

Asymptotic behavior of mean fixation times for a single mutant in the Moran process with frequency-independent fitnesses

Rosângela A. Pires¹
Armando G. M. Neves²

¹Departamento de Matemática, Universidade Federal de Minas Gerais
r04assis@gmail.com

²Departamento de Matemática, Universidade Federal de Minas Gerais
aneves@mat.ufmg.br

January 2, 2023

Abstract

We derive asymptotic formulas in the limit when population size N tends to infinity for mean fixation times (conditional and unconditional) for a single individual of one type in a population with two types and governed by the Moran process. We consider only the case in which the fitness of the two types of individuals do not depend on the population frequencies. It will turn out that our formulas are much more accurate than the ones found by Antal and Scheuring (Bull Math Biol 68(8):1923–1944, 2006).

Keywords: Markov chains, Asymptotic analysis, Birth-death processes

1 Introduction

In this paper we will be concerned with the Moran process [17] in its simpler setting: a population with fixed size equal to N haploid individuals which reproduce asexually. There are two types of individuals, that we call A and

B, distinguished e.g. by different alleles at a single locus. Time is considered to be discrete. At each time step two random processes happen: one individual produces a descendant having the same type as itself (*no-mutation hypothesis*) and an individual dies, being replaced by the descendant of the reproducing individual. The reproduction and death lotteries are *independent*. We consider the death lottery to be uniform, but the reproduction lottery is such that fitter individuals reproduce more frequently. In this paper we will suppose that the fitness of an individual depends only on its type A or B.

Taylor et al. [20] extended the Moran process for two types of individuals to the Evolutionary Game Theory [15] setting in which fitnesses depend not only on individuals' types, but also on their population frequencies. The Moran process has been also extended to three or more types [11] and to populations structured by graphs [9, 18, 5]. We will not treat any of these extensions in the present paper.

Another popular stochastic model for the same situation is the Wright-Fisher process [12, 21], in which at each time step the whole population is replaced by a new offspring generation, which composition depends on the parent generation. From a mathematical point of view, both Wright-Fisher and Moran processes are discrete-time Markov chains with a finite state space. Due to the no-mutation hypothesis, the states in which the population is either all-A, or all-B are *absorbing*. All other states are *transient*. Due to the existence of absorbing states, and also to the finiteness of the population, it can be shown [1, 13] that in both processes, after a long enough time, the population state will be, with probability 1, either all-A, or all-B. This mathematical phenomenon is called *fixation*. Important quantities to be calculated are, for given initial composition of the population, the fixation probability of either type, the mean times for unconditional fixation, and for conditional fixation of either A or B individuals.

The transition probability from any transient state to any other state in the Wright-Fisher process is positive. On the contrary, the only non-zero transition probabilities in the Moran process are between states in which the number of A individuals differ at most by 1. As a consequence, for the Moran process there exist exact explicit formulas for quantities such as fixation probabilities [18], see (5), and mean unconditional, see (13), or conditional fixation times [2, 3], see (6), (8). These formulas also hold when fitnesses depend on population frequencies. On the other hand, calculations of the same quantities in the Wright-Fisher model must rely on approximations,

such as the diffusion approximation [10]. For the extension of the Moran process to three or more types we also do not know exact formulas for fixation probabilities or times. The same for structured populations, except for some highly symmetric situations [5, 14, 16].

The possibility of some exact calculations is an advantage of the Moran process over Wright-Fisher's. Despite that, although exact and explicit, the alluded formulas for the Moran process are unwieldy in the sense that they do not immediately display qualitative and quantitative features of the fixation probabilities or times. As an example, they do not show how mean fixation times depend on the population size N . Some works have been devoted to developing simpler ways to understand them. Antal and Scheuring [3] calculated – in the game-theoretic Moran process – the asymptotic behavior when the population size N tends to infinity both for fixation probabilities and for conditional mean fixation times for a single individual of either type. de Souza et al. [8] studied the shapes of the graphs for fixation probabilities in all scenarios of the game-theoretic Moran process, and also provided asymptotic formulas for the fixation probabilities for any population frequency of A individuals. Chalub and Souza [6] also have asymptotic formulas for the Moran process fixation probabilities, but restricted to what they call *regular families of suitable birth-death processes*. This class includes the Moran process, but not in the strong-selection regime considered here and also in [3] and [8].

This paper continues the work in [8] and derives asymptotic formulas in the limit $N \rightarrow \infty$ for the mean fixation times, both conditional and unconditional, but restricted to the case of frequency-independent fitnesses and initial condition of a single A or a single B individual (and $N - 1$ individuals of the other type), i.e. the situation in which a mutant appears at a homogeneous population. We want to know how long it takes, in average, for such mutants either to fixate, if they do so, or for the population to become uniform again.

