigma_1 - \sigma_2)z_0/\alpha\}}{1 - \exp\{-2(\rho_1 - \rho_2 + \sigma_1 - \sigma_2)/\alpha\}}. \tag{66}$$

In the case of two age classes ($d = 2$), let $(p_1, p_2) = (f_1, f_2) = (f, 1 - f)$ with $0 < f_2 = 1 - f \leq f = f_1 < 1$ so that $1/2 \leq f < 1$. Then, we have

$$w_1 = \frac{1}{p_1 + 2p_2} = \frac{1}{2 - f}, \tag{67}$$

while

$$w_2 = \frac{p_2}{p_1 + 2p_2} = \frac{1 - f}{2 - f}. \tag{68}$$

Note that w_1 increases from $2/3$ to 1 as f increases from $1/2$ to 1, while w_2 decreases from $1/3$ to 0. Moreover, we have

$$\alpha = w_1^2\left(\frac{1}{f_1} - \frac{1}{f_0}\right) + w_2^2\left(\frac{1}{f_2} - \frac{1}{f_1}\right) = \frac{3 - 2f}{(2 - f)^2}. \tag{69}$$

Note that α increases as f increases from $1/2$ to 1 but very little, going from $8/9$ to 1.

Now suppose that there is only one individual of type 1 in age class 1 at time step $\tau = 0$ in a population of total size N, that is, $x_{1,1}(0) = f/s$ and $x_{1,2}(0) = 0$ with $s = 1/N$. In this case, we have

$$z_0 = w_1 x_{1,1}(0) = \frac{s}{(2 - f)f} = z_0(s). \tag{70}$$

This is the fixation probability for type 1 under neutrality, which decreases from $(4/3)s$ to s as f increases from $1/2$ to 1. Under selection, the fixation probability is given by

$$u(z_0(s)) = \frac{1 - \exp\left\{-2(\rho_1 - \rho_2 + \sigma_1 - \sigma_2)\frac{s}{(2-f)f\alpha}\right\}}{1 - \exp\{-2(\rho_1 - \rho_2 + \sigma_1 - \sigma_2)/\alpha\}}, \tag{71}$$

from which

$$U(f) = \lim_{s\to 0} \frac{1}{s} u(z_0(s)) = \frac{2(\rho_1 - \rho_2 + \sigma_1 - \sigma_2)\frac{1}{(2-f)f\alpha}}{1 - \exp\{-2(\rho_1 - \rho_2 + \sigma_1 - \sigma_2)/\alpha\}} \tag{72}$$

by applying L'Hôpital's rule.

In the following subsections, we consider the fixation probability for a single mutant arising in age class 1 in three situations: selection only on fertility, selection only on survival, and selection on fertility and survival. In each situation, we study $u(z_0(s))/s$, which corresponds to the fixation probability times the population size N, as a function of the fraction of the population in age class 1, given by f, for some particular values of $s = 1/N$ and in the limit as $N \to +\infty$.

5.1 Selection on fertility

First, consider fertility differences in favor of type 1 with $\rho_{1,1} = \rho_{1,2} = 1$, $\rho_{2,1} = \rho_{2,2} = 0$ and $\sigma_{2,1} = \sigma_{1,1} = 0$. Then the coefficient of fertility selection for type 1 is

$$\rho_1 = w_1 = \frac{1}{2-f}, \tag{73}$$

while $\rho_2 = \sigma_2 = \sigma_1 = 0$. Note that ρ_1 increases from $2/3$ to 1 as f increases from $1/2$ to 1. Moreover, in this situation, the fixation probability for type 1 is

$$u(z_0(s)) = \frac{1 - \exp\left\{-\frac{2s}{f(3-2f)}\right\}}{1 - \exp\left\{-\frac{2(2-f)}{3-2f}\right\}}, \tag{74}$$

which leads to

$$U(f) = \lim_{s\to 0} \frac{1}{s} u(z_0(s)) = \frac{\frac{2}{f(3-2f)}}{1 - \exp\left\{-\frac{2(2-f)}{3-2f}\right\}}. \tag{75}$$

The derivative of $U(f)$ with respect to f, $U'(f)$, is given by

$$\frac{2\exp\left\{\frac{4}{3-2f}\right\}\left(\exp\left\{\frac{4}{3-2f}\right\}\left(8f^2 - 18f + 9\right) + \exp\left\{\frac{2f}{3-2f}\right\}\left(-8f^2 + 20f - 9\right)\right)}{\left(\exp\left\{\frac{4}{3-2f}\right\} - \exp\left\{\frac{2f}{3-2f}\right\}\right)^2 f^2(2f-3)^3}. \tag{76}$$

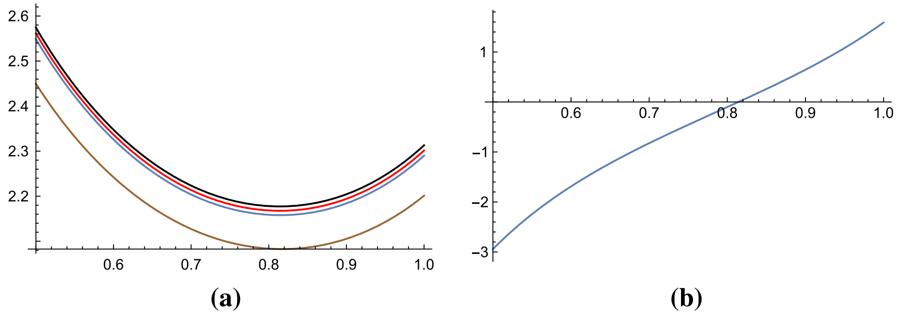


Fig. 1 Selection on fertility: **a** Fixation probability for type 1 divided by the selection parameter s. The brown curve is for $s = 1/20$, the blue one for $s = 1/100$, the red one for $s = 1/200$, and the black one for $U(f)$, that is, for the limit case as $s \to 0$. **b** Derivative of $U(f)$ with respect to f, $U'(f)$ (color figure online)

It is difficult to find algebraically the critical points of $U(f)$. Figure 1b shows that $U'(f)$ has only one zero from negative values to positive values as f increases. Therefore, it is the only critical point of $U(f)$, and actually the global minimum point of $U(f)$. This is confirmed by a rigorous mathematical analysis in "Appendix D". Notice that $U'(f)$ is an increasing function, so that $U(f)$ is a convex function.

5.2 Selection on survival

In the case of survival probabilities in favor of type 1 with $\sigma_{1,1} = 1$, $\sigma_{2,1} = 0$ and $\rho_{1,1} = \rho_{1,2} = \rho_{2,1} = \rho_{2,2} = 0$, we have a coefficient of viability selection for type 1 given by

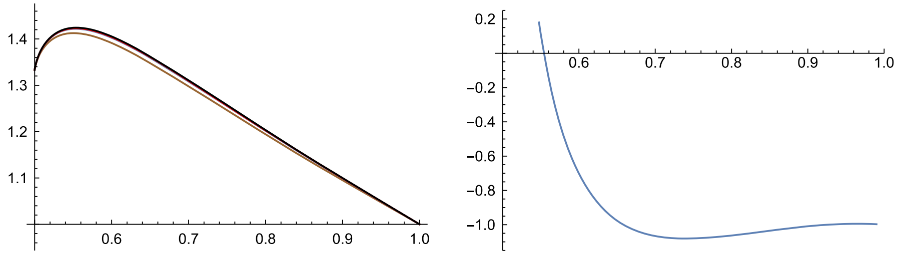
$$\sigma_1 = w_2 C_1 = \frac{2f-1}{2-f} \ln\left(\frac{f}{2f-1}\right), \tag{77}$$

while $\rho_1 = \rho_2 = \sigma_2 = 0$. Note that

$$C_1 = \frac{2f-1}{1-f} \ln\left(\frac{f}{2f-1}\right) \tag{78}$$

increases from 0 to 1 as f increases from $1/2$ to 1, while w_2 decreases from $1/3$ to 0. As a result, σ_1 is concave for f between $1/2$ and 1, and takes the value 0 at $f = 1/2$ and $f = 1$. On the other hand, in this situation, the fixation probability for type 1 is

$$u(z_0(s)) = \frac{1 - \exp\left\{-\frac{2s}{f(3-2f)}\left((2f-1)\ln\left(\frac{f}{2f-1}\right)\right)\right\}}{1 - \exp\left\{-\frac{2(2-f)}{3-2f}\left((2f-1)\ln\left(\frac{f}{2f-1}\right)\right)\right\}}, \tag{79}$$



(a) **(b)**

Fig. 2 Selection on viability: **a** Fixation probability for type 1 divided by the selection parameter s. The brown curve is for $s = 1/20$, the blue one for $s = 1/100$, the red one for $s = 1/200$, and the black one for $U(f)$, that is, for the limit case as $s \to 0$. **b** Derivative of $U(f)$ with respect to f, $U'(f)$ (color figure online)

which gives

$$U(f) = \lim_{s\to 0} \frac{1}{s} u(z_0(s)) = \frac{2(2f-1)\ln\left(\frac{f}{2f-1}\right)}{(3-2f)f\left(1-\left(\frac{f}{2f-1}\right)^{-\frac{2(2-f)(2f-1)}{3-2f}}\right)}. \tag{80}$$

The expression of $U'(f)$ can be found in "Appendix E" (Eq. (216)). It is difficult to find algebraically the critical points of $U(f)$. Figure 2b shows that $U'(f)$ has only one zero from positive values to negative values as f increases, so it is the only critical point of $U(f)$ in the interval $(1/2, 1)$, and actually the global maximum point. Notice that $U'(f)$ is first decreasing and then increasing on $(1/2, 1)$, so that $U(f)$ has an inflection point in $(1/2, 1)$.

