MORAN PROCESS AND WRIGHT-FISHER PROCESS FAVOR LOW VARIABILITY

JAN RYCHTÁŘ AND DEWEY T. TAYLOR

Abstract. We study evolutionary dynamics in finite populations. We assume the individuals are one of two competing genotypes, A or B. The genotypes have the same average fitness but different variances and/or third central moments. We focus on two frequency-independent stochastic processes: (1) Wright-Fisher process and (2) Moran process. Both processes have two absorbing states corresponding to homogeneous populations of all A or all B. Despite the fact that types A and B have the same average fitness, both stochastic dynamics differ from a random drift. In both processes, the selection favors A replacing B and opposes B replacing A if the fitness variance for A is smaller than the fitness variance for B. In the case the variances are equal, the selection favors A replacing B and opposes B replacing A if the third central moment of A is larger than the third central moment of B. We show that these results extend to structured populations and other dynamics where the selection acts at birth. We also demonstrate that the selection favors a larger variance in fitness if the selection acts at death.

1. Introduction

Variability is omnipresent in nature. Individuals vary in their abilities (Mesterton-Gibbons and Sherratt, 2011) and environmental conditions fluctuate in time and space (Frank and Slatkin, 1990; Childs et al., 2010; Schreiber, 2012; Evans et al., 2015). How individuals should optimize their behavior in a variable and unpredictable environment is addressed by bet-hedging theory (Cohen, 1966). Bet-hedging is defined as a strategy that reduces the temporal variance in fitness at the expense of a lowered arithmetic mean fitness (Ripa et al., 2009). Olsson et al. (2009) describe three main bet-hedging strategies: (1) conservative bet-hedging, i.e. play-it-safe (Seger and Brockmann, 1987), Philippi and Seger...
(1989), (2) diversified bet-hedging, i.e. don’t put all your eggs in one basket (Cohen 1966) and (3) adaptive coin flipping, i.e. choose a strategy at random from a fixed distribution (Cooper and Kaplan 1982).

As a motivational example, already discussed in Gillespie (1974), consider a bird that can lay all of its $M$ eggs in one nest or distribute the eggs evenly across $k$ nests. Assume that a nest (with all eggs) is destroyed by a predator with probability $p$. Which strategy should a bird choose? Traditionally, the answer would be the one that yields the higher average fitness. However, in both cases, the mean number of surviving eggs is $M(1 - p)$. If the eggs are distributed evenly, the variance of surviving eggs is $M^2p(1 - p)/k$, which is smaller than $M^2p(1 - p)$, the variance if all eggs are in one nest. Gillespie (1974) then used a diffusion approximation of the Wright-Fisher process to conclude that distributing eggs evenly is a better strategy.

The optimal behavior evolves through natural selection. Natural selection is studied through evolutionary dynamics. The traditional approach is to consider infinite populations and use the replicator dynamics (Taylor and Jonker 1978). However, under the replicator dynamics, only the individuals’ average fitness matter. Therefore, the replicator dynamics cannot distinguish the above two types of bird behavior. While there are many dynamics to consider in infinite populations (Hofbauer and Sigmund 1998), including continuous versions of the Wright-Fisher process (Chalub and Souza 2014), it is plausible to assume that if there is a difference between the above two types, the difference will manifest itself more clearly in the finite populations and stochastic dynamics.