The results in this paper are stronger in one sense than those of Antal and Scheuring [3], because they are more precise and mathematically fully rigorous, as will be further explained. We also include asymptotic formulas for the unconditional times, which do not appear in [3]. But their results apply to the game-theoretic setting in which fitnesses depend on population frequencies. We avoid here these more complicated cases, postponing their treatment to another work.

The paper is organized as follows. In Section 2 we define the Moran

process for two types of individuals, introduce all the notation and terminology pertaining to it and display without proof the exact formulas on which the rest of the paper is based. Having defined the necessary terms, we also describe the results of this paper. In Section 3 we derive the asymptotic formulas for mean conditional fixation times and for mean absorption (i.e. unconditional fixation) times and compare them with the result of the exact formulas. Some overall discussion is finally provided in Section 4. Appendix A is devoted to the statement of purely mathematical results and their proofs.

2 Moran process: definitions and formulas

We start with the following definition [18]:

Definition 1. *A birth-death process is a discrete-time Markov chain with finite state space $S = \{0, 1, \dots, N\}$, in which states 0 and N are absorbing, transitions between states i and j with $|i - j| \geq 2$ have probability 0 and, for $i \in \{1, 2, \dots, N - 1\}$, transitions $i \rightarrow i \pm 1$ have non-zero probabilities.*

It turns out that states $1, 2, \dots, N - 1$ in a birth-death process are all transient and that fixation of the chain in one of the absorbing states occurs with total probability [1, 13]. If $X_n \in S$ denotes the state at time n , we define the relevant non-zero transition probabilities

$$\alpha_i = P(X_{n+1} = i + 1 \mid X_n = i) \quad (1)$$

and

$$\beta_i = P(X_{n+1} = i - 1 \mid X_n = i). \quad (2)$$

As states 0 and N are absorbing, $\alpha_0 = \beta_N = 0$.

We let π_i denote the probability that a birth-death process is absorbed at state N , given that it started in state i . Of course, the corresponding probability for fixation at state 0 is $1 - \pi_i$. Letting, for $i = 1, 2, \dots, N$,

$$r_i = \frac{\alpha_i}{\beta_i}, \quad (3)$$

it can be shown, see e.g. chapter 6 in [18], that

$$\pi_1 = \frac{1}{1 + \sum_{j=1}^{N-1} \prod_{k=1}^j r_k^{-1}}, \quad (4)$$

and, for $i = 2, \dots, N$

$$\pi_i = \frac{1 + \sum_{j=1}^{i-1} \prod_{k=1}^j r_k^{-1}}{1 + \sum_{j=1}^{N-1} \prod_{k=1}^j r_k^{-1}}. \quad (5)$$

We define now the mean times we will deal with in this paper:

Definition 2. *The mean absorption time (or mean unconditional fixation time) t_i for a birth-death process is the expectation of the random time it takes to the Markov chain, starting at state i , to reach either of the absorbing states.*

The mean conditional fixation time at state N with initial condition i , denoted t_i^N , is the expectation of the random time it takes for the chain to go from state i to state N , conditioned that it is absorbed at N . Analogously, t_i^0 is the mean conditional time for fixation at state 0 starting from i .

As for fixation probabilities, exact formulae for the above defined times may be derived. Antal and Scheuring [3] do the derivation for the mean conditional fixation times. In this paper's notation, their formulae are

$$t_i^0 = \frac{1}{1 - \pi_i} (Q_N - Q_i) - Q_N, \quad (6)$$

where

$$Q_i = \sum_{n=1}^{i-1} \left(\prod_{k=1}^n r_k^{-1} \right) \sum_{j=1}^n \frac{1 - \pi_j}{\alpha_j} \left(\prod_{\ell=1}^j r_\ell \right), \quad (7)$$

and

$$t_i^N = \frac{1}{\pi_i} (S_0 - S_i) - S_0, \quad (8)$$

with

$$S_i = \sum_{n=i+1}^{N-1} \left(\prod_{k=1}^{n-1} r_k^{-1} \right) \sum_{j=n}^{N-1} \frac{\pi_j}{\alpha_j} \left(\prod_{\ell=1}^j r_\ell \right). \quad (9)$$

With some work in (6) with $i = N - 1$ and (8) with $i = 1$, it is possible to show that the mean conditional fixation times t_1^N and t_{N-1}^0 are exactly equal [3]. This is a consequence of a symmetry in conditional fixation times proved by Taylor et al. [19]. We may thus define

$$t_{fix} = t_{N-1}^0 = t_1^N. \quad (10)$$

As some of our results are exactly for t_{fix} , we write a simpler formula for it:

$$t_{fix} = \sum_{n=1}^{N-1} \frac{\pi_n}{\alpha_n} \left(\prod_{m=1}^n r_m \right) \sum_{k=n}^{N-1} \left(\prod_{\ell=1}^k r_\ell^{-1} \right). \quad (11)$$