5.3 Selection on survival and fertility

Finally, with survival and fertility in favor of type 1 in the form $\rho_{1,1} = \rho_{1,2} = \sigma_{1,1} = 1$ and $\rho_{2,1} = \rho_{2,2} = \sigma_{2,1} = 0$, we have a coefficient of selection for type 1 given by

$$\rho_1 + \sigma_1 = w_1 + w_2 C_1 = \frac{1}{2-f} + \frac{2f-1}{2-f}\ln\left(\frac{f}{2f-1}\right), \tag{81}$$

while $\rho_2 + \sigma_2 = 0$. Then, the fixation probability for type 1 is

$$u(z_0(s)) = \frac{1-\exp\left\{-\frac{2s}{f(3-2f)}\left(1+(2f-1)\ln\left(\frac{f}{2f-1}\right)\right)\right\}}{1-\exp\left\{-\frac{2(2-f)}{3-2f}\left(1+(2f-1)\ln\left(\frac{f}{2f-1}\right)\right)\right\}}, \tag{82}$$

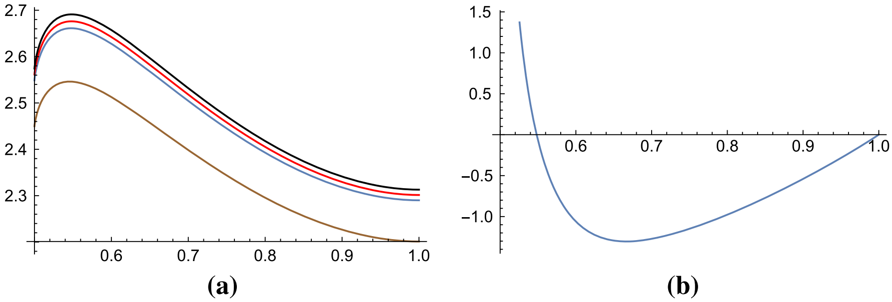


Fig. 3 Selection on fertility and survival: **a** fixation probability for type 1 divided by the selection parameter s. The brown curve is for $s = 1/20$, the blue one for $s = 1/100$, the red one for $s = 1/200$, and the black one for $U(f)$, that is, for the limit case as $s \to 0$. **b** Derivative of $U(f)$ with respect to f, $U'(f)$ (color figure online)

which gives

$$U(f) = \lim_{s\to 0} \frac{1}{s} u(z_0(s)) = \frac{2\left((2f-1)\ln\left(\frac{f}{2f-1}\right)+1\right)}{(3-2f)f\left(1-\exp\left\{-\frac{2(f-2)\left((2f-1)\ln\left(\frac{f}{2f-1}\right)+1\right)}{2f-3}\right\}\right)}. \tag{83}$$

The expression of $U'(f)$ can be found in "Appendix E" (Eq. (217)). Figure 3b shows the curve of $U'(f)$. We can see that $U'(f)$ is first decreasing and then increasing on $(1/2, 1)$, so that $U(f)$ has an inflection point in $(1/2, 1)$ like in the case with selection on survival only. The main difference with this case is that the increase after the inflection point is much more pronounced.

6 Discussion

We have considered a haploid population in discrete time with a fixed age-class structure in which the numbers of individuals in the different age classes are kept constant from one time step to the next one. This model was introduced in Felsenstein (1971), studied later on in Emigh (1979a, b) and used recently in Li et al. (2016) to show the effects of life histories on the fixation probability.

We have assumed weak frequency-dependent selection with fertility parameters and survival probabilities, besides mutation probabilities, all of order N^{-1} for n types of individuals, where N is the total population size. For the array of frequencies of the different types in the different age classes, we have checked the conditions in Ethier and Nagylaki (1980) to ascertain convergence of this discrete-time Markov chain with two timescales to a continuous-time diffusion process in the limit of a large population size. On a short timescale, the frequency of each type converges to the same value in each age class. This value corresponds to a weighted average of the type frequencies in the age classes with weights given by the stationary probability distribution of a backward transition matrix under neutrality. These weights correspond to the relative reproductive values of the different age classes in the neutral model. Moreover, on a longer timescale, the weighted averages of the type frequencies change according to a diffusion process analogous to a Wright–Fisher diffusion for a well-mixed population.

In order to establish Ethier and Nagylaki's (1980) conditions, we have obtained the rates of convergence of the moments of a multivariate Wallenius' non-central hypergeometric distribution to the moments of a multinomial distribution in the limits of a weak bias and a large sample size.

We have used the ascertained diffusion approximation to get the fixation probability for a mutant type. In the particular case of two types in two age classes with constant selection parameters and in the absence of mutation, the probability of ultimate fixation of a single mutant introduced in age class 1 can be expressed as a function of the frequency of age class 1. Under neutrality, this function is always decreasing as the frequency of age class 1 increases from $1/2$ to 1. We have shown that this function stops decreasing beyond some threshold frequency in the case of a fertility-advantageous mutant, while it starts decreasing only beyond some threshold frequency in the case of a viability-advantageous mutant. These patterns can be explained by the facts that the coefficient of selection for a fertility-advantageous mutant always increases with the frequency of age class 1, while the coefficient of selection for a viability-advantageous mutant increases when the frequency of age class 1 is close enough to $1/2$ and decreases when this frequency is close enough to 1. This is the case because the former coefficient is given by the relative reproductive value of age class 1, while the latter coefficient is given by the relative reproductive value of age class 2 times some coefficient that depends on the population age-class structure. It is noted that the relative reproductive value of age class 1 increases with the frequency of age class 1, while it is the opposite for the relative reproductive value of age class 2. Moreover, the population-structure coefficient increases with an increase in the frequency of age class 1. Such an increase enhances the strength of viability selection since it diminishes the proportion of individuals in age class 1 surviving to age class 2. In the case of selection on both fertility and survival, the coefficient of selection is a linear combination of the corresponding coefficients for fertility only and for survival only. In all cases, the decrease of the fixation probability under neutrality as the frequency of age class 1 increases can be stopped and reversed by selection. This occurs when the increase in frequency of age class 1 is accompanied by a large enough increase in the coefficient of selection.

Our analytical results in the limit of a large population size confirm and explain similar patterns obtained by numerical iterations based on matrix analysis for small population sizes and by computer simulations for larger population sizes in Li et al. (2016). Notice that a first-order approximation of the fixation probability can be obtained by considering small perturbations on the neutral genealogical process (Soares and Lessard 2019), but a diffusion approximation is a more general result with a wide range of possible applications. It could be used to study not only fixation probabilities, but also many other features of the population such as stationary states.

Since our diffusion approximation has been obtained in the general case of frequency-dependent selection, it could allow us to study evolutionary games in age-class-structured populations, e.g., to find conditions for cooperation to be favored by selection over defection with respect to the fixation probability in the absence of mutation (see, e.g., Nowak et al. 2004) or with respect to the mean abundance in the stationary state in the presence of recurrent mutation (see, e.g., Nowak et al. 2010; Kroumi and Lessard 2015). Assumptions on the population structure could also be lessened. We could consider, for instance, that the size of each age class is not fixed from one time step to the next one, or that the population size varies over time as in Huang et al. (2015) for a well-mixed population. We could also extend the model to any class-structured population as considered for an infinite population in Lessard and Soares (2018). Another possible extension would be to study a diploid population as in Emigh and Pollak (1979) in the absence of mutation. Some of these extensions such as a non-fixed age-class structure would require a much more recondite analysis, but they are not out of reach.

Acknowledgements This research was supported in part by NSERC of Canada, Grant No. 8833, and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 0968-13-7. We are also grateful to two anonymous referees for helpful comments to improve the manuscript.

Appendix A: hard selection

With hard selection for reproduction in a population in initial state $\mathbf{x}$, an offspring is produced by a parent of type $j = 1, \ldots, n$ in age class $k = 1, \ldots, d$ with probability

$$\frac{(1 + r_{j,k}(\mathbf{x}))x_{j,k} f_k}{1 + \sum_{j'=1}^{n} \sum_{l=1}^{d} r_{j',l}(\mathbf{x})x_{j',l} f_l}. \tag{84}$$

This assumes that an individual of type j in age class k has a relative fertility $1+r_{j,k}(\mathbf{x})$ with respect to all individuals in the population. Afterwards, there is mutation to type $i = 1, \ldots, n$ with probability $u_{j,i,k}(\mathbf{x})$, so that

$$P_i(\mathbf{x}) = \sum_{k=1}^{d} \sum_{j=1}^{n} \frac{(1 + r_{j,k}(\mathbf{x}))x_{j,k} f_k u_{j,i,k}(\mathbf{x})}{1 + \sum_{j'=1}^{n} \sum_{l=1}^{d} x_{j',l} r_{j',l}(\mathbf{x}) f_l} \tag{85}$$

is the probability for the offspring to be of type $i = 1, \ldots, n$. This probability can be written in the form

$$P_i(\mathbf{x}) = \sum_{k=1}^{d} x_{i,k} f_k + \frac{1}{N} \sum_{k=1}^{d} \left(\tilde{\mu}_{i,k}(\mathbf{x}) + x_{i,k}\tilde{\rho}_{i,k}(\mathbf{x})\right) f_k + o(N^{-1}) \tag{86}$$

with $No(N^{-1}) \to 0$ uniformly in $\mathbf{x}$ as $N \to +\infty$, where $\tilde{\mu}_{i,k}(\mathbf{x})$ is defined in (11) and

$$\tilde{\rho}_{i,k}(\mathbf{x}) = \rho_{i,k}(\mathbf{x}) - \sum_{l=1}^{d} \sum_{j=1}^{n} \rho_{j,l}(\mathbf{x})x_{j,l} f_l. \tag{87}$$

Here, the conditional expected frequency of type i in the next cohort in age class k is given by

$$E_{\mathbf{x}}[x_{i,k}(1)] = (M\mathbf{x}_i)_k + \frac{1}{N}\phi_{i,k}(\mathbf{x}) + o(N^{-1}) \tag{88}$$

with

$$\phi_{i,1}(\mathbf{x}) = \sum_{k=1}^{d} \left(\tilde{\mu}_{i,k}(\mathbf{x}) + x_{i,k}\tilde{\rho}_{i,k}(\mathbf{x})\right) f_k \tag{89}$$

and

$$\phi_{i,k}(\mathbf{x}) = x_{i,k-1}C_{k-1}\tilde{\sigma}_{i,k-1}(\mathbf{x}) \tag{90}$$

with $\tilde{\sigma}_{i,k-1}(\mathbf{x})$ defined in (19) for $k = 2, \ldots, d$, where

$$M = \begin{pmatrix} f_1 & f_2 & f_3 & \ldots & f_d \\ 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & 1 & 0 \end{pmatrix} \tag{91}$$

is a backward transition matrix under neutrality. Note that this matrix coincide with the backward transition matrix in the case of soft selection if $p_k = f_k$ for $k = 1, \ldots, d$.