There are many evolutionary dynamics in finite populations (Schaffer 1988; Kandori et al. 1993; Fogel et al. 1998; Traulsen and Hauert 2009). Two of the best-studied dynamics are the Moran process (Moran 1962) and the Wright-Fisher process (Fisher 1923; Wright 1931). The Moran process applies to populations with asynchronous reproduction and overlapping generations. It has been extensively studied recently, see for example Bürger (2000); Nowak et al. (2004); Fudenberg et al. (2006); Hauert and Imhof (2012); Chalub and Souza (2017). In each step, one individual is chosen for reproduction with probability proportional to its fitness. The offspring is of the same type as their parent and replaces a randomly chosen individual. In contrast, the Wright-Fisher process describes populations with synchronous reproduction and non-overlapping generations (Ewens 2004; Imhof and Nowak 2006; Durrett 2008). Here all individuals reproduce at the same time. They generate a pool of offsprings proportional to their fitness. The next generation is then chosen from the pool at random. For both processes, the total population size is kept constant.
For the rest of the paper, we consider a population of $N$ individuals consisting of two competing haploid genotypes, $A$ and $B$. We assume that the genotypes have the same average fitness $\mu$ but differ in the variances, $\nu_A$ and $\nu_B$, and/or third central moments $s_A$ and $s_B$. There are two absorbing states of the evolutionary dynamics, corresponding to all $A$ or all $B$. Let $\rho_{AB}$ denote the probability that a single $A$ individual among $N-1$ individuals of genotype $B$ will generate a lineage that will take over the whole population. For neutral drift the corresponding fixation probability would be $1/N$ \citep{Moran1962, Karlin2014}. We say that the selection favors $A$ replacing $B$ if $\rho_{AB} > 1/N$ and opposes $A$ replacing $B$ if $\rho_{AB} < 1/N$ \citep{Taylor2004, Imhof2006}. We show that, for both the Wright-Fisher and the Moran processes, the selection favors $A$ replacing $B$ (and opposes $B$ replacing $A$) if

$$\frac{\nu_B - \nu_A}{N\mu^2} - \frac{s_B - s_A}{N^2\mu^3} > 0.$$ \hfill (1)

This extends the result of \cite{Gillespie1974} who considered only the term $\frac{\nu_B - \nu_A}{N}$ and only the Wright-Fisher process.

2. Model

We consider a population of $N$ individuals consisting of two competing haploid genotypes, $A$ and $B$. Let $N_A$ denote the number and $\pi_A = N_A/N$ be the proportion of $A$ individuals in the population at a given time.

Fitness of genotype $A$ individuals will be independent, identically distributed random variables $F_A$. We assume that the fitness is at least 1 almost surely (to avoid extinction), has a mean $\mu_A = E[F_A]$, a variance $\nu_A = E[(F_A - E[F_A])^2]$ and a third central moment $s_A = E[(F_A - E[F_A])^3]$. Notation for genotype $B$ is analogous.

We will assume that $\nu_A = \nu_B$. We also assume that the fitness does not depend on $\pi_A$ (the process is frequency-independent) and that the probability distributions of $F_A$ and $F_B$ stay constant over time.

Let $S_k(X)$ denote the sum of $k$ independent identically distributed i.i.d. random variables $X$.

We consider two kinds of evolutionary dynamics: (1) the Wright-Fisher process for synchronous reproduction with non-overlapping generations and (2) the Moran process for asynchronous reproduction with overlapping generations.

For the Wright-Fisher process, we assume the following sequence of events. (1) every individual in the original generation contributes their fitness share to the pool of potential offspring, (2) all of the individuals die, and finally (3) a new set of $N$ individuals is determined by a sequence of $N$ random drawings (without replacement) from the pool of potential offspring. Every $A$ individual
in the original population contributes $F_A$ to the pool; the total contribution from all $A$ individuals together will thus be $S_{N_A}(F_A)$. Similarly, $B$ individuals contribute collectively $S_{N_B}(F_B)$. Consequently, the number of $A$ individuals in the new generation follows a hypergeometric distribution with $N$ draws out of $S_{N_A}(F_A) + S_{N_B}(F_B)$ candidates from which $S_{N_A}(F_A)$ are genotype $A$ and $S_{N_B}(F_B)$ are genotype $B$. The mean number (given $\pi_A$ and specific fitnesses of all individuals in the original population) of $A$ individuals in the new generation is thus $N\pi_A^\prime$, where

$$
\pi_A^\prime = \frac{S_{N_A}(F_A)}{S_{N_A}(F_A) + S_{N_B}(F_B)}.
$$

It follows that $\pi_A^\prime$ is the mean proportion of genotype $A$ in the new generation. The actual proportion will vary and can be between 0 and 1. The expected proportion of genotype $A$ in the next generation given the proportion $\pi_A$ in the original population, $E[\pi_A^\prime | \pi_A]$, is the expected value of $\pi_A^\prime$ taken over all possible realizations of the fitnesses of individuals in the population with $N_A = \pi_A N$ of $A$ individuals.