The derivation of the formulae for the mean absorption times is similar. These formulae appear e.g. in [2] and, translated to our notation, are

$$t_1 = \pi_1 \sum_{j=1}^{N-1} \left(\prod_{k=1}^j r_k^{-1} \right) \sum_{i=1}^j \frac{1}{\alpha_i} \left(\prod_{\ell=1}^i r_\ell \right), \quad (12)$$

and, for $i = 2, 3, \dots, N-1$,

$$t_i = \sum_{j=i}^{N-1} \left(\prod_{\ell=1}^j r_\ell^{-1} \right) \sum_{k=1}^j \frac{1}{\alpha_k} \left(\prod_{m=1}^k r_m \right) - t_1 \sum_{j=i}^{N-1} \left(\prod_{\ell=1}^j r_\ell^{-1} \right). \quad (13)$$

The Moran process is a special case of birth-death process. The state $i \in S$ is identified as the number of A individuals in the population. Of course, the number of B individuals in a population at state i is just $N-i$. In the more general case [20, 18, 8], the fitnesses of A and B individuals may be calculated through a pay-off matrix and may depend on their population frequencies. In the simpler context of this paper, we suppose that type-A individuals have fitness $r > 0$, whereas B individuals have fitness 1. Accordingly, parameter r will be called *the relative fitness of type-A individuals*. The Moran process probability for drawing an A individual for reproduction when the population is at state i is defined as

$$\frac{ir}{ir + N - i}.$$

If $r > 1$, A individuals are fitter than B. If $0 < r < 1$, it is the reverse. The important case $r = 1$ is known as the *neutral* Moran process.

With the above formula, the reproduction lottery favors fitter individuals, as remarked in Section 1. On the other hand, the death lottery will be considered uniform. Then, the probability of drawing a B individual for death in state i is

$$\frac{N-i}{N}.$$

The transition $i \rightarrow i+1$ happens only if an A is drawn for reproduction and a B drawn for death. Its probability α_i , defined in (1), is thus the product

of the above probabilities:

$$\alpha_i = \frac{i(N-i)r}{N(ir+N-i)} . \quad (14)$$

An analogous reasoning gives for the $i \rightarrow i-1$ transition

$$\beta_i = \frac{i(N-i)}{N(ir+N-i)} . \quad (15)$$

The ratio r_i defined in (3) becomes $r_i = r$, i.e. frequency-independent. We say that (14) and (15) define the Moran process with *frequency-independent* fitnesses. The products and sums in (5) can then be easily calculated and the fixation probability becomes

$$\pi_i = \begin{cases} \frac{1-r^{-i}}{1-r^{-N}}, & \text{if } r \neq 1 \\ \frac{i}{N}, & \text{if } r = 1 \end{cases} . \quad (16)$$

Frequency independence of r_i causes many simplifications to arise in the formulas for the mean absorption and fixation times, too.

Explicitly, we get from (14) that

$$\frac{1}{\alpha_j} = N \left(\frac{1}{N-j} + \frac{1}{rj} \right) . \quad (17)$$

Using this with (16) in (12) we have, for $r \neq 1$, a simpler formula for the mean absorption time of an A mutant in a B population:

$$t_1 = \frac{N}{r} \frac{1-r^{-1}}{1-r^{-N}} \sum_{k=1}^{N-1} r^{-k} \sum_{j=1}^k \left(\frac{r}{N-j} + \frac{1}{j} \right) r^j .$$

Exchanging the summation order in the above formula, we get, after some simple manipulations,

$$t_1 = \frac{N(1+r)}{r(1-r^{-N})} H_{N-1} - \frac{Nr^{-N}}{r(1-r^{-N})} \sum_{j=1}^{N-1} \frac{r^j}{j} - \frac{Nr^{-N}}{1-r^{-N}} \sum_{j=1}^{N-1} \frac{r^j}{N-j} , \quad (18)$$

where

$$H_n = \sum_{i=1}^n \frac{1}{i} \quad (19)$$

are the *harmonic numbers*. Replacing r by r^{-1} in (18) we obtain a formula for the mean absorption time for a B mutant in an A population:

$$t_{N-1} = \frac{N(1+r)}{(1-r^N)} H_{N-1} - \frac{Nr}{(r^{-N}-1)} \sum_{j=1}^{N-1} \frac{r^{-j}}{j} - \frac{N}{r^{-N}-1} \sum_{j=1}^{N-1} \frac{r^{-j}}{N-j}. \quad (20)$$