Under the additional assumptions that $f_d > 0$ and $\gcd\{k : 1 \leq k \leq d,\ f_k > 0\} = 1$, the stochastic matrix M in (91) is necessarily irreducible and aperiodic. Its stationary distribution $\mathbf{w} = (w_1, \ldots, w_d)^T$ is given by

$$w_k = \frac{1}{\bar{k}} \sum_{l=k}^{d} f_l \tag{92}$$

for $k = 1, \ldots, d$ with

$$\bar{k} = \sum_{k=1}^{d} k f_k \tag{93}$$

being the mean age of reproduction in the neutral model.

Appendix B: approximations for a multivariate Wallenius' non-central hypergeometric distribution

Definition of Wallenius' distribution

Consider an urn model for R balls of n different colors and weights in a biased sampling. Let m_i be the number of balls of color i with weight ω_i for $i = 1, \ldots, n$, so that $\mathbf{m} = (m_1, m_2, \ldots, m_n)^T \in \mathbb{N}^n$, $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_n)^T \in \mathbb{R}^n$, and $R = \sum_{i=1}^{n} m_i$. Let us sample r balls without replacement and count the number of balls of each color. The result of the experiment is a random vector $(Y_1, \ldots, Y_n)$ with state space

$$\mathrm{S} = \left\{ \mathbf{y} = (y_1, \ldots, y_n)^T \in \mathbb{N}^n \ : \ \sum_{i=1}^{n} y_i = r \right\}, \tag{94}$$

and mass function

$$p(\mathbf{y}, \boldsymbol{\omega}) = \left(\prod_{i=1}^{n} \binom{m_i}{y_i} \right) \int_0^1 \left(\prod_{k=1}^{n} (1 - t^{\omega_k / D(\boldsymbol{\omega})})^{y_k} \right) dt, \tag{95}$$

where

$$D(\boldsymbol{\omega}) = \boldsymbol{\omega}^T (\mathbf{m} - \mathbf{y}) = \sum_{i=1}^{n} \omega_i (m_i - y_i). \tag{96}$$

This probability distribution is represented by $mwnchypg(r, R, \mathbf{m}, \boldsymbol{\omega})$. It was first studied by Wallenius (1963) in the univariate case and by Chesson (1976) in the multivariate case.

Approximation of the mass function under weak bias

We want to approximate the distribution $mwnchypg(r, R, \mathbf{m}, \boldsymbol{\omega})$ around the point $\boldsymbol{\omega} = \mathbf{1}$. This is what is needed to study the first-order effects of differences in viability from one age class to the next that are of order N^{-1} as $N \to +\infty$. Note that

$$f(\mathbf{y}, \boldsymbol{\omega}, t) = \prod_{i=1}^{n} (1 - t^{\omega_i / D(\boldsymbol{\omega})})^{y_i}, \tag{97}$$

for $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_k, \ldots, \omega_n)^T \in \mathbb{R}^n$, $0 < t < 1$ and $\mathbf{y} \in \mathrm{S}$, has partial derivative with respect to ω_k given by

$$\begin{aligned} &\frac{\partial f}{\partial \omega_k}(\mathbf{y}, \boldsymbol{\omega}, t) \\ &\quad = \sum_{j=1}^{n} y_j \left(\frac{\omega_j (m_k - y_k) - D(\boldsymbol{\omega}) \delta_{jk}}{D(\boldsymbol{\omega})^2} \right) \left(\frac{t^{\omega_j / D(\boldsymbol{\omega})} \ln(t)}{1 - t^{\omega_j / D(\boldsymbol{\omega})}} \right) \left(\prod_{i=1}^{n} (1 - t^{\omega_i / D(\boldsymbol{\omega})})^{y_i} \right). \end{aligned} \tag{98}$$

This implies that (see, e.g., Godement 2005, p. 36)

$$\frac{\partial p}{\partial \omega_k}(\mathbf{y}, \boldsymbol{\omega}) = \left(\prod_{i=1}^{n} \binom{m_i}{y_i} \right) \frac{\partial}{\partial \omega_k} \left(\int_0^1 f(\mathbf{y}, \boldsymbol{\omega}, t) dt \right)$$

$$= \left(\prod_{i=1}^{n} \binom{m_i}{y_i} \right) \int_0^1 \frac{\partial f}{\partial \omega_k}(\mathbf{y}, \boldsymbol{\omega}, t) dt. \tag{99}$$

Moreover,

$$\frac{\partial p}{\partial \omega_k}(\mathbf{y}, \boldsymbol{\omega}) = \left(\prod_{i=1}^{n} \binom{m_i}{y_i} \right) \sum_{j=1}^{n} y_j \frac{\omega_j (m_k - y_k) - D(\boldsymbol{\omega}) \delta_{jk}}{D(\boldsymbol{\omega})^2} I_j(\mathbf{y}, \boldsymbol{\omega}) \tag{100}$$

with

$$I_j(\mathbf{y}, \boldsymbol{\omega}) = \int_0^1 \frac{t^{\omega_j / D(\boldsymbol{\omega})} \ln(t)}{1 - t^{\omega_j / D(\boldsymbol{\omega})}} \left(\prod_{i=1}^{n} (1 - t^{\omega_i / D(\boldsymbol{\omega})})^{y_i} \right) dt \tag{101}$$

for $j = 1, \ldots, n$. This uses the fact that

$$t^{\omega_i / D(\boldsymbol{\omega})} = \exp\left\{ \frac{\omega_i}{D(\boldsymbol{\omega})} \ln(t) \right\}. \tag{102}$$

At $\boldsymbol{\omega} = \mathbf{1}$, we have

$$\frac{\partial p}{\partial \omega_k}(\mathbf{y}, \mathbf{1}) = \left(\prod_{i=1}^{n} \binom{m_i}{y_i} \right) \left(\frac{r m_k - R y_k}{(R - r)^2} \right) I(\mathbf{y}, \mathbf{1}) \tag{103}$$

with

$$I(\mathbf{y}, \mathbf{1}) = \int_0^1 t^{(R-r)^{-1}} \ln(t) (1 - t^{(R-r)^{-1}})^{r-1} dt.$$

With the change of variable $x = t^{(R-r)^{-1}}$, the above integral becomes

$$I(\mathbf{y}, \mathbf{1}) = (R - r)^2 \int_0^1 \ln(x) x^{(R-r)} (1 - x)^{r-1} dx = (R - r)^2 \left. \frac{\partial B}{\partial z}(z, u) \right|_{(R-r+1, r)}, \tag{104}$$

where

$$B(z, u) = \int_0^1 x^{z-1} (1 - x)^{u-1} dx = \frac{\Gamma(z) \Gamma(u)}{\Gamma(z + u)} \tag{105}$$

and

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \tag{106}$$

are respectively the beta and gamma functions (see, e.g., Stegan 1964, pp. 255–258).

The partial derivative of the beta function with respect to the first variable is given by

$$\frac{\partial B}{\partial z}(z,u)=\Gamma(u)\frac{\Gamma(z+u)\Gamma'(z)-\Gamma'(z+u)\Gamma(z)}{(\Gamma(z+u))^2}=B(z,u)\left(\psi(z)-\psi(z+u)\right) \tag{107}$$

with

$$\psi(z)=\frac{d}{dz}\ln(\Gamma(z))=\frac{\Gamma'(z)}{\Gamma(z)}. \tag{108}$$

This function verifies

$$\psi(m_1)-\psi(m_2)=\sum_{k=m_2}^{m_1-1}\frac{1}{k} \tag{109}$$

for integers m_1 and m_2 satisfying $m_1 > m_2 \geq 1$. Using the fact that $\Gamma(m)=(m-1)!$ for $m \in \mathbb{N}$, we find

$$\begin{aligned}\left.\frac{\partial B}{\partial z}(z,u)\right|_{(R-r+1,r)} &= \frac{\Gamma(R-r+1)\Gamma(r)}{\Gamma(R+1)}\left(\psi(R-r+1)-\psi(R+1)\right)\\ &= -\frac{(R-r)!(r-1)!}{R!}S_{r,R}\end{aligned} \tag{110}$$

with

$$S_{r,R}=\sum_{k=R-r+1}^{R}\frac{1}{k}. \tag{111}$$

Replacing Eq. (110) in Eq. (104), we get

$$\begin{aligned}I(\mathbf{y},\mathbf{1}) &= (R-r)^2\left.\frac{\partial B}{\partial z}(z,u)\right|_{(R-r+1,r)}\\ &= -(R-r)^2\frac{(R-r)!(r-1)!}{R!}S_{r,R}\\ &= -\frac{(R-r)^2 S_{r,R}}{\binom{R}{r}r}.\end{aligned} \tag{112}$$

Then Eq. (103) can be written as

$$\frac{\partial p}{\partial \omega_k}(\mathbf{y},\mathbf{1})=p(\mathbf{y},\mathbf{1})\left(y_k\frac{R}{r}-m_k\right)S_{r,R} \tag{113}$$

with

$$p(\mathbf{y},\mathbf{1})=\frac{\prod_{i=1}^{n}\binom{m_i}{y_i}}{\binom{R}{r}} \tag{114}$$

being the mass function of a multivariate hypergeometric distribution with parameters given by $(r, R, \mathbf{m})$, represented by $mhypg(r, R, \mathbf{m})$.