For the Moran process, we again assume that every individual in the original generation contributes their fitness share to the pool of potential offspring. However, only one individual dies and is replaced by a new individual. A particular individual dies with probability $1/N$. The new individual is determined by a single random drawing from the pool of potential offspring. Given $\pi_A$, the expected probability that the new individual will be genotype $A$ is given by $E[\pi_A^\prime | \pi_A]$, where, as before $\pi_A^\prime$ is given by (2) and $E[\pi_A^\prime | \pi_A]$ is the expected value of $\pi_A^\prime$ taken over all possible realizations of the fitnesses of individuals in the population with $N_A = \pi_A N$ of $A$ individuals. We note that for the Moran process we only need $F_A$ and $F_B$ to be positive almost surely (instead of at least 1 almost surely).

Now, let $\pi_{A,B}$, be an initial proportion of genotype $A$ in a population. We define $\pi_{A,n+1}$, the expected proportion of genotype $A$ in the $(n + 1)^{th}$ step (a replacement of all individuals for the Wright-Fisher process or a replacement of a single individual for the Moran process). Either one of the processes eventually reaches one of the two absorbing states (either all $A$ or all $B$) and we are interested in the fixation probability $\rho_{AB}$ (given the initial proportion $\pi_{A,0}$). When a process reaches the absorbing state, there are either $N$ individuals of genotype $A$ (with probability $\rho_{AB}$) or 0 individuals of genotype $A$, i.e. the expected proportion of genotype $A$ is $\rho_{AB}$. It follows that $\rho_{AB} = \lim_{n \to \infty} \pi_{A,n}$.

Before proceeding further, we would like to make two remarks.

First, we can consider an alternative procedure for the Wright-Fisher process and make the drawing of the individuals for the new generation be a drawing with replacement. In that case, the number of $A$ individuals would follow a binomial distribution $B(N, \pi_A^\prime)$. The mean number of $A$ individuals would be
Figure 1. Fixation probabilities for the Wright-Fisher process (left) and the Moran process (right). Darker areas correspond to higher fixation probabilities; 0.1 corresponds to a random drift. For \( N = 10 \), we considered genotypes \( G_i; i = 1, \ldots, 11 \) that have a fitness \( i + 1 \) with probability \( 1/i \) and 1 with probability \( 1 - 1/i \). The average fitness of \( G_i \) is 2, the variance is \( i - 1 \). For every pair of \( i, j \), we run \( 10^5 \) simulations starting with a single \( G_i \) individual among \( G_j \) individuals. Note that the individuals’ fitness is sometimes more than double the expected value.

the same. The variance of that number would be slightly larger; specifically

\[ N\pi_A'(1 - \pi_A') \text{ as opposed to } N\pi_A'(1 - \pi_A') \left( \frac{S_{N_A}(F_A) + s_{N_A}(F_B) - N}{S_{N_A}(F_A) + s_{N_A}(F_B) - 1} \right); \]  
but otherwise the calculations would not be affected. In this case, the fitness does not have to be at least 1 anymore and we only need to assume that it is positive almost surely.

Second, we assume that the total contribution of \( A \) individuals into the offspring pool is \( S_{N_A}(F_A) \) instead of \( N_A F_A \). This seems to better mimic the individual variation in fitness; it also scales more appropriately since \( \text{var}(S_{N_A}(F_A)) = N_A \text{var}(F_A) \) as opposed to \( \text{var}(N_A F_A) = N_A^2 F_A \).

3. Results

For either process, the selection favors \( A \) replacing \( B \) and opposes \( B \) replacing \( A \) if

\[
\frac{v_B - v_A}{\mu^2} - \frac{1}{N} \cdot \frac{s_B - s_A}{\mu^3} > 0. \tag{3}
\]

If \( F_A < 2\mu \) and \( F_B < 2\mu \) almost surely, then the result holds for all population sizes \( N \). Otherwise, the result holds only for the populations that are large enough. In the later case, when \( v_A \neq v_B \), we can neglect the contribution of the term \( \frac{1}{N} \cdot \frac{s_B - s_A}{\mu^3} \), and the result says that the selection favors \( A \) replacing \( B \) and opposes \( B \) replacing \( A \) if \( v_A < v_B \).

The fixation probabilities obtained by simulations are shown in Figure 1 when \( v_A \neq v_B \) and Figure 2 when \( v_A = v_B \).