For t_{fix} , (11) becomes similarly

$$\begin{aligned} t_{fix} &= \frac{N(1+r)}{(r-1)(1-r^{-N})} [(1+r^{-N})H_{N-1}] \\ &\quad - \frac{Nr^{-N}}{(r-1)(1-r^{-N})} \left[r \sum_{j=1}^{N-1} \frac{r^j}{N-j} + \sum_{j=1}^{N-1} \frac{r^j}{j} \right] \\ &\quad - \frac{N}{(r-1)(1-r^{-N})} \left[r \sum_{j=1}^{N-1} \frac{r^{-j}}{N-j} + \sum_{j=1}^{N-1} \frac{r^{-j}}{j} \right]. \end{aligned} \quad (21)$$

The corresponding formulas for the neutral case $r = 1$ may be derived directly from (12) and (11). The summations can be trivially performed. We state here, for completeness, the results for mean absorption and mean conditional fixation times in the *neutral* Moran process:

$$\begin{aligned} t_1 &= t_{N-1} = NH_{N-1} \\ &= N \log(N-1) + N\gamma + \frac{1}{2} + O(\frac{1}{N^2}) \end{aligned} \quad (22)$$

and

$$t_{fix} = N(N-1), \quad (23)$$

where the asymptotic estimate (22) is consequence of a classical result due to Euler, see Lemma 1.

The main results of this paper are formulas such as the above two ones, displaying explicitly how the mean times for the *non-neutral* Moran process depend on the population size N and on the relative fitness r of A individuals. They appear as Theorems 1, 2 and 3 in the following section.

3 Mean conditional fixation and mean absorption times

We start with the mean fixation time of a single individual – a mutant – of either type in a population with $N - 1$ individuals of the other type *condi-*

tioned to fixation of the mutant's type. Although the fixation probabilities of a single A and of a single B individual may be quite different, their mean conditional fixation times are exactly equal [19]. This quantity, denoted as t_{fix} , is exactly given by (21) and its asymptotic behavior when the population size N tends to infinity is calculated in the following result:

Theorem 1. *Let r be the relative fitness of A individuals in the non-neutral Moran process with frequency-independent fitnesses.*

If $r > 1$,

$$\begin{aligned} t_{fix} = & N \left(\frac{r+1}{r-1} \right) [\log(N-1) + \gamma + \log(1-r^{-1})] + \\ & + \frac{(r+1)(r-3)}{2(r-1)^2} + O\left(\frac{1}{N}\right). \end{aligned} \quad (24)$$

If $r \in (0, 1)$,

$$\begin{aligned} t_{fix} = & N \left(\frac{r+1}{1-r} \right) [\log(N-1) + \gamma + \log(1-r)] + \\ & + \frac{(r+1)(1-3r)}{2(1-r)^2} + O\left(\frac{1}{N}\right). \end{aligned} \quad (25)$$

Proof. The right-hand side of (21) contains four sums and also the harmonic numbers H_n . All of these are estimated in the results in Appendix A. We may obtain both (24) and (25) by just using these results, taking the care to include in the remainders terms which are $O(1/N)$ or smaller.

Alternatively, the formula for the case $r \in (0, 1)$ may also be obtained from the formula for $r > 1$ by using the fact that a single A mutant with fitness $r \in (0, 1)$ is equivalent in terms of conditional fixation time to a single B mutant with fitness $\bar{r} = 1/r > 1$. Using (24) with r replaced by \bar{r} , we get (25). \square

Accuracy of the estimates in Theorem 1 can be checked in Fig. 1, in which the vertical axis represents the ratio between the numerically evaluated t_{fix} given by exact formula (11) and the asymptotic estimates (24) and (25). This figure should be compared with the analogous Fig. 5 in the paper by Antal and Scheuring [3]. The scale of the vertical axes in our figures is of order 10^4 times smaller than theirs, indicating a much larger accuracy.