The second partial derivatives of the mass function of a $mhypg(R, r, \mathbf{m})$ random vector are given by

$$\begin{aligned}\frac{\partial}{\partial\omega_l}\frac{\partial p}{\partial\omega_k}(\mathbf{y},\boldsymbol{\omega})=&\left(\prod_{i=1}^{n}\binom{m_i}{y_i}\right)\sum_{j=1}^{n}y_j\left(\frac{\partial}{\partial\omega_l}\left(\frac{\omega_j(m_k-y_k)-D(\boldsymbol{\omega})\delta_{jk}}{D(\boldsymbol{\omega})^2}\right)I_j(\mathbf{y},\boldsymbol{\omega})\right)\\&+y_j\left(\frac{\omega_j(m_k-y_k)-D(\boldsymbol{\omega})\delta_{jk}}{D(\boldsymbol{\omega})^2}\frac{\partial}{\partial\omega_l}\left(I_j(\mathbf{y},\boldsymbol{\omega})\right)\right)\end{aligned} \tag{115}$$

where

$$I_j(\mathbf{y},\boldsymbol{\omega})=\int_0^1\frac{t^{\omega_j/D(\boldsymbol{\omega})}\ln(t)}{1-t^{\omega_j/D(\boldsymbol{\omega})}}f(\mathbf{y},\boldsymbol{\omega},t)dt. \tag{116}$$

We have that

$$\begin{aligned}\frac{\partial}{\partial\omega_l}\left(\frac{\omega_j(m_k-y_k)-D(\boldsymbol{\omega})\delta_{jk}}{D(\boldsymbol{\omega})^2}\right)&=\frac{\delta_{lj}D(\boldsymbol{\omega})-2\omega_l(m_l-y_l)}{(D(\boldsymbol{\omega}))^3}+\frac{\delta_{jk}}{(D(\boldsymbol{\omega}))^2}\\&=\frac{(\delta_{lj}+\delta_{jk})D(\boldsymbol{\omega})-2\omega_l(m_l-y_l)}{(D(\boldsymbol{\omega}))^3},\end{aligned} \tag{117}$$

$$\frac{\partial}{\partial\omega_l}\left(\frac{t^{\omega_j/D(\boldsymbol{\omega})}}{1-t^{\omega_j/D(\boldsymbol{\omega})}}\right)=\frac{t^{\omega_j/D(\boldsymbol{\omega})}\ln(t)\left(D(\boldsymbol{\omega})\delta_{lj}-\omega_j(m_l-y_l)\right)}{(D(\boldsymbol{\omega})(1-t^{\omega_j/D(\boldsymbol{\omega})}))^2} \tag{118}$$

and

$$\frac{\partial}{\partial\omega_l}\left(f(\mathbf{y},\boldsymbol{\omega},t)\right)=\sum_{j=1}^{n}y_j\left(\frac{\omega_j(m_l-y_l)-D(\boldsymbol{\omega})\delta_{jl}}{D(\boldsymbol{\omega})^2}\right)\frac{t^{\omega_j/D(\boldsymbol{\omega})}\ln(t)}{1-t^{\omega_j/D(\boldsymbol{\omega})}}f(\mathbf{y},\boldsymbol{\omega},t). \tag{119}$$

Then

$$\frac{\partial}{\partial\omega_l}\frac{\partial p}{\partial\omega_k}(\mathbf{y},\boldsymbol{\omega})=\left(\prod_{i=1}^{n}\binom{m_i}{y_i}\right)\sum_{j=1}^{n}y_j\frac{(\delta_{lj}+\delta_{jk})D(\boldsymbol{\omega})-2\omega_l(m_l-y_l)}{(D(\boldsymbol{\omega}))^3}I_j(\mathbf{y},\boldsymbol{\omega})$$

$$+ y_j \left(\frac{\omega_j(m_k - y_k) - D(\boldsymbol{\omega})\delta_{jk}}{D(\boldsymbol{\omega})^2} J_{lj}(\mathbf{y}, \boldsymbol{\omega}) \right) \tag{120}$$

with

$$\begin{aligned} J_{lj}(\mathbf{y}, \boldsymbol{\omega}) =& \frac{\partial}{\partial \omega_l} I_j(\mathbf{y}, \boldsymbol{\omega}) \\ =& \int_0^1 \frac{t^{\omega_j/D(\boldsymbol{\omega})}(\ln(t))^2}{(D(\boldsymbol{\omega})(1 - t^{\omega_j/D(\boldsymbol{\omega})}))^2} f(\mathbf{y}, \boldsymbol{\omega}, t) \\ & \cdot \Bigg(D(\boldsymbol{\omega})\delta_{lj} - \omega_j(m_l - y_l) \\ & + \sum_{k=1}^{n} y_k \frac{(\omega_k(m_l - y_l) - D(\boldsymbol{\omega})\delta_{kl})\, t^{\omega_k/D(\boldsymbol{\omega})}}{1 - t^{\omega_k/D(\boldsymbol{\omega})}} \Bigg) dt. \end{aligned} \tag{121}$$

Approximation of the k-th moments under weak bias

According to Eq. (113), with $(H_1, \ldots, H_n)$ representing a $mhypg(R, r, \mathbf{m})$ random vector, we have

$$\begin{aligned} E\left[\left(\frac{Y_i}{r}\right)^k\right] =& E\left[\left(\frac{H_i}{r}\right)^k\right] \\ & + \sum_{l=1}^{n} (\omega_l - 1) \sum_{\mathbf{y} \in S} \left(\frac{y_i}{r}\right)^k p(\mathbf{y}, \mathbf{1}) \left(y_l \frac{R}{r} - m_l \right) S_{r,R} + o(|\boldsymbol{\omega} - \mathbf{1}|), \end{aligned} \tag{122}$$

where

$$\sum_{\mathbf{y} \in S} \left(\frac{y_i}{r}\right)^k p(\mathbf{y}, \mathbf{1}) \left(y_l \frac{R}{r} - m_l \right) = E[H_i^k H_l] \frac{R}{r^k} - \frac{m_l}{r^k} E[H_i^k] \tag{123}$$

and

$$\frac{o(|\boldsymbol{\omega} - \mathbf{1}|)}{|\boldsymbol{\omega} - \mathbf{1}|} \to 0 \tag{124}$$

as $|\boldsymbol{\omega} - \mathbf{1}| \to 0$ with $| \cdot |$ denoting the Euclidean norm. Actually, the convergence is uniform in $\boldsymbol{\omega}$ as can be shown by induction on the number of balls removed, r, at least for $k = 1, 2, 3, 4$ (Soares 2019). Therefore, we have

$$E\left[\left(\frac{Y_i}{r}\right)^k\right] = E\left[\left(\frac{H_i}{r}\right)^k\right]$$

$$+ S_{r,R} \sum_{l=1}^{n} (\omega_l - 1) \left(E[H_i^k H_l] \frac{R}{r^{k+1}} - \frac{m_l}{r^k} E[H_i^k] \right) + o(|\boldsymbol{\omega} - \mathbf{1}|) \tag{125}$$

uniformly in $\boldsymbol{\omega}$ for the k-th moments.

For the first moments, for instance, we have

$$E\left[\frac{Y_i}{r}\right] = E\left[\frac{H_i}{r}\right] + S_{r,R} \sum_{l=1}^{n} (\omega_l - 1) \left(E[H_i H_l] \frac{R}{r^2} - \frac{m_l}{r} E[H_i] \right) + o(|\boldsymbol{\omega} - \mathbf{1}|) \tag{126}$$

uniformly in $\boldsymbol{\omega}$, where

$$E[H_i] = r \frac{m_i}{R} \tag{127}$$

and

$$E[H_i H_l] = \begin{cases} \frac{m_l m_i r(r-1)}{R(R-1)} & \text{if} \quad i \neq l, \\ \frac{m_i (m_i - 1) r(r-1)}{R(R-1)} + \frac{m_i r}{R} & \text{if} \quad i = l, \end{cases} \tag{128}$$

for $i, l = 1, \ldots, n$. This yields

$$E\left[\frac{Y_i}{r}\right] = \frac{m_i}{R} + \frac{m_i}{R} \frac{R(R-r)}{r(R-1)} S_{r,R} \left(\omega_i - 1 - \sum_{l=1}^{n} (\omega_l - 1) \frac{m_l}{R} \right) + o(|\boldsymbol{\omega} - \mathbf{1}|) \tag{129}$$

uniformly in $\boldsymbol{\omega}$ for $i = 1, \ldots, n$.

Approximation of the covariances under weak bias

According to Eq. (113), we have

$$E[Y_i Y_j] = E[H_i H_j] + \sum_{l=1}^{n} (\omega_l - 1) \sum_{\mathbf{y} \in S} y_i y_j p(\mathbf{y}, \mathbf{1}) \left(y_l \frac{R}{r} - m_l \right) S_{r,R} + o(|\boldsymbol{\omega} - \mathbf{1}|), \tag{130}$$

where

$$\sum_{\mathbf{y} \in S} y_i y_j p(\mathbf{y}, \mathbf{1}) \left(y_l \frac{R}{r} - m_l \right) = E[H_i H_j H_l] \frac{R}{r} - m_l E[H_i H_j]. \tag{131}$$

Therefore,

$$E\left[\frac{Y_i}{r}\frac{Y_j}{r}\right] = E\left[\frac{H_i}{r}\frac{H_j}{r}\right] + S_{r,R}\sum_{l=1}^{n}(\omega_l - 1)\left(E[H_i H_j H_l]\frac{R}{r^3} - \frac{m_l}{r^2}E[H_i H_j]\right) + o(|\boldsymbol{\omega} - \mathbf{1}|) \tag{132}$$

for $i, j = 1, \ldots, n$, where $o(|\boldsymbol{\omega} - \mathbf{1}|)$ is actually uniform in $\boldsymbol{\omega}$ as can be proved by induction on the number of balls removed, r (Soares 2019). Here, $E[H_i H_j]$ is given by (128), while

$$E[H_i H_j H_l] = \begin{cases} m_i m_j m_l \frac{r(r-1)(r-2)}{R(R-1)(R-2)} & \text{if } i \neq j,\ j \neq l \text{ and } l \neq i, \\ m_i m_j \frac{r(r-1)}{R(R-1)}\left(1 + (m_i - 1)\frac{r-2}{R-2}\right) & \text{if } i \neq j \text{ and } l = i, \\ m_i \frac{r}{R}\left[1 + (m_i - 1)\frac{r-1}{R-1}\left(3 + (m_i - 2)\frac{r-2}{R-2}\right)\right] & \text{if } i = j = l. \end{cases} \tag{133}$$

This gives an approximation for the covariances

$$Cov\left[\frac{Y_i}{r}, \frac{Y_j}{r}\right] = E\left[\frac{Y_i}{r}\frac{Y_j}{r}\right] - E\left[\frac{Y_i}{r}\right]E\left[\frac{Y_j}{r}\right] \tag{134}$$

for $i, j = 1, \ldots, n$.