The result is proved as follows. For the Wright-Fisher process, \( \pi_{A,n+1} = E[\pi_A'|\pi_{A,n}] \). For the Moran process, \( E[\pi_A'|\pi_{A,n}] \) is the expected probability that an \( A \) individual is chosen for reproduction. In a random drift, the expected
probability is \( \pi_A \). Consequently, if \( \mathbb{E}[\pi_A^\prime|\pi_{A,n}] > \pi_A \), the \( A \) individual is chosen for reproduction more likely than in the random drift and so \( \pi_{A,n+1} > \pi_{A,n} \) (because in random drift, the expected proportion stays constant).

To compare either process to a random drift, we only need to study the sign of \( \mathbb{E}[\pi_A^\prime|\pi_A] - \pi_A \). If the sign is positive for all \( \pi_A \), the series \( \pi_{A,n} \) is increasing and so \( \pi_{A,0} < \pi_{A,1} < \pi_{A,2} < \ldots \), \( \lim_{n \to \infty} \pi_{A,n} = \rho_{AB} \). By setting \( \pi_{A,0} = 1/N \), we get that the selection favors \( A \) replacing \( B \). By setting \( \pi_{A,0} = 1 - 1/N \), we get that the selection will oppose \( B \) replacing \( A \). By a similar argument, if the sign of \( \mathbb{E}[\pi_A^\prime|\pi_A] - \pi_A \) is negative for all \( \pi_A \), the selection will oppose \( A \) replacing \( B \) and favor \( B \) replacing \( A \).

As shown in the Appendix [A] using the bivariate Taylor expansion of \( \mathbb{E}[\pi_A^\prime|\pi_A] \), we have \( \mathbb{E}[\pi_A^\prime|\pi_A] - \pi_A = E_3 + O\left( N^{-1}\right) \), where

\[
E_3 = \pi_A (1 - \pi_A) \left( \frac{1}{N} \cdot \frac{\nu_B - \nu_A}{\mu^2} - \frac{1}{N^2} \cdot \frac{s_B - s_A}{\mu^2} \right).
\]  

(4)

The approximation by \( E_3 \) is guaranteed to be better than a simple \( \mathbb{E}[\pi_A^\prime|\pi_A] - \pi_A = 0 + O(N^{-1}) \) only if \( 0 < F_A < 2\mu_A \) and \( 0 < F_B < 2\mu_B \) almost surely. For \( N \) large enough, by the central limit theorem, the approximation by \( E_3 \) works even if the conditions are not met, see the Appendix for more details. However, when the conditions are not met, the second order approximation \( \mathbb{E}[\pi_A^\prime|\pi_A] - \pi_A \approx E_2 \)
where
\[ E_2 = \pi_A (1 - \pi_A) \left( \frac{1}{N} \cdot \frac{\mu_B - \mu_A}{\sigma^2} \right), \]  
(5)

may be better for intermediate \( N \), see Appendix.

We note that the calculations of \( E_3 \) and \( E_2 \) do not require \( \mu_A = \mu_B \), but if \( \mu_A \neq \mu_B \), the formulas are not very elegant.

4. Discussion

We studied the Wright-Fisher process and the Moran process in order to investigate the dynamics of strategies with variable payoffs in finite populations. We compared the probability \( \rho_{AB} \) that a single genotype \( A \) individual takes over a population of genotype \( B \) individuals with the corresponding probability under a neutral drift, which is \( 1/N \). We saw that the selection favors \( A \) replacing \( B \) (and prevents \( B \) replacing \( A \)), if
\[ \frac{\mu_B - \mu_A}{N \mu^2} - \frac{s_B - s_A}{N^2 \mu^3} > 0. \]  
(6)

In particular, for \( N \) large enough, the selection favors the genotype with a lower variance in fitness. When the two genotypes have the same variance, it is better to have a larger third central moment.