We remind that our result is a special case of theirs, which was obtained for the more difficult case of the Moran process with frequency-dependent

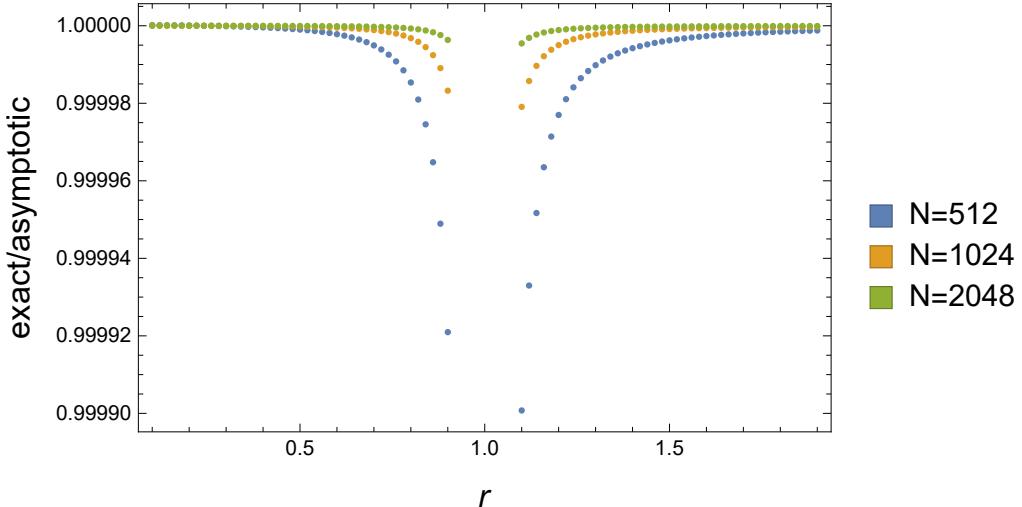


Figure 1: Ratio between t_{fix} evaluated by the exact formula (21) and its asymptotic approximations given by (24) and (25) as a function of r for population values $N = 512$, $N = 1024$ and $N = 2048$.

fitness, not treated in this paper. Despite that, their asymptotic estimate was carried out only to the dominant $O(N \log N)$ term, and thus, our estimate, containing $O(N)$ and $O(1)$ terms, is much more precise.

The next situation we consider is the mean absorption time t_1 when a single A individual appears in a population with $N - 1$ B individuals. It may happen that the final state is either all-A, or all-B and t_1 is the mean time it takes for one of them to happen. The result is given by

Theorem 2. *If $r > 1$,*

$$t_1 = \left(\frac{r+1}{r} \right) N \log(N-1) + \left[\left(\frac{r+1}{r} \right) \gamma + \log(1-r^{-1}) \right] N + \left(\frac{r+1}{2r} - \frac{1}{r(r-1)} \right) + O\left(\frac{1}{N}\right). \quad (26)$$

If $r \in (0, 1)$,

$$t_1 = \left[-\frac{\log(1-r)}{r} \right] N + \frac{r}{1-r} + O\left(\frac{1}{N}\right). \quad (27)$$

Proof. As in Theorem 1, it suffices to use the results in Appendix A. \square

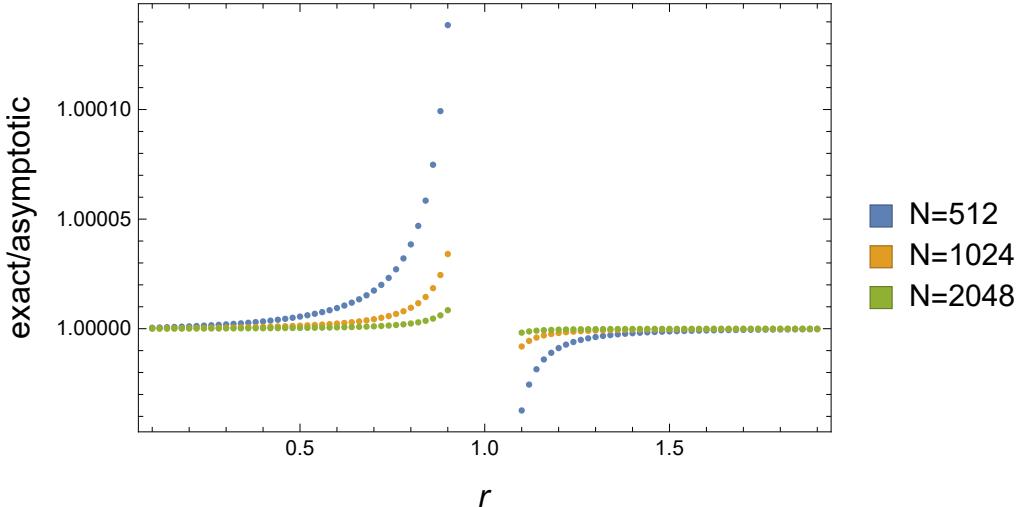


Figure 2: Ratio between t_1 evaluated by the exact formula (18) and its asymptotic approximations given by (26) and (27) as a function of r for population values $N = 512$, $N = 1024$ and $N = 2048$.