Approximations in the case of a large sample size in a large population

Consider that the weight associated to a ball of color i is in the form

$$\omega_i - 1 = \frac{f}{R}\sigma_i \tag{135}$$

with $0 < f < 1$ and σ_i constant for $i = 1, \ldots, n$. As $R \to +\infty$, let us keep constant the ratios

$$\beta = \frac{r}{R} \tag{136}$$

and

$$x_i = \frac{m_i}{R} \tag{137}$$

for $i = 1, \ldots, n$. Then, defining

$$X_i = \frac{Y_i}{r} = \frac{Y_i}{\beta R} \tag{138}$$

and using (129) for its first moment, we get

$$E[X_i] = x_i + \frac{f}{R} x_i \left(\frac{1-\beta}{\beta}\right) S \left(\sigma_i - \sum_{l=1}^{n} \sigma_l x_l\right) + o(R^{-1}), \tag{139}$$

where

$$S = \lim_{R \to +\infty} S_{\beta R, R} = \ln\left(\frac{1}{1-\beta}\right). \tag{140}$$

This gives

$$E[X_i] = x_i + \frac{x_i f}{R} C \tilde{\sigma}_i + o(R^{-1}), \tag{141}$$

where

$$C = \frac{1-\beta}{\beta} \ln\left(\frac{1}{1-\beta}\right) = \frac{R-r}{r} \ln\left(\frac{R}{R-r}\right) \tag{142}$$

and

$$\tilde{\sigma}_i = \sigma_i - \sum_{l=1}^{n} \sigma_l x_l. \tag{143}$$

On the other hand, using (128) and (133), it can be checked that we have the following limits as $R \to +\infty$ with β and x_i for $i = 1, \ldots, n$ being kept fixed :

$i \neq j, l \neq i, j$	$E[H_i H_j H_l]\frac{R}{r^3} - E[H_i H_j]\frac{m_l}{r^2} \to -\frac{2(1-\beta)x_i x_j x_l}{\beta}$	(144)
$i \neq j, l = i$	$E[H_i^2 H_j]\frac{R}{r^3} - E[H_i H_j]\frac{m_i}{r^2} \to \frac{(1-\beta)x_i(1-2x_i)x_j}{\beta}$	
$i = j, l \neq i$	$E[H_i^2 H_l]\frac{R}{r^3} - E[H_i^2]\frac{m_l}{r^2} \to -\frac{2(1-\beta)x_i^2 x_l}{\beta}$	
$i = j = l$	$E[H_i^3]\frac{R}{r^3} - E[H_i^2]\frac{m_i}{r^2} \to \frac{2(1-\beta)(1-x_i)x_i^2}{\beta}$	

Therefore, (144) and (132) lead to

$$E[X_i X_j] = E\left[\frac{H_i H_j}{r^2}\right] + \frac{f}{R}\left(\frac{1-\beta}{\beta}\right) x_i x_j S \left(\sigma_i + \sigma_j - 2\sum_{l=1}^{n} \sigma_l x_l\right) + o(R^{-1}). \tag{145}$$

On the other hand, owing to (129) and (144), we have

$$E[X_i]E[X_j] = E\left[\frac{H_i}{r}\right]E\left[\frac{H_j}{r}\right] + \frac{f}{R}\left(\frac{1-\beta}{\beta}\right)x_i x_j S\left(\sigma_i + \sigma_j - 2\sum_{l=1}^{n}\sigma_l x_l\right) + o(R^{-1}). \quad (146)$$

Therefore, for the covariance, we get

$$Cov[X_i, X_j] = Cov\left[\frac{H_i}{r}, \frac{H_j}{r}\right] + o(R^{-1}), \quad (147)$$

where

$$Cov\left[\frac{H_i}{r}, \frac{H_j}{r}\right] = \frac{1}{R}x_i(\delta_{ij} - x_j)\left(\frac{1-\beta}{\beta}\right) + o(R^{-1}). \quad (148)$$

As for the third moment, we have

$$E\left[\left(\frac{Y_i}{r}\right)^3\right] = \frac{1}{r^3}\left(\frac{3m_i^{(2)}r^{(2)}}{R^{(2)}} + \frac{m_i^{(3)}r^{(3)}}{R^{(3)}} + \frac{m_i r}{R}\right) + \frac{R(R-r)}{r^3}S_{r,R}\left(\frac{m_i}{R^{(2)}} + \frac{6m_i^{(2)}(r-1)}{R^{(3)}} + \frac{3m_i^{(3)}(r-1)^{(2)}}{R^{(4)}}\right) \cdot \left(\omega_i - \sum_{l=1}^{n}\omega_l \frac{m_l}{R}\right) + o(|\boldsymbol{\omega} - \mathbf{1}|), \quad (149)$$

with the notation $r^{(k)} = r(r-1)\cdots(r-k+1)$, where

$$\frac{3m_i^{(2)}r^{(2)}}{r^3 R^{(2)}} = \frac{3x_i(x_i - 1/R)(\beta - 1/R)}{\beta^2 R(1 - 1/R)} = o(1), \quad (150)$$

$$\frac{m_i^{(3)}r^{(3)}}{r^3 R^{(3)}} = \frac{x_i(x_i - 1/R)(x_i - 2/R)\beta(\beta - 1/R)(\beta - 2/R)}{\beta^3(1 - 1/R)(1 - 2/R)} = x_i^3 + o(1), \quad (151)$$

$$\frac{m_i r}{r^3 R} = \frac{x_i}{(\beta R)^2} = o(1), \quad (152)$$

$$\frac{m_i}{r R^{(2)}} = \frac{x_i}{\beta R(R-1)} = o(1), \quad (153)$$

$$\frac{6m_i^{(2)}(r-1)}{r R^{(3)}} = \frac{6x_i(x_i - 1/R)(\beta - 1/R)}{\beta R(1 - 1/R)(1 - 2/R)} = o(1), \quad (154)$$

$$\frac{3m_i^{(3)}(r-1)^{(2)}}{r R^{(4)}} = \frac{3x_i(x_i - 1/R)(x_i - 2/R)(\beta - 1/R)(\beta - 2/R)}{\beta(1 - 1/R)(1 - 2/R)(1 - 3/R)} = 3x_i^3\beta + o(1) \quad (155)$$

and

$$\frac{R(R-r)}{r^2}S_{r,R}=\frac{(1-\beta)}{\beta^2}S_{\beta R,R}=\frac{(1-\beta)}{\beta^2}\ln\left(\frac{1}{1-\beta}\right)+o(1). \tag{156}$$

This yields

$$\begin{aligned}E\left[X_i^3\right]&=x_i^3+\frac{3x_i^3f(1-\beta)}{R\beta}\ln\left(\frac{1}{1-\beta}\right)\left(\sigma_i-\sum_{l=1}^{n}\sigma_l x_l\right)+o(R^{-1})\\&=x_i^3+\frac{3x_i^3f}{R}C\tilde{\sigma}_i+o(R^{-1}).\end{aligned} \tag{157}$$

Similarly, for the fourth moment, we have

$$\begin{aligned}E\left[\left(\frac{Y_i}{r}\right)^4\right]&=\frac{1}{r^4}\left(\frac{7m_i^{(2)}r^{(2)}}{R^{(2)}}+\frac{6m_i^{(3)}r^{(3)}}{R^{(3)}}+\frac{m_i^{(4)}r^{(4)}}{R^{(4)}}+\frac{m_ir}{R}\right)\\&\quad+\frac{R(R-r)}{r^5}S_{r,R}\left(\frac{m_ir}{R^{(2)}}+\frac{14m_i^{(2)}r^{(2)}}{R^{(3)}}+\frac{18m_i^{(3)}r^{(3)}}{R^{(4)}}+\frac{4m_i^{(4)}r^{(4)}}{R^{(5)}}\right)\\&\quad\cdot\left(\omega_i-\sum_{l=1}^{n}\omega_l\frac{m_l}{R}\right)+o(|\boldsymbol{\omega}-\mathbf{1}|),\end{aligned} \tag{158}$$

where

$$\frac{7m_i^{(2)}r^{(2)}}{r^4R^{(2)}}=\frac{7x_i(x_i-1/R)\beta(\beta-1/R)}{\beta^4R(R-1)}=o(1), \tag{159}$$

$$\frac{6m_i^{(3)}r^{(3)}}{r^4R^{(3)}}=\frac{6x_i(x_i-1/R)(x_i-2/R)\beta(\beta-1/R)(\beta-2/R)}{\beta^4R(1-1/R)(1-2/R)}=o(1), \tag{160}$$

$$\begin{aligned}\frac{m_i^{(4)}r^{(4)}}{r^4R^{(4)}}&=\frac{x_i(x_i-1/R)(x_i-2/R)(x_i-3/R)\beta(\beta-1/R)(\beta-2/R)(\beta-3/R)}{\beta^4(1-1/R)(1-2/R)(1-3/R)}\\&=x_i^4+o(1),\end{aligned} \tag{161}$$

$$\frac{m_ir}{r^4R}=\frac{x_i}{\beta^3R^3}=o(1), \tag{162}$$

$$\frac{m_ir}{r^3R^{(2)}}=\frac{x_i}{\beta^2R^2(R-1)}=o(1), \tag{163}$$

$$\frac{14m_i^{(2)}r^{(2)}}{r^3R^{(3)}}=\frac{14x_i(x_i-1/R)(\beta-1/R)}{\beta^2(R-1)(R-2)}=o(1), \tag{164}$$

$$\frac{18m_i^{(3)}r^{(3)}}{r^3R^{(4)}}=\frac{18x_i(x_i-1/R)(x_i-2/R)(\beta-1/R)(\beta-2/R)}{\beta^2R(1-1/R)(1-2/R)(1-3/R)}=o(1) \tag{165}$$

and

$$\frac{4m_i^{(4)}r^{(4)}}{r^3R^{(5)}} = \frac{4x_i(x_i-1/R)(x_i-2/R)(x_i-3/R)(\beta-1/R)(\beta-2/R)(\beta-3/R)}{\beta^2(1-1/R)(1-2/R)(1-3/R)(1-4/R)}$$
$$= 4x_i^4\beta + o(1). \tag{166}$$

This yields

$$E\left[X_i^4\right] = x_i^4 + \frac{4x_i^4 f(1-\beta)}{R\beta}\ln\left(\frac{1}{1-\beta}\right)\left(\sigma_i - \sum_{l=1}^{n}\sigma_l x_l\right) + o(R^{-1})$$
$$= x_i^4 + \frac{4x_i^4 f}{R}C\tilde{\sigma}_i + o(R^{-1}). \tag{167}$$

More generally, we expect

$$E[X_i^t] = x_i^t + \frac{t x_i^t f}{R}C\tilde{\sigma}_i + o(R^{-1}) \tag{168}$$

for every integer $t \geq 1$.