Olofsson et al. (2009) created an individual-based model accounting for both ecological and evolutionary forces. In their model, a female’s reproductive strategy is characterized by three parameters that govern how a female should distribute the mass of the different propagules in a clutch within and between years. They demonstrated that the optimal strategy is a combination of conservative and diversified bet-hedging and adaptive coin flipping, which means a variation in egg size both within clutches and between years. We note that their result (that the optimal strategy can be variable) is not in a contradiction with our result. We demonstrated that the fitness of the optimal strategy should have as low variation as possible. If the environment is variable, a “constant” strategy will have a variable fitness and it is conceivable that a “variable” strategy may lower the variation in fitness. We note that the between-generation variance was further studied in Starrfelt and Kokko (2012). Furthermore, Schreiber (2015) provided a unifying theory for within- and between-generation variations of the fitness that incorporated the original result of Gillespie (1974).

Formula (6) extends the result of Gillespie (1974) in two ways: (1) it accounts more precisely for higher order moments and (2) it extends the result to Moran process as well.
There is a pattern in \((6)\) suggesting a possible extension to even higher moments. In fact, parts of our proof could easily adapt to higher order approximations. An approximation of \(E[\pi_A'|\pi_A]\) by a polynomial of any degree is not difficult as seen in the Appendix, see also Rice and Papadopoulos (2009) and Rice (2015). However, our calculations rely on the additive property of second and third central moments and this additive property does not extend to higher order central moments. Still, because the higher order approximations contain \(N^{-4}\), a third degree approximation in our paper is likely enough. Also, let us mention that while Gillespie (1974) uses only second order diffusion approximation, that method allowed for a conclusion that genotype \(A\) does better when \(\mu_A - \nu_A / N\) is larger than \(\mu_B - \nu_B / N\), i.e. when the means are not necessarily same.

Our results and method could be extended to many other kinds of dynamics. The core of our argument was the approximation of \(E[\pi_A'|\pi_A] - \pi_A\) by \(E_3\) from \((4)\). This approximation most readily applies to the Wright-Fisher process where \(E[\pi_A'|\pi_A]\) is the expected proportion of genotype \(A\) in the next generation. However, we were able to relate the sign of \(E[\pi_A'|\pi_A] - \pi_A\) to the sign of \(\pi_{A,n+1} - \pi_{A,n}\) for the Moran process as well. This kind of reasoning could be generalized and applied to a wider range of processes as seen below.

The Moran process can be broken down into a birth event, a death event and a selection. There are four main updating mechanisms considered for the Moran processes, depending on the order of birth and death and whether the selection acts at birth or at death (Masuda 2009; Pattni et al. 2015; Pattni, 2017). In our paper we considered birth-death with selection at birth (often denoted Bd where the order of letters specifies that birth comes first and the capital B specifies that the selection acts at birth), also called the invasion process (Lieberman et al. 2005). As seen in this kind of a process, if an expression like \(E[\pi_A'|\pi_A]\) is the expected probability of genotype \(A\) being selected for the reproduction and if

\[
E[\pi_A'|\pi_A] > \pi_A, \text{ for all } \pi_A \in (0, 1), \tag{7}
\]

then the selection favors \(A\) replacing \(B\). Consequently, our argument extends without a change to the Bd-Moran process on a structured population; more specifically a process where one individual is chosen for reproduction proportionally to their fitness and the offspring then replaces one of parent’s neighbors (possibly even the parent) chosen at random.

Our argument also stands for the dB-Moran process, i.e. a death-birth updating with selection at birth where first an individual dies at random and is replaced by an offspring of the remaining individuals (from the whole population or just the neighborhood) proportionally to their fitness. We note that if we consider a structured population and a replacement by a neighbors’ offspring,
The meanings of $N$, $\pi_A$ and $\pi'_A$ need to be restricted only to the neighborhood (in particular, we may not be able to take $N$ as large as we need to, unless we consider special kinds of graphs). A little more care would have to be taken for the Db-Moran process, where an individual dies proportionally to the inverse of their fitness (and is replaced by an offspring of the remaining individuals which is selected uniformly at random). Since $\text{var}(1/X) = \text{var}(X)/(EX)^4$ (Selman, 2018), the selection favors $A$ if $\upsilon_A > \upsilon_B$ as illustrated in Figure 3. Similar result holds for the bD-Moran process.