The reader should notice that the absorption time when the population has a single A individual is $O(N \log N)$ if $r > 1$, i.e. the mutant is fitter than the rest of the population, whereas it is only $O(N)$ if $0 < r < 1$, i.e. the mutant is less fit than the rest of the population. These are reasonable results, because there is an $O(1)$ probability, use (16) with $i = 1$, that the mutation is fixated if $r > 1$, resulting in a longer absorption time of the same order as t_{fix} when $r > 1$. On the other hand, if $0 < r < 1$, the fixation probability for the mutant is $O(r^N)$, very small, and the most probable outcome is extinction of the mutant type. This will usually happen when the first mutant or mutants are drawn in the death lottery, which takes a time $O(N)$.

Fig. 2 is analogous to Fig. 1 and displays the high accuracy of the asymptotic results in Theorem 2.

At last, we quote here the results for the mean absorption time t_{N-1} for the situation of a single B mutant in a population with $N - 1$ A individuals. Their proof is again a simple consequence of Appendix A, or of Theorem 2 by noticing that a single B with fitness r has the same mean absorption time as an A with fitness $\bar{r} = 1/r$.

Theorem 3. *If $r > 1$,*

$$t_{N-1} = [-r \log(1 - r^{-1})] N + \frac{1}{r-1} + O\left(\frac{1}{N}\right). \quad (28)$$

If $0 < r < 1$,

$$\begin{aligned} t_{N-1} = & (r+1) N \log(N-1) + [(r+1)\gamma + \log(1-r)] N \\ & + \left(\frac{r+1}{2} - \frac{r^2}{1-r} \right) + O\left(\frac{1}{N}\right). \end{aligned} \quad (29)$$

As the results for t_{N-1} are obtainable from the analogous results for t_1 , we found it unnecessary to illustrate their accuracy in a figure such as the previous ones.

4 Discussion

The Moran process is an important stochastic model for the evolution of finite populations. In such populations, if mutations can be neglected, genetic traits will be fixated in the population. The Moran process with the *no mutation hypothesis* exhibits the fixation phenomenon. But for how long can we expect that mutations at a certain trait will not occur? Of course the larger population sizes are, the longer fixation times will be, whereas mutations will appear more frequently. We must be more quantitatively explicit on fixation times if we want to use a hypothesis of no-mutations.

Although exact and explicit, formulas (5), (11) and (12) – used in this paper – are not able to answer questions such as how will the mean absorption time change if population is doubled? Or how will the mean fixation time change if the fitness of A individuals increases by 10%?

The main reasons for such inadequacy is that all of these formulas involve sums of products in a complicated way. In particular, the population size N appears as a summation limit.

Antal and Scheuring [3] were first in answering such quantitative questions both for fixation probabilities and for mean fixation times in the limit when population size N tends to infinity. Moreover, they did so in the more general setting in which fitnesses of the individuals depend on their population frequencies.

Their approach involved transforming the product appearing in the alluded formulas in exponentials of sums of logarithms. This procedure was

adopted again by Chalub and Souza [6], which defined the *fitness potential* further explored in [7]. One of us, with de Souza and Ferreira [8], also expanded the work of Antal and Scheuring on the asymptotic behavior of fixation probabilities when the population size N tends to infinity. In doing so, we noticed the lack of complete mathematical rigor in the method of Antal and Scheuring. At a certain point, they approximate a sum by an integral, without taking into account the errors produced in doing so. Following [6], these errors were called *continuation errors* and we were able to show [8] that they must be taken into account if accurate asymptotic formulas for the fixation probabilities are to be obtained.

If the fitnesses of types A and B do not depend on the population frequencies, the fixation probabilities are given by the simple formula (16), in which the asymptotic behavior when $N \rightarrow \infty$ is evident. The same cannot be said of formulas (18), (20) and (21) for the mean fixation times. This difference justifies the present work. Although we need not deal with continuation errors and fitness potentials here, we still feel the necessity of displaying explicitly how the mean fixation times depend on N and r when population size N is large. Moreover, as displayed in comparing Figs. 1 and 2 with analogous figures in [3], inclusion of sub-leading terms in the formulas of our Theorems 1, 2 and 3 improves a lot the accuracy of the results. The reader should also notice that our formulas also contain the asymptotic order of the remainder terms.

For the more general game-theoretic Moran process, in which fitnesses depend on the population frequencies, we should use formulas such as (11) and (12) instead of the simpler ones cited above. This will certainly involve consideration of continuation errors and will be left to a future work.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

RAP received scholarships from CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil) and CAPES. AGMN is partially funded by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Brazil.

A Some auxiliary results and their proofs

Our first result here is a classical consequence of the Euler-Maclaurin formula. It appears e.g. as formula (31) in [4].