Finally, for the fourth central moment, we have

$$\begin{aligned}
E[(X_i - E[X_i])^4] &= E[X_i^4] - 4E[X_i^3]E[X_i] + 6E[X_i^2]E[X_i]^2 - 3E[X_i]^4 \\
&= x_i^4 + \frac{4x_i^4 f}{R}C\tilde{\sigma}_i + o(R^{-1}) \\
&\quad - 4\left(x_i^3 + \frac{3x_i^3 f}{R}C\tilde{\sigma}_i + o(R^{-1})\right)\left(x_i + \frac{x_i f}{R}C\tilde{\sigma}_i + o(R^{-1})\right) \\
&\quad + 6\left(x_i^2 + \frac{2x_i^2 f}{R}C\tilde{\sigma}_i + o(R^{-1})\right)\left(x_i + \frac{x_i f}{R}C\tilde{\sigma}_i + o(R^{-1})\right)^2 \\
&\quad - 3\left(x_i + \frac{x_i f}{R}C\tilde{\sigma}_i + o(R^{-1})\right)^4 \\
&= \left(x_i^4 - 4x_i^4 + 6x_i^4 - 3x_i^4\right) \\
&\quad + (4x_i^4 - 16x_i^4 + 24x_i^4 - 12x_i^4)\frac{f}{R}C\tilde{\sigma}_i + o(R^{-1}) \\
&= o(R^{-1}).
\end{aligned} \tag{169}$$

Appendix C: conditions for a diffusion approximation

Condition I

Using (15), we have

$$E_{\mathbf{x}}[\mathbf{w}^T\mathbf{x}_i(1)] = \sum_{k=1}^{d} w_k E_{\mathbf{x}}[\mathbf{x}_i(1)] = \sum_{k=1}^{d} w_k (M\mathbf{x}_i)_k + \frac{1}{N}\sum_{k=1}^{d} w_k \phi_{i,k}(\mathbf{x}) + o(N^{-1}) \tag{170}$$

with

$$\sum_{k=1}^{d} w_k (M\mathbf{x}_i)_k = \mathbf{w}^T M\mathbf{x}_i = \mathbf{w}^T\mathbf{x}_i \tag{171}$$

owing to the fact that $\mathbf{w}^T M = \mathbf{w}^T$. Then, using (44), we have

$$E_{\mathbf{x}}[\Delta z_i] = E_{\mathbf{x}}[\mathbf{w}^T\mathbf{x}_i(1) - \mathbf{w}^T\mathbf{x}_i] = \frac{1}{N}\sum_{k=1}^{d} w_k \phi_{i,k}(\mathbf{z}\mathbf{1}^T + \mathbf{y}) + o(N^{-1}) \tag{172}$$

as $N \to +\infty$.

Condition II

We have

$$E_{\mathbf{x}}[(\Delta z_i)(\Delta z_j)] = E_{\mathbf{x}}[z_i(1)z_j(1)] - z_i E_{\mathbf{x}}[z_j(1)] - z_j E_{\mathbf{x}}[z_i(1)] + z_i z_j \tag{173}$$

with

$$E_{\mathbf{x}}[z_i(1)z_j(1)] = Cov_{\mathbf{x}}[z_i(1), z_j(1)] + E_{\mathbf{x}}[z_i(1)]E_{\mathbf{x}}[z_j(1)]. \tag{174}$$

Owing to condition I, we have

$$\begin{aligned} E_{\mathbf{x}}[z_i(1)]E_{\mathbf{x}}[z_j(1)] - z_i E_{\mathbf{x}}[z_j(1)] - z_j E_{\mathbf{x}}[z_i(1)] + z_i z_j &= E_{\mathbf{x}}[(\Delta z_i)]E_{\mathbf{x}}[(\Delta z_j)] \\ &= o(N^{-1}). \end{aligned} \tag{175}$$

Therefore,

$$E_{\mathbf{x}}[(\Delta z_i)(\Delta z_j)] = Cov_{\mathbf{x}}[z_i(1), z_j(1)] + o(N^{-1}). \tag{176}$$

Since the variables $x_{i,k}(1)$ and $x_{j,l}(1)$ are independent as soon as $l \neq k$, we have

$$
\begin{aligned}
Cov_{\mathbf{x}}[z_i(1), z_j(1)] &= Cov_{\mathbf{x}}\left[\sum_{k=1}^{d} w_k x_{i,k}(1), \sum_{k=1}^{d} w_k x_{j,k}(1)\right] \\
&= \sum_{k=1}^{d} w_k^2 Cov_{\mathbf{x}}[x_{i,k}(1), x_{j,k}(1)]. \qquad (177)
\end{aligned}
$$

The vector $(x_{1,1}(1), \ldots, x_{n,1}(1))$ having a multinomial probability distribution with parameters $N_1, P_1(\mathbf{x}), \ldots, P_n(\mathbf{x})$, the covariances of the components are given by

$$
Cov_{\mathbf{x}}[x_{i,1}(1), x_{j,1}(1)] = -\frac{1}{N_1} P_i(\mathbf{x}) P_j(\mathbf{x}) = -\frac{1}{N_1}(M\mathbf{x}_i)_1 (M\mathbf{x}_j)_1 + o(N^{-1}) \qquad (178)
$$

and

$$
\begin{aligned}
Cov_{\mathbf{x}}[x_{i,1}(1), x_{i,1}(1)] =& \frac{1}{N_1} P_i(\mathbf{x})(1 - P_i(\mathbf{x})) \\
=& \frac{1}{N_1}(M\mathbf{x}_i)_1 (1 - (M\mathbf{x}_i)_1) + o(N^{-1}) \qquad (179)
\end{aligned}
$$

for $i, j = 1, \ldots, n$ with $i \neq j$, using Eq. (10). Therefore, by (15), we have

$$
Cov_{\mathbf{x}}[x_{i,1}(1), x_{j,1}(1)] = \frac{1}{N f_1}(M\mathbf{x}_i)_1 \left(\delta_{ij} - (M\mathbf{x}_j)_1\right) + o(N^{-1}) \qquad (180)
$$

with $\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ii} = 1$, for $i, j = 1, \ldots, n$. For $k = 2, \ldots, d$, and with reference to "Appendix B" (with $R = N_{k-1}$ and $\beta = f_k / f_{k-1}$), we have

$$
\begin{aligned}
Cov_{\mathbf{x}}[x_{i,k}(1), x_{j,k}(1)] &= \frac{1}{N} x_{i,k-1}(\delta_{ij} - x_{j,k-1}) \left(\frac{1}{f_k} - \frac{1}{f_{k-1}}\right) + o(N^{-1}) \\
&= \frac{1}{N}(M\mathbf{x}_i)_k (\delta_{ij} - (M\mathbf{x}_j)_k) \left(\frac{1}{f_k} - \frac{1}{f_{k-1}}\right) + o(N^{-1}). \qquad (181)
\end{aligned}
$$

Therefore, condition II holds with

$$
a_{i,j}(\mathbf{z}, \mathbf{y}) = \sum_{k=1}^{d} w_k^2 ((M\mathbf{y}_i)_k + z_i)(\delta_{ij} - (M\mathbf{y}_j)_k - z_j) \left(\frac{1}{f_k} - \frac{1}{f_{k-1}}\right) \qquad (182)
$$

using the convention that $1/f_0 = 0$.

Condition III

We start with

$$\begin{aligned}
&E_{\mathbf{x}}[(\Delta z_i)^4] - E_{\mathbf{x}}[(z_i(1) - E_{\mathbf{x}}[z_i(1)])^4] \\
&\quad = E_{\mathbf{x}}[(z_i(1) - z_i)^4] - E_{\mathbf{x}}[(z_i(1) - E_{\mathbf{x}}[z_i(1)])^4] \\
&\quad = E_{\mathbf{x}}[z_i(1)^4] - 4z_i E_{\mathbf{x}}[z_i(1)^3] + 6z_i^2 E_{\mathbf{x}}[z_i(1)^2] - 4z_i^3 E_{\mathbf{x}}[z_i(1)] + z_i^4 \\
&\qquad - \Big(E_{\mathbf{x}}[z_i(1)^4] - 4E_{\mathbf{x}}[z_i(1)]E_{\mathbf{x}}[z_i(1)^3] + 6E_{\mathbf{x}}[z_i(1)]^2 E_{\mathbf{x}}[z_i(1)^2] \\
&\qquad - 4E_{\mathbf{x}}[z_i(1)]^3 E_{\mathbf{x}}[z_i(1)] + E_{\mathbf{x}}[z_i(1)]^4\Big) \\
&\quad = -4z_i E_{\mathbf{x}}[z_i(1)^3] + 6z_i^2 E_{\mathbf{x}}[z_i(1)^2] - 4z_i^3 E_{\mathbf{x}}[z_i(1)] + z_i^4 \\
&\qquad + 4E_{\mathbf{x}}[z_i(1)]E_{\mathbf{x}}[z_i(1)^3] - 6E_{\mathbf{x}}[z_i(1)]^2 E_{\mathbf{x}}[z_i(1)^2] + 3E_{\mathbf{x}}[z_i(1)]^4 \\
&\quad = 4E_{\mathbf{x}}[z_i(1)^3](E_{\mathbf{x}}[z_i(1)] - z_i) + 6E_{\mathbf{x}}[z_i(1)^2](z_i^2 - E_{\mathbf{x}}[z_i(1)]^2) \\
&\qquad - 4z_i^3 E_{\mathbf{x}}[z_i(1)] + z_i^4 + 3E_{\mathbf{x}}[z_i(1)]^4. \qquad (183)
\end{aligned}$$

From conditions I and II, we have

$$E_{\mathbf{x}}[\Delta z_i] = b_i(\mathbf{z}, \mathbf{y}) N^{-1} + o\left(N^{-1}\right), \qquad (184a)$$

$$E_{\mathbf{x}}[(\Delta z_i)^2] = a_{i,i}(\mathbf{z}, \mathbf{y}) N^{-1} + o\left(N^{-1}\right), \qquad (184b)$$

from which

$$\begin{aligned}
&E_{\mathbf{x}}[(\Delta z_i)^4] - E_{\mathbf{x}}[(z_i(1) - E_{\mathbf{x}}[z_i(1)])^4] \\
&\quad = 4E_{\mathbf{x}}[z_i(1)^3](b_i(\mathbf{z}, \mathbf{y})N^{-1}) + 6(z_i^2 + 2z_i b_i(\mathbf{z}, \mathbf{y})N^{-1} \\
&\qquad + a_{i,i}(\mathbf{z}, \mathbf{y})N^{-1})(-2z_i b_i(\mathbf{z}, \mathbf{y})N^{-1}) \\
&\qquad - 4z_i^3(z_i + b_i(\mathbf{z}, \mathbf{y})N^{-1}) + z_i^4 + 3(z_i^4 + 4z_i^3 b_i(\mathbf{z}, \mathbf{y})N^{-1}) + o\left(N^{-1}\right) \\
&\quad = 4b_i(\mathbf{z}, \mathbf{y})N^{-1}(E_{\mathbf{x}}[z_i(1)^3] - z_i^3) + o\left(N^{-1}\right). \qquad (185)
\end{aligned}$$