Extending the analysis to other kinds of updating rules, such as the imitation dynamics (Traulsen et al., 2006) is also possible, although in this case the analysis may rely much less on the formula (6). Under the imitation dynamics, a pair of individuals is selected at random and an individual with lower fitness is replaced by an individual with a larger fitness. Consequently, $\pi_{A,n+1} > \pi_{A,n}$ if $P(F_A - F_B > 0) > 1/2$. In particular, if both $F_A$ and $F_B$ follow a (truncated) normal distribution with the same mean but different variances, the variances will still have no effect on the fixation probability since $F_A - F_B$ will follow a normal distribution with a mean 0 and variance $\upsilon_A + \upsilon_B$, i.e. $P(F_A - F_B > 0) = 1/2$ as in the random drift. On the other hand, if $\upsilon_A \approx 0$ and $F_B$ follows a distribution with mean $\mu$ but such that $P(\mu > F_B) > 1/2$, then selection favors $A$ replacing $B$. Yet, if $F_B$ follows a distribution that is skewed to the other side, i.e. $P(\mu < F_B) > 1/2$, then selection favors $B$ replacing $A$. Note that in both cases $\upsilon_B > 0$, i.e. the results seem to be more about the relationships of the third central moments (the smaller the better). We note that the imitation dynamics can be equipped with many kinds of updating rules such as the Fermi rule (Traulsen et al., 2006).
and the analysis under those rules may not be as straightforward as the one shown above. It would also be interesting to see how our analysis extends to the dynamics where the population size is not constant; for example to the dynamics investigated in (Czuppon and Traulsen 2018; McAvoy et al. 2018; Park et al., 2019). Extending our analysis to frequency dependent fitness is the next natural step. (Nishimura and Stephens (1997) already demonstrated the negative effect of a large payoff variance on the evolution of cooperation in the Iterated Prisoner’s Dilemma game. It is therefore likely that strategies with low payoff variances are less prone to invasion by strategies with large variances in any finite and/or structured population. However, exact results may be difficult to obtain. While the Moran process is now relatively well understood (Taylor et al. 2004; Lieberman et al. 2005; Allen and Tarnita 2014; Allen and Nowak 2014; Pattini et al. 2015; Pattini 2017; Pattini et al. 2018; Schimit et al. 2019), the introduction of variances makes the calculations intractable and some sort of approximation is necessary (Traulsen et al. 2007; Hofbauer and Sandholm 2007; Wallace and Young, 2015). The Wright-Fisher process is hard to analyze on its own (Imhof and Nowak 2006). Moreover, as Ripa et al. (2009) point out, frequency dependence makes bet-hedging difficult to define and to study properly. Yet, it would be very interesting to see how the payoff variance affects the fixation probabilities in structured populations.

Acknowledgements

The authors would like to thank the handling editor and two anonymous reviewers for their suggestions and comments that helped to improve the manuscript.

References


Here we show that if $\mu_B = \mu_A$, then
\[
\mathbb{E} [\pi'_A | \pi_A] \approx \pi_A - \pi_A (1 - \pi_A) \left( \frac{\nu_A - \nu_B}{N \mu^2} + \frac{s_A - s_B}{N^2 \mu^3} \right) + O(N^{-3})
\]
where
\[
\pi'_A = \frac{S_{N_A}(F_A)}{S_{N_A}(F_A) + S_{N_B}(F_B)}
\]
and $S_k(X)$ is a sum of $k$ iid random variables $X$.

Let us start by an auxiliary lemma.

**Lemma 1.** Let $Z_i$, $i = 1, \ldots, n$, be independent random variables with $\mathbb{E} Z_i = 0$. Then,
\[
\mathbb{E} \left[ \left( \sum_{i=1}^{n} Z_i \right)^2 \right] = \sum_{i=1}^{n} \mathbb{E} [Z_i^2] \tag{10}
\]
\[
\mathbb{E} \left[ \left( \sum_{i=1}^{n} Z_i \right)^3 \right] = \sum_{i=1}^{n} \mathbb{E} [Z_i^3]. \tag{11}
\]