Lemma 1. *If $n \in \mathbb{N}$, then*

$$H_n = \log n + \gamma + \frac{1}{2n} + O\left(\frac{1}{n^2}\right), \quad (30)$$

where γ is the Euler-Mascheroni constant.

Lemma 2. *For $z > 1$ we have*

$$\sum_{k=1}^{N-1} \frac{z^{-k}}{k} = -\log(1 - z^{-1}) + O\left(\frac{z^{-N}}{N}\right). \quad (31)$$

Proof. Notice that $\sum_{k=1}^{\infty} \frac{x^k}{k}$ is just the Maclaurin series for $-\log(1-x)$, which converges if $0 < x < 1$. Thus, for $z > 1$,

$$\sum_{k=1}^{N-1} \frac{z^{-k}}{k} = -\log(1 - z^{-1}) - \sum_{k=N}^{\infty} \frac{z^{-k}}{k}.$$

But

$$\sum_{k=N}^{\infty} \frac{z^{-k}}{k} < \frac{1}{N} \sum_{k=N}^{\infty} z^{-k} = \frac{1}{N} \frac{z^{-N}}{1 - z^{-1}}.$$

□

Theorem 4.

$$\sum_{j=1}^{N-1} \frac{r^j}{j} = \begin{cases} \frac{r^N}{N} \left[\frac{1}{r-1} + O\left(\frac{1}{N}\right) \right], & \text{if } r > 1 \\ -\log(1 - r) + O\left(\frac{r^N}{N}\right), & \text{if } 0 < r < 1 \end{cases}. \quad (32)$$

Proof. We start with the case $r > 1$. Reversing the summation order, we get

$$\begin{aligned} \sum_{j=1}^{N-1} \frac{r^j}{j} &= r^N \sum_{k=1}^{N-1} \frac{r^{-k}}{N-k} = \frac{r^N}{N} \sum_{k=1}^{N-1} \frac{r^{-k}}{1 - \frac{k}{N}} \\ &= \frac{r^N}{N} \left[\sum_{k=1}^{N-1} r^{-k} + \sum_{k=1}^{N-1} \left(\frac{1}{1 - \frac{k}{N}} - 1 \right) r^{-k} \right]. \end{aligned}$$

The first sum in the above expression is equal to $\frac{r^{-1}-r^{-N}}{1-r^{-1}} = \frac{1}{r-1} + O(r^{-N})$. In order to prove the estimate for $r > 1$, we must show that the second sum is $O(1/N)$. In fact,

$$\frac{1}{1-\frac{k}{N}} - 1 = \frac{k}{N} + \left(\frac{k}{N}\right)^2 \frac{1}{1-\frac{k}{N}}.$$

Thus

$$\sum_{k=1}^{N-1} \left(\frac{1}{1-\frac{k}{N}} - 1 \right) r^{-k} = \frac{1}{N} \sum_{k=1}^{N-1} k r^{-k} + \frac{1}{N^2} \sum_{k=1}^{N-1} \frac{k^2}{1-\frac{k}{N}} r^{-k}.$$

For $k \in \{1, 2, \dots, N-1\}$ we have $1 - \frac{k}{N} \geq \frac{1}{N}$. We may also extend the summations up to ∞ and it results that

$$\begin{aligned} \sum_{k=1}^{N-1} \left(\frac{1}{1-\frac{k}{N}} - 1 \right) r^{-k} &< \frac{1}{N} \sum_{k=1}^{\infty} k r^{-k} + \frac{1}{N} \sum_{k=1}^{\infty} k^2 r^{-k} \\ &= O\left(\frac{1}{N}\right). \end{aligned}$$

With this we prove the estimate for $\sum_{j=1}^{N-1} \frac{r^j}{j}$ when $r > 1$.

The case $0 < r < 1$ is a simple consequence of Lemma 2 with $z = r^{-1}$. \square

Corollary 1.

$$\sum_{j=1}^{N-1} \frac{r^{-j}}{j} = \begin{cases} -\log(1-r^{-1}) + O\left(\frac{r^{-N}}{N}\right), & \text{if } r > 1 \\ \frac{r^{-N}}{N} \left[\frac{r}{1-r} + O\left(\frac{1}{N}\right) \right], & \text{if } 0 < r < 1 \end{cases}. \quad (33)$$

Proof. It suffices to define $\bar{r} = 1/r$, rewrite the sum in the left-hand side in terms of \bar{r} and use Theorem 4 to produce the estimates. \square

Theorem 5.