But we have

$$E_{\mathbf{x}}[z_i(1)^3] - z_i^3 = o(1) \qquad (186)$$

since $z_i(1) = \sum_{k=1}^{d} w_k x_{i,k}(1)$ with $x_{i,k}(1), x_{i,l}(1), x_{i,m}(1)$ being conditionally independent for k, l, m all different, and satisfying, for $t = 1, 2, 3$,

$$\begin{aligned}E_{\mathbf{x}}[x_{i,1}(1)^t] &= P_i(\mathbf{x})^t + o(1)\\ &= \left(\sum_{l=1}^{d} p_l x_{i,l} + o(1)\right)^t + o(1)\\ &= \left(\sum_{l=1}^{d} p_l x_{i,l}\right)^t + o(1)\\ &= ((M\mathbf{x}_i)_1)^t + o(1),\end{aligned}\tag{187}$$

since $N_1 x_{i,1}(1)$ has a conditional binomial distribution of parameters N_1 and $P_i(\mathbf{x})$ given in (9), and

$$E_{\mathbf{x}}[x_{i,k}(1)^t] = x_{i,k-1}^t + o(1) = ((M\mathbf{x}_i)_k)^t + o(1)\tag{188}$$

for $k = 2, \ldots, d$, as shown in "Appendix B". As a matter of fact, for some coefficients c_2 and c_3, we have

$$\begin{aligned}E_{\mathbf{x}}[z_i(1)^3] &= E_{\mathbf{x}}\left[\left(\sum_{k=1}^{d} w_k x_{i,k}(1)\right)^3\right]\\ &= \sum_{k=1}^{d} w_k^3 E_{\mathbf{x}}\left[\left(x_{i,k}(1)\right)^3\right] + c_2 \sum_{k=1}^{d} \sum_{l=1,l\neq k}^{d} w_k^2 w_l E_{\mathbf{x}}\left[\left(x_{i,k}(1)\right)^2\right] E_{\mathbf{x}}\left[x_{i,l}(1)\right]\\ &\quad + c_3 \sum_{k=1}^{d} \sum_{l>k}^{d} \sum_{m>l}^{d} w_k w_l w_m E_{\mathbf{x}}\left[x_{i,k}(1)\right] E_{\mathbf{x}}\left[x_{i,l}(1)\right] E_{\mathbf{x}}\left[x_{i,m}(1)\right]\\ &= \sum_{k=1}^{d} w_k^3 \left((M\mathbf{x}_i)_k\right)^3 + c_2 \sum_{k=1}^{d} \sum_{l=1,l\neq k}^{d} w_k^2 w_l \left((M\mathbf{x}_i)_k\right)^2 (M\mathbf{x}_i)_l\\ &\quad + c_3 \sum_{k=1}^{d} \sum_{l>k}^{d} \sum_{m>l}^{d} w_k w_l w_m (M\mathbf{x}_i)_k (M\mathbf{x}_i)_l (M\mathbf{x}_i)_m + o(1)\\ &= \left(\sum_{k=1}^{d} w_k (M\mathbf{x}_i)_k\right)^3 + o(1)\\ &= \left(\sum_{k=1}^{d} w_k x_{i,k}\right)^3 + o(1)\\ &= z_i^3 + o(1).\end{aligned}\tag{189}$$

Therefore,

$$E_{\mathbf{x}}[(\Delta z_i)^4] = E_{\mathbf{x}}[(z_i(1) - E_{\mathbf{x}}[z_i(1)])^4] + o(N^{-1}). \tag{190}$$

Furthermore,

$$\begin{aligned}
E_{\mathbf{x}}\left[(z_i(1) - E_{\mathbf{x}}[z_i(1)])^4\right] &= E_{\mathbf{x}}\left[\left(\sum_{k=1}^{d} w_k(x_{i,k}(1) - E_{\mathbf{x}}[x_{i,k}(1)])\right)^4\right] \\
&\le E_{\mathbf{x}}\left[\sum_{k=1}^{d} w_k \left(x_{i,k}(1) - E_{\mathbf{x}}[x_{i,k}(1)]\right)^4\right] \\
&= \sum_{k=1}^{d} w_k E_{\mathbf{x}}\left[\left(x_{i,k}(1) - E_{\mathbf{x}}[x_{i,k}(1)]\right)^4\right].
\end{aligned} \tag{191}$$

The inequality is obtained by using the Jensen inequality for the power of order 4 with the values $a_k = x_{i,k}(1) - E_{\mathbf{x}}[x_{i,k}(1)]$ and the positive probabilities w_k for $k = 1, \ldots, d$.

For the first age class ($k = 1$) whose size is $N_1 = f_1 N$, we get

$$\begin{aligned}
&E_{\mathbf{x}}\left[\left(x_{i,1}(1) - E_{\mathbf{x}}[x_{i,1}(1)]\right)^4\right] \\
&= \frac{P_i(\mathbf{x})(1 - P_i(\mathbf{x}))}{N_1^2}\left(3P_i(\mathbf{x})(1 - P_i(\mathbf{x})) + \frac{1 - 6P_i(\mathbf{x}) + 6P_i(\mathbf{x})^2}{N_1}\right) \\
&= o\left(N^{-1}\right).
\end{aligned} \tag{192}$$

On the other hand, for the following age classes ($k = 2, \ldots, d$), we have

$$E_{\mathbf{x}}\left[\left(x_{i,k}(1) - E_{\mathbf{x}}[x_{i,k}(1)]\right)^4\right] = o\left(N^{-1}\right) \tag{193}$$

as shown in "Appendix B" (Eq. (169) with $X_i = x_{i,k}(1)$ and $R = N_{k-1} = f_{k-1}N$).

Condition IV

We have

$$E_{\mathbf{x}}\left[\left(y_{i,k}(1) - y_{i,k}\right)\right] = E_{\mathbf{x}}\left[\left(x_{i,k}(1) - x_{i,k}\right)\right] + o\left(1\right) \tag{194}$$

by condition I. Moreover, for the first age class ($k = 1$), we have

$$E_{\mathbf{x}}\left[x_{i,1}(1)\right] = \sum_{k=1}^{d} p_k x_{i,k} + o\left(1\right), \tag{195}$$

from which

$$c_{i,1}(\mathbf{z},\mathbf{y}) = \sum_{k=1}^{d} p_k x_{i,k} - x_{i,1}, \tag{196}$$

while for the following age classes ($k = 2, \ldots, d$), we have

$$E_{\mathbf{x}}\left[x_{i,k}(1)\right] = x_{i,k-1} + o\left(N^{-1}\right), \tag{197}$$

from which

$$c_{i,k}(\mathbf{z},\mathbf{y}) = x_{i,k-1} - x_{i,k}. \tag{198}$$

Therefore, we get condition IV with

$$c_{i,k}(\mathbf{z},\mathbf{y}) = (M\mathbf{x}_i)_k - x_{i,k} = (M\mathbf{y}_i)_k - y_{i,k}, \tag{199}$$

where M is the matrix defined in (14).

Condition V

According to (47), we have

$$Var_{\mathbf{x}}\left[\Delta y_{i,k}\right] = Var_{\mathbf{x}}\left[\Delta x_{i,k}\right] + Var_{\mathbf{x}}\left[\Delta z_i\right] - Cov_{\mathbf{x}}\left[\Delta x_{i,k}, \Delta z_i\right]. \tag{200}$$

Owing to (181) and (180), we have that

$$Var_{\mathbf{x}}\left[\Delta x_{i,k}\right] = o(1), \tag{201}$$

and to (48a) and (48b), that

$$Var_{\mathbf{x}}\left[\Delta z_i\right] = o(1). \tag{202}$$

On the other hand, the random variables $x_{i,k}(1)$ and $x_{i,l}(1)$ for $l \neq k = 1, \ldots, d$ and $i = 1, \ldots, n$ are conditionally independent so that

$$Cov_{\mathbf{x}}\left[\Delta x_{i,k}, \Delta z_i\right] = w_k Cov_{\mathbf{x}}\left[x_{i,k}(1), x_{i,k}(1)\right] = o(1), \tag{203}$$

owing to (181) and (180).

Condition VI

Owing to (199), the recurrence system of equations (50) becomes

$$Y_{i,k}(t+1,\mathbf{z},\mathbf{y}) = (M\mathbf{Y}_i(t,\mathbf{z},\mathbf{y}))_k \tag{204}$$

with

$$\mathbf{Y}_i(t, \mathbf{z}, \mathbf{y}) = (Y_{i,1}(t, \mathbf{z}, \mathbf{y}), \ldots, Y_{i,d}(t, \mathbf{z}, \mathbf{y}))^T. \tag{205}$$

Then

$$\mathbf{Y}_i(t+1, \mathbf{z}, \mathbf{y}) = M^{t+1}\mathbf{Y}_i(0, \mathbf{z}, \mathbf{y}) = M^{t+1}\mathbf{y}_i \tag{206}$$

for all integers $t \geq 0$. According to Eq. (28), we have

$$\lim_{t\to+\infty} M^t = \mathbf{1}\mathbf{w}^{\mathbf{T}} \tag{207}$$

entrywise. Moreover,

$$\mathbf{w}^T\mathbf{y}_i = \mathbf{w}^T(\mathbf{x}_i - z_i\mathbf{1}) = \mathbf{w}^T\left(\mathbf{x}_i - (\mathbf{w}^T\mathbf{x}_i)\mathbf{1}\right) = \mathbf{w}^T\mathbf{x}_i - (\mathbf{w}^T\mathbf{x}_i)\mathbf{w}^T\mathbf{1} = 0, \tag{208}$$

since $\mathbf{w}^T\mathbf{1} = 1$. Therefore,

$$\lim_{t\to+\infty} \mathbf{Y}_i(t, \mathbf{z}, \mathbf{y}) = \mathbf{0} \tag{209}$$

uniformly in $\mathbf{z} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n \times \mathbb{R}^d$ for $i = 1, \ldots, n$ by linearity.