*Proof.* We will prove (11) by induction. The proof of (10) is analogous. The statement clearly holds for $n = 1$. Now assume that the statement holds for $n \geq 1$, we will show it holds for $n + 1$.
\[
\mathbb{E} \left[ \left( \sum_{i=1}^{n+1} Z_i \right)^3 \right] = \mathbb{E} \left[ \left( \sum_{i=1}^{n} Z_i \right)^3 \right] + \mathbb{E} \left[ 3 \left( \sum_{i=1}^{n} Z_i \right)^2 Z_{n+1} \right] + \mathbb{E} \left[ 3 \left( \sum_{i=1}^{n} Z_i \right) Z_{n+1}^2 \right] + \mathbb{E} \left[ Z_{n+1}^3 \right] \tag{12}
\]
\[
= \sum_{i=1}^{n} \mathbb{E} [Z_i^3] + 3 \mathbb{E} \left[ \left( \sum_{i=1}^{n} Z_i \right)^2 \right] \mathbb{E} Z_{n+1} + 3 \mathbb{E} \left[ \sum_{i=1}^{n} Z_i \right] \mathbb{E} [Z_{n+1}^2] + \mathbb{E} \left[ Z_{n+1}^3 \right] \tag{13}
\]
\[
= \sum_{i=1}^{n} \mathbb{E} [Z_i^3] + 3 \mathbb{E} \left[ \left( \sum_{i=1}^{n} Z_i \right)^2 \right] \cdot 0 + 3 \cdot 0 \cdot \mathbb{E} [Z_{n+1}^2] + \mathbb{E} \left[ Z_{n+1}^3 \right] \tag{14}
\]
\[
= \sum_{i=1}^{n+1} \mathbb{E} [Z_i^3]. \tag{15}
\]

We note that Lemma 1 cannot be easily generalized to the fourth or higher powers.
The bivariate $n^{th}$ order Taylor polynomial of a function $g(x, y)$ around the point $(x_0, y_0)$ is given by

$$T_n(x, y) = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \frac{1}{(i+j)!} \left( \frac{\partial^{i+j}g(x_0, y_0)}{\partial x^i \partial y^j} \right) (x-x_0)^i (y-y_0)^j. \quad (16)$$

When $g(x, y) = x/y$, we get $\frac{\partial g}{\partial y} = (-1)^j y^{-(j+1)}$, $\frac{\partial^{i+j}g}{\partial x^i \partial y^{j+1}} = (-1)^j y^{-(j+1)}$, and $\frac{\partial^{i+j}g}{\partial x^i \partial y^j} = 0$ whenever $i \geq 2$. Consequently, the Taylor polynomial will become

$$T_n(x, y) = \frac{x_0}{y_0} + \sum_{k=1}^{n} \left( \frac{1}{y_0^k} (x-x_0)(y-y_0)^{k-1} + \frac{1}{y_0^{k+1}} (y-y_0)^k \right). \quad (17)$$

Now, let $X$ and $Y$ be random variables. When we take the expected value of $(17)$ applied to $x_0 = EX, y_0 = EY$, we get

$$E[T_n(X, Y)] = \frac{EX}{EY} + \sum_{k=1}^{n} \left( \frac{(-1)^{k-1}}{(EY)^k} E[(X - EX)(Y - EY)^{k-1}] + \frac{(-1)^k}{(EY)^{k+1}} E[(Y - EY)^k] \right). \quad (18)$$

Below, we will use the additive property of second and third central moments, specifically Lemma 1 above, to evaluate $ET_2$ and $ET_3$. Since the additive property does not extend to higher degree central moments, there does not seem to be an immediate simple generalization to approximations by a polynomial of fourth degree or higher.

Let us denote $X_A = S_{N_A}(F_A)$ and $X_B = S_{N_B}(F_B)$. In order to substitute into $(18)$, we will use $X = X_A, Y = X_A + X_B$. We have $E[X - EX] = 0 = E[Y - EY]$. Also,

$$EX = E\left[S_{N_A}(F_A)\right] = N_A \mu \quad (19)$$
$$EY = E\left[X_A + X_B\right] = N_A \mu + N_B \mu = N \mu. \quad (20)$$

By Lemma 1

$$E[(Y - EY)^2] = N_A v_A + N_B v_B \quad (21)$$
$$E[(Y - EY)^3] = N_A s_A + N_B s_B. \quad (22)$$