$$\sum_{j=1}^{N-1} \frac{r^j}{N-j} = \begin{cases} -\log(1-r^{-1}) r^N + O\left(\frac{1}{N}\right), & \text{if } r > 1 \\ \frac{1}{N} \frac{r}{1-r} + O\left(\frac{1}{N^2}\right), & \text{if } 0 < r < 1 \end{cases}. \quad (34)$$

Proof. We replace the summation index j by $k = N - j$, getting

$$\sum_{j=1}^{N-1} \frac{r^j}{N-j} = r^N \sum_{k=1}^{N-1} \frac{r^{-k}}{k}.$$

For $r > 1$ the result follows by using Lemma 2 with $z = r$. For $0 < r < 1$, we define $\bar{r} = 1/r$ and use the first of the estimates in (32). \square

For completeness, we state a simple consequence of Theorem 5 also appearing in the formulas for the mean fixation times:

Corollary 2.

$$\sum_{j=1}^{N-1} \frac{r^{-j}}{N-j} = \begin{cases} \frac{1}{N} \frac{1}{r-1} + O\left(\frac{1}{N^2}\right), & \text{if } r > 1 \\ -\log(1-r) r^{-N} + O\left(\frac{1}{N}\right), & \text{if } 0 < r < 1 \end{cases}. \quad (35)$$

The proof goes as in Corollary 1.

References

- [1] Linda J. S. Allen. *An introduction to stochastic processes with applications to biology*. Chapman & Hall/CRC, Boca Raton, FL, 2011.
- [2] Philipp Altrock. *Selection intensity and the time to fixation in evolutionary systems*. PhD thesis, Christian-Albrechts-Universität zu Kiel, 05 2011.
- [3] Tibor Antal and István Scheuring. Fixation of Strategies for an Evolutionary Game in Finite Populations. *B. Math. Biol.*, 68(8):1923–1944, 2006.
- [4] Tom M. Apostol. An Elementary View of Euler’s Summation Formula. *Am. Math. Mon.*, 106:409 – 418, 1999.
- [5] M. Broom and J. Rychtář. An analysis of the fixation probability of a mutant on special classes of non-directed graphs. *Proc. R. Soc. A*, 464:2609–2627, 2008.
- [6] Fabio A. C. C. Chalub and Max O. Souza. Fixation in large populations: a continuous view of a discrete problem. *J. Math. Biol.*, 72(1):283–330, 2016.
- [7] Fabio A. C. C. Chalub and Max O. Souza. Fitness potentials and qualitative properties of the wright-fisher dynamics. *Journal of Theoretical Biology*, 457:57–65, 2018.

- [8] Evandro P. de Souza, Eliza M. Ferreira, and Armando G. M. Neves. Fixation probabilities for the Moran process in evolutionary games with two strategies: graph shapes and large population asymptotics. *J. Math. Biol.*, 78(6):1033 – 1065, 2019.
- [9] E. Lieberman, C. Hauert, and M. A. Nowak. Evolutionary dynamics on graphs. *Nature*, 433:312 – 316, 2005.
- [10] Warren J. Ewens. *Mathematical population genetics. I. , Theoretical introduction*. Interdisciplinary applied mathematics. Springer, New York, 2004.
- [11] Eliza M Ferreira and Armando G M Neves. Fixation probabilities for the Moran process with three or more strategies: general and coupling results. *J. Math. Biol.*, 81:277–314, 2020.
- [12] R. A. Fisher. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford, 1930.
- [13] Charles M. Grinstead and J. Laurie Snell. *Introduction to Probability*. AMS, 2003.
- [14] C. Hadjichrysanthou, M. Broom, and J. Rychtář. Evolutionary games on star graphs under various updating rules. *Dyn. Games Appl.*, 1:386, 2011.
- [15] J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge, 1998.
- [16] T. Monk, P. Green, and M. Paulin. Martingales and fixation probabilities of evolutionary graphs. *Proc. R. Soc. A*, 470:20130730, 2014.
- [17] P. A. P. Moran. Random processes in genetics. *P. Camb. Philos. Soc.*, 54(1):60–71, 1958.
- [18] M. Nowak. *Evolutionary Dynamics*. The Belknap Press of Harvard University Press, Cambridge, Massachusetts, and London, England, 1 edition, 2006.
- [19] C. Taylor, Y. Iwasa, and M. A. Nowak. A symmetry of fixation times in evolutionary dynamics. *J. Theor. Biol.*, pages 243, 245–251, 2006.

- [20] Christine Taylor, Drew Fudenberg, Akira Sasaki, and Martin A. Nowak. Evolutionary game dynamics in finite populations. *B. Math. Biol.*, 66(6):1621–1644, 2004.
- [21] S. Wright. Evolution in mendelian populations. *Genetics*, 16:97–159, 1931.