Appendix D: analysis of $U(f)$ with selection on fertility only

With the change of variable $f = (3-y)/2$, we get

$$V(y) = U(f(y)) = \frac{\frac{4}{(3-y)y}}{1 - \exp\left\{-\frac{(1+y)}{y}\right\}} \tag{210}$$

for $1 < y < 2$. The derivative of $V(y)$ is given by

$$V'(y) = \frac{4e^{\frac{1}{y}+1}\left(-2(y-1)y + e^{\frac{1}{y}+1}(2y-3)y + 3\right)}{\left(e^{\frac{1}{y}+1} - 1\right)^2 (y-3)^2 y^3}. \tag{211}$$

In particular, we have

$$V'(1) = \frac{e^2\left(3 - e^2\right)}{\left(e^2 - 1\right)^2} < 0, \quad V'(2) = \frac{e^{3/2}\left(2e^{3/2} - 1\right)}{2\left(e^{3/2} - 1\right)^2} > 0. \tag{212}$$

Owing to the intermediate value theorem, there exists a real number $y^* \in (1, 2)$ such that $V'(y^*) = 0$. Furthermore, the second derivative of $V(y)$ given by

$$V''(y) = \frac{4e^{\frac{1}{y}+1} A(y)}{\left(e^{\frac{1}{y}+1} - 1\right)^3 (y-3)^3 y^5} \tag{213}$$

is a positive function with

$$A(y) = -6e^{\frac{2}{y}+2}((y-3)y+3)y^2 + e^{\frac{1}{y}+1}(y(y(6y(2y-5)+5)+42)-9) \\ -9-30y+11y^2+12y^3-6y^4 \tag{214}$$

being negative for $1 < y < 2$ (see below). Then the critical point y^* is unique, and it is actually the global minimum point of $V(y)$. Therefore, $f^* = (3 - y^*)/2$ is the only critical point of $U(f)$, and it is actually the global minimum point of $U(f)$.

It remains to prove that $A(y) < 0$ for $1 < y < 2$. Note that

$$\begin{aligned} &-6e^{\frac{2}{y}+2}((y-3)y+3)y^2 + e^{\frac{1}{y}+1}(y(y(6y(2y-5)+5)+42)-9) \\ &\quad = e^{\frac{1}{y}+1}\left(-6y^2 e^{\frac{1}{y}+1}((y-3)y+3) - 9 + 42y + 5y^2 - 30y^3 + 12y^4\right) \\ &\quad = e^{\frac{1}{y}+1}\left(-6y^2 e^{\frac{1}{y}+1}((y-3)y+3) + y^2\left(5 - 30y + 12y^2\right) - 9 + 42y\right) \\ &\quad = e^{\frac{1}{y}+1}\left(-6y^2((y-3)y+3)\left(e^{\frac{1}{y}+1} - 4\right) - 12y^4 + 42y^3 - 67y^2 + 42y - 9\right) \end{aligned} \tag{215}$$

is negative, since $e^{\frac{1}{y}+1} - 4 > e^2 - 4 > 0$, while $-12y^4 + 42y^3 - 67y^2 + 42y - 9 < 0$ and $(y-3)y + 3 > 0$, both polynomials having no real roots. Note also that the polynomial $-9 - 30y + 11y^2 + 12y^3 - 6y^4$ in (214) has only two real roots, namely $y \approx -1.38072$ and $y \approx -0.281155$. Moreover, since it goes to $-\infty$ as $y \to +\infty$, it is necessarily negative for $1 < y < 2$.

Appendix E: derivative of $U(f)$ with selection on survival

With selection on survival only, we have

$$U'(f) \\ = \frac{2A\left((2f-3)\left(4f^2(A-2) + 3A + f(14-4E) - 7\right)L + 2fPL^2 + (A-1)(3-2f)^2\right)}{f^2(2f-3)^3(A-1)^2}, \tag{216}$$

while with selection on both survival and fertility, we have

$U'(f)$

$$= \frac{2C\left(L\left(QB - 8f^3 + 32f^2 - 36f + 21\right) + 2(f-1)\left((6f-9)B - 8f + 15\right) + 2fPL^2\right)}{f^2(2f-3)^3\,(B-1)^2} \tag{217}$$

with

$$P = 8f^3 - 28f^2 + 34f - 11, \tag{218a}$$

$$Q = 8f^3 - 20f^2 + 18f - 9, \tag{218b}$$

$$L = \ln\left(\frac{f}{2f-1}\right), \tag{218c}$$

$$\ln A = \frac{2(f-2)(2f-1)}{2f-3}\ln\left(\frac{f}{2f-1}\right) \tag{218d}$$

and

$$B = \exp\left\{\frac{2(f-2)\left((2f-1)\ln\left(\frac{f}{2f-1}\right)+1\right)}{2f-3}\right\}. \tag{218e}$$

References

Auger P, Bravo de la Parra R, Poggiale JC, Sànchez E, Nguyen Huu T (2008) Aggregation of variables and applications to population dynamics. In: Magal P, Ruan S (eds) Structured population models in biology and epidemiology, vol 1936. Lecture notes in mathematics. Springer, Berlin, pp 209–263

Chalub FA, Souza MO (2009) From discrete to continuous evolution models: a unifying approach to drift-diffusion and replicator dynamics. Theor Popul Biol 76:268–277

Chalub FA, Souza MO (2014) The frequency-dependent Wright–Fisher model: diffusive and non-diffusive approximations. J Math Biol 68:1089–1133

Champagnat N, Ferrière R, Méléard S (2006) Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models. Theor Popul Biol 69:297–321

Champagnat N, Ferrière R, Méléard S (2008) From individual stochastic processes to macroscopic models in adaptive evolution. Stoch Models 24:2–44

Charlesworth B (1980) Evolution in age-structured populations, vol 1. Cambridge studies in mathematical biology. Cambridge University Press, Cambridge

Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. Nat Rev Genet 10:195–205

Chesson J (1976) A non-central multivariate hypergeometric distribution arising from biased sampling with application to selective predation. J Appl Probab 13:795–797

Crow JF, Kimura M (1970) An introduction to population genetics theory. Harper and Row, New York

Cushing JM (1998) An introduction to structured population dynamics, CBMS-NSF regional conference series in applied mathematics, vol 71. Society for Industrial and Applied Mathematics, Philadelphia

Emigh TH (1979a) The dynamics of finite haploid populations with overlapping generations. I. Moments, fixation probabilities and stationary distributions. Genetics 92:323–337

Emigh TH (1979b) The dynamics of finite haploid populations with overlapping generations. II. The diffusion approximation. Genetics 92:339–351

Emigh TH, Pollak E (1979) Fixation probabilities and effective population numbers in diploid populations with overlapping generations. Theor Popul Biol 15:86–107

Etheridge A (2011) Some mathematical models from population genetics. Springer, Heidelberg

Ethier SN (1976) A class of degenerate diffusion processes occurring in population genetics. Commun Pure Appl Math 29:483–493

Ethier SN, Nagylaki T (1980) Diffusion approximations of Markov chains with two time scales and applications to population genetics. Adv Appl Probab 12:14–49

Ewens WJ (2004) Mathematical population genetics 1: theoretical introduction, 2nd edn. Springer, New York

Felsenstein J (1971) Inbreeding and variance effective numbers in populations with overlapping generations. Genetics 68:581–597

Fisher RA (1930) The genetical theory of natural selection. Clarendon, Oxford

Godement R (2005) Analysis II: differential and integral calculus, Fourier series, holomorphic functions. Springer, Berlin

Huang W, Hauert C, Traulsen A (2015) Stochastic game dynamics under demographic fluctuations. Proc Natl Acad Sci USA 112:9064–9069

Jones OR, Scheuerlein A, Salguero-Gómez R, Camarda CG, Schaible R, Casper BB, Dahlgren JP, Ehrlén J, García MB, Menges ES, Quintana-Ascencio PF, Caswell H, Baudisch A, Vaupel JW (2014) Diversity of ageing across the tree of life. Nature 505:169–173

Karlin S (1982) Classification of selection-migration structures and conditions for a protected polymorphism. In: Hecht MK, Wallace B, Prance GT (eds) Evolutionary biology. Plenum, New York, pp 61–204

Karlin S, Taylor HM (1975) A first course in stochastic processes. Academic Press, New York

Kebir A, Miled SB, Hbid ML, Bravo de La Parra R (2010) Effects of density dependent sex allocation on the dynamics of a simultaneous hermaphroditic population: modelling and analysis. J Theor Biol 263:521–529

Kebir A, Fefferman NH, Miled SB (2015) Understanding hermaphrodite species through game theory. J Math Biol 71:1505–1524

Kimura M (1964) Diffusion models in population genetics. J Appl Probab 1:177–232

Kroumi D, Lessard S (2015) Strong migration limit for games in structured populations: applications to dominance hierarchy and set structure. Games 6:318–346

Leslie PH (1945) On the use of matrices in certain population mathematics. Biometrika 33:183–212

Lessard S (2009) Diffusion approximations for one-locus multi-allele kin selection, mutation and random drift in group-structured populations: a unifying approach to selection models in population genetics. J Math Biol 59:659–696

Lessard S, Soares CD (2018) Frequency-dependent growth in class-structured populations: continuous dynamics in the limit of weak selection. J Math Biol 77:229–259

Li XY, Kurokawa S, Giaimo S, Traulsen A (2016) How life history can sway the fixation probability of mutants. Genetics 203:1297–1313

Marvà M, Moussaoui A, Bravo de la Parra R, Auger P (2013) A density-dependent model describing age-structured population dynamics using hawk-dove tactics. J Differ Equ Appl 19:1022–1034

Marvà M, San Segundo F (2018) Age-structure density-dependent fertility and individuals dispersal in a population model. Math Biosci 300:157–167

Nagylaki T (1980) The strong-migration limit in geographically structured populations. J Math Biol 9:101–114

Nowak MA, Sasaki A, Taylor C, Fudenberg D (2004) Emergence of cooperation and evolutionary stability in finite populations. Nature 428:646–650

Nowak MA, Tarnita CE, Antal T (2010) Evolutionary dynamics in structured populations. Philos Trans R Soc B 365:19–30

Soares CD (2019) Évolution dans des populations structurées en classes. Ph.D. Université de Montréal, Montréal

Soares CD, Lessard S (2019) First-order effect of frequency-dependent selection on fixation probability in an age-structured population with application to a public goods game. Theor Popul Biol. https://doi.org/10.1016/j.tpb.2019.05.001

Stegan IA (1964) Handbook of mathematical functions: with formulas, graphs and mathematical tables. Dover, New York

Wallenius KT (1963) Biased sampling; the non-central hypergeometric probability distribution. Ph.D. Stanford University, Stanford