Finally,

$$E\left[(X - EX)(Y - EY)\right] = E\left[(X_A - EX_A)(X_A - EX_A)\right] + E\left[(X_A - EX_A)(X_B - EX_B)\right]$$

$$= N_A v_A + E\left[X_A - EX_A\right] E\left[X_B - EX_B\right] = N_A v_A \quad (23)$$

and
16 MORAN PROCESS AND WRIGHT-FISHER PROCESS FAVOR LOW VARIABILITY

Figure 4. Errors in approximation of $1/y$ by Taylor polynomials $P_n(y)$ at $y_0$. The approximation by a higher degree polynomial is better only on $(0, 2y_0)$ and worse on $(2y_0, \infty)$.

\[
E[(X - EX)(Y - EY)] =
\text{(26)}
\]

\[
= E[(X_A - EX_A)^3] + 2E[(X_A - EX_A)^2(X_B - EX_B)] + E[(X_A - EX_A)(X_B - EX_B)^2]
\text{(27)}
\]

\[
= N_A s_A + 2E[(X_A - EX_A)^2]E[X_B - EX_B] + E[X_A - EX_A]E[(X_B - EX_B)^2]
\text{(28)}
\]

\[
= N_A s_A.
\text{(29)}
\]

The formula (8) now follows from (18).

Let us conclude the section by demonstrating that an approximation by a higher degree polynomial is not always better. Consider a function $1/y$ of just one variable $y$. The Taylor polynomials are $P_n(y) = \frac{1}{y_0} \sum_{k=0}^{n} (-1)^k \left( \frac{y-y_0}{y_0} \right)^k$. The approximation of $1/y$ by $P_{n+1}(y)$ for positive $y$ is better than the approximation by $P_n(y)$ if and only if $y - y_0 < y_0$, i.e. on $(0, 2y_0)$, see Figure 4.

Similarly, in a bivariate setting, when the distribution of the fitness is such that $Y < 2EY = 2N\mu$ almost surely for all $N_A + N_B = N$ (in particular when $F_A < 2\mu$ and $F_B < 2\mu$ almost surely), then the third degree approximation is better than the second degree approximation which is better than the constant approximation. If the condition does not hold, both second and third degree approximation still seem fine (see for example Rice and Papadopoulos (2009) and Rice (2015)) but the approximation by a second degree may actually be better as shown in Figure 5.

We cannot guarantee $Y < 2EY$ and thus the goodness of the approximation (8) for all $N$. However, we have a specific random variable $Y = X_A + X_B$ and we can prove that the probability of $Y > 2EY$ approaches 0 as $N$ goes to $\infty$, see Lemma 2 below. Consequently, for large $N$, the approximation (8) still works well.
Lemma 2. Let \( F_A \) and \( F_B \) be random variables with finite means and variances that are positive almost surely. Let \( X_A = S_{N_A}(F_A) \), \( X_B = S_{N_B}(F_B) \) and \( Y = X_A + X_B \). Then the probability of \( Y > 2N\mu \) converges to 0 as \( N = N_A + N_B \) goes to \( \infty \).

Proof. We have

\[
P(Y \leq 2N\mu) \geq P\left(X_A < \left(1.5N_A + 0.5N_B\right)\mu\right) \cdot P\left(X_B < \left(0.5N_A + 1.5N_B\right)\mu\right) \tag{30}
\]

\[
= P\left(X_A - N_A\mu < \frac{N_A + N_B}{2}\mu\right) \cdot P\left(X_B - N_B\mu < \frac{N_A + N_B}{2}\mu\right) \tag{31}
\]

\[
= P\left(\frac{X_A - N_A\mu}{\sqrt{N_A}} < \frac{N_A + N_B}{2\sqrt{N_A}}\mu\right) \cdot P\left(\frac{X_B - N_B\mu}{\sqrt{N_B}} < \frac{N_A + N_B}{2\sqrt{N_B}}\mu\right). \tag{32}
\]

We just need to prove that both factors in (32) approach 1 as \( N_A + N_B \) approaches infinity. By the central limit theorem, there is \( N_0 \) such that both factors approach 1 if \( N_A > N_0 \) and \( N_B > N_0 \). Also, for any \( N_A < N_0 \), \( (X_A - N_A)/\sqrt{N_A} \) is a random variable with finite mean and variance and so the first factor of (32) converges to 1 for all \( N_A < N_0 \) as \( N_B \to \infty \). Similarly, the second factor of (32) converges to 1 for all \( N_B < N_0 \) as \( N_A \to \infty \). \( \square